

Analisis Status Gizi Anak Menggunakan Metode Klustering pada Dataset Anthropometri

Joang Ipmawati^{1)*}, Ilham Unggara²⁾

¹⁾²⁾Universitas Nahdlatul Ulama Yogyakarta

Jl. Siliwangi (Ring Road), Banyuraden, Gamping, Sleman, D.I. Yogyakarta, Indonesia

¹⁾joang@unu-jogja.ac.id

²⁾ilham@unu-jogja.ac.id

Article history:

Received 01 Des 2024;
Revised 03 Des 2024;
Accepted 05 Des 2024;
Available online 27 Des 2024

Keywords:

Anthropometr
K-means
Klustering
Stunting
Status Gizi

Abstrak

Stunting merupakan salah satu masalah kesehatan masyarakat yang serius di Indonesia, memengaruhi pertumbuhan fisik dan kognitif anak-anak. Penelitian ini bertujuan untuk mengelompokkan status gizi anak-anak berdasarkan data antropometri untuk mengidentifikasi kelompok risiko tinggi stunting. Dataset yang digunakan terdiri dari 120.000 entri dengan variabel umur (bulan), tinggi badan (cm), jenis kelamin, dan status gizi, yang diperoleh dari sumber sekunder. Metode penelitian menggunakan algoritma K-Means untuk klusterisasi data dengan jumlah kluster optimal ditentukan melalui metode Elbow dan Silhouette Score. Proses analisis melibatkan tahap preprocessing, klusterisasi, dan validasi serta evaluasi eksternal. Hasil penelitian menunjukkan bahwa data dapat dikelompokkan ke dalam empat kluster dengan karakteristik yang berbeda. Kluster 0 dan Kluster 2 didominasi oleh anak usia muda (0–35 bulan) dengan rata-rata tinggi badan masing-masing 74.34 cm dan 73.13 cm, yang mencerminkan kelompok risiko tinggi stunting. Sebaliknya, Kluster 1 dan Kluster 3 mencakup anak-anak dengan pertumbuhan optimal, dengan rata-rata tinggi badan di atas 100 cm. Analisis korelasi menunjukkan hubungan signifikan antara tinggi badan, umur, dan status gizi, mendukung pentingnya intervensi gizi pada kelompok risiko tinggi. Penelitian ini memberikan kontribusi penting dalam upaya pencegahan stunting melalui identifikasi kelompok risiko tinggi secara lebih terarah. Temuan ini relevan untuk mendukung program kesehatan masyarakat di Indonesia, khususnya dalam merancang intervensi berbasis data untuk meningkatkan status gizi anak-anak. Pendekatan klusterisasi berbasis machine learning yang digunakan membuktikan efektivitasnya dalam memetakan pola pertumbuhan anak, sehingga dapat digunakan untuk mendukung perencanaan kebijakan yang lebih efisien.

I. PENDAHULUAN

Kesehatan anak, khususnya status gizi, merupakan indikator penting dalam menilai kualitas pembangunan suatu negara. Masalah gizi seperti stunting, wasting, dan kelebihan berat badan memengaruhi jutaan anak di seluruh dunia. Berdasarkan data global, lebih dari 149 juta anak mengalami stunting, yang menunjukkan kekurangan gizi kronis dan dampaknya pada tumbuh kembang anak [1]. Stunting tidak hanya berdampak pada tinggi badan anak tetapi juga berpengaruh terhadap kemampuan kognitif, kesehatan mental, dan produktivitas di masa depan. Masalah ini sering kali disebabkan oleh berbagai faktor demografi seperti usia, jenis kelamin, lokasi geografis, serta faktor antropometri seperti tinggi badan, berat badan, dan lingkaran lengan.

Data antropometri telah lama digunakan sebagai alat penting dalam memantau pertumbuhan anak. Pengukuran fisik tubuh, seperti tinggi badan, berat badan, dan lingkaran lengan, menyediakan informasi esensial untuk menilai status gizi anak secara individual maupun populasi [2]. Indikator-indikator ini memberikan gambaran tentang pola pertumbuhan anak dan membantu mendeteksi masalah gizi sejak dini. Namun, pendekatan analisis konvensional sering kali hanya terbatas pada evaluasi kategori sederhana, seperti pengelompokan anak dalam kategori normal, kurang gizi, atau stunting berdasarkan standar yang telah ditentukan. Pendekatan ini tidak mampu mengeksplorasi

* Corresponding author

pola-pola baru yang tersembunyi dalam data antropometri, sehingga peluang untuk memberikan rekomendasi intervensi yang lebih spesifik sering kali terlewatkan.

Untuk mengatasi keterbatasan pendekatan konvensional, diperlukan metode analisis yang lebih canggih yang dapat mengeksplorasi pola tersembunyi dalam data. Salah satu pendekatan yang menjanjikan adalah machine learning, khususnya metode unsupervised learning seperti klustering. Klustering merupakan salah satu metode dalam data mining untuk mengelompokkan data ke dalam beberapa klaster sehingga data dalam satu klaster memiliki tingkat kemiripan yang maksimum dan data antar klaster memiliki kemiripan yang minimum [3]. Salah satu metode klasterisasi yang paling umum digunakan dalam analisis data adalah algoritma K-Means. Algoritma ini bekerja dengan membagi data menjadi beberapa klaster berdasarkan kemiripan karakteristik. Prosesnya melibatkan penentuan centroid sebagai pusat klaster, pengelompokan data berdasarkan jarak ke centroid, dan pengulangan proses ini hingga hasil klasterisasi mencapai stabilitas [4]. Keunggulan utama K-Means adalah kemampuannya dalam menangani dataset yang besar dan kompleks, sehingga cocok untuk diterapkan pada data antropometri anak. Metode ini memungkinkan pembagian anak ke dalam kelompok yang memiliki karakteristik serupa, seperti kelompok anak dengan risiko stunting tinggi, sehingga dapat membantu pembuat kebijakan dalam merancang intervensi yang lebih efektif.

