



Kategorisasi Tingkat Gizi Balita Menggunakan *K-Means*: Studi Kasus Puskesmas Desa Karangsembung

Fajar Tri Wahyuni¹, Nahila Shofie Kirana², Natasya Syafila ‘Ashifa³, Putri Ella Nazila⁴

^{1, 2, 3, 4}Sains Data, Institut Teknologi Telkom Purwokerto

¹2211110009@ittelkom-pwt.ac.id, ²2211110023@ittelkom-pwt.ac.id, ³2211110028@ittelkom-pwt.ac.id, ⁴2211110047@ittelkom-pwt.ac.id

Corresponding author email: ¹2211110009@ittelkom-pwt.ac.id

Abstract: The toddler years are a crucial period for children's growth and development, so ensuring they receive adequate nutrition is essential to prevent health issues such as stunting, wasting, and being underweight. It is important to assess the nutritional status of children under five in a specific area. In this case, the dataset on toddlers in Karangsembung Village necessitates being analyzed to gain insights. This study aims to cluster the nutritional status of toddlers in Karangsembung Village based on their weight and height. The research involved 97 toddlers aged 1-36 months, obtained from the medical records of the Karangsembung Community Health Center. The *K-Means* was used and the number of clusters was determined using the Elbow method. The Elbow graph indicated that 2 and 4 clusters are the optimal *k* clusters, with a Silhouette Score of 0.488 for $k=4$ and 0.567 for $k=2$. However, based on the nutritional status of the Community Health Center, 4 categories were used for clustering with $k=4$. As a result, cluster 0 consisted of 23 toddlers, cluster 1 consisted of 35, cluster 2 consisted of 19 toddlers, and cluster 3 consisted of 20 toddlers. The *K-Means* results of the clusters were then compared with the nutritional calculations from the "TemanSiGizi" website, managed by the Cakung Community Health Center. The comparison revealed a discrepancy between the cluster results and the nutrition category, as most toddlers in each cluster had normal nutrition, leading to insignificant differences in the member profiles between the clusters. The conclusion, this clustering method isn't appropriate for this dataset.

Keywords: *K-means*, nutritional status, toddlers, clustering, silhouette score

Abstrak: Usia balita adalah masa emas tumbuh kembang anak, sehingga gizi yang memadai sangat penting untuk mencegah masalah kesehatan seperti *stunting*, *wasting*, dan *underweight*. Klasterisasi status gizi sangat penting untuk kategori tumbuh kembang balita di suatu daerah. Terkait hal ini, data tumbuh kembang di Desa Karangsembung perlu diklasterisasi sehingga memberikan *insight*. Penelitian ini bertujuan mengelompokkan status gizi balita di Desa Karangsembung berdasarkan berat dan tinggi badan. Data penelitian berjumlah 97 balita berusia 1-36 bulan yang diambil dari catatan medis Puskesmas Karangsembung. Metode yang digunakan adalah *K-Means*, dengan penentuan jumlah *cluster* menggunakan metode *Elbow*. Posisi siku grafik *Elbow* menunjukkan 2 dan 4 *cluster* adalah *k* optimal. Validasi *cluster* dengan *Silhouette Score* sebesar 0.488 untuk $k=4$ dan sebesar 0.567 untuk $k=2$. Namun, sesuai status gizi dari Puskesmas yang ada 4 kategori maka untuk pengelompokan menggunakan $k=4$. Hasilnya, *cluster* 0 terdiri 23 balita, *cluster* 1 terdiri 35, *cluster* 2 terdiri 19 balita, dan *cluster* 3 terdiri 20 balita. Kemudian, anggota *cluster* hasil *K-Means* dibandingkan dengan perhitungan kalkulator gizi yang tersedia pada laman "TemanSiGizi" yang dikelola oleh Puskesmas Cakung. Hasil perbandingan menunjukkan adanya ketidaksesuaian hasil *cluster* dengan kategori gizi karena setiap *cluster* mayoritas adalah balita dengan gizi normal, sehingga profil anggota antara *cluster* tidak berbeda signifikan.

Kata kunci: status gizi, balita, clustering, *K-means*, silhouette score

I. PENDAHULUAN

Masa balita, yang juga disebut sebagai masa keemasan, merupakan periode penting dalam pembentukan dan perkembangan manusia. Pada fase ini, balita sangat rentan terhadap gangguan pertumbuhan dan perkembangan yang mungkin terjadi. Pada masa ini, terjadi pembentukan dasar-dasar kemampuan indra, berpikir, berbicara, dan perkembangan intelektual yang intensif. Awal pertumbuhan moral juga terjadi pada masa ini, yang akan berdampak pada perilaku anak di masa depan. Oleh karena itu, pemenuhan gizi yang memadai menjadi sangat penting untuk menjamin balita tumbuh dan berkembang dengan optimal [1].

Status gizi balita merupakan salah satu indikator penting dalam menentukan kualitas kesehatan dan perkembangan anak. Pada masa emas pertumbuhan, kekurangan gizi dapat berdampak negatif pada



perkembangan fisik dan kognitif anak. Kekurangan gizi pada balita dapat menyebabkan berbagai masalah kesehatan, seperti *stunting*, *wasting*, dan *underweight*. Oleh karena itu, pemantauan dan penanganan masalah gizi pada balita harus menjadi prioritas utama bagi fasilitas kesehatan. Upaya ini juga penting untuk memastikan bahwa balita dapat mencapai potensi pertumbuhan dan perkembangan mereka secara optimal. Namun, di Indonesia, banyak faktor yang menyebabkan malnutrisi pada balita. Malnutrisi, atau kurang gizi, memiliki dampak yang sangat luas. Selain meningkatkan angka kesakitan dan kematian, malnutrisi juga mempengaruhi aspek psikososial dan perkembangan intelektual [2]. Kekurangan gizi dapat dialami oleh siapa saja di masyarakat, namun balita adalah kelompok yang paling rentan karena memerlukan nutrisi yang tinggi untuk tumbuh dan berkembang. Kekurangan gizi pada balita menggambarkan status pertumbuhan yang diukur melalui berat badan dan tinggi badan sesuai usia, menggunakan indeks seperti BB/U (berat badan menurut umur), TB/U (tinggi badan menurut umur), dan BB/TB (berat badan menurut tinggi badan) [3].

