



# Evaluasi Jumlah *Cluster* pada Algoritma K-Means++ Menggunakan *Silhouette* dan *Elbow* dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita

Muhammad Raqib Syahkur<sup>1\*</sup>, Dedy Hartama<sup>2</sup>, Solikhun<sup>3</sup> 

<sup>1,3</sup> Teknik Informatika, STIKOM Tunas Bangsa Pematangsiantar, Pematangsiantar, Indonesia

<sup>2</sup> Sistem Informasi, STIKOM Tunas Bangsa Pematangsiantar, Pematangsiantar, Indonesia

## ARTICLE INFO

### Article history:

Received August 03, 2024

Accepted October 16, 2024

Available online October 25, 2024

### Kata Kunci:

Gizi Balita, Algoritma, K-Means++, Silhouette, Elbow, DBI

### Keywords:

Toddler Nutrition, Algorithms, K-Means++, Silhouettes, Elbows, DBI



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

Copyright © 2024 by Author. Published by Universitas Pendidikan Ganesha.

## ABSTRAK

Status gizi balita merupakan indikator penting dalam menentukan kualitas kesehatan masyarakat. Ketidakseimbangan gizi pada balita dapat menyebabkan masalah kesehatan serius, seperti stunting dan peningkatan risiko penyakit kronis dikemudian hari. Tujuan penelitian ini untuk menganalisis metode pengelompokan data gizi balita melalui penerapan algoritma K-Means++ yang lebih presisi, serta evaluasi jumlah *cluster* yang optimal dengan metode elbow dan silhouette, serta validasi dengan nilai DBI dalam mengelompokkan gizi balita. Penelitian ini merupakan penelitian kuantitatif dengan subjek data status gizi balita yang diperoleh dari platform Kaggle. Data dianalisis menggunakan algoritma K-means ++ untuk menentukan titik centroid awal secara lebih presisi dibandingkan algoritma K-Means klasik. Evakuasi jumlah *cluster* dilakukan menggunakan metode Silhouette dan Elbow, sedangkan validasi hasil *clustering* dilakukan dengan Davies-Bouldin Index (DBI). Hasil analisis menunjukkan metode Silhouette menghasilkan jumlah *cluster* terbaik pada K sama dengan 2 dengan rata-rata nilai Silhouette mendekati 1. Sementara itu, metode Elbow menunjukkan titik siku pada K sama dengan 5 yang artinya jumlah *cluster* optimal berada pada titik tersebut. Validasi menggunakan DBI menunjukkan bahwa nilai DBI untuk metode Silhouette dengan K sama dengan 2 adalah 0,682471249, sedangkan nilai DBI untuk metode Elbow dengan K sama dengan 5 adalah 1,238458579. Nilai DBI yang lebih rendah pada K sama dengan 2 menunjukkan bahwa *clustering* tersebut memiliki kohesi internal yang kuat dan pemisahan yang jelas antar *cluster*, sehingga ditetapkan K=2 yang diperoleh dengan metode Silhouette sebagai hasil *clustering* yang optimal.

## ABSTRACT

The nutritional status of toddlers is a crucial indicator in determining the quality of public health. Dietary imbalances in toddlers can lead to serious health problems, such as stunting and an increased risk of chronic diseases later in life. Therefore, clustering toddler nutritional data is essential for analyzing and determining appropriate interventions. This quantitative study uses nutritional status data of toddlers obtained from the Kaggle platform. The data were analyzed using the K-Means++ algorithm to determine initial centroids more precisely than the classical K-Means algorithm. The number of clusters was evaluated using the Silhouette and Elbow methods, while clustering results were validated using the Davies-Bouldin Index (DBI). The analysis showed that the Silhouette method identified the optimal number of clusters at K=2, with an average Silhouette score close to 1. Meanwhile, the Elbow method indicated an optimal cluster number at K=5. DBI validation revealed a DBI value of 0.682471249 for Silhouette with K=2 and 1.238458579 for Elbow with K=5. The lower DBI value at K=2 indicates strong internal cohesion and clear separation between clusters, making K=2 from the Silhouette method the optimal clustering result.

## 1. PENDAHULUAN

Status gizi balita merupakan salah satu indikator penting dalam menentukan kualitas kesehatan masyarakat (Anggie Nauli, 2021; Wahyuni & Fithriyana, 2020). Gizi yang tidak seimbang pada balita dapat menyebabkan berbagai masalah kesehatan, seperti stunting, gizi buruk, hingga risiko penyakit kronis di kemudian hari (Emmaria et al., 2024; Isni & Dinni, 2020; Suriany Simamora & Kresnawati, 2021; Yunita et al., 2020). Pengelompokan data gizi balita berdasarkan karakteristik tertentu sangat penting untuk mempermudah analisis dan penentuan intervensi yang tepat (Chabibah et al., 2020; Putri et al., 2021; Z et al., 2024). Upaya peningkatan SDM yang berkualitas dimulai dengan cara penanganan pertumbuhan anak sebagai bagian dari keluarga dengan asupan gizi dan perawatan yang baik (Maulidah et al., 2019). Dengan lingkungan keluarga yang sehat, maka hadirnya infeksi menular ataupun penyakit masyarakat lainnya dapat dihindari (Agustina et al., 2020; Fauzan et al., 2021). Di tingkat masyarakat seperti faktor lingkungan yang

\*Corresponding author.

E-mail addresses: [muhammadraqib0852@gmail.com](mailto:muhammadraqib0852@gmail.com) (Muhammad Raqib Syahkur)

higenis, asupan makanan, pola asuh terhadap anak, dan pelayanan kesehatan seperti imunisasi sangat menentukan dalam membentuk anak yang tahan gizi buruk (Mirham Nurul Hairunis et al., 2018). Keadaan gizi yang baik merupakan salah satu faktor penting dalam upaya mencapai derajat kesehatan yang optimal.