Penelitian ini berfokus pada penerapan algoritma K-Means untuk mengklasifikasikan status gizi anak menggunakan data antropometri. Terdapat tiga tujuan utama dalam penelitian ini, yaitu: mengidentifikasi pola status gizi anak berdasarkan atribut antropometri seperti tinggi badan, berat badan, dan umur; mengevaluasi sejauh mana teknik klasterisasi K-Means efektif dalam mengelompokkan anak-anak berdasarkan status gizi mereka; serta menyusun rekomendasi intervensi gizi yang dapat dimanfaatkan oleh pembuat kebijakan atau tenaga kesehatan dalam menangani kelompok anak yang berisiko.

Penelitian ini memberikan kontribusi signifikan pada pengembangan ilmu pengetahuan dan aplikasi praktis di bidang kesehatan masyarakat. Dengan memanfaatkan pendekatan berbasis data, penelitian ini menawarkan solusi inovatif untuk memahami dinamika status gizi anak, yang dapat membantu mendukung pencapaian target global dalam pengurangan stunting serta diharapkan dapat memperkuat dasar ilmiah bagi pembuat kebijakan dalam merancang strategi intervensi gizi yang lebih efektif dan efisien. Dengan menggunakan data antropometri yang dianalisis melalui metode klasterisasi, penelitian ini tidak hanya memberikan wawasan baru tetapi juga memperkuat upaya peningkatan kualitas kesehatan anak di tingkat nasional maupun global.

II. TINJAUAN PUSTAKA

Data antropometri, seperti tinggi badan, berat badan, dan umur, merupakan alat penting dalam mengevaluasi status gizi anak. Data ini membantu mengidentifikasi risiko stunting dan kategori status gizi seperti normal, stunted, dan severely stunted. Penelitian sebelumnya menunjukkan bahwa pengukuran antropometri yang akurat dapat secara signifikan meningkatkan pengambilan keputusan kesehatan masyarakat, terutama di daerah seperti Indonesia, di mana prevalensi stunting tetap tinggi pada 24,4% di kalangan balita [5]. Di Indonesia, pengukuran antropometri menjadi standar dalam survei kesehatan seperti SSGI (Kementerian Kesehatan RI, 2022). Studi [6] dan [7] di Indonesia juga menunjukkan bahwa data antropometri dapat digunakan untuk mengevaluasi pertumbuhan anak secara lokal, terutama untuk mendeteksi risiko malnutrisi di wilayah tertentu. Secara keseluruhan, pengumpulan dan analisis sistematis data antropometri sangat penting untuk intervensi kesehatan yang efektif [8] dan perencanaan kebijakan di tingkat nasional dan regional [9].

Metode klustering adalah metode unsupervised learning yang banyak digunakan untuk mengelompokkan data berdasarkan karakteristiknya. Algoritma K-Means, yang berbasis centroid, merupakan salah satu metode yang paling populer karena efisiensinya dalam menangani dataset besar. Dalam konteks kesehatan, K-Means digunakan untuk mengidentifikasi kelompok anak dengan risiko gizi buruk atau malnutrisi.

Penelitian di Indonesia telah secara efektif menggunakan K-Means untuk menganalisis berbagai indikator kesehatan, seperti malnutrisi dan stunting di kalangan balita [10]. Hasil klustering membantu mengelompokkan anak-anak berdasarkan kategori gizi, sehingga dapat memberikan prioritas pada kelompok risiko tinggi. Selain itu, K-Means telah diterapkan pada data pasien diabetes, mengungkapkan profil faktor risiko yang berbeda di antara pasien [11]. Efektivitas algoritma dalam mengidentifikasi kasus stunting juga disorot, dengan prevalensi signifikan 37,2% dilaporkan di Indonesia [12]. Secara keseluruhan, K-Means berfungsi sebagai alat penting untuk memprioritaskan intervensi pada kelompok berisiko tinggi, sehingga meningkatkan strategi kesehatan masyarakat.

Pemilihan jumlah klaster (k) yang optimal merupakan aspek penting dalam analisis klustering. Metode Elbow dan Silhouette score adalah dua pendekatan evaluasi yang sering digunakan. Metode Elbow mengevaluasi total dalam-klaster dengan *Within-Cluster Sum of Squares* (WCSS) metrik yang digunakan dalam analisis klustering, khususnya dalam algoritma K-Means, untuk mengukur seberapa baik data dikelompokkan dalam klaster., sedangkan Silhouette Score mengukur pemisahan antar klaster [13]. Di Indonesia, penelitian Amalia dan Arianto menunjukkan bahwa evaluasi dengan metode ini memberikan hasil yang signifikan dalam mengoptimalkan klustering data balita [14]. Metode Elbow dapat diterapkan untuk menentukan jumlah kluster optimal dalam analisis data kesehatan, dengan hasil yang mendukung penggunaan kombinasi kedua metode untuk meningkatkan

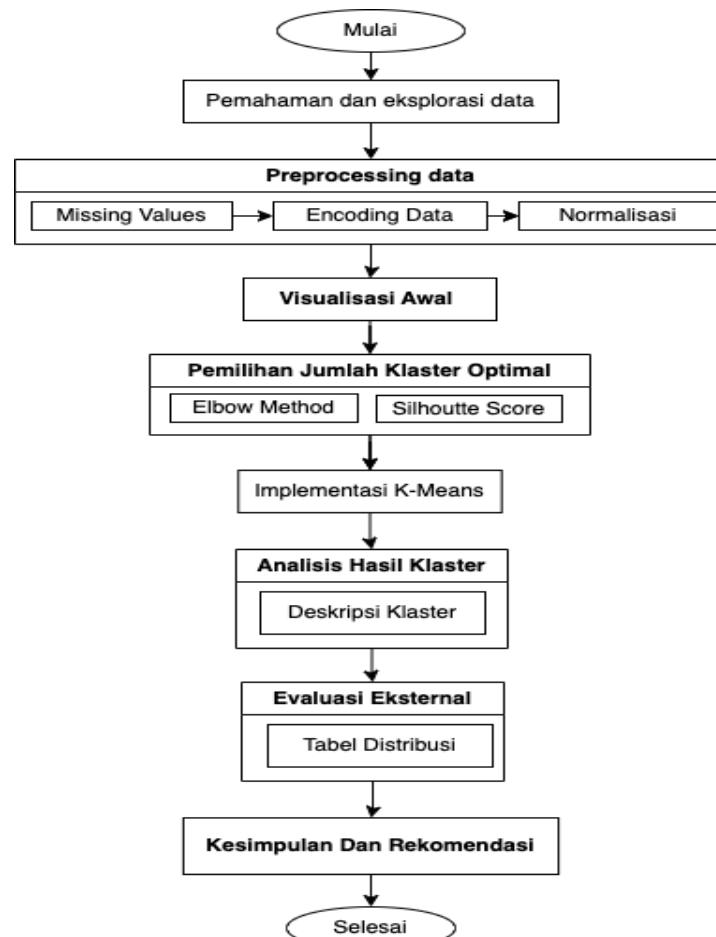
akurasi analisis klustering dalam dataset kesehatan yang kompleks [15]. Secara keseluruhan, integrasi teknik evaluasi ini telah menjadi praktik standar dalam penelitian pengelompokan, memastikan hasil analisis data yang lebih akurat dan bermakna [16].