Antropometri, yang secara harfiah berarti ukuran tubuh, sering digunakan untuk mengukur status gizi anak. Prosedur ini sangat sederhana, karena tidak memerlukan tenaga ahli, dan menghasilkan data yang tepat serta akurat. Selain itu, antropometri dapat mendeteksi atau menggambarkan riwayat gizi di masa lampau, menjadikannya alat yang berguna dalam pemantauan gizi [4]. Berdasarkan standar antropometri yang ditetapkan oleh WHO, terdapat beberapa parameter yang sering digunakan untuk menentukan status gizi anak. Parameter tersebut meliputi berat badan menurut usia (BB/U), panjang badan atau tinggi badan menurut usia (PB/U atau TB/U), berat badan menurut panjang badan atau tinggi badan (BB/PB atau BB/TB), serta indeks massa tubuh menurut usia (IMT/U) [5].

Global Food Security Summit (GFSS) yang diselenggarakan pada tahun 2023 menyoroti pentingnya penanganan gizi anak. Deputi Bidang Koordinasi Pangan dan Agribisnis Kementerian Koordinator Bidang Perekonomian, Dida Gareda, menyampaikan beberapa poin penting. Salah satu komitmen Indonesia adalah mengurangi prevalensi gizi buruk pada anak dari 10,2% pada tahun 2018 menjadi kurang dari 7% pada tahun 2024. Selama dua dekade terakhir, Indonesia telah berhasil menurunkan angka *stunting* pada balita dari lebih dari 40% pada tahun 2000 menjadi di bawah 30% pada tahun 2022. Pencapaian ini menunjukkan kemajuan yang signifikan, namun masih banyak yang perlu dilakukan untuk mencapai target yang lebih baik di masa depan. Berbagai metode telah dikembangkan untuk mengidentifikasi masalah gizi pada balita, mulai dari pendekatan konvensional hingga penggunaan teknologi canggih [6]. Metode konvensional melibatkan pengukuran fisik, seperti tinggi badan, berat badan, dan lingkar kepala. Pengelompokan data berdasarkan metrik-metrik ini sangat penting karena memungkinkan identifikasi status gizi balita dengan lebih akurat, membantu mengklasifikasikan balita ke dalam kategori, seperti gizi baik, kurang gizi, atau gizi buruk. Pengelompokan ini juga membantu dalam melakukan pengawasan terhadap perkembangan balita dari waktu ke waktu dan menilai efektivitas program intervensi gizi. Selain itu, pengelompokan tersebut tidak hanya membantu dalam identifikasi dan pemantauan status gizi balita, tetapi juga dalam perancangan dan pelaksanaan program intervensi yang lebih efektif dan efisien. Hal ini sangat penting untuk mencapai target pengurangan prevalensi gizi buruk dan *stunting* yang telah ditetapkan oleh pemerintah Indonesia. Dengan berkembangnya teknologi, metode *machine learning* seperti *K-Means clustering* dapat digunakan untuk menganalisis data gizi balita dengan lebih efektif. Pengelompokan data menggunakan *K-Means* memungkinkan identifikasi pola-pola tersembunyi yang mungkin tidak terlihat dengan analisis tradisional serta membantu mendeteksi anomali atau *outlier* yang dapat menunjukkan masalah kesehatan serius. Selain itu, pengelompokan ini memfasilitasi pemahaman yang lebih baik tentang distribusi status gizi balita, memungkinkan segmentasi populasi ke dalam kelompok yang lebih



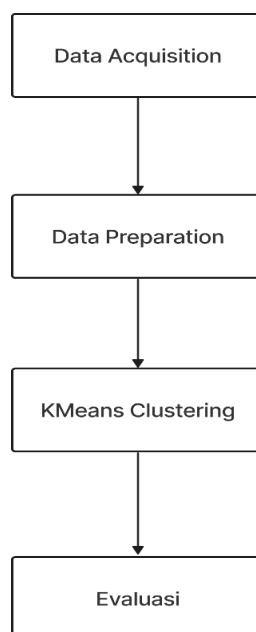
homogen. Hal ini sangat bermanfaat untuk merancang program intervensi gizi yang lebih tepat sasaran dan efisien, serta mengalokasikan sumber daya kesehatan dengan lebih efektif. Dengan menyederhanakan analisis data besar, pengelompokan juga memudahkan visualisasi dan pemahaman data, baik oleh peneliti maupun pembuat kebijakan.