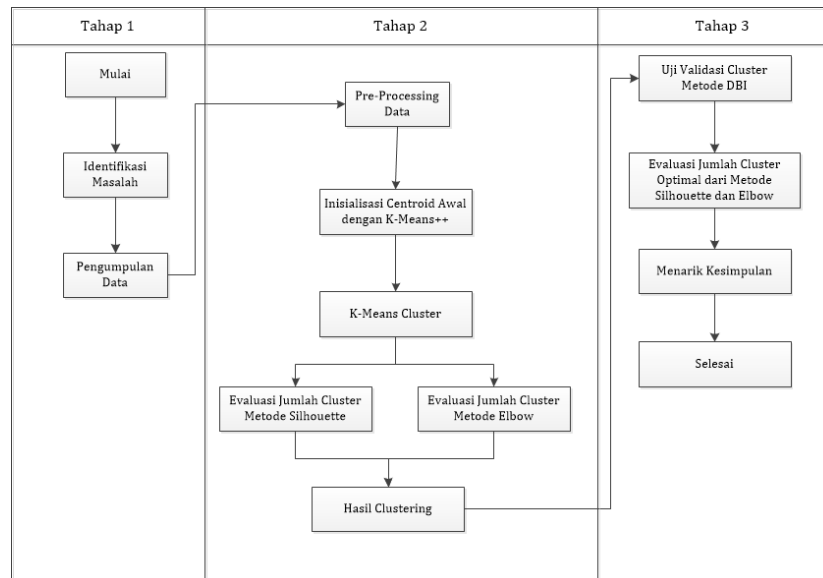
Salah satu masalah utama dalam pengelompokan data adalah menentukan jumlah *cluster* yang optimal (Syahfitri et al., 2023). Ketidaktepatan dalam penentuan jumlah *cluster* dapat menghasilkan pengelompokan yang tidak mencerminkan kondisi sebenarnya (Maori & Evanita, 2023; Siregar, 2020). Masalah ini sering kali menghambat upaya dalam menganalisis dan merancang program intervensi gizi yang efektif (Devi Sari et al., 2023). Status gizi dipengaruhi oleh konsumsi makanan dan penggunaan zat-zat gizi di dalam tubuh. Bila tubuh memperoleh cukup zat-zat gizi dan digunakan secara efisien akan tercapai status gizi optimal yang memungkinkan pertumbuhan fisik, perkembangan otak, kemampuan kerja dan kesehatan secara umum pada tingkat setinggi mungkin (Fernanda et al., 2021; Purwaningsih et al., 2019). Kurang gizi pada masa balita dapat menimbulkan gangguan tumbuh kembang secara fisik, mental, sosial, dan intelektual yang sifatnya menetap dan terus dibawa sampai anak menjadi dewasa (Bayu Lokananta et al., 2024; Fauzan et al., 2021). Secara lebih spesifik, kekurangan gizi dapat menyebabkan keterlambatan pertumbuhan badan, lebih penting lagi keterlambatan perkembangan otak, dan dapat pula terjadinya penurunan atau rendahnya daya tahan tubuh terhadap penyakit infeksi. Status gizi yang tidak optimal pada anak-anak dan remaja dapat berdampak negatif pada pertumbuhan dan perkembangan fisik, kognitif, dan sosial.

Upaya pencegahan dan penanganan masalah gizi, penelitian ini akan menggunakan metode K-Means untuk melakukan pengelompokan data berdasarkan atribut umur (5-18 tahun), berat badan, dan tinggi badan. Pengelompokan data menjadi penting karena memungkinkan identifikasi pola-pola tertentu dalam status gizi. Algoritma K-Means++ merupakan salah satu metode *clustering* yang sering diterapkan karena kemampuannya dalam memilih titik centroid awal yang lebih akurat (Nugraha et al., 2024). Hal ini membantu mengatasi salah satu masalah klasik K-Means, yaitu memilih titik pusat *cluster* secara acak yang dapat mempengaruhi hasil akhir *clustering* (Adhitama et al., 2020; Chaerudin et al., 2021). Penelitian ini menggunakan dua metode evaluasi jumlah *cluster*, yaitu metode Silhouette dan Elbow untuk menilai keakuratan dan kualitas jumlah *cluster* (Khairunnas et al., 2023). *Davies-Bouldin Index* (DBI) adalah indeks validasi yang digunakan untuk mengukur kualitas hasil *clustering* dalam menentukan jumlah *cluster* yang optimal (Bau et al., 2023; Suraya & Wijayanto, 2022). Nilai DBI digunakan sebagai validasi kualitas jumlah *cluster* yang dihasilkan dari metode Silhouette dan Elbow, di mana jumlah *cluster* dengan nilai DBI terendah dianggap sebagai hasil *cluster* yang optimal (Setiyawati et al., 2023; Sidik et al., 2023). Melalui membagi populasi menjadi kelompok-kelompok tertentu berdasarkan kriteria status gizi, seperti gizi buruk, gizi kurang, gizi baik, gizi lebih, dan obesitas, akan mempermudah para pemangku kebijakan dan tenaga kesehatan dalam merancang dan mengimplementasikan program-program gizi yang tepat sasaran.

Beberapa penelitian sebelumnya telah dilakukan untuk pengelompokan gizi menggunakan metode K-Means. Penerapan metode K-Means dalam pengelompokan status gizi anak dan remaja. Hasil penelitian menunjukkan keberhasilan algoritma K-Means dalam mengelompokkan status gizi, dengan mayoritas data masuk dalam kategori "Gizi Baik (Normal)", meskipun terdapat beberapa data yang termasuk kategori "Gizi Buruk (*Severely Thinnes*)" dan "Gizi Kurang (*Thinness*)" (Tholabah et al., 2024). Algoritma K-Means dalam mengelompokkan data balita, penelitian ini berhasil mengelompokkan data balita menjadi tiga klaster, yang kemudian diuji menggunakan aplikasi RapidMiner (Bayu Lokananta et al., 2024). Metode K-Means dapat diterapkan untuk mempermudah pengelompokan status gizi balita dalam jumlah besar. Penelitian lain cenderung menggunakan algoritma K-Means dalam pengelompokan data tanpa melakukan evaluasi dalam menentukan jumlah *cluster* terbaik (Bayu Lokananta et al., 2024; Tholabah et al., 2024). Keterbatasan ini berpotensi menghasilkan *cluster* yang kurang akurat karena pemilihan centroid awal pada K-Means bersifat acak. Perbedaan dengan penelitian terdahulu adalah penelitian ini menerapkan algoritma K-Means++ yang mampu menentukan titik centroid awal dengan lebih presisi, dibandingkan dengan algoritma K-Means klasik yang menentukan centroid awal secara acak sehingga membuat hasil *cluster* menjadi kurang presisi. Selain itu penelitian ini juga melakukan evaluasi jumlah *cluster* dengan metode Silhouette dan Elbow, serta memvalidasi hasil *cluster* terbaik yang dihasilkan kedua metode dengan nilai DBI yang dapat memastikan bahwa jumlah *cluster* yang dihasilkan lebih optimal. Tujuan penelitian ini untuk menganalisis metode pengelompokan data gizi balita melalui penerapan algoritma K-Means++ yang lebih presisi, serta evaluasi jumlah *cluster* yang optimal dengan metode elbow dan silhouette, serta validasi dengan nilai DBI. Hasil pengelompokan yang lebih akurat dari penelitian diharapkan dapat digunakan untuk membantu tenaga kesehatan dalam menganalisis kondisi gizi balita secara lebih efektif, sehingga mendukung pengambilan keputusan yang lebih tepat dalam program intervensi gizi.

