

Analisis Klusterisasi Stunting Pada Balita Menggunakan Algoritma K-Medoids Untuk Mengidentifikasi Faktor Dominan

Leonardo Saragih*, Nanda Sabrina Pasaribu, Novi Karlianti Harefa, Tajrin

Sains dan Teknologi, Sistem Informasi, Universitas Prima Indonesia, Medan, Indonesia

Email: ^{1,*}Leonardosaragih15000@gmail.com, ²nandapasaribu396@gmail.com, ³novikarliantiharefa133@gmail.com, ⁴tajrin@unprimdn.ac.id

Email Penulis Korespondensi: Leonardosaragih15000@gmail.com

Submitted 05-06-2025; Accepted 30-06-2025; Published 30-06-2025

Abstrak

Penelitian ini bertujuan untuk mengelompokkan desa di Kabupaten Langkat berdasarkan faktor risiko stunting pada balita menggunakan metode K-Medoids. Data yang dianalisis meliputi persentase balita stunting dan jumlah balita dari 277 desa, dengan fokus pada 100 balita di Puskesmas Sawit Seberang. Hasil analisis menunjukkan bahwa desa dapat dikelompokkan menjadi tiga kluster yang menunjukkan bahwa jumlah balita tidak selalu berkorelasi langsung dengan tingkat stunting. Faktor lain seperti akses layanan kesehatan dan kesadaran gizi turut berperan dalam menentukan risiko stunting. Temuan ini memberikan gambaran yang lebih komprehensif untuk merancang intervensi yang lebih efektif dan terarah. Disarankan penambahan variabel lain dan penggunaan metode analisis berbeda untuk penelitian selanjutnya guna meningkatkan akurasi dan manfaat hasil klusterisasi. Hasil penelitian ini diharapkan dapat membantu pemerintah dalam merancang program berbasis kluster untuk pencegahan dan penanganan stunting secara lebih optimal di Kabupaten Langkat.

Kata Kunci: Kluster Stunting; Balita; Algoritma K-Medoids; Identifikasi Faktor Dominan.

Abstract

This study aims to group villages in Langkat Regency based on stunting risk factors in toddlers using the K-Medoids method. The data analyzed includes the percentage of stunted toddlers and the number of toddlers from 277 villages, focusing on 100 toddlers at the Sawit Seberang Health Center. The analysis results show that villages can be divided into three clusters, indicating that the number of toddlers is not always directly related to the level of stunting. Other factors, such as access to health services and nutritional awareness, also play a role in determining the risk of stunting. These findings provide a more comprehensive picture for designing more effective and targeted interventions. It is recommended that other variables be added and different analysis methods be used for further research to improve the accuracy and benefits of clustering results. The results of this study are expected to assist the government in designing cluster-based programs for more optimal prevention and handling of stunting in Langkat Regency.

Keywords: Stunting Cluster; Toddler; K-Medoids Algorithm; Identification of Dominant Factors.

1. PENDAHULUAN

Survei Kesehatan Indonesia (SKI) 2023 yang menggabungkan Survei Status Gizi Balita Indonesia (SSGI) dan Riset Kesehatan Dasar (Riskesmas) memberikan informasi terkini tentang prevalensi stunting di Indonesia [1]. Secara umum, prevalensi stunting di Indonesia menunjukkan tren penurunan dibandingkan tahun-tahun sebelumnya. Namun, angka penurunan ini belum mencapai tingkat yang diinginkan, tercatat misalnya di provinsi Sumatera Utara prevalensi stunting sebesar 4,5% walaupun presentase ini tidak setinggi provinsi lainnya [2]. Kementerian kesehatan mengungkapkan bahwa salah satu penyebab lambatnya penurunan angka stunting adalah kurangnya model implementasi yang efektif dari program yang telah ditetapkan. Ia menekankan bahwa masalah eksekusi di lapangan perlu diperbaiki agar program pencegahan stunting dapat berjalan lebih optimal [3].