Penelitian terdahulu telah menerapkan metode *K-Means* untuk mengelompokkan status gizi balita. Sebagai contoh, penelitian di Banjar Titih menggunakan variabel umur, jenis kelamin, dan berat badan dalam pengelompokkan status gizi balita [7]. Studi lain juga menggunakan algoritma *K-Means* dengan variabel berat badan, tinggi badan, dan lingkar kepala, menghasilkan lima kelompok yaitu gizi buruk, gizi kurang, gizi baik, gizi lebih, dan obesitas [8]. Selain itu, penelitian di Desa Karang Songo juga menerapkan metode ini dan membandingkan hasil pengelompokkan dengan tabel Growth Chart, dan menemukan bahwa algoritma *K-Means* memiliki tingkat akurasi tertentu dalam menentukan status gizi balita [9]. Selain menggunakan metode *Growth Chart*, status gizi balita juga dapat diukur melalui perhitungan di laman TemanSiGizi yang dikelola oleh Puskesmas Kecamatan Cakung dan dapat diakses di <https://temansigizi.com/>. Dalam penelitian ini, kalkulator gizi digunakan untuk menentukan status gizi balita.

Penelitian ini bertujuan untuk menerapkan metode *K-Means clustering* dalam mengidentifikasi kategori status gizi balita, khususnya di Desa Karangsembung Kabupaten Kebumen. Melalui penerapan metode ini, diharapkan dapat diperoleh hasil yang lebih akurat dan efisien dalam mengelompokkan balita berdasarkan status gizi. Identifikasi yang tepat akan sangat berguna dalam menentukan intervensi gizi yang sesuai untuk masing-masing kelompok balita. Selain itu, hasil dari penelitian ini diharapkan dapat menjadi acuan bagi kebijakan kesehatan dalam upaya meningkatkan status gizi balita. Dengan demikian, diharapkan dapat tercipta generasi yang lebih sehat dan cerdas di masa depan.

II. METODE PENELITIAN

Metodologi penelitian ini terdiri dari beberapa tahapan yang ditunjukkan oleh Gambar 1 berikut.



Gambar 1. Metodologi Penelitian



1. *Data Acquisition*

Data acquisition dimulai dari tahap pengumpulan data yang diperoleh dari lima posyandu di Puskesmas Karangsambung, yaitu Posyandu Melati, Posyandu Kantil, Posyandu Kenanga, Posyandu Ngudi Rahayu, dan Posyandu Mawar. Data yang dikumpulkan mencakup informasi umur, jenis kelamin, berat badan, dan tinggi badan balita. Pengumpulan data dilakukan guna memastikan ketersediaan data yang representatif dan memadai untuk analisis. Instrumen pengukuran standar, seperti timbangan digital dan alat pengukur tinggi badan digunakan untuk memastikan akurasi pengukuran fisik yang diperlukan dalam evaluasi status gizi balita.

2. *Data Preparation*

Setelah pengumpulan data, tahap *data preparation* dimulai dengan *preprocessing*. Langkah awal adalah mengidentifikasi dan menghapus kolom yang tidak relevan, seperti nomor urut balita untuk memfokuskan analisis pada variabel yang lebih signifikan. Data yang digunakan mencakup data balita di rentang usia 0-36 bulan dengan total data 97. Selanjutnya, dilakukan pemeriksaan terhadap keberadaan nilai yang hilang, data yang tidak valid, atau duplikasi data. Pada dataset ini tidak ada nilai yang hilang sehingga tidak perlu penanganan nilai hilang. Langkah selanjutnya normalisasi data yaitu mengubah nilai pada skala 0-1 untuk menyeragamkan rentang nilai data.

3. Pengelompokan dengan *K-Means*

Pada tahap ini, penggunaan algoritma pengelompokan data (*clustering*) menjadi penting, karena merupakan metode yang efektif untuk mengelompokkan data berdasarkan kesamaan tertentu. Salah satu metode *clustering* yang populer adalah *K-Means*, yang terkenal karena kemampuannya untuk mengklasifikasikan data dalam jumlah besar dengan cepat dan efisien. Dalam menentukan hasil *clustering*, *K-Means* sangat bergantung pada penentuan awal pusat *cluster*. Oleh karena itu, pemilihan pusat *cluster* yang tepat pada tahap awal sangat krusial untuk memastikan hasil yang akurat dan representatif. Pada penelitian ini, metode yang digunakan adalah metode *clustering* menggunakan algoritma *K-Means clustering* dan menentukan nilai optimasi menggunakan metode *Elbow*. Metode *Elbow* digunakan untuk menentukan jumlah *k-cluster* terbaik pada proses *clustering*. Pengujian ini bertujuan untuk menentukan jumlah *cluster* yang paling optimal dalam mengelompokkan balita berdasarkan kesamaan karakteristik mereka, seperti berat badan dan tinggi badan. Setiap iterasi dilakukan hingga konvergensi tercapai, di mana posisi *centroid cluster* stabil [10]. Dalam penelitian ini, metode *K-Means* diterapkan untuk mengelompokkan data balita ke dalam empat *cluster*, dengan tujuan menentukan jumlah *cluster* yang sesuai dengan kategori gizi [11]. Analisis deskriptif dilakukan untuk memahami distribusi dan karakteristik utama dari dataset sebelum melanjutkan ke tahap analisis yang lebih mendalam [12]. Untuk mengukur status gizi anak pada usia 0-36 bulan dapat digunakan panduan grafik WHO 2006. Penggunaan grafik ini dibedakan berdasarkan jenis kelamin anak yang dapat menggunakan indikator Berat Badan menurut Umur (BB/U), Tinggi Badan menurut Umur (TB/U), atau Berat Badan menurut Tinggi Badan (BB/TB). Status gizi anak dikategorikan sebagai “Sangat Kurang” apabila nilai standar deviasi BB/U kurang dari -3. Status gizi “Kurang” jika nilai standar deviasi berada antara -3 hingga kurang dari -2. Kategori status gizi “Normal” ditentukan oleh nilai standar deviasi antara -2 hingga +2. Sedangkan status gizi “Lebih” dikategorikan apabila nilai standar deviasi lebih dari +2 [13]. Perhitungan ini dapat dilakukan melalui perhitungan di laman TemanSiGizi yang dikelola oleh Puskesmas Kecamatan Cakung dan dapat diakses di <https://temansigizi.com/>.