## 2. METODE

Penelitian ini menggunakan pendekatan studi kuantitatif yang bertujuan untuk menghasilkan hasil berdasarkan perhitungan matematis dalam analisis data. Proses perhitungan ini didasarkan pada penerapan metode *machine learning* dengan algoritma K-Means++ yang dilengkapi dengan evaluasi jumlah *cluster* optimal menggunakan metode Silhouette dan Elbow, serta validasi DBI. Alur penelitian ini dapat dilihat secara keseluruhan pada Gambar 1.



Gambar 1. Alur Penelitian

Berdasarkan Gambar 1 menjelaskan bahwa tahap awal penelitian ini dimulai dengan tahap Identifikasi Masalah, di mana penelitian ini merumuskan masalah utama dalam mengelompokkan gizi balita untuk analisis kesehatan yang mendalam. Setelah itu, penelitian dilanjutkan ke tahap Pre-Processing Data. Data yang digunakan dalam penelitian ini diambil dari platform Kaggle. Dataset yang didapat dari platform Kaggle tertera pada Tabel 1.

Tabel 1. Dataset Mentah

No	Id Balita	Jenis Kelamin	Umur Bulan/Tahun	Berat	Tinggi	Bmi
1	B1	P	24	5.8	0.65	13.73
2	B2	P	24	5.5	0.59	15.8
3	B3	L	28	6.7	0.715	13.11
...	...	...	...	...	...	...
...	...	...	...	...	...	...
498	B498	L	11	8.1	74.5	23.6
499	B499	P	37	10.7	65.4	47.3
500	B500	L	25	9.9	66.8	18.7

Berdasarkan Tabel 1 menjelaskan bahwa Dataset yang didapat dari platform kaggle terdiri dari variabel ID Balita, Jenis Kelamin, Umur, Berat, Tinggi dan BMI. Data yang diperoleh dibersihkan sebelum diterapkan pada algoritma K-Means++. Langkah pertama adalah penghapusan data outlier yang menunjukkan nilai ekstrem. Penghapusan data outlier dilakukan untuk memastikan bahwa analisis yang dilakukan tidak terdistorsi oleh data yang tidak valid. Selanjutnya Transformasi Data dengan mengubah data yang bersifat karakter menjadi numerik. Kemudian, dilakukan Normalisasi Data ke rentang 0 hingga 1, sehingga setiap data memiliki nilai yang seimbang dalam proses *clustering*. Selanjutnya memeriksa Missing Values untuk menjaga kelengkapan data. Data yang sudah dibersihkan untuk dilakukan penelitian tertera pada Tabel 2.

**Tabel 2.** Dataset Penelitian

No.	Id Balita	Jenis Kelamin (X1)	Umur (X2)	Berat (X3)	Tinggi (X4)	Bmi (X5)
1	B1	0	0.4	0.369231	0.326531	0.307141
2	B2	0	0.4	0.346154	0.204082	0.360123
3	B3	1	0.48	0.438462	0.459184	0.291272
...	...	...	...	...	...	...
198	B198	0	0.76	0.846154	0.530612	0.501664
199	B199	0	0.8	0.769231	0.510204	0.469926
200	B200	1	0.08	0.692308	0.489796	0.43614

Berdasarkan **Tabel 2** menjelaskan bahwa dataset yang dipakai pada penelitian memiliki atribut Jenis Kelamin sebagai X1, Umur sebagai X2, Berat sebagai X3, Tinggi sebagai X4 dan BMI sebagai X5. Tahap Evaluasi Jumlah *Cluster* Optimal dengan menerapkan algoritma K-Means++ untuk menginisialisasi titik centroid awal dengan lebih presisi (Mohammad Ferdiansyah & Umi Chotijah, 2024). Algoritma K-Means++ ialah varian dari algoritma K-Means yang dirancang untuk meningkatkan proses pemilihan centroid awal dengan tujuan meningkatkan konvergensi dan kualitas pengelompokan (Laili et al., 2023). Pemilihan titik pusat *cluster* pertama pada algoritma K-Means++ dilakukan secara random. Titik data yang akan dipilih sebagai centroid berikutnya dipilih berdasarkan probabilitas yang proporsional terhadap jarak kuadratnya dari centroid yang sudah ada, yaitu titik yang lebih jauh dari centroid awal memiliki kemungkinan yang lebih tinggi untuk dipilih sebagai centroid baru. Probabilitas tiap data sebagai centroid. Berdasarkan persamaan (1),  $P(x_i)$  adalah probabilitas data  $x_i$  untuk dipilih sebagai centroid baru.  $D(x_i)$  adalah jarak antara data  $x_i$  dengan centroid terdekat yang sudah ada.  $n$  adalah jumlah data. Hasil titik centroid awal menggunakan algoritma K-Means++ dapat dilihat pada **Tabel 3**.

**Tabel 3.** Hasil Titik *Centroid* Awal dari K-Means++

<i>Centroid</i>	ID Balita	Probabilitas Tertinggi
C1	B5	(Dipilih secara acak)
C2	B109	0.013361161
C3	B18	0.019603734
C4	B107	0.019327844
C5	B16	0.019603734
C6	B7	0.017068318
C7	B102	0.014174452
C8	B57	0.014100645
C9	B22	0.014969013

Berdasarkan **Tabel 3** bahwa semua titik centroid awal telah dipilih berdasarkan perhitungan algoritma K-Means++. Selanjutnya algoritma pengelompokan dilanjutkan dengan langkah-langkah standar K-Means (Ramdani & Safadila, 2022). Pada tahap ini, menggunakan Elbow untuk mengevaluasi jumlah *cluster* dengan melakukan perbandingan persentase antar *cluster* yang membentuk siku pada grafik (Amelia et al., 2022; Juanita & Cahyono, 2024). Jumlah *cluster* yang optimal ditandai dengan jarak yang semakin besar dari titik siku (Pangestu et al., 2022). Perbandingan dapat diperoleh dengan menghitung nilai *Sum of Squared Error* (SSE) untuk setiap jumlah *cluster*. Berdasarkan persamaan  $x_i^{(j)}$  merupakan data ke- $i$  dalam *cluster* ke- $j$  dan  $c_j$  adalah centroid dari *cluster* ke- $j$ , serta  $k$  adalah jumlah *cluster*. Selain itu, metode Silhouette juga digunakan untuk mengevaluasi jumlah *cluster* optimal dengan menghitung nilai Silhouette global sebagai ukuran kualitas *clustering*. Nilai Silhouette ini dapat diperoleh melalui dua nilai, yaitu nilai Silhouette untuk data ke- $i$  dan nilai Silhouette global. Kedua nilai ini digunakan untuk mengevaluasi seberapa data data atau *cluster* terhadap angka 1. Untuk mendapatkan nilai Silhouette dari data ke- $i$ , terdapat dua koefisien utama, yaitu  $a$  dan  $b$ . Koefisien  $a$  adalah rata-rata jarak antar data di dalam *cluster*. Nilai  $a$  digunakan untuk menilai seberapa jauh data tersebut berbeda dengan *cluster* yang tempat data itu berada.