Studi yang dilakukan oleh Pratama menunjukkan bahwa data antropometri seperti tinggi badan dan berat badan dapat digunakan untuk mendeteksi risiko gizi buruk pada anak-anak [17]. Di Indonesia menemukan bahwa klustering berbasis data antropometri tidak hanya membantu mengidentifikasi kelompok risiko tinggi, tetapi juga memberikan wawasan tentang distribusi gizi di wilayah tertentu. Penggunaan data ini dalam program kesehatan masyarakat, seperti yang dicontohkan oleh Kementerian Kesehatan RI [18], menunjukkan bahwa data antropometri dapat digunakan untuk perencanaan intervensi yang lebih terarah. Dengan algoritma seperti K-Means, pola pertumbuhan dapat diidentifikasi dengan lebih efisien.

Berdasarkan hasil penelitian yang disebutkan sebelumnya, penelitian ini menggunakan algoritma K-Means untuk menganalisis data antropometri guna mengelompokkan anak-anak berdasarkan status gizi mereka. Dengan memanfaatkan metode ini, penelitian diharapkan dapat memberikan wawasan baru tentang pola pertumbuhan anak di Indonesia dan mendukung upaya pengurangan stunting.

III. METODE

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksploratif menggunakan metode *unsupervised learning*. Algoritma utama yang digunakan adalah K-Means, yang bertujuan untuk mengelompokkan data anak berdasarkan karakteristik antropometri, seperti tinggi badan, umur, dan jenis kelamin. Berikut ini adalah tahapan penelitian seperti yang tertera pada gambar 1.



Gambar 1 Tahapan Penelitian

Gambar 1 menjelaskan tahapan penelitian yang akan dilaksanakan, mulai dari tahap pemahaman dan eksplorasi data hingga tahap kesimpulan dan rekomendasi.

A. Pemahaman Dan Eksplorasi Dataset

Tahap ini bertujuan untuk memahami struktur dataset, mengidentifikasi variabel yang relevan, serta mendeteksi nilai kosong (*missing values*) atau pencilan (*outliers*). Eksplorasi data adalah langkah awal yang penting untuk memastikan data siap digunakan dalam analisis lanjutan. Dataset yang digunakan merupakan

dataset sekunder yang diperoleh dari website <https://www.kaggle.com/code/hendrimardani/stunting> yang terdiri dari 120.999 entri dengan variabel: umur (bulan), jenis kelamin, tinggi badan (cm), dan status gizi. Untuk informasi awal dataset dilihat pada tabel 1.

TABEL 1
INFORMASI DATASET

No	Atribut		
	Nama	Jenis	Deskripsi
1	Umur	Numerik	usia anak dalam satuan bulan, Rentang Nilai: Dari 0 bulan hingga 60 bulan
2	Jenis Kelamin	Categorical	jenis kelamin anak berisi perempuan atau laki-laki
3	Tinggi Badan	Numerik	tinggi badan anak dalam satuan centimeter
4	Status Gizi	Categorical	status gizi berdasarkan tinggi badan terhadap usia (<i>Height-for-Age z-score</i>) menggunakan klasifikasi standar dengan nilai Severely Stunted, Stunted, Normal, Tinggi

B. Preprocessing data

Preprocessing dilakukan untuk meningkatkan kualitas dataset sehingga data siap untuk digunakan [19]. Tahapan preprocessing meliputi:

1. Penanganan Nilai Kosong menggunakan metode imputasi, seperti nilai rata-rata atau modus, untuk menjaga konsistensi data
2. Encoding Variabel Kategorikal. Variabel seperti jenis kelamin dan status gizi dikonversi menjadi numerik menggunakan metode Label Encoding agar dapat diproses oleh algoritma klustering
3. Normalisasi Data berupa data numerik seperti umur dan tinggi badan dinormalisasi menggunakan z-score standardization untuk memastikan skala yang seragam, sehingga algoritma klustering dapat bekerja optimal.

C. Visualisasi Data

Pada tahapan ini digunakan untuk memahami distribusi data meliputi tinggi badan, berat badan, dan umur. Scatterplot merupakan fungsi untuk menggambarkan sebaran data dalam bentuk koordinat kartesian sehingga dapat menunjukkan kedekatan antara data yang merupakan hubungan karakteristik yang serupa. Grafik ini juga dapat memberikan pengetahuan mengenai arah pergerakan data menurut kluster yang ditetapkan. Titik yang dipetakan merupakan titik X dan Y yang dikomposisikan dari atribut yang didefinisikan pada dataset. Menunjukkan jangkauan data yang jelas serta gambaran tentang titik maksimum dan titik minimum dapat diketahui dengan jelas. Selain itu dapat menunjukkan hubungan positif maupun negative antara data.