Dalam hal prevalensi stunting, Indonesia saat ini berada di peringkat keempat secara global, jauh di bawah target Organisasi Kesehatan Dunia (WHO). Bila tinggi badan seorang anak berada di bawah anak-anak seusianya, kondisi tersebut disebut stunting [4]. Di Indonesia, Dampak kekurangan gizi kronis pada pertumbuhan dan perkembangan anak tetap menjadi masalah yang signifikan. Anak dengan sindrom ini akan lebih pendek dari anak-anak seusianya. Selain mengganggu perkembangan fisik anak, stunting juga dapat berdampak pada produktivitas, kekebalan tubuh, dan perkembangan kognitif di masa mendatang [5].

Proses pengambilan data set yang dihasilkan secara algoritmik dari database atau manajemen statistik dikenal sebagai penambahan data [6]. Salah satu tekniknya, yaitu *clustering*, digunakan untuk menganalisis data besar dan mengidentifikasi pola baru yang bermanfaat. Clustering, regresi, asosiasi, dan klasifikasi merupakan beberapa metode yang digunakan dalam data mining untuk mengungkap informasi atau pola yang tersembunyi. K-medoids merupakan salah satu pengelompokan data mining dalam metode clustering. Ini merupakan teknik clustering yang mencoba mengelompokkan data ke dalam beberapa kelompok (*cluster*) berdasarkan kualitas yang serupa [7]. Partitioning Around Medoids (PAM) adalah nama lain untuk K-Medoids, berbeda dari metode clustering sebelumnya seperti K-Means karena memusatkan cluster di sekitar objek data asli, atau medoid.

Berdasarkan penelitian yang dilakukan oleh Melisa dkk., 2024 dengan judul "Implementasi Data Mining Untuk Klusterisasi Stunting Gizi Pada Balita Di Puskesmas Sigambal Menggunakan Metode K-Medoids Dan K-Means" dengan mempertimbangkan jarak terdekat dari centroid atau medoid, pendekatan K-Medoids secara efektif mengklasifikasikan balita menurut faktor-faktor yang berkontribusi terhadap stunting gizi. Pengetahuan mendalam tentang tren dan faktor

risiko yang memengaruhi stunting gizi diperoleh dari analisis ini, menunjukkan betapa kedua kelompok itu berbeda secara signifikan[8].

Penelitian Fadzly dkk.,2024 dengan judul “Klasterisasi kabupaten dan kota di Jawa barat dalam kasus gizi buruk menggunakan algoritma k-means dan k- medoids” Penelitian ini secara efektif membagi data anak kurang gizi di Provinsi Jawa Barat menjadi tiga kelompok menggunakan algoritma K-Means dan K-Medoids. Ketiga kelompok tersebut membentuk kelompok ideal, yaitu kelompok 1 (tinggi), kelompok 2 (sedang), dan kelompok 3 (rendah)[9].

Wiwid Wahyudi dkk.,2023 dengan judul “Implementasi data mining untuk klasifikasi stunting pada balita di Surabaya menggunakan metode K-Medoids” dalam penerapan metode PAM (*Partitioning Around Medoids*) Dua klaster, yaitu klaster 0 dan klaster 1, dikumpulkan di Puskesmas untuk menentukan penyuluhan stunting. Klaster 0 terdiri dari sembilan Puskesmas, sedangkan klaster 1 terdiri dari dua puluh tiga Puskesmas. Hasil uji klaster kemudian menunjukkan angka 0,272, yang menunjukkan klaster tersebut baik karena nilai DBI berada pada kisaran 0[10].