$m_j$  adalah jumlah titik data pada *cluster*  $j$  dan  $d(x_i^j, x_r^j)$  adalah jarak data  $x_i^j$  dengan data ke- $r$  di *cluster*  $j$ . Berdasarkan persamaan (3), Koefisien  $b$  adalah rata-rata jarak terkecil dari data tersebut ke semua data dari *cluster* lain yang berbeda. Nilai  $b$  bertujuan untuk mengevaluasi seberapa besar ketidakpastian

suatu data terhadap *cluster* lain. Semakin besar nilai  $b$ , maka semakin tidak sesuai atau jelek data tersebut terhadap *cluster* lain.  $m_n$  adalah jumlah titik data dalam *cluster*  $n$  ( $n \neq j$ ) dan  $d(x_i^j, x_r^n)$  adalah jarak data  $x_i^j$  dengan data ke- $r$  dalam *cluster*  $n$ . Selanjutnya adalah mencari nilai Silhouette global yang dapat menentukan kualitas jumlah *cluster*. Silhouette global diperoleh dengan perhitungan nilai rata-rata dari seluruh nilai Silhouette untuk setiap data dalam dataset dengan  $n$  adalah jumlah total data dalam dataset dan  $s(i)$  adalah nilai Silhouette untuk data ke- $i$ . Nilai Silhouette global ini memberikan gambaran keseluruhan tentang kualitas *clustering*. Nilai Silhouette global yang lebih tinggi menunjukkan jumlah *cluster* yang optimal (Paembonan & Abduh, 2021). Penelitian ini dilanjutkan dengan melakukan validasi hasil *clustering* menggunakan DBI. DBI adalah metode evaluasi internal yang mengukur kualitas *clustering* berdasarkan nilai kohesi dan separasi. Kohesi merupakan tingkat keterhubungan data terhadap titik pusat *cluster* yang diikutinya. Persamaan *Sum of Square Within Cluster* (SSW) digunakan dalam menghitung kohesi *cluster* ke- $i$  dan dinyatakan,  $m_i$  adalah jumlah data pada *cluster* ke- $i$ ,  $c_i$  adalah titik pusat *cluster* ke- $i$  dan  $d(x_j, c_i)$  merupakan jarak setiap data ke centroid. Separasi didasarkan pada jarak antar centroid dari masing-masing *cluster*. Persamaan *Sum of Square Between Cluster* (SSB) digunakan untuk mengukur separasi antar *cluster*,  $d(c_i, c_j)$  merupakan jarak antara centroid *cluster* ke- $i$  dan *cluster* ke- $j$ . Kemudian dilakukan pengukuran rasio ( $R_{ij}$ ) untuk membandingkan *cluster* ke- $i$  dan *cluster* ke- $j$ . *Cluster* yang baik memiliki nilai kohesi serendah mungkin dan separasi setinggi mungkin. Rasio yang telah dihitung digunakan untuk mendapatkan nilai DBI.  $K$  ialah jumlah total *cluster*. Nilai DBI yang lebih rendah menunjukkan pengelompokan yang lebih baik (Kamila et al., 2019). Penelitian ini menarik kesimpulan berdasarkan nilai DBI serta evaluasi dari metode Silhouette dan Elbow, di mana nilai DBI yang rendah menunjukkan bahwa *cluster* yang dihasilkan cukup baik dan metode tersebut lebih baik dalam mengevaluasi jumlah *cluster* yang optimal.

### 3. HASIL DAN PEMBAHASAN

#### Hasil

Penelitian ini mengimplmentasikan K-Means++ untuk menetapkan titik pusat *cluster* awal yang lebih akurat, dilanjutkan dengan proses *clustering* K-Means dengan jumlah *cluster*  $K=2$  hingga  $K=9$ . Hasil dari setiap *cluster* dievaluasi menggunakan metode Silhouette dan Elbow untuk menentukan *cluster* optimal dan diuji validasinya dengan menggunakan nilai DBI. Hasil centroid awal yang diperoleh dari algoritma K-Means++ pada  $K=2$  hingga  $K=9$  dapat dilihat pada Tabel 4.

**Tabel 4.** Titik Centroid Awal

Centroid	Nama Balita	X1	X2	X3	X4	X5
C1	B5	0	0.48	0.453846	0.489796	0.287177
C2	B109	1	0.86	0.692308	0.081633	0.866906
C3	B18	1	0.48	0.7	0.938776	0.24213
C4	B107	1	0	0.923077	0.367347	0.696954
C5	B16	0	0.54	0.707692	0.020408	1
C6	B7	1	0.38	0.284615	0.071429	0.392117
C7	B102	0	1	0.507692	0.897959	0.180701
C8	B57	0	0.62	1	0.489796	0.628103
C9	B22	0	0.38	0.846154	1	0.275403

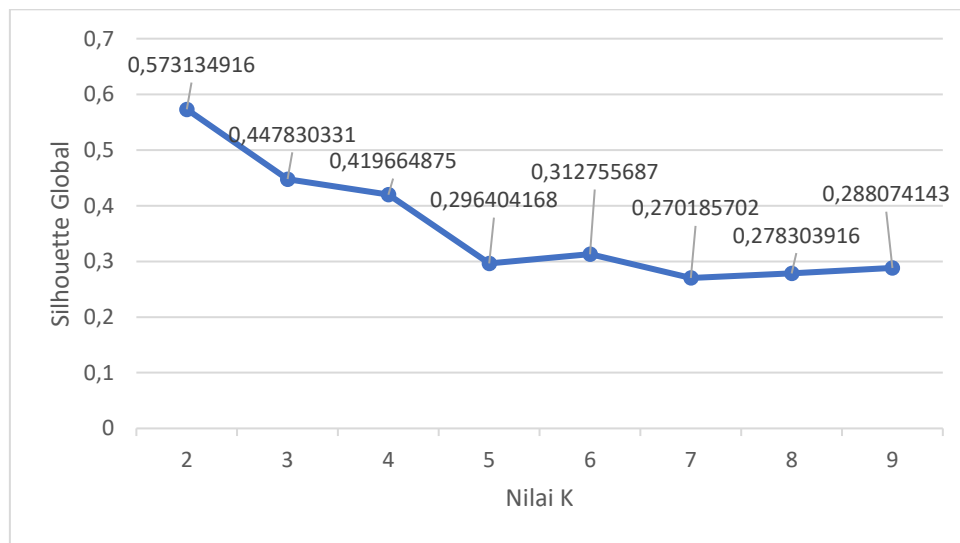
Pada Tabel 4 terlihat bahwa titik centroid awal telah ditetapkan dengan menerapkan K-Means++. Tahap selanjutnya adalah melakukan proses *clustering* menggunakan algoritma K-Means untuk setiap nilai  $K$  dari  $K=2$  hingga  $K=9$ . Dengan memanfaatkan titik centroid awal yang diperoleh dari Tabel 4. Hasil dari proses *clustering* untuk masing-masing nilai  $K$ , mulai dari  $K=2$  hingga  $K=9$  dapat dilihat pada Tabel 5 yang memberikan gambaran mengenai pembagian kelompok data sesuai dengan jumlah *cluster* yang berbeda.