D. Pemilihan Jumlah Kluster optimal

Untuk menentukan jumlah kluster terbaik, penelitian ini menggunakan metode Elbow dan Silhouette Score. Metode ini membantu dalam menentukan jumlah kluster optimal untuk klustering *k-means*. Metode Elbow digunakan untuk mengukur *inertia* untuk berbagai jumlah kluster, memanfaatkan nilai WCSS (*Within-Cluster Sum of Squares*), yaitu jumlah total kuadrat jarak setiap titik data ke centroid kluster yang seperti ditunjukkan pada persamaan 1[13].

$$WCSS = \sum_{i=1}^k x^i - y^i \quad (1)$$

dimana,

- k : jumlah kluster
- x^i : titik data ke-i
- y^i : centroid kluster ke i

Silhouette Score digunakan mengukur seberapa baik kluster terpisah satu sama lain. Nilai Silhouette Score berkisar dari -1 hingga 1, dengan nilai lebih tinggi menunjukkan kluster yang lebih baik. Perhitungannya seperti ditunjukkan pada persamaan 2 berikut ini :

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

dimana,

- $S(i)$: Silhouette Score untuk titik data -i
- $a(i)$: Jarak rata-rata antara titik data -i dan semua titik dalam klasternya sendiri.
- $b(i)$: Jarak rata-rata antara titik data -i dan semua titik dalam kluster terdekat (kluster berbeda yang paling dekat).

E. Implementasi Algoritma K-Means

K-Means adalah algoritma unsupervised learning yang digunakan untuk melakukan klustering atau pengelompokan data ke dalam beberapa kelompok (clusters). Tujuan utama K-Means adalah untuk meminimalkan jarak antara data dalam satu cluster dengan centroidnya (pusat cluster) [20]. Tahapan dalam Algoritma K-Means :

1. Tentukan jumlah cluster centroid secara acak dari data.
2. Setiap data dialokasikan ke cluster berdasarkan jarak Euclidean terdekat antara data dan centroid cluster.
3. Hitung ulang centroid setiap cluster sebagai rata-rata dari semua data yang berada dalam cluster tersebut.
4. Ulangi langkah 2 dan 3 hingga centroid tidak berubah lagi atau mencapai iterasi maksimum.

Fungsi objektif yang diminimalkan dalam K-Means seperti pada persamaan 3 berikut ini:

$$J = \sum_{i=1}^k \sum_{j=1}^n ||x_j^{(i)} - \mu_i||^2 \quad (3)$$

dimana,

- J : fungsi objektif yang ingin diminimalkan
 k : jumlah klaster
 n : jumlah data
 $x_j^{(i)}$: data ke-j yang termasuk dalam klaster ke-i
 μ_i : centroid dari cluster ke - i
 $||x_j^{(i)} - \mu_i||^2$: jarak Euclidean antara data $x_j^{(i)}$ dan centroid μ_i

F. Analisis Hasil Klustering

Deskripsi Kluster pada bagian ini menggunakan analisis statistik deskriptif terhadap setiap klaster untuk menggambarkan karakteristik data dalam klaster yang terbentuk. Mengidentifikasi pola atau ciri khas dari setiap klaster berdasarkan atribut yang dianalisis serta memberikan pemahaman numerik mengenai klaster, seperti rata-rata, median, standar deviasi, dan rentang nilai untuk atribut utama.

Visualisasi hasil klustering menggunakan scatter plot dilakukan untuk mempermudah interpretasi pola data. Kegunaannya yaitu untuk menampilkan distribusi data dalam dimensi yang relevan, seperti hubungan antara umur, tinggi badan, dan klaster serta mengidentifikasi kemungkinan overlap antar klaster atau pemisahan yang jelas di antara kelompok data.

G. Evaluasi Eksternal

Digunakan untuk menginterpretasikan dan memvalidasi hasil klustering dengan membandingkan pola klaster terhadap atribut yang relevan di dataset, seperti distribusi status gizi. Evaluasi eksternal tidak hanya bagian dari hasil, tetapi juga berfungsi sebagai alat validasi dan interpretasi. Validasi eksternal ini memastikan klaster yang terbentuk relevan terhadap informasi atau kategori lain di dataset

1. Tabel Distribusi status gizi tiap klaster.
2. Heatmap untuk membantu memvisualisasikan proporsi status gizi pada masing-masing klaster [21].

H. Kesimpulan dan rekomendasi

Hasil akhir dari proses analisis, menghubungkan hasil dengan implikasi praktis atau kebijakan.

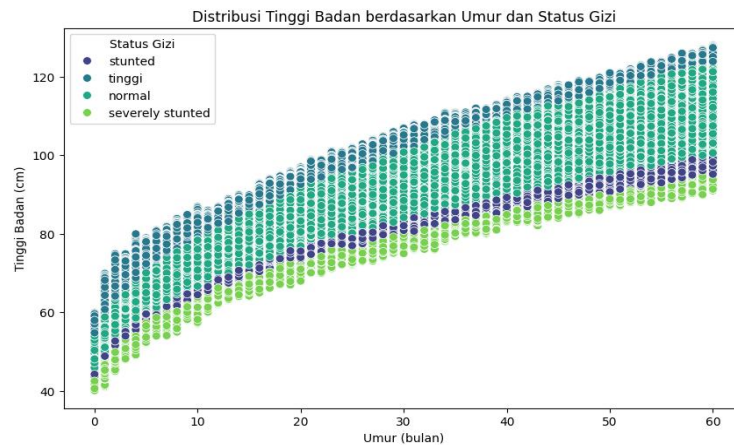
IV. HASIL

A. Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini mencakup 120.000 data anak, dengan variabel utama berupa umur (bulan), tinggi badan (cm), jenis kelamin, dan kategori status gizi. Sebelum proses klustering, dilakukan eksplorasi awal untuk memahami distribusi data. Rata-rata tinggi badan anak adalah 85 cm, dengan standar deviasi 15 cm, sedangkan rata-rata umur anak adalah 25 bulan dengan standar deviasi 10 bulan.

Dari data kategori status gizi, 40% anak dikategorikan normal, 30% stunted, 20% severely stunted, dan 10% tinggi. Distribusi jenis kelamin hampir seimbang, dengan 51% laki-laki dan 49% perempuan

Gambar 2 merupakan visualisasi awal dataset menggunakan scatter plot atau histogram. Terlihat pada gambar Umur (bulan) sebagai sumbu X dan Tinggi Badan (cm) sebagai sumbu Y. Pewarnaan didasarkan pada kolom Status Gizi, sehingga setiap kategori status gizi akan memiliki warna berbeda.