4. Evaluasi

Pada tahap hasil evaluasi, metode *clustering* yang telah diterapkan, yaitu *K-Means*, dievaluasi untuk memastikan akurasi dan keandalan pengelompokan data balita berdasarkan karakteristik mereka. Evaluasi dilakukan menggunakan *Silhouette Score* yang merupakan metode kalkulasi statistik untuk memperkirakan jumlah *cluster* optimal dalam sebuah dataset [14]. Nilai *Silhouette Score* yang tinggi menunjukkan bahwa *clustering* lebih baik dan lebih kohesif, sementara nilai yang rendah menunjukkan bahwa terdapat tumpang tindih antar kelompok [15]. Sebuah metrik yang mengukur seberapa mirip suatu data dengan *cluster* yang menjadi anggotanya dibandingkan dengan *cluster* lainnya, dengan nilai berkisar antara -1 hingga 1, di mana nilai lebih tinggi menunjukkan kualitas pengelompokan yang lebih baik. Dalam penelitian ini, berbagai nilai *k* hasil grafik metode *Elbow*. Cara kerja *Elbow* melakukan klusterisasi nilai *Within-Cluster Sum of Square* (WCSS) untuk setiap *k* dan mencari titik "*Elbow/ siku*".

III. HASIL DAN PEMBAHASAN

Pembahasan hasil penelitian ini dijelaskan dalam beberapa poin penting mulai dari *preprocessing* sampai hasil *clustering*. Penjelasan berikut meliputi hasil normalisasi, hasil inisialisasi *cluster* menggunakan *Elbow*, dan hasil *clustering*.

3.1. Tahap Normalisasi

Tabel 1. Dataset Awal dan Hasil Normalisasi

Balita	Berat Badan (BB)	Normalisasi BB	Tinggi Badan (TB)	Normalisasi TB	Usia	Normalisasi Usia	JK
Balita 1	11.30	0.60	97.00	1.00	32	0.88	0
Balita 2	15.80	1.00	95.00	0.95	31	0.85	0
Balita 3	10.50	0.52	89.00	0.80	31	0.85	0
Balita 4	13.30	0.77	94.00	0.92	30	0.82	0
Balita 5	12.30	0.68	84.00	0.70	23	0.62	0
...
Balita 93	10.00	0.48	74.50	0.45	20	0.54	1
Balita 94	9.00	0.39	77.00	0.51	26	0.71	0
Balita 95	5.60	0.09	63.00	0.17	16	0.42	1
Balita 96	10.10	0.49	90.00	0.82	29	0.80	0
Balita 97	6.50	0.17	61.00	0.12	16	0.42	1

Pada Tabel 1 merupakan hasil proses normalisasi angka pada variabel berat badan, tinggi badan, dan Usia(bulan) menghasilkan nilai yang berada pada rentang 0 sampai dengan 1 menggunakan sebuah *function* pustaka *sklearn.preprocessing* yaitu *MinMaxScaler()*. Variabel JK memiliki nilai direntang 0 dan 1 sehingga tidak dilakukan normalisasi. Nilai 0 pada JK mewakili balita berjenis kelamin laki-laki, sedangkan nilai 1 mewakili balita berjenis kelamin perempuan. Normalisasi *MinMax* adalah teknik untuk mengubah nilai data menjadi rentang antara 0 dan 1 serta memastikan setiap fitur memiliki bobot yang sama dalam analisis [16] menggunakan persamaan (1) berikut:

$$\text{Nilai Normalisasi} = \frac{(\text{Nilai Asli} - \text{Nilai Minimum})}{(\text{Nilai Maksimum} - \text{Nilai Minimum})} \quad (1)$$



Berdasarkan persamaan (1), perhitungan dapat dilakukan untuk setiap atribut dengan mengambil contoh pada data Balita 1. Nilai minimum pada atribut berat badan adalah 4.55 dan nilai maksimum adalah 15.8. Nilai minimum pada atribut tinggi badan adalah 56 dan nilai maksimumnya adalah 97. Sehingga dapat diperoleh nilai normalisasi berdasarkan persamaan (1) untuk data Balita 1 sebagai berikut:

$$\text{Normalisasi } BB = \frac{(11.3 - 4.55)}{(15.8 - 4.55)} = 0.6$$

$$\text{Normalisasi } TB = \frac{(97 - 56)}{(97 - 56)} = 1$$

$$\text{Normalisasi } Usia = \frac{(32 - 1)}{(36 - 1)} = 0.85$$

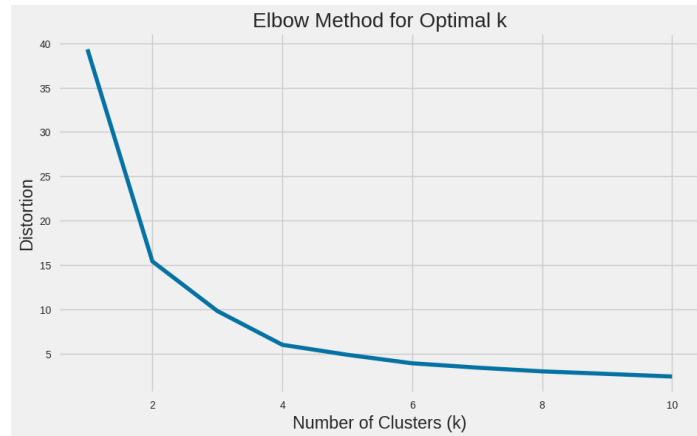
Dalam penelitian ini, diterapkan pada berat badan, tinggi badan, dan usia balita untuk menghilangkan perbedaan skala. Misalnya, nilai berat badan, tinggi badan, dan usia asli diubah menjadi nilai antara 0 dan 1 berdasarkan nilai minimum dan maksimum dalam dataset. Hasil normalisasi, ditunjukkan pada kolom "Normalisasi BB", "Normalisasi TB", dan "Normalisasi Usia" berisi nilai-nilai yang telah disesuaikan sesuai dengan metode *MinMax*. Proses ini membantu algoritma *K-Means clustering* berfungsi lebih efektif dalam mengelompokkan balita berdasarkan kategori gizi.

3.2. Inisialisasi Jumlah *Cluster*

Setelah data dinormalisasi, langkah berikutnya adalah menentukan titik pusat awal untuk *cluster*. Untuk menentukan jumlah *cluster*, digunakan metode *Elbow*. Metode *Elbow* merupakan salah satu teknik yang dapat digunakan untuk menentukan jumlah *cluster* terbaik dengan mengamati persentase masing-masing *cluster* yang membentuk pola siku pada titik tertentu. Metode ini biasanya disajikan dalam bentuk grafik untuk memudahkan identifikasi titik siku. Nilai k pada kombinasi siku dengan *K-means* menggambarkan hubungan antara jumlah *cluster* dan penurunan *error* [11]. Metode *Elbow* digunakan untuk menentukan jumlah *cluster* (c) optimal dengan menghitung *Sum of Square Error* (SSE) dari setiap *cluster*. Rumus yang digunakan untuk perhitungan *Elbow* didefinisikan dengan persamaan (2) berikut.

$$SSE = \sum_{k=1}^k \sum_{x_1 \in s_k} \|X_1 - C_k\|_2^2 \quad (2)$$

Berdasarkan persamaan (2), Jumlah *cluster* yang optimal ditentukan dengan mencari titik di grafik SSE terhadap jumlah *cluster* di mana terjadi perubahan signifikan yang membentuk sudut siku. Nilai c dalam kombinasi *Elbow* dan *K-Means* mencerminkan plot yang menunjukkan pengurangan *error*. Dengan meningkatnya nilai c, grafik SSE cenderung menurun secara bertahap hingga mencapai nilai k yang stabil. SSE merupakan metrik umum yang digunakan untuk mencari jumlah *cluster* c yang optimal. Metode ini menguji berbagai nilai SSE dengan jumlah *cluster* yang berbeda dan mencari titik di grafik *Elbow* di mana perbedaan SSE terbesar terjadi untuk menentukan jumlah *cluster* terbaik [17]. Grafik nilai-nilai *Elbow* dari dataset penelitian ini terlihat pada Gambar 2 di bawah ini.



Gambar 2. Grafik *Elbow*

Berdasarkan analisis yang ditunjukkan Gambar 2, titik *Elbow* yang paling menonjol terletak pada $k=2$ dan $k=4$, yang menunjukkan bahwa dua dan empat *cluster* adalah jumlah optimal untuk mengelompokkan data berdasarkan status gizi balita. Penentuan jumlah *cluster* ini penting dalam menentukan jumlah kelompok optimal. Setelah itu, proses pengelompokkan dilakukan dengan menggunakan algoritma *K-Means*. *K-Means* ini akan bekerja mengelompokkan data berdasarkan kesamaan karakter data.

3.3. Evaluasi *Clustering* dengan *K-Means*

Clustering menggunakan *K-Means* dengan k sesuai hasil grafik *Elbow*. Sebelum menganalisis hasil *cluster*, perlu memvalidasi *cluster* yang terbentuk dengan metode *Silhouette Score*. Rumus yang digunakan untuk perhitungan *Silhouette Score* didefinisikan dengan persamaan (3) berikut.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3)$$

Berdasarkan persamaan (3), di mana $a(i)$ adalah jarak rata-rata antara i dengan anggota lain dalam *cluster* yang sama, dan $b(i)$ adalah rata-rata jarak antara i dengan anggota *cluster* terdekat berikutnya. Nilai yang diperoleh dari metode *Silhouette Score* berkisar antara -1 hingga 1. Jika nilai *Silhouette Score* mendekati 1, hasil pengelompokkan data lebih baik. Sebaliknya, jika nilai tersebut mendekati -1, hasil pengelompokkan data menjadi lebih buruk [18].

Berdasarkan Tabel 2 di bawah ini, jumlah *cluster* dipilih berdasarkan nilai *Silhouette* yang lebih rendah untuk menyesuaikan hasil dengan perhitungan Indeks Massa Tubuh (IMT) yang sudah ada. Meskipun *Silhouette Score* yang lebih tinggi menunjukkan pemisahan yang lebih jelas antar *cluster*, hasil perhitungan IMT dalam konteks kesehatan menunjukkan bahwa jumlah $k=4$ memiliki *Silhouette Score* yang lebih rendah namun lebih mencerminkan kategori atau kelompok balita yang telah diakui secara medis, yaitu status gizi: normal, lebih, kurang, dan sangat kurang. Dengan mengikuti hasil perhitungan IMT ini, analisis data sejalan dengan pemahaman dan praktik kesehatan yang sudah diakui [19].