**Tabel 5.** Hasil Clustering  $K=2$  hingga  $K=9$

K	Data Cluster								
	1	2	3	4	5	6	7	8	9
K=2	108	98	-	-	-	-	-	-	-
K=3	108	28	64	-	-	-	-	-	-
K=4	108	25	46	21	-	-	-	-	-

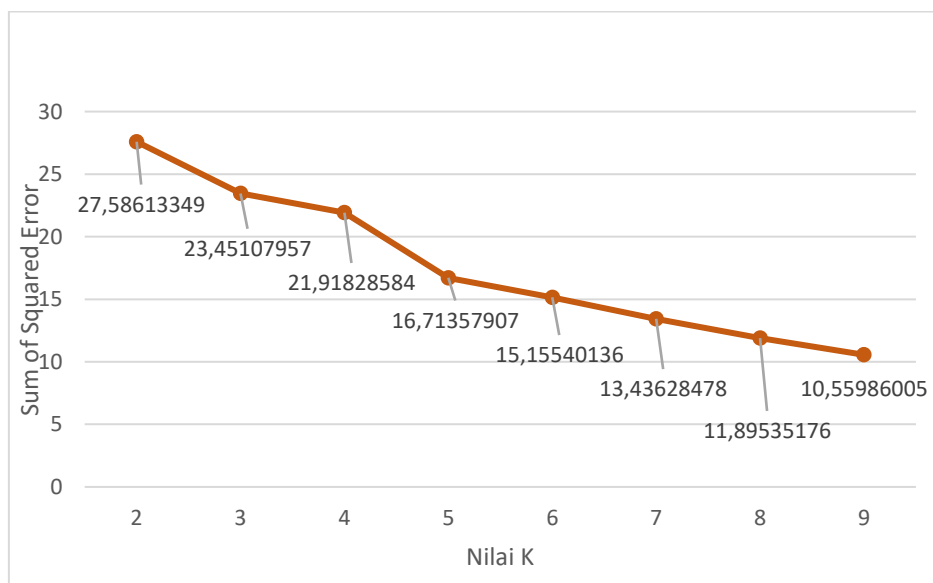
K	Data Cluster								
	1	2	3	4	5	6	7	8	9
K=5	69	25	46	21	39	-	-	-	-
K=6	69	22	30	19	39	21	-	-	-
K=7	57	22	30	19	21	21	30	-	-
K=8	26	22	30	19	17	21	27	38	-
K=9	19	22	30	19	15	21	36	18	20

Berdasarkan Tabel 5, hasil clustering telah diperoleh menggunakan algoritma K-Means dengan jumlah cluster K=2 hingga K=9. Namun, hasil ini masih belum mengindikasikan jumlah cluster yang paling optimal untuk pembagian data. Tahap selanjutnya adalah mengevaluasi setiap hasil clustering dari K=2 hingga K=9 dengan metode Silhouette. Ringkasan nilai Silhouette Global untuk setiap jumlah cluster K dapat dilihat dalam bentuk grafik pada Gambar 2.



Gambar 2. Grafik Silhouette

Berdasarkan Gambar 2, grafik Silhouette menunjukkan bahwa jumlah cluster optimal adalah 2 cluster. Hal ini ditunjukkan oleh nilai rata-rata Silhouette tertinggi yang diperoleh saat menggunakan 2 cluster sebesar 0,573134916. Nilai ini mengindikasikan bahwa jumlah cluster optimal adalah 2 berdasarkan nilai rata-rata Silhouette yang mendekati 1 menunjukkan kualitas clustering yang lebih baik.



Gambar 3. Grafik Elbow

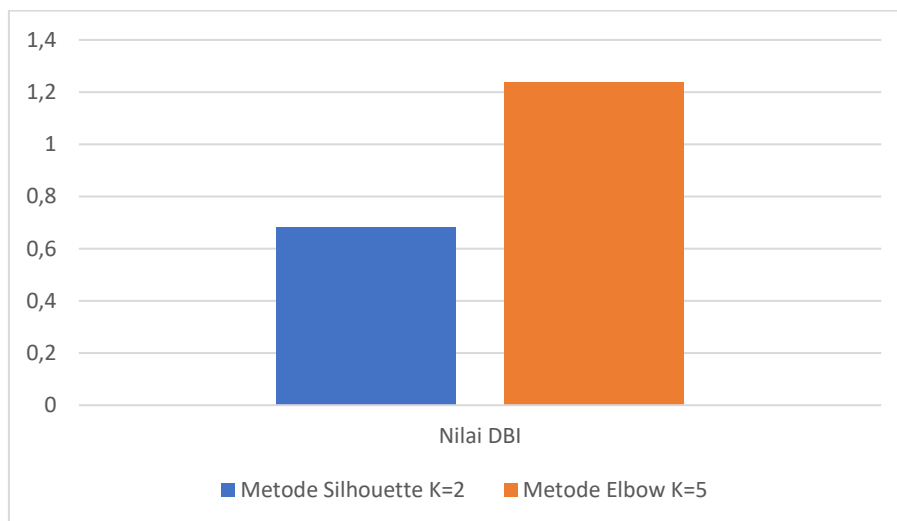


Gambar 3 menampilkan grafik Sum of Squared Error (SSE) yang digunakan untuk menentukan jumlah *cluster* yang optimal berdasarkan perubahan nilai SSE. Dalam analisis SSE, jika nilai SSE meningkat, kualitas pembentukan *cluster* menurun. Sebaliknya, semakin rendah nilai SSE, semakin baik kualitas *cluster* yang dihasilkan. Dalam grafik SSE tersebut terlihat bahwa saat jumlah *cluster* K=2, nilai SSE berada pada titik tertinggi. Ketika jumlah *cluster* K=3, nilai SSE menunjukkan penurunan, yang berlanjut hingga K=9. Berdasarkan pola penurunan ini, terbentuk sebuah "siku" atau titik perubahan signifikan pada K=5, di mana penurunan SSE tidak lagi sebesar penurunan sebelumnya pada jumlah *cluster* yang lebih besar. Pada titik ini, terdapat selisih nilai SSE sebesar 5,20470677 dibandingkan dengan SSE pada jumlah *cluster* sebelumnya. Pada titik K=6 hingga K=10 terlihat nilai SSE lebih stabil, menunjukkan bahwa penambahan *cluster* tidak lagi memberikan peningkatan signifikan dalam pengurangan SSE. Untuk menilai kualitas hasil *clustering* dalam penelitian ini, digunakan metode validasi *cluster* yang dapat menunjukkan baik atau kurang baiknya suatu pengelompokan. Metode yang digunakan untuk mengevaluasi hasil *clustering* dari metode Silhouette dan elbow adalah DBI. Nilai DBI digunakan sebagai acuan untuk menentukan kualitas *clustering*. Suatu *cluster* dianggap optimal jika memiliki nilai DBI yang minimal. Validasi *cluster* dengan DBI dapat dilihat pada Tabel 6.