Gambar 2 Hasil visualisasi awal dataset

B. Hasil Preprocessing data

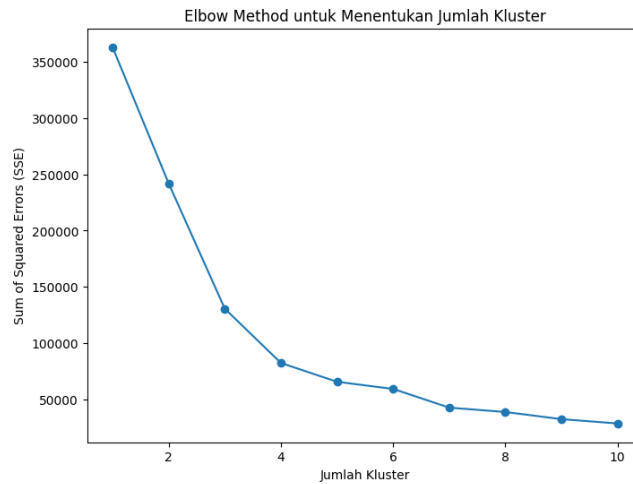
Tahapan preprocessing menghasilkan dataset yang lebih bersih dan konsisten. Variabel numerik seperti umur dan tinggi badan dinormalisasi menggunakan metode z-score, dengan rata-rata 0 dan standar deviasi 1. Variabel jenis kelamin dan status gizi dikonversi menjadi nilai numerik untuk mempermudah analisis algoritma K-Means. Hasil dari preprocessing data dapat dilihat pada tabel 2

TABEL 2
DATA HASIL NORMALISASI

No	Umur (bulan)	Tinggi Badan (cm)	Jenis Kelamin	Status Gizi
0	1.716.855	-2.546.885	1.008.341	stunted
1	1.716.855	-1.846.735	1.008.341	tinggi
2	1.716.855	-2.415.598	1.008.341	normal
3	1.716.855	-2.378.336	1.008.341	normal
4	1.716.855	-2.653.727	1.008.341	severely stunted
...
120994	1.697.076	0.690400	0.991728	normal
120995	1.697.076	0.557460	0.991728	stunted
120996	1.697.076	1.886.868	0.991728	normal
120997	1.697.076	1.360.885	0.991728	normal
120998	1.697.076	1.222.164	0.991728	normal

C. Pemilihan Jumlah Kluster Optimal

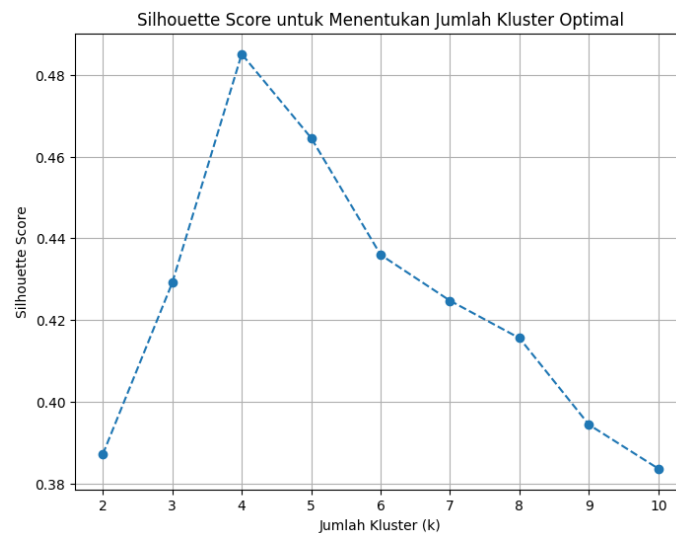
Untuk menentukan jumlah kluster terbaik, penelitian ini menggunakan metode Elbow dan Silhouette Score. Metode ini membantu dalam menentukan jumlah kluster optimal untuk klustering k-means. Gambar 3 merupakan hasil perhitungan menggunakan metode elbow. Pada gambar, nilai WCSS menurun tajam saat jumlah kluster bertambah dari k = 1 hingga k = 4, penurunan WCSS mulai melambat, membentuk pola yang menyerupai "siku" atau elbow point. Penurunan WCSS setelah k = 4 menjadi lebih kecil, menunjukkan bahwa menambahkan lebih banyak kluster tidak memberikan manfaat signifikan dalam mengurangi variabilitas dalam kluster. dari hasil penggunaan metode elbow diperoleh K Optimal (k = 4)



Gambar 3 Hasil Perhitungan Metode Elbow

Gambar 4 adalah hasil dari perhitungan silhouette score, terlihat puncak tertinggi silhouette score mencapai nilai maksimum sekitar 0.48 ketika jumlah kluster (k) adalah 4. Selanjutnya terjadi penurunan setelah k=4. Berdasarkan metode elbow dan silhouette score, penelitian ini akan menggunakan k = 4.

Jumlah Kluster	Silhouette Score
2	0,3871
3	0,4291
4	0,4851
5	0,4646
6	0,4359
7	0,4247
8	0,4155
9	0,3944
10	0,3836



Gambar 4 Hasil Perhitungan Silhouette Score

D. Analisis Hasil Klustering dengan K-Means

Proses klustering menggunakan algoritma K-Means menghasilkan pembagian data ke dalam empat kluster yang ditentukan berdasarkan hasil evaluasi menggunakan Elbow Method. Hasil klustering dianalisis untuk memahami karakteristik setiap kluster berdasarkan atribut utama, yaitu umur, tinggi badan, dan jenis kelamin. Hasil dari klustering dilihat pada Tabel 3.

TABEL 3
HASIL KLASSTERING

Umur (bulan)	Jenis Kelamin	Tinggi Badan (cm)	Status Gizi	Cluster
0	0	44.591.973	stunted	0
0	0	56.705.203	tinggi	0
0	0	46.863.358	normal	0
0	0	47.508.026	normal	0
0	0	42.743.494	severely stunted	0
...
60	1	98.03.00	stunted	1
60	1	121.03.00	normal	1
60	1	112.02.00	normal	1
60	1	109.08.00	normal	1

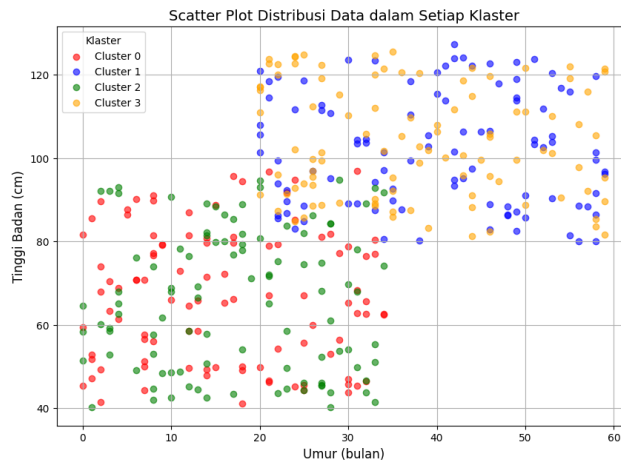
Hasil statistik deskriptif untuk setiap kluster ditampilkan dalam Tabel 4, yang mencakup rata-rata, median dari atribut umur dan tinggi badan.