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif, yaitu mengukur, menganalisis, dan menarik kesimpulan tentang faktor-faktor yang berhubungan dengan stunting dengan menggunakan data numerik dan teknik statistik. Informasi yang digunakan dalam penelitian ini diperoleh dari setiap puskesmas di wilayah Langkat yang menyediakan data frekuensi yang dimoderasi untuk penelitian ini. Untuk mengelompokkan balita menggunakan metode k-medoids, data diolah dengan membagi persentase balita yang menderita stunting menjadi dua kelompok, yaitu kelompok tertinggi dan terendah. Hal ini memungkinkan identifikasi dan penentuan faktor-faktor utama yang berkontribusi terhadap stunting[11].

2.1 Sumber Data

Penelitian ini menggunakan data resmi stunting tahun 2023 dari Platform e-Monev Konvergensi Stunting Kementerian Dalam Negeri (<https://konvergensi.bangda.kemendagri.go.id/emonev/DashPrev>), dimana tahap awal pengolahan data meliputi: mengakses platform untuk memfilter data Provinsi Sumatera Utara → Kabupaten Langkat, mengunduh data dalam format asli kemudian mengkonversinya secara manual ke format Excel, serta menyimpan file hasil konversi sebagai Kab.Langkat.csv.

2.2 Populasi dan Sampel

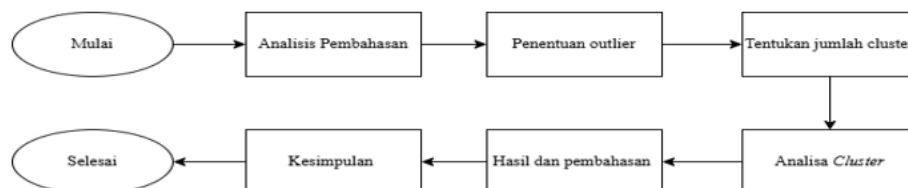
Populasi mencakup anak atau individu di puskesmas wilayah kabupaten langkat yang memenuhi kriteria penelitian. Pada konteks stunting biasanya dilakukan dengan teknik seperti pemilihan sampel secara acak, stratifikasi, atau klaster untuk menggambarkan populasi yang menunjukkan bagaimana pengetahuan gizi ibu mempengaruhi kejadian stunting pada anak[12].

2.3 Populasi dan Sampel

Instrumen yang digunakan dalam penelitian ini melibatkan beberapa metode untuk mengumpulkan dan menganalisis data dengan menggunakan DBI dengan bantuan *tools rapidminer* dalam evaluasi dan pengujian data.

2.4 Prosedur Penelitian

Tahap pelaksanaan penelitian dapat dilihat pada bagan alir berikut:



Gambar 1. Alur Penelitian

Analisis *Cluster* merupakan teknik analisis data yang bertujuan untuk mengelompokkan data ke dalam beberapa kelompok (*cluster*) berdasarkan fitur dominan dan karakteristik yang sebanding, *Flowchart* ini menunjukkan alur kerja untuk melakukan analisis.

- Mulai: dimulainya proses analisis.
- Analisis Permasalahan: Langkah awal ini melibatkan pemahaman mendalam tentang permasalahan atau tujuan yang ingin dicapai melalui analisis *cluster*.
- Penentuan Outlier: Identifikasi dan penanganan data outlier yang dapat memengaruhi hasil analisis. Outlier dapat dihilangkan, diubah, atau dianalisis secara terpisah.
- Tentukan Jumlah *Cluster*: Menetapkan jumlah *cluster* ideal menggunakan metode seperti *elbow*, *silhouette*, dan celah statistik.
- Analisa *Cluster*: Pengelompokan data dengan algoritma *clustering* (misalnya K-means, *hierarchical clustering*, atau DBSCAN) dan analisis karakteristik setiap *cluster*.

- f. Hasil dan Pembahasan: Penyajian hasil dalam bentuk visualisasi dan narasi, serta interpretasi temuan.
- g. Kesimpulan: Merangkum temuan utama dan menjawab pertanyaan penelitian.
- h. Selesai: Proses analisis *cluster* berakhir.