Tabel 6. Validasi Hasil Clustering

Metode Evaluasi	Cluster Optimal	Nilai DBI
Metode Silhouette	K=2	0.682471249
Metode Elbow	K=5	1.238458579

Pada Tabel 6 terlihat bahwa nilai DBI dari hasil *clustering* K-Means++ menggunakan metode Silhouette pada K=2 mencapai rata-rata nilai DBI sebesar 0,682471249 yang lebih rendah, dibandingkan dengan nilai DBI menggunakan metode Elbow pada K=5 sebesar 1,238458579. Nilai DBI yang lebih rendah pada metode Silhouette menunjukkan bahwa hasil *clustering* k-Means++ menghasilkan kualitas *cluster* yang lebih baik dibandingkan dengan metode Elbow.



Gambar 3. Grafik Nilai DBI

Gambar 3 menunjukkan grafik nilai DBI yang memperlihatkan bahwa metode Silhouette menghasilkan nilai DBI lebih kecil dengan kualitas pengelompokan terbaik pada jumlah *cluster* K=2. Analisis *clustering* dalam penelitian ini menunjukkan bahwa jumlah *cluster* optimal adalah K=2 dengan nilai rata-rata Silhouette sebesar 0,573134916. Nilai ini mendekati nilai maksimal 1 dan menunjukkan kualitas *clustering* yang tinggi. Pada K=2, pengelompokan yang dihasilkan memiliki nilai kohesi internal yang kuat dan pemisahan yang jelas antar *cluster*. Evaluasi SSE pada metode Elbow menunjukkan titik "siku" atau perubahan signifikan terjadi pada K=5. Setelah K=5, penambahan jumlah *cluster* tidak menghasilkan penurunan SSE yang substansial yang mengindikasikan bahwa peningkatan jumlah *cluster* tidak selalu meningkatkan kualitas *clustering* secara signifikan. Metode DBI digunakan untuk memvalidasi kualitas *cluster* dengan melihat nilai DBI yang paling rendah pada *cluster* K optimal dari metode Silhouette dan Elbow. Nilai DBI pada K optimal yang didapatkan dengan metode silhouette yaitu K=2 mendapatkan nilai DBI sebesar 0,682471249, lebih rendah dibandingkan pada K optimal metode Elbow yaitu K=5 dengan nilai

DBI yang mencapai 1,238458579. Nilai DBI yang lebih rendah pada metode Silhouette mengindikasikan bahwa *clustering* dengan jumlah *cluster* K=2 memiliki kualitas terbaik dan lebih optimal dibandingkan dengan jumlah *cluster* lainnya. Berdasarkan analisis ini, jumlah *cluster* K=2 dengan *clustering* K-Means++ menggunakan metode Silhouette dinilai paling optimal untuk mengelompokkan gizi balita secara lebih efektif dalam menghasilkan pengelompokan yang presisi dan relevan untuk tujuan penelitian.

## Pembahasan

Berdasarkan analisis ini, jumlah *cluster* K=2 dengan *clustering* K-Means++ menggunakan metode Silhouette dinilai paling optimal untuk mengelompokkan gizi balita secara lebih efektif dalam menghasilkan pengelompokan yang presisi dan relevan untuk tujuan penelitian. K-means merupakan metode data mining yang melakukan proses pemodelan tanpa supervisi (*unsupervised*) dan melakukan pengelompokan data dengan sistem partisi. Penelitian ini mengintegrasikan evaluasi menggunakan metode Silhouette dan Elbow, yang tidak dilakukan pada penelitian sebelumnya, sehingga menghasilkan analisis yang lebih sistematis dalam menentukan jumlah *cluster* optimal. Metode DBI yang diterapkan sebagai validasi tambahan memberikan jaminan bahwa kualitas *clustering* yang dihasilkan lebih terukur dibandingkan metode evaluasi tunggal. Penelitian ini memiliki beberapa kelebihan, diantaranya adalah penerapan algoritma K-Means++ yang mampu menentukan titik centroid awal dengan lebih presisi dibandingkan K-Means klasik, serta penggunaan evaluasi sistematis melalui metode Silhouette dan Elbow yang diperkuat dengan validasi DBI. Algoritma X-means terdiri dari beberapa operasi berulang hingga eksekusi berakhir, dasar proses dari algoritma x-means adalah memodifikasi struktur *cluster* k-means dengan membagi data ke dalam minimal dua *cluster* dilanjutkan dengan pembagian dua kluster untuk setiap kluster yang telah terbentuk (dua *cluster* lokal untuk setiap *cluster*) (Al-Fahmi et al., 2023; Bau et al., 2023). Jika menggunakan 3 *cluster* maka masing-masing *cluster* akan dibagi ke dalam dua lokal *cluster* dari pusat *cluster* awalnya.

Hasil penelitian ini menunjukkan perbaikan yang signifikan dibandingkan dengan penelitian sebelumnya yang hanya menggunakan algoritma K-Means klasik tanpa melakukan evaluasi jumlah *cluster* optimal. Selain itu penelitian sebelumnya juga tidak memanfaatkan algoritma K-Means++ yang mampu menentukan titik centroid awal dengan lebih presisi (Bayu Lokananta et al., 2023; Tholabah et al., 2024). Dibandingkan dengan penelitian yang menghasilkan tiga *cluster* dengan jumlah data yang relatif merata, penelitian ini menunjukkan bahwa pembagian data menjadi dua *cluster* memberikan kualitas *clustering* yang lebih tinggi berdasarkan nilai Silhouette dan validasi DBI (Bayu Lokananta et al., 2023). Penelitian ini memiliki keterbatasan seperti kurangnya perbandingan dengan algoritma *clustering* lainnya dan cakupan dataset masih terbatas pada wilayah tertentu, sehingga hasilnya mungkin belum dapat digeneralisasi untuk populasi yang lebih luas. Implikasi penelitian ini dalam meningkatkan akurasi pengelompokan data gizi balita dan dapat menjadi acuan untuk analisis praktis bagi lembaga kesehatan atau pemerintah dalam menyusun kebijakan berbasis data.