Tabel 2. Hasil *Silhouette Score*

<i>Silhouette Score</i>	
4 Cluster	0,488
2 Cluster	0,567



3.4 Analisis Hasil *Clustering* ($k=4$)

Pengelompokan dengan *K-Means* dengan jumlah $k=4$ menghasilkan 4 kelompok data. Nama 4 *cluster* berturut-turut adalah *cluster 0*, *cluster 1*, *cluster 2*, *cluster 3*. Anggota setiap kelompok adalah *cluster 0* sebanyak 23 balita, *cluster 1* sebanyak 35 balita, *cluster 2* sebanyak 19 balita, dan *cluster 3* sebanyak 20 balita. Tabel 3 dibawah ini memperlihatkan beberapa contoh data balita beserta *cluster* yang dihasilkan oleh model *K-Means*.

Tabel 3. Hasil *Cluster*

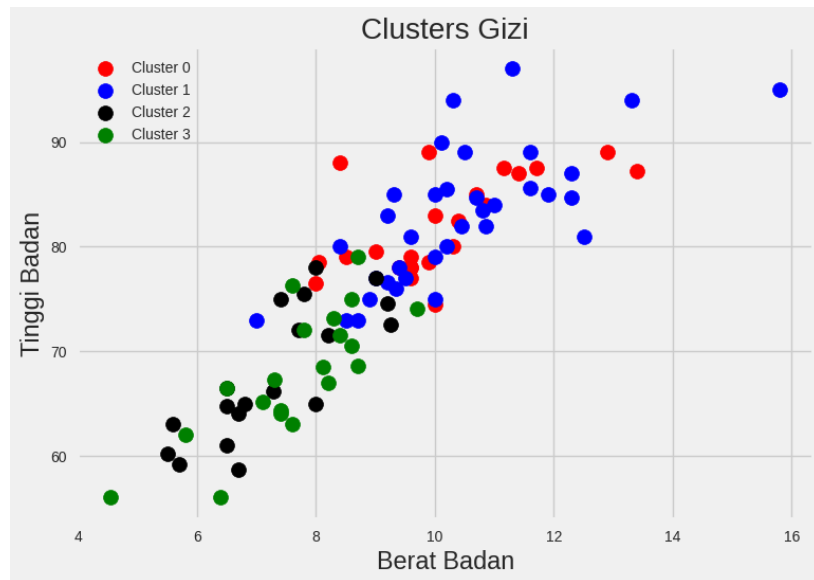
Balita	Berat Badan	Tinggi Badan	Usia	JK	Cluster
Balita 1	0.60	1.00	0.88	0	1
Balita 2	1.00	0.95	0.85	0	1
Balita 3	0.52	0.80	0.85	0	1
Balita 4	0.77	0.92	0.82	0	1
Balita 5	0.68	0.70	0.62	0	1
...
Balita 92	0.48	0.45	0.54	1	0
Balita 93	0.39	0.51	0.71	0	1
Balita 94	0.09	0.17	0.42	1	2
Balita 95	0.49	0.82	0.80	0	1
Balita 96	0.17	0.12	0.42	1	2

Untuk analisis lebih dalam, hasil *cluster* dicari rata-rata nilainya untuk setiap kelompok. Tabel 4 berikut menunjukkan pusat empat *cluster* hasil analisis klasterisasi pada data berat badan dan tinggi badan, yang masing-masing merepresentasikan kondisi gizi berbeda [20]. *Cluster 0* dengan rata-rata berat badan 0.493527, tinggi badan 0.634358, dan usia 0.714286. *Cluster 1*, dengan rata-rata berat badan 0.519356, tinggi badan 0.654774, dan usia 0.720000. Sementara itu, *Cluster 2*, dengan rata-rata berat badan 0.242713, tinggi badan 0.289730, dan usia 0.288722. Serta *Cluster 3*, dengan rata-rata berat badan 0.274533, tinggi badan 0.292683, dan usia 0.218571. Nilai-nilai pusat *cluster* ini merupakan nilai hasil normalisasi tiap variabel, sehingga bukan nilai-nilai yang menunjukkan nilai riil berat badan, tinggi badan, maupun usia.

Tabel 4. Pusat *cluster*

Hasil <i>Cluster</i>	Berat Badan	Tinggi Badan	Usia
<i>Cluster 0</i>	0.493527	0.634358	0.714286
<i>Cluster 1</i>	0.519356	0.654774	0.720000
<i>Cluster 2</i>	0.242713	0.289730	0.288722
<i>Cluster 3</i>	0.274533	0.292683	0.218571

Pusat *cluster* untuk setiap status gizi divisualisasikan dalam grafik dua dimensi yang terlihat pada Gambar 3. Berdasarkan *scatter plot* yang disajikan pada Gambar 3 di atas, *cluster* yang teridentifikasi menunjukkan pola distribusi data mengenai status gizi balita dalam empat kelompok. *Cluster 0* (warna merah) terdiri dari 23 balita, sementara *Cluster 1* (warna biru) merupakan kelompok terbesar dengan 35 balita. Jumlah balita dalam *Cluster 2* (warna hitam) adalah 19 serta *Cluster 3* (warna hijau) terdiri dari 20 balita. Visualisasi *cluster* tersebut menunjukkan bahwa pemisahan data secara tegas tidak terlihat, anggota satu *cluster* dengan *cluster* lain terlihat saling tumpang tindih. Maka, analisis lebih lanjut mengenai karakteristik spesifik dari masing-masing *cluster* akan membantu dalam memahami faktor-faktor yang mempengaruhi status gizi. Pemilihan metode yang lebih tepat perlu dilakukan untuk mengelompokkan data yang benar-benar mempunyai karakter mirip dalam satu kelompok yang sama sehingga satu kelompok dengan lainnya menunjukkan karakter yang benar-benar berbeda. Hal ini bisa memberikan informasi yang mendalam dan berharga bagi *stakeholder* data.