2.5 Data Mining

Data *Mining* adalah praktik penggunaan data untuk menemukan pola atau hubungan antar kelompok. Dengan menggunakan hasil proses data mining, keputusan harus dibuat. Data Mining adalah teknik untuk memeriksa dan mengevaluasi data dengan banyak catatan. Ini digunakan untuk melihat bagaimana data yang sebanding berhubungan satu sama lain dan untuk mengkompilasi data sehingga penilaian dapat dibuat[13].

2.6 Jarak Antar Objek

Terdapat sejumlah formula untuk mengukur jarak antara dua objek data yang memiliki atribut numerik. Jarak *Euclidean* adalah cara yang paling populer untuk mengukur jarak. Bila Anda ingin menentukan jarak terpendek (jarak lurus) antara dua tempat, Anda memanfaatkan jarak *Euclidean*. Persamaan untuk menghitung jarak *Euclidean* adalah sebagai berikut[14]:

$$d(x_i, x_j) = \sqrt{\sum_{p=1}^q (x_{ip} - x_{jp})^2} \quad (1)$$

Dimana:

$d(x_i, x_j)$: Jarak *Euclidean* data pengamatan ke- i dengan data pengamatan ke- j dimana $i, j, = 1, 2, 3, \dots, n$.

x_{ip} : data pengamatan ke- i pada variabel ke- p .

x_{jp} : data pengamatan ke- j pada variabel ke- p Dimana $p = 1, 2, 3, \dots, q$.

2.7 Klasterisasi

Analisis kluster (*cluster analysis*) adalah metode statistik multivariat yang bertujuan mengelompokkan objek atau variabel ke dalam *cluster* berdasarkan karakteristik yang dimilikinya. Metode ini termasuk dalam teknik *unsupervised learning* karena dapat digunakan untuk mengidentifikasi pola alami dalam data yang belum berlabel. Prinsip dasarnya adalah mengelompokkan objek-objek yang memiliki kemiripan tinggi ke dalam *cluster* yang sama[15]. Algoritma K-Medoids merupakan salah satu metode clustering yang lebih robust terhadap outlier dibanding K-Means. Algoritma ini bekerja dengan memilih medoid sebagai pusat cluster dan meminimalkan jarak total antar objek dan medoid yang terpilih [17][18][19]. Langkah-langkah dari algoritma K-Medoids adalah sebagai berikut:

- a. Tentukan Jumlah *cluster*.
- b. Pilih objek secara acak sebagai medoids (titik pusat *cluster*).
- c. Dengan menggunakan persamaan (2), dapatkan jarak Euclidean antara setiap item dan setiap medoid[16]:

$$d(x_{ip}, 0_{mp}) = \sqrt{(x_{i1} - 0_{m1})^2 + (x_{i2} - 0_{m2})^2 + \dots + (x_{iq} - 0_{mq})^2} \quad (2)$$

Dengan $d(x_{ip}, 0_{mp})$ adalah pemisahan, di mana $m=1, 2, \dots, K$ dan $p=1, 2, \dots, q$, antara data ke- i pada variabel ke- p dan medoid ke- m pada variabel ke- p .

- d. Tentukan fungsi objektif dengan menetapkan setiap objek ke dalam *cluster* yang bersesuaian dengan medoid terdekat yaitu dengan cara menjumlahkan seluruh kedekatan (jarak paling minimum) dari semua objek ke setiap medoids[17].
- e. Pilih objek observasi non-representatif 0_h (non-medoid) secara acak.
- f. Tentukan jarak Euclidean setiap objek ke setiap non-medoid menggunakan persamaan (3) sebagai panduan:

$$d(x_{ip}, 0_{hp}) = \sqrt{(x_{i1} - 0_{h1})^2 + (x_{i2} - 0_{h2})^2 + \dots + (x_{iq} - 0_{hq})^2} \quad (3)$$