## 4. SIMPULAN

Berdasarkan hasil analisis evaluasi jumlah *cluster* menggunakan metode Silhouette dan Elbow pada algoritma *clustering* k-Means++ dalam mengelompokkan gizi balita dapat diperoleh kesimpulan bahwa jumlah *cluster* optimal adalah K=2. Jumlah *cluster* optimal K=2 yang didapatkan menggunakan metode Silhouette juga mendapatkan nilai DBI paling rendah, dibandingkan dengan jumlah *cluster* optimal yang diperoleh melalui metode Elbow. Direkomendasikan penelitian selanjutnya memperluas cakupan dataset, menambahkan variabel atau atribut pendukung lainnya, serta membandingkan algoritma lain untuk meningkatkan kualitas hasil *clustering* secara keseluruhan.

## 5. DAFTAR PUSTAKA

- Adhitama, R., Burhanuddin, A., & Ananda, R. (2020). Penentuan Jumlah *Cluster* Ideal Smk Di Jawa Tengah Dengan Metode X-Means *Clustering* Dan K-Means *Clustering*. *JIKO (Jurnal Informatika Dan Komputer)*, 3(1), 1–5. <https://doi.org/10.33387/jiko>.
- Agustina, R., Mandala, Z., & Liyola, R. (2020). Kadar Ferritin dengan Status Gizi Pasien Thalassemia  $\beta$  Mayor Anak di RSAM Bandar Lampung Pendahuluan. *Jurnal Ilmiah Kesehatan Sandi Husada*, 1(1), 219–224. <https://doi.org/10.35816/jiskh.v10i2.252>.
- Al-Fahmi, B. M., Rahmawati, E., & Sagirani, T. (2023). Penerapan K-Means *Clustering* Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 9(2), 141–149. <https://doi.org/10.25077/teknosi.v9i2.2023.141-149>.
- Amelia, D., Padilah, T. N., & Jamaludin, A. (2022). Optimasi Algoritma K-Means Menggunakan Metode Elbow



- dalam Pengelompokan Penyakit Demam Berdarah Dengue (DBD) di Jawa Barat. *Jurnal Ilmiah Wahana Pendidikan*, 8(11), 207–215. <https://doi.org/10.5281/zenodo.6831380>.
- Anggie Nauli, H. (2021). Analisis Deskriptif Phbs Dan Status Gizi Masyarakat Upaya Peningkatan Status Kesehatan Melalui Pendekatan Kolaboratif Berbagai Pemangku Kepentingan. *Abdi Dosen : Jurnal Pengabdian Pada Masyarakat*, 5(1), 49. <https://doi.org/10.32832/abdidos.v5i1.832>.
- Bau, M. K. D., Setyawan, Y., & Jatipaningrum, M. T. (2023). Perbandingan Metode Algoritma K-Means Dan K-Medoids Pada Pengelompokan Kabupaten/Kota Di Provinsi Nusa Tenggara Timur Berdasarkan Dimensi Indeks Pembangunan Manusia Tahun 2020. *Jurnal Statistika Industri Dan Komputasi*, 8(1), 48–57. <https://doi.org/10.34151/statistika.v8i1.4418>.
- Bayu Lokananta, R., Yuana, H., & Dwi Puspitasari, W. (2023). 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>.
- 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>.
- Chabibah, N., Khanifah, M., & Kristiyanti, R. (2020). Pengaruh Pemberian Modifikasi Edukasi Booklet Gizi Balita Dan Cooking Class Terhadap Pengetahuan Dan Pola Pemberian Makan Balita. *Jurnal Kebidanan Indonesia*, 11(2), 47. <https://doi.org/10.36419/jkebin.v11i2.372>.
- Chaerudin, A., Murdiansyah, D. T., & Imrona, M. (2021). Implementation of K-Means++ Algorithm for Store Customers Segmentation Using Neo4j. *Indonesia Journal on Computing (Indo-JC)*, 6(1), 53–60. <https://doi.org/10.34818/INDOJC.2021.6.1.547>.
- Devi Sari, Sri Achadi Nugraheni, & Mohammad Zen Rahfiludin. (2023). Bagaimana Kontribusi Intervensi Gizi Sensitif dalam Upaya Penurunan Stunting? : Literature Review. *Media Publikasi Promosi Kesehatan Indonesia (MPPKI)*, 6(5), 885–895. <https://doi.org/10.56338/mpkvi.v6i6.3416>.
- Emmaria, R., Sinaga, K., Manurung, B., Tobing, R. A. L., Lubis, R. D., & Sihombing, P. A. (2024). Edukasi Penanganan Dan Pencegahan Gizi Kurang Pada Balita Di Wilayah Kerja Puskesmas Rantang Kota Medan Tahun 2024. *Jurnal Pengabdian Kolaborasi Dan Inovasi IPTEKS*, 2(4), 1309–1314. <https://doi.org/10.59407/jpki2.v2i4.1168>.
- Fauzan, M. A., Nurmalarisari, Y., & Anggunan, A. (2021). Hubungan Status Gizi dengan Prestasi Belajar. *Jurnal Ilmiah Kesehatan Sandi Husada*, 10(1), 105–111. <https://doi.org/10.35816/jiskh.v10i1.517>.
- Fernanda, C., Gifari, N., Mulyani, E., Nuzrina, R., & Ronitawati, P. (2021). Hubungan Asupan, Status Gizi, Aktivitas Fisik, Tingkat Stres dan Siklus Menstruasi Atlet Bulutangkis. *Sport and Nutrition Journal*, 3(1), 1–14. <https://doi.org/10.15294/spnj.v3i1.41133>.
- Isni, K., & Dinni, S. M. (2020). Pelatihan Pengukuran Status Gizi Balita Sebagai Upaya Pencegahan Stunting Sejak Dini Pada Ibu Di Dusun Randugunting, Sleman, Diy. *Panrita Abdi - Jurnal Pengabdian Pada Masyarakat*, 4(1), 60. <https://doi.org/10.20956/pa.v4i1.7299>.
- Juanita, S., & Cahyono, R. D. (2024). K-Means Clustering With Comparison of Elbow and Silhouette Methods for Medicines Clustering Based on User Reviews. *Jurnal Teknik Informatika (JUTIF)*, 5(1), 283–289. <https://doi.org/10.52436/1.jutif.2024.5.1.1349>.
- Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119. <https://doi.org/10.24014/rmsi.v5i1.7381>.
- Khairunnas, M. A., Jamaludin, A., & Adam, R. I. (2023). Pengaruh Pendapatan Orang Tua terhadap Hasil Belajar Siswa Menggunakan Algoritma K-Means Clustering. *Jurnal Pendidikan Tambusai*, 7(3), 31434–31444. <https://doi.org/10.31004/jptam.v7i3.12130>.
- Laili, U. F., Umatin, C., & Ridwanulloh, M. U. (2023). Analisis Potensial Drop Out Mahasiswa Dengan K-Means++ Clustering Dalam Upaya Peningkatan Kualitas IAIN Kediri. *Paedagogia: Jurnal Kajian, Penelitian Dan Pengembangan Kependidikan*, 14(2), 145–153. <https://doi.org/10.31764/paedagogia.v14i2.14077>.
- Maori, N. A., & Evanita, E. (2023). Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 14(2), 277–288. <https://doi.org/10.24176/simet.v14i2.9630>.
- Mirham Nurul Hairunis, Salimo, H., & Dewi, Y. L. R. (2018). Hubungan Status Gizi dan Stimulasi Tumbuh Kembang dengan Perkembangan Balita. *Sari Pediatri*, 20(3), 146 – 51. <https://doi.org/10.14238/sp20.3.2018.146-51>.
- Mohammad Ferdiansyah, & Umi Chotijah. (2024). Implementasi Algoritme K-Means++ Untuk Clustering Penjualan Bahan Bangunan. *Jurnal Ilmiah Teknik Informatika Dan Komunikasi*, 4(1), 181–193. <https://doi.org/10.55606/juitik.v4i1.767>.
- Nugraha, R. P., Laxmi, G. F., & Riana, F. (2024). Penerapan K-Means ++ Untuk Pengelompokan Mahasiswa