Gambar 3. Hasil Pusat *Clustering*

3.5 Perbandingan Hasil Pengelompokan *K-Means* dan Kalkulator Gizi

Hasil pengelompokan *K-Means* pada Tabel 2 kemudian dibandingkan dengan pengelompokan yang diperoleh menggunakan kalkulator gizi. Perhitungan dilakukan dengan memasukkan berat badan, tinggi badan, usia, dan jenis kelamin setiap balita. Hasil perbandingan ke-97 data balita ditunjukkan pada Tabel 5.

Tabel 5. Perbandingan Hasil *Cluster K-Means* dan Kalkulator Gizi

Balita	<i>K-Means</i>	Kalkulator Gizi
Balita 1	1	Normal
Balita 2	1	Normal
Balita 3	1	Kurang
Balita 4	1	Normal
Balita 5	1	Normal
...
Balita 93	0	Normal
Balita 94	1	Normal
Balita 95	2	Sangat Kurang
Balita 96	1	Kurang
Balita 97	2	Sangat Kurang

Berdasarkan Tabel 5, diperoleh bahwa metode *K-Means* menghasilkan empat *cluster* yaitu *cluster* 0, *cluster* 1, *cluster* 2, dan *cluster* 3, sedangkan kalkulator gizi mengelompokkan ke dalam empat *cluster* yaitu "Sangat Kurang", "Kurang", "Normal", dan "Lebih". Pada *cluster* 0 berdasarkan kalkulator gizi menunjukkan ada 3 balita yang dikategorikan sangat kurang, 3 balita kategori kurang, 16 balita kategori normal, dan 1 balita kategori lebih. *Cluster* 1 menunjukkan 3 balita kategori sangat kurang, 13 balita kategori kurang, dan 19 balita normal. *Cluster* 2 menunjukkan 3 balita sangat kurang, 2 balita kurang, dan 14 balita normal. *Cluster* 3 menunjukkan ada 1 balita yang berkategori sangat kurang, 4 balita kategori kurang, 14 balita normal, dan 1 balita kategori lebih. Dari perbandingan tersebut, tiap *cluster* memiliki kategori secara medis yang berbeda-beda dengan mayoritasnya adalah normal. Sehingga *cluster* yang ada tidak bisa dikategorikan sebagai data yang berbeda.



IV. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan kepada 97 data balita dengan menggunakan algoritma *K-Means* pada *python* dapat disimpulkan bahwa:

1. Pada klasterisasi tingkat status gizi balita dengan menggunakan data berat badan, tinggi badan, dan usia balita di Puskesmas Karangsembung Kabupaten Kebumen diperoleh jumlah *cluster* optimal berdasarkan *Elbow* adalah 2 dan 4. Hasil validasi dengan *Silhouette Score*, $k=2$ dengan nilai *Silhouette Score* tertinggi yaitu 0,567, sedangkan pada $k=4$, *Silhouette Score* hanya 0.488.
2. Berdasarkan status gizi nasional, klasterisasi menggunakan $k=4$. Hasil pusat *cluster* pertama [0.493527; 0.634358; 0.714286] ditandai dengan angka 0, *cluster* kedua dengan pusat [0.519356; 0.654774; 0.720000] ditandai dengan angka 1, *cluster* ketiga dengan pusat [0.242713; 0.289730; 0.288722] ditandai dengan angka 2, dan *cluster* keempat dengan pusat [0.274533; 0.292683; 0.218571] ditandai dengan nomor 3. Pada hasil klasterisasi ini, walaupun *Silhouette Score* hanya 0.488 menunjukkan bahwa pengelompokan data telah menghasilkan struktur yang sesuai dengan perhitungan Indeks Massa Tubuh (IMT) balita.
3. Distribusi anggota kelompok yaitu *cluster* 0, mencakup 23.71% dari total data. Kelompok kedua, atau *cluster* 1, mencakup 36,08% dari total data. Kelompok ketiga, atau *cluster* 2, mencakup 19.58% dari total data. Kelompok keempat, atau *cluster* 3, mencakup 20.61% dari total data.
4. Hasil *clustering* status gizi pada Puskesmas Karangsembung menggunakan *K-Means* menghasilkan empat *cluster* berdasarkan berat badan, tinggi badan, dan usia balita. Namun, hasil ini tidak sesuai dengan kriteria gizi medis dan tidak menunjukkan perbedaan yang jelas antar data. Setiap *cluster* memuat berbagai kategori gizi, dengan mayoritas balita berada pada kategori normal. Oleh karena itu, metode *K-Means* tidak tepat untuk mengkategorisasi gizi balita, karena tidak dapat mencerminkan perbedaan kategori gizi secara medis dengan akurat.