Dengan $d(x_{ip}, 0_{hp})$ adalah pemisahan antara data ke- i dari variabel p dan non-medoid ke- h : dimana $p=1, 2, \dots, q$ dan $h=1, 2, \dots, n-K$

- g. Tentukan fungsi objektif dan tetapkan setiap objek ke *cluster* yang sesuai dengan non-medoid terdekat yaitu dengan cara menjumlahkan seluruh kedekatan (jarak paling minimum) dari semua objek ke setiap *non-medoids*[18].
- h. Kurangi fungsi objektif non-medoid dari fungsi objektif medoid untuk menentukan perbedaan fungsi objektif [19].
- i. Jika nilai fungsi objektif kurang dari nol, substitusikan medoid o dengan non-medoid o [20].
- j. Pengelompokan selesai jika didapatkan fungsi objektif lebih dari nol atau sudah tidak terdapat perubahan objek representatif.

3. HASIL DAN PEMBAHASAN.

3.1 Penerapan K-Medoids

Metode K-Medoids merupakan salah satu algoritma dalam klasterisasi yang digunakan untuk mengelompokkan data ke dalam beberapa klaster berdasarkan tingkat kesamaan antar data. Berbeda dengan K-Means yang menggunakan rata-rata (mean) sebagai pusat klaster, K-Medoids menggunakan medoid, yaitu objek aktual dari dataset yang memiliki jarak paling

kecil terhadap anggota lainnya dalam klaster tersebut. Hal ini membuat K-Medoids lebih tahan terhadap outlier dan noise, sehingga cocok digunakan untuk data seperti prevalensi stunting yang memiliki variasi cukup besar antar wilayah.

Dalam konteks ini, K-Medoids digunakan untuk mengelompokkan desa-desa di Kabupaten Langkat berdasarkan persentase stunting balita. Tujuan dari penerapan metode ini adalah untuk:

- Mengidentifikasi desa-desa dengan kategori stunting rendah, sedang, atau tinggi.
- Menjadi dasar pengambilan keputusan dalam menentukan prioritas intervensi program kesehatan masyarakat.

Sebelum melakukan proses klusterisasi, data terlebih dahulu diolah:

3.1.1 Persentase Stunting

Persentase stunting dihitung menggunakan rumus berikut:

$$\text{Persentase Stunting} = \frac{\text{Stunting Pendek} + \text{Stunting Sangat Pendek}}{\text{Jumlah Balita}} \times 100\%$$

Berdasarkan data dari Kabupaten Langkat, berikut adalah hasil perhitungan persentase stunting untuk beberapa desa (contoh):

Tabel 1. Persentase Stunting

Nama Desa	Jumlah Balita	Stunting Pendek	Stunting Sangat Pendek	Persentase Stunting (%)
Batu Jong Jong	160	0	3	$\frac{0 + 3}{160} \times 100\% = 1.9\%$
Lau Damak	162	0	6	$\frac{0 + 6}{162} \times 100\% = 3.7\%$
Timbang Lawan	306	7	17	$\frac{7 + 17}{306} \times 100\% = 7.8\%$
Bukit Lawang	229	18	14	$\frac{18 + 14}{229} \times 100\% = 14.0\%$
Timbang Jaya	278	33	35	$\frac{33 + 35}{278} \times 100\% = 24.5\%$
Alur Gadung	360	32	45	$\frac{32 + 45}{360} \times 100\% = 21.4\%$
Jaring Halus	339	39	48	$\frac{39 + 48}{339} \times 100\% = 25.7\%$

3.1.2 Normalisasi Data (Mix-Max Scalling)

Normalisasi data menggunakan metode Min-Max Scaling untuk menyamakan skala variabel, dimana nilai prevalensi stunting dan jumlah balita diubah ke rentang 0-1. Sebagai contoh, normalisasi prevalensi Desa Batu Jong Jong (1.9%) menghasilkan nilai 0.074 dengan rumus:

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}$$

Gunakan huruf kecil dan abjad untuk penomoran list.