- Berpotensi Drop Out ( Studi Kasus : Universitas Ibn Khaldun Bogor ). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(3), 3493–3500. <https://doi.org/10.36040/jati.v8i3.9738>.
- Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi *Clustering* Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2), 48. [https://doi.org/10.51557/pt\\_jiit.v6i2.659](https://doi.org/10.51557/pt_jiit.v6i2.659).
- Pangestu, B. A., Kristiawan, N. A., & Sulistiyowati, N. (2022). *Clustering* Obat Untuk Menentukan Pola Pemasaran Efektif di Apotek Amarta Sehat. *Jurnal Ilmiah Wahana Pendidikan*, 8(16), 115–126. <https://doi.org/10.5281/zenodo.7058995>.
- Purwaningsih, K. A., Weta, I. W., & Aryani, P. (2019). Asupan zat gizi dan status gizi anak vegetarian dan nonvegetarian kelas 3-6 sekolah dasar di bhaktivedanta dharma school. *E-Journal Medika*, 8(1), 23–26.
- Putri, V. H., Sitoayu, L., & Ronitawati, P. (2021). Pengaruh media AR Book terhadap peningkatan pengetahuan dan sikap gizi seimbang pada anak usia sekolah. *Aceh Nutrition Journal*, 118–127. <https://doi.org/10.30867/action.v6i2.380>.
- Ramdani, C., & Safadila, N. (2022). Analisis Data Akademis dengan Menerapkan Algoritme K-Means dan K-Means++. *LEDGER: Journal Informatic and Information Technology*, 1(4), 155–160. <https://doi.org/10.20895/LEDGER.V1I4.918>.
- Setiyawati, N., Bangkalang, D. H., & Purnomo, H. D. (2023). Comparison of K-Means & K-Means++ *Clustering* Models using Singular Value Decomposition (SVD) in Menu Engineering. *JOIV: International Journal on Informatics Visualization*, 7(3), 871–877. <https://doi.org/10.30630/joiv.7.3.1053>.
- Sidik, R., Suarna, N., & Rinaldi Dikananda, A. (2023). Analisa Data Set Peminatan Siswa Menggunakan Algoritma K-Means Dengan Optimize Parameter Di Sekolah Menengah Kejuruan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(2), 1197–1203. <https://doi.org/10.36040/jati.v7i2.6335>.
- Siregar, I. S. (2020). Hubungan Pengetahuan Dengan Sikap Ibu Tentang Pemberian Makanan Pendamping Asi (Mp-Asi) Di Wilayah Puskesmas Binjai Estate Tahun 2020. *Jurnal Health Reproductive*, 5(2), 8–16. <https://doi.org/10.51544/jrh.v5i2.1623>.
- Suraya, G. R., & Wijayanto, A. W. (2022). Comparison of Hierarchical *Clustering*, K-Means, K-Medoids, and Fuzzy C-Means Methods in Grouping Provinces in Indonesia according to the Special Index for Handling Stunting. *Indonesian Journal of Statistics and Its Applications*, 6(2), 180–201. <https://doi.org/10.29244/ijsa.v6i2p180-201>.
- Suriany Simamora, R., & Kresnawati, P. (2021). Pemenuhan Pola Makan Gizi Seimbang Dalam Penanganan Stunting Pada Balita Di Wilayah Puskesmas Kecamatan Rawalumbu Bekasi. *Jurnal Bidang Ilmu Kesehatan*, 11(1), 34–45. <https://doi.org/10.52643/jbik.v11i1.1345>.
- Syahfitri, N., Budianita, E., Nazir, A., & Afrianty, I. (2023). Pengelompokan Produk Berdasarkan Data Persediaan Barang Menggunakan Metode Elbow dan K-Medoid. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(3), 1668–1675. <https://doi.org/10.30865/klik.v4i3.1525>.
- Tholabah, M., Nugroho, R. A., & Mu'min, S. (2024). Penerapan Metode K-Means dalam Pengelompokan Status Gizi Anak dan Remaja. *Nusantara Computer and Design Review*, 2(1), 7–14. <https://doi.org/10.55732/ncdr.v2i1.1209>.
- Wahyuni, D., & Fithriyana, R. (2020). Pengaruh Sosial Ekonomi Dengan Kejadian Stunting Pada Balita Di Desa Kualu Tambang Kampar. *PREPOTIF: Jurnal Kesehatan Masyarakat*, 4(1), 20–26. <https://doi.org/10.31004/prepotif.v4i1.539>.
- Yunita, D., Luthfi, A., & Erlinawati. (2020). Hubungan Pemberian Stimulasi Dini Dengan Perkembangan Motorik Pada Balita Di Desa Tanjung Berulak Wilayah Kerja Puskesmas Kampar Tahun 2019. *Jurnal Kesehatan Tambusai*, 1(2), 61–68. <https://doi.org/10.32668/jkep.v3i2.205>.
- Z, M. Z. R., Muttakin, F., Zarnelly, & Permana, I. (2024). Sistem Pakar Untuk Mendiagnosa Gizi Buruk Pada Balita Menggunakan Metode Forward Chaining. *KLIK: Kajian Ilmiah Informatika Dan Komputer*, 4(5), 2567–2666. <https://doi.org/10.30865/klik.v4i5.1776>.