TABEL 4
 HASIL STATISTIK DESKRIPTIF

Kluster	Umur (bulan) Mean	Umur (bulan) Median	Tinggi Badan (cm) Mean	Tinggi Badan (cm) Median	Proporsi Jenis Kelamin
0	13,96	14	74,34	76,04	100% Laki-laki
1	43,57	44	100,74	99,07	100% Perempuan
2	13,85	13	73,13	74,07	100% Perempuan
3	43,55	44	101,14	100,01	100% Laki-laki

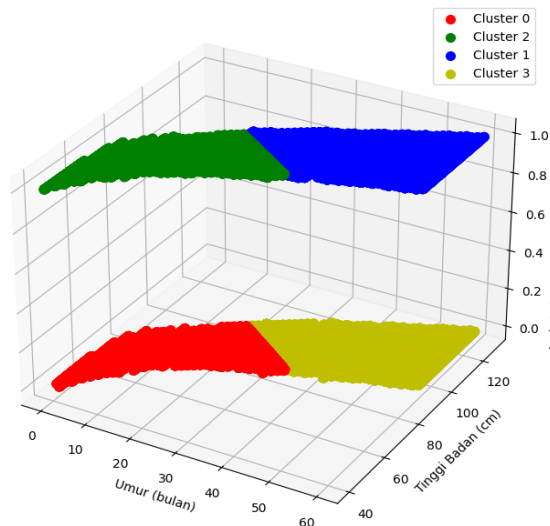
Berdasarkan tabel 4 setiap kluster memiliki karakteristik yang berbeda. Kluster 0 merupakan kelompok anak laki-laki dengan pertumbuhan lebih rendah daripada kluster yang lain. Kluster 1 menunjukkan kelompok anak perempuan dengan pertumbuhan optimal. Kluster 2 mencerminkan kelompok anak perempuan dengan risiko pertumbuhan terhambat dan kluster 3 mencerminkan kelompok anak laki-laki dengan pertumbuhan optimal.

Untuk memvisualisasikan hubungan antara umur, tinggi badan, dan kluster yang terbentuk digunakan scatter plot, terlihat pada Gambar 5 dan juga hasil visualisasi 3D pada Gambar 6.



Gambar 5 Scatter Plot Data Setiap Kluster

Visualisasi 3D Clustering (Umur, Tinggi Badan, Jenis Kelamin)



Gambar 6 Visualisasi Klustering

E. Evaluasi Eksternal

Dilakukan untuk memvalidasi hasil klustering yang telah terbentuk dengan atribut yang tidak digunakan dalam proses klustering, yaitu status gizi. Tujuan utama dari evaluasi ini untuk memastikan bahwa kluster yang terbentuk merepresentasikan pola nyata dalam kategori status gizi anak-anak, seperti normal, stunted, severely

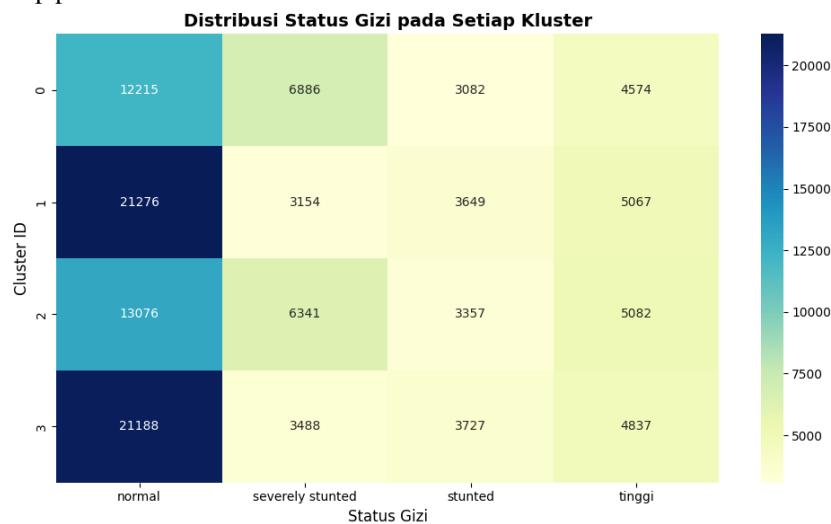
stunted, dan tinggi. Hasil evaluasi eksternal menunjukkan distribusi status gizi di setiap kluster seperti yang ditampilkan pada Tabel 5.

TABEL 5
 DISTRIBUSI STATUS GIZI DALAM SETIAP KLUSTER

Kluster	Normal (%)	Severely Stunted (%)	Stunted (%)	Tinggi (%)
0	43,60%	24,60%	11,00%	16,30%
1	50,20%	7,40%	8,60%	12,00%
2	42,10%	20,40%	11,40%	16,40%
3	49,20%	8,10%	8,70%	11,40%

Kluster 0 dan Kluster 2 memiliki proporsi signifikan anak-anak dalam kategori stunted dan severely stunted, yang menunjukkan kelompok dengan risiko tinggi stunting. Dan Kluster 1 dan Kluster 3 didominasi oleh anak-anak dalam kategori normal dan tinggi, mencerminkan kelompok dengan pertumbuhan optimal.

Untuk memberikan gambaran visual yang lebih jelas, distribusi status gizi dalam setiap kluster divisualisasikan menggunakan heatmap pada Gambar 7.