REFERENSI

1. L. Khulafa'ur Rosidah and S. Harsiwi, “Hubungan Status Gizi Dengan Perkembangan Balita Usia 1-3 Tahun (Di Posyandu Jaan Desa Jaan Kecamatan Gondang Kabupaten Nganjuk),” *Jurnal Kebidanan Dharma Husada Kediri*, vol. 6, no. 2, pp. 24–36, 2017, doi: <https://doi.org/10.35890/jkdh.v6i1.48>.
2. K. J. S. Risal, A. Bamahry, and I. K. A. B., “Analisis Luaran Pasien Malnutrisi yang Mendapat Terapi Gizi di RS Ibnu Sina Makassar Tahun 2015-2016,” *UMI Medical Journal*, vol. 4, no. 1, 2019, doi: <https://doi.org/10.33096/umj.v4i1.47>.
3. S. P. Hastoety *et al.*, “Disparitas Balita Kurang Gizi di Indonesia,” *Media Penelitian dan Pengembangan Kesehatan*, vol. 28, no. 3, pp. 201–210, Dec. 2018, doi: 10.22435/mpk.v28i3.219.
4. N. Fidiatoro and T. Setiadi, “Model Penentuan Status Gizi Balita di Puskesmas,” *Jurnal Sarjana Teknik Informatika*, vol. 1, no. 1, pp. 367–373, 2013.
5. S. Permana Ratumanan, Achadiyani, and A. Feinisa Khairani, “Metode Antropometri Untuk Menilai Status Gizi : Sebuah Studi Literatur,” *HIJP: Health Information Jurnal Penelitian*, vol. 15, 2023, [Online]. Available: <https://myjurnal.poltekkes-kdi.ac.id/index.php/hijp>
6. Kementerian Koordinator Bidang Perekonomian, “Global Food Security Summit 2023: Indonesia Serukan Komitmen untuk Mengurangi Prevalensi Gizi Buruk pada Anak,” 2023. [Online]. Available: www.ekon.go.id
7. N. Komang Sri Julyantari, I. Komang Budiarta, and N. Made Dewi Kansa Putri, “Implementasi K-Means Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih),” *Jurnal Janitra Informatika dan Sistem Informasi*, vol. 1, no. 2, pp. 92–101, 2021, doi: 10.25008/janitra.



8. H. Sulastri, H. Mubarak, and S. Sefia Iasha, “Implementasi Algoritma Machine Learning Untuk Penentuan Cluster Status Gizi Balita,” *JURTI*, vol. 5, no. 2, pp. 184–191, 2021, doi: <http://dx.doi.org/10.30872/jurti.v5i2.6779>.
9. W. M. P. Duhita, “Clustering Menggunakan Metode K-Means Untuk Menentukan Status Gizi Balita,” *Jurnal Informatika*, vol. 15, no. 2, pp. 160–174, 2015, doi: <https://doi.org/10.30873/ji.v15i2.598>.
10. F. Dikarya and S. Muharni, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Universitas Terbaik di Dunia,” *Jurnal Informatika*, vol. 22, no. 2, pp. 124–131, 2022, doi: <https://doi.org/10.30873/ji.v22i2.3324>.
11. N. A. Maori and Evanita, “Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering,” *Jurnal SIMETRIS*, vol. 14, no. 2, pp. 277–287, 2023, doi: <https://doi.org/10.24176/simet.v14i2.9630>.
12. I. Griha Tofik Isa, Zulkarnaini, L. Novianti, F. Elfaladonna, and S. Agustri, “Exploratory Data Analysis (EDA) dalam Dataset Penerimaan Mahasiswa Baru Universitas XYZ Palembang,” *Smart Comp*, vol. 12, no. 3, pp. 600–609, 2023, doi: 10.30591/smartcomp.v12i3.4125.
13. A. H. Al-Rahmad and I. Fadillah, *Penilaian Status Gizi dan Pertumbuhan Balita*, 1st ed. Aceh: Jurusan Gizi Politeknik Kesehatan Kemenkes Aceh, 2023. Accessed: Jul. 06, 2024. [Online]. Available: <https://www.halodoc.com/artikel/ini-yang-dimaksud-status-gizi-anak-dan-cara-tepat-menilainya>
14. A. H. Yusup and W. Maharani, “Pembangunan Model Prediksi Kepribadian Berdasarkan Tweet dan Kategori Kepribadian Big Hive Dengan Metode Agglomerative Hierarchical Clustering,” *Jurnal Telekomunikasi Elektro Komputasi & Informatika*, vol. 1, no. 1, pp. 44–50, 2021.
15. A. Rizqi Mulyawan, D. Gunawan, H. Basri, S. Alfarizi, and N. Ichsan, “Penerapan K-Medoids Clustering dan Silhouette Method untuk Strategi Pemasaran Program Donasi pada Lembaga Amil Zakat,” *Information System For Educators And Professionals*, vol. 8, no. 1, pp. 107–118, 2023, doi: <http://dx.doi.org/10.12928/jstie.v1i1.2552>.
16. I. Permana and F. Nur Salisah, “Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation,” *IJIRSE: Indonesian Journal of Informatic Research and Software Engineering*, vol. 2, no. 1, pp. 67–72, 2022, doi: 10.57152/ijirse.v2i1.311.
17. A. Putri Riani, A. Voutama, and T. Ridwan, “Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow,” *Januari*, vol. 6, no. 1, pp. 164–172, 2023, [Online]. Available: <https://ojs.trigunadharma.ac.id/index.php/jsk/index>
18. F. Zahra, A. Khalif, and B. N. Sari, “Pengelompokan Tingkat Kemiskinan di Setiap Provinsi di Indonesia Menggunakan Algoritma K-Medoids,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, pp. 2830–7062, 2024, doi: 10.23960/jitet.v12i2.4199.
19. Peraturan Menteri Kesehatan Republik Indonesia, “Standar Antropometri Anak,” 2, 2020
20. A. Setiawan, “Implementasi Algoritma K-Means Dengan Menggunakan Metode Elbow dan Silhouette Pada Klasterisasi Tingkat Gizi Balita di Puskesmas Margoyoso Tahun 2020,” Universitas Lampung, Bandar Lampung, 2023. [Online]. Available: <http://digilib.unila.ac.id/id/eprint/77587>