Nilai Prevalensi : 1,9

$X_{\text{min}} = 0.0\%$, $X_{\text{max}} = 25,7\%$

$$X_{\text{norm}} = \frac{1.9 - 0.0}{25.7 - 0.0} = 0.074$$

3.1.3 Hitung Jarak Euclidean

Jarak Euclidean antara setiap desa dengan medoid dihitung untuk pembentukan cluster, misalnya jarak Desa Timbang Lawan ke medoid Batu Jong Jong adalah 0.230 berdasarkan rumus:

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

Contoh Perhitungan Jarak Desa “Timbang Lawan” ke Medoid 1 (Batu Jong Jong):

a. Data Timbang Lawan (Normalisasi):

$$\text{Prevalensi: } \frac{7.8}{25.7} = 0.303$$

$$\text{Jumlah Balita: } \frac{306}{\max(\text{Balita})} = 0.17 \text{ (asumsi normalisas)}$$

b. Medoid 1 (Batu Jong Jong)

Prevalensi: 0.074

Jumlah Balita: 0.064

c. Jarak Euclidean:

$$d = \sqrt{(0.303 - 0.074)^2 + (0.17 - 0.064)^2} \approx 0.254$$

3.2 Hasil Analisis Data

3.2.1 Data Stunting Kabupaten Langkat

Data ini berisi informasi kasus stunting di 277 desa di Kabupaten Langkat, mencakup jumlah balita, kasus stunting pendek, dan stunting sangat pendek. Prevalensi stunting bervariasi antar desa, dengan Bukit Lawang memiliki persentase tertinggi (14%) dan Limau Mungkur serta Perkebunan Damar Condong melaporkan 0% kasus stunting sangat pendek. Desa Timbang Lawan mencatat jumlah kasus stunting sangat pendek tertinggi (17 kasus) dari 306 balita. Data ini membantu mengidentifikasi desa dengan beban stunting berat, seperti Damar Condong dengan prevalensi 8.1%. Informasi ini berguna untuk program intervensi gizi dan kesehatan balita di wilayah prioritas:

No	ID BPS	ID Dagri	Nama Desa	Jumlah Balita	Stunting Pendek	Stunting Sangat Pendek	Prevalensi (%)	
0	1	1213010001	1205012014	Batu Jong Jong	160	0	3	1.9
1	2	1213010002	1205012002	Lau Damak	162	0	6	3.7
2	3	1213010003	1205012003	Timbang Lawan	306	7	17	7.8
3	4	1213010004	1205012004	Sampe Raya	275	0	2	0.7
4	5	1213010005	1205012009	Bukit Lawang	229	18	14	14.0
...
272	273	1213201004	1205232004	Perkebunan Perapen	75	0	1	1.3
273	274	1213201005	1205232005	Limau Mungkur	124	0	0	0.0
274	275	1213201006	1205232003	Perkebunan Damar Condong	146	0	0	0.0
275	276	1213201007	1205232002	Damar Condong	111	4	5	8.1
276	277	1213201008	1205232006	Serang Jaya Hilir	142	1	1	1.4

277 rows x 8 columns

Gambar 2. Data Stunting Kabupaten Langkat

3.2.2 Hasil Data Persentase Stunting

Data ini menunjukkan persentase stunting di 10 desa Kabupaten Langkat, dengan prevalensi bervariasi dari 0.7% hingga 14%. Desa Bukit Lawang memiliki tingkat stunting tertinggi (14%) dengan 18 kasus stunting pendek dan 14 kasus sangat pendek dari 229 balita. Sebaliknya, Sampe Raya mencatat prevalensi terendah (0.7%) dengan hanya 2 kasus stunting sangat pendek dari 275 balita. Beberapa desa seperti Perkebunan Turangi menunjukkan angka signifikan (8.4%) dengan 11 kasus sangat pendek. Data ini mengindikasikan perlunya intervensi khusus di desa-desa dengan prevalensi tinggi seperti Bukit Lawang dan Perkebunan Turangi:

Data Persentase Stunting:

No	Nama Desa	Jumlah Balita	Stunting Pendek	Stunting Sangat Pendek	Prevalensi (%)
0	Batu Jong Jong	160	0	3	1.9
1	Lau Damak	162	0	6	3.7
2	Timbang Lawan	306	7	17	7.8
3	Sampe Raya	275	0	2	0.7
4	Bukit Lawang	229	18	14	14.0
5	Perkebunan Bungara	256	0	4	1.6
6	Pekan Bahorok	236	1	7	3.4
7	Empus	186	3	9	6.5
8	Perkebunan Turangi	178	4	11	8.4
9	Simpang Pulau Rambung	161	0	2	1.2

Gambar 3. Hasil Data Persentase Stunting

3.2.3 Hasil Normalisasi

Data ini menampilkan hasil normalisasi dari prevalensi stunting dan jumlah balita di 10 desa Kabupaten Langkat. Nilai normalisasi prevalensi berkisar antara 0.027 hingga 0.545, dengan Bukit Lawang memiliki nilai tertinggi (0.545) yang mencerminkan tingkat stunting paling serius. Sementara itu, normalisasi jumlah balita menunjukkan distribusi populasi, di mana Timbang Lawan memiliki proporsi terbesar (0.173). Hasil normalisasi ini memudahkan perbandingan antar desa untuk analisis lebih lanjut:

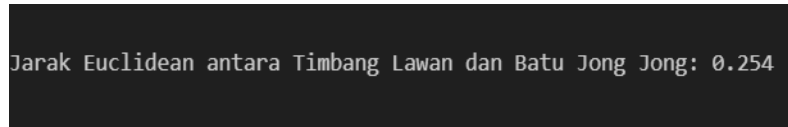
Data Setelah Normalisasi:

No	Nama Desa	Prevalensi (%)	Prevalensi Norm	Jumlah Balita	Jumlah Balita Norm
0	Batu Jong Jong	1.9	0.073930	160	0.063718
1	Lau Damak	3.7	0.143969	162	0.065217
2	Timbang Lawan	7.8	0.303502	306	0.173163
3	Sampe Raya	0.7	0.027237	275	0.149925
4	Bukit Lawang	14.0	0.544747	229	0.115442
5	Perkebunan Bungara	1.6	0.062257	256	0.135682
6	Pekan Bahorok	3.4	0.132296	236	0.120690
7	Empus	6.5	0.252918	186	0.083208
8	Perkebunan Turangi	8.4	0.326848	178	0.077211
9	Simpang Pulau Rambung	1.2	0.046693	161	0.064468

Gambar 4. Hasil Normalisasi

3.2.4 Hasil Jarak Euclidean

Jarak Euclidean sebesar 0.254 antara Timbang Lawan dan Batu Jong Jong menunjukkan tingkat kemiripan atau perbedaan berdasarkan data yang dinormalisasi sebelumnya. Nilai ini mengindikasikan bahwa kedua desa memiliki perbedaan signifikan dalam prevalensi stunting dan jumlah balita setelah diproses secara statistik. Analisis jarak seperti ini berguna untuk mengelompokkan desa dengan karakteristik serupa dalam penanganan stunting:

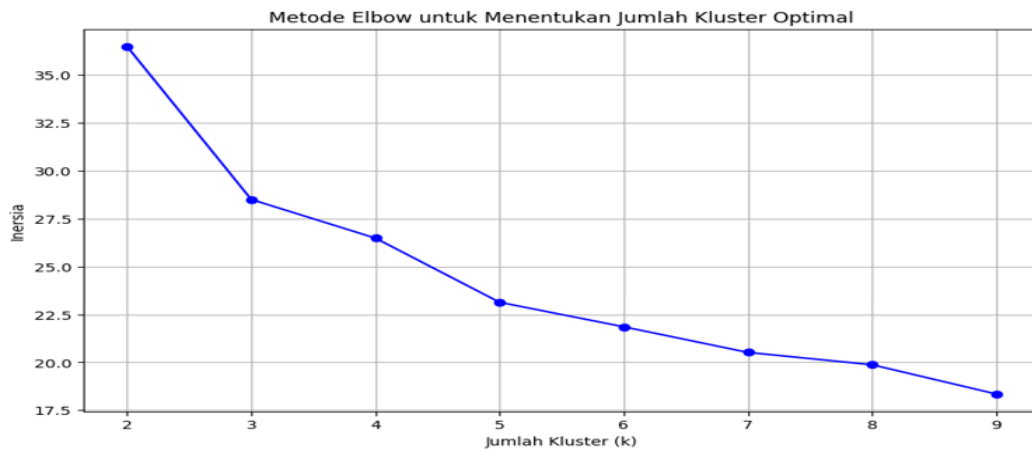


Gambar 5. Hasil Jarak Euclidean

3.2.5 Pemilihan Jumlah Kluster Optimal

a. Metode Elbow

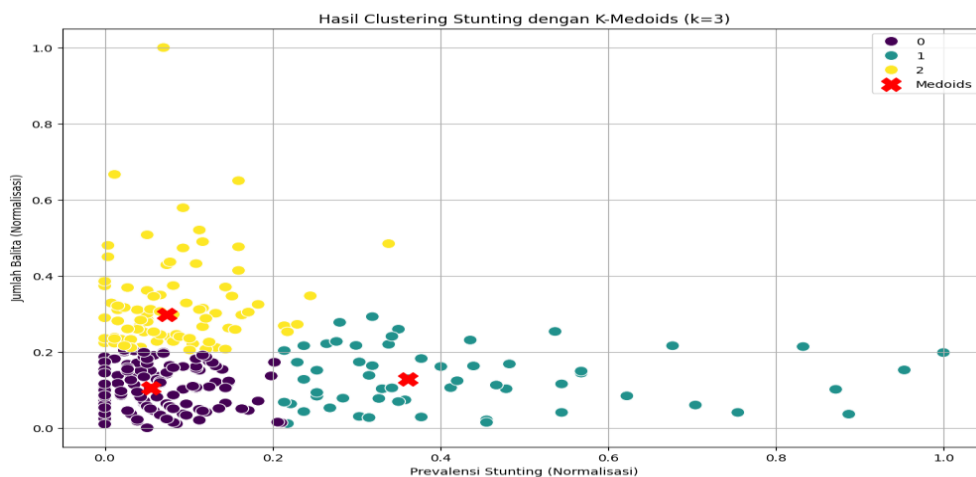
Grafik Elbow menunjukkan hubungan antara jumlah kluster (k) dengan inersia (variasi dalam kluster) untuk menentukan jumlah kelompok optimal. Inersia menurun tajam dari $k=2$ hingga $k=4$, kemudian penurunannya melambat setelah $k=4$, membentuk pola seperti siku (elbow). Titik optimal kemungkinan berada di $k=4$, di mana penambahan kluster tidak lagi memberikan penurunan inersia yang signifikan. Hasil ini menyarankan 4 kluster sebagai pilihan terbaik untuk pengelompokan data stunting di Kabupaten Langkat:



Gambar 6. Metode Elbow

b. Silhouette Score

Grafik clustering K-Medoids dengan $k=3$ menghasilkan Silhouette Score 0.410, menunjukkan tingkat pemisahan kluster yang cukup baik meski belum optimal. Grafik visualisasi menampilkan tiga kelompok desa yang terdistribusi berdasarkan jumlah balita (sumbu X) dan prevalensi stunting (sumbu Y) dalam bentuk nilai normalisasi