Gambar 7 Heatmap Distribusi Status Gizi

V. PEMBAHASAN

A. Analisis Kluster Berdasarkan Umur, Tinggi Badan, dan Jenis Kelamin

Kluster yang dihasilkan oleh algoritma K-Means berhasil membagi dataset menjadi empat kelompok berdasarkan pola pertumbuhan anak, dengan mempertimbangkan variabel umur, tinggi badan, dan jenis kelamin. Hasil klustering menunjukkan bahwa terdapat perbedaan signifikan dalam distribusi atribut di antara kluster. Kluster 0 dan Kluster 2, yang didominasi oleh anak-anak usia muda (0–35 bulan), memiliki rata-rata tinggi badan yang lebih rendah dibandingkan kluster lain. Sebaliknya, Kluster 1 dan Kluster 3, yang mencakup anak-anak usia lebih tua (20–60 bulan), menunjukkan pola pertumbuhan yang lebih optimal.

B. Identifikasi Risiko Stunting

Analisis statistik deskriptif dan visualisasi mengindikasikan bahwa kluster 0 didominasi oleh anak laki-laki usia muda dengan rata-rata tinggi badan 74.34 cm. Kluster ini mencerminkan kelompok dengan risiko stunting yang tinggi. Kluster 2 didominasi oleh anak perempuan usia muda dengan rata-rata tinggi badan 73.13 cm, lebih rendah dibandingkan Kluster 0. Anak-anak dalam kluster ini menunjukkan risiko severe stunting yang lebih tinggi. Pola ini konsisten dengan penelitian sebelumnya, yang menunjukkan bahwa anak-anak usia dini (terutama perempuan) lebih rentan terhadap masalah gizi seperti stunting karena fase pertumbuhan yang kritis pada usia ini.

C. Hubungan Tinggi Badan, Umur, dan Status Gizi

Distribusi data menunjukkan bahwa tinggi badan anak-anak sangat berhubungan dengan status gizi mereka. Anak-anak dengan tinggi badan <80 cm sebagian besar berada dalam kategori stunted atau severely stunted. Anak-anak dalam kategori normal atau tinggi memiliki tinggi badan rata-rata >100 cm dan tersebar di Kluster 1 dan Kluster 3. Selain itu, hubungan antara umur dan status gizi juga terlihat jelas. Anak-anak usia tua dalam Kluster 1 dan Kluster 3, yang memiliki tinggi badan rata-rata tertinggi (100.74 cm dan 101.14 cm), menunjukkan pola pertumbuhan yang lebih sehat.

D. Pengaruh Jenis Kelamin pada Pertumbuhan

Jenis kelamin memainkan peran penting dalam distribusi klaster. Klaster 0 dan Klaster 3 hanya mencakup anak laki-laki, sementara Klaster 1 dan Klaster 2 didominasi oleh perempuan. Hal ini mengindikasikan bahwa perbedaan jenis kelamin mungkin memengaruhi pola pertumbuhan, terutama dalam kelompok risiko stunting. Anak perempuan dalam Klaster 2 memiliki risiko severe stunting yang lebih tinggi dibandingkan anak laki-laki dalam Klaster 0. Perbedaan ini dapat dikaitkan dengan faktor biologis seperti kebutuhan nutrisi yang berbeda atau akses sumber daya yang tidak merata di komunitas tertentu.

E. Rekomendasi Berdasarkan Temuan

Anak-anak dalam Klaster 0 dan Klaster 2 memerlukan perhatian khusus untuk mencegah dampak jangka panjang dari stunting. Program nutrisi tambahan dan pendidikan kesehatan bagi orang tua harus menjadi prioritas. Berikan intervensi nutrisi (asupan protein, zat besi, dan vitamin) serta selalu pantau lebih intensif kondisi kesehatan anak-anak di klaster ini.

Anak-anak dalam Klaster 1 dan Klaster 3 perlu dipantau secara berkala untuk memastikan pertumbuhan yang berkelanjutan. Anak-anak dalam klaster ini mungkin tidak membutuhkan intervensi mendesak, tetapi mereka tetap perlu dimonitor untuk mencegah risiko overweight atau gangguan pertumbuhan lain.

VI. KESIMPULAN

Untuk menganalisis pola pertumbuhan anak berdasarkan atribut umur, tinggi badan, dan jenis kelamin, serta mengidentifikasi kelompok risiko stunting melalui metode klustering algoritma K-Means. Data antropometri dikelompokkan menjadi empat klaster. Klaster 0 dan Klaster 2 didominasi oleh anak-anak usia muda (0–35 bulan) dengan tinggi badan rata-rata yang lebih rendah, sedangkan Klaster 1 dan Klaster 3 mencakup anak-anak usia tua (20–60 bulan) dengan tinggi badan rata-rata yang lebih tinggi.

Klaster 0 dan Klaster 2 menunjukkan risiko stunting tertinggi dengan Klaster 0 merupakan Anak laki-laki usia muda dengan rata-rata tinggi badan 74.34 cm. Klaster 2 merupakan anak perempuan usia muda dengan rata-rata tinggi badan lebih rendah, yaitu 73.13 cm, menunjukkan risiko severe stunting yang lebih tinggi dibandingkan klaster lain. Anak-anak dalam kedua klaster ini memerlukan perhatian khusus melalui intervensi nutrisi dan kesehatan.

Klaster 1 dan Klaster 3 didominasi oleh anak-anak dengan kategori status gizi Normal dan Tinggi, menunjukkan pertumbuhan optimal. Anak dalam klaster ini memiliki rata-rata tinggi badan masing-masing 100.74 cm dan mencerminkan kondisi pertumbuhan yang sehat. Klaster menunjukkan perbedaan pola pertumbuhan antara anak laki-laki dan perempuan. Anak perempuan dalam Klaster 2 memiliki risiko lebih tinggi untuk stunting dibandingkan anak laki-laki dalam Klaster 0.

DAFTAR PUSTAKA

- [1] W. Lestari, I. Samidah, and F. Diniarti, 'Determinant Analysis of Stunting Incidence in Toddlers in Lubuklinggau City', *Jurnal EduHealth*, vol. 14, no. 3, pp. 1141–1153, Aug. 2023, doi: 10.54209/jurnaleduhealth.v14i3.2509.
- [2] R. M. Kusuma and R. A. Hasanah, 'Antropometri Pengukuran Status Gizi Anak Usia 24-60 Bulan Di Kelurahan Bener Kota Yogyakarta Reni Merta Kusuma, Rizki Awalunisa Hasanah', 2018.
- [3] E. Virantika, K. Kusnawi, and J. Ipmawati, 'Evaluasi Hasil Pengujian Tingkat Clusterisasi Penerapan Metode K-Means Dalam Menentukan Tingkat Penyebaran Covid-19 di Indonesia', *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 3, p. 1657, Jul. 2022, doi: 10.30865/mib.v6i3.4325.
- [4] E. Nurjannah, M. Nasution, and R. Muti'ah, 'Data Mining Clustering Analysis of Child Growth and Development Using the K-Means Method', *sinkron*, vol. 8, no. 3, pp. 1909–1919, Jul. 2024, doi: 10.33395/sinkron.v8i3.13817.
- [5] A. Rosari *et al.*, 'Developing a Nutritional Assessment Tool for Toddlers Using Anthropometry and IoT Technology', *International Journal of Advanced Health Science and Technology*, vol. 4, no. 2, pp. 67–71, Apr. 2024, doi: 10.35882/ijahst.v4i2.319.
- [6] A. Ali, 'Clustering Data Antropometri Balita Untuk Menentukan Status Gizi Balita Di Kelurahan Jumput Rejo Sukodono Sidoarjo', *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 7, no. 3, pp. 395–407, Dec. 2020, doi: 10.35957/jatisi.v7i3.530.
- [7] S. S. Nagari and L. Inayati, 'Implementation Of Clustering Using K-Means Method To Determine Nutritional Status', *Jurnal Biometrika dan Kependudukan*, vol. 9, no. 1, p. 62, Jun. 2020, doi: 10.20473/jbk.v9i1.2020.62-68.
- [8] M. Razzaq, S. Masood, S. Bakhsh, F. A. Raza, N. Azam, and R. Asif, 'Assessment of Nutritional Status Using Composite Index of Anthropometric Failure Among Children Aged < 5 Years and its Correlates',

- Pakistan Journal of Public Health*, vol. 14, no. Special.NI, pp. 139–143, Aug. 2024, doi: 10.32413/pjph.v14iSpecial.ni.1323.
- [9] Agus Widyawati and Titih Huriah, ‘Comparison of the Effectiveness of Portable and Manual Digital Anthropometrics in Early Detection of Stunting in Children Literature Review’, *International Journal For Multidisciplinary Research*, vol. 6, no. 4, Aug. 2024, doi: 10.36948/ijfmr.2024.v06i04.26986.
- [10] M. Ratuanik *et al.*, ‘Classification of villages in Tanimbar Islands based on stunting service packages using the K-Means Algorithm’, 2024.
- [11] A. Simanjuntak and M. S. Hasibuan, ‘Application of PCA and K-Means Clustering Methods to Identify Diabetes Mellitus Patient Groups Based on Risk Factors’, *Prisma Sains : Jurnal Pengkajian Ilmu dan Pembelajaran Matematika dan IPA IKIP Mataram*, vol. 11, no. 4, p. 1002, Oct. 2023, doi: 10.33394/j-ps.v11i4.9263.
- [12] I. P. Sari, Al-Khowarizmi, O. K. Sulaiman, and D. Apdilah, ‘Implementation of Data Classification Using K-Means Algorithm in Clustering Stunting Cases’, *Journal of Computer Science, Information Technology and Telecommunication Engineering*, vol. 4, no. 2, Sep. 2023, doi: 10.30596/jcositte.v4i2.15765.
- [13] D. A. Setiady and H. Leong, ‘Implementation Of K-Means Algorithm Elbow Method And Silhouette Coefficient For Rainfall Classification’, *Proxies : Jurnal Informatika*, vol. 4, no. 1, pp. 18–25, Aug. 2024, doi: 10.24167/proxies.v4i1.12433.
- [14] M. F. Amalia and D. B. Arianto, ‘Implementasi Algoritma K-Means Clustering Dalam Klusterisasi Kabupaten/Kota Provinsi Jawa Barat Berdasarkan Faktor Pemicu Stunting Pada Balita’, *SIMKOM*, vol. 9, no. 1, pp. 36–46, Jan. 2024, doi: 10.51717/simkom.v9i1.356.
- [15] C. A. Sugianto and K. K. Wanaziana, ‘Optimasi Algoritma K-Means Menggunakan Metode Elbow Pada Data Penerima Program Keluarga Harapan (PKH)’, 2024. [Online]. Available: <https://e-journal.unper.ac.id/index.php/informatics>
- [16] V. A. Permadi, S. P. Tahalea, and R. P. Agusdin, ‘K-Means And Elbow Method For Cluster Analysis Of Elementary School Data’, *PROGRES PENDIDIKAN*, vol. 4, no. 1, pp. 50–57, Jan. 2023, doi: 10.29303/prospek.v4i1.328.
- [17] I. S. Pratama, ‘Prevalensi Malnutrisi Berdasarkan Indeks Antropometri Tunggal dan Komposit pada Anak Usia Sekolah di SDN Cibodas’, *Unram Medical Journal*, vol. 7, no. 2, p. 6, Jun. 2018, doi: 10.29303/jku.v7i2.175.
- [18] Kementerian Kesehatan RI, *Buku Saku Hasil Survei Status Gizi Indonesia (SSGI) 2022*. Jakarta: Kementerian Kesehatan RI, 2022.
- [19] K. Kusnawi, J. Ipmawati, B. P. Asadulloh, A. Aminuddin, F. F. Abdulloh, and M. Rahardi, ‘Leveraging Various Feature Selection Methods for Churn Prediction Using Various Machine Learning Algorithms’, *JOIV: International Journal on Informatics Visualization*, vol. 8, no. 2, p. 897, May 2024, doi: 10.62527/joiv.8.2.2453.
- [20] A. Saputra and R. Yusuf, ‘Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM’, *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 4, pp. 1346–1361, Jul. 2024, doi: 10.57152/malcom.v4i4.1516.
- [21] F. L. Dennig *et al.*, ‘The Categorical Data Map: A Multidimensional Scaling-Based Approach’, *Cornell University*, Apr. 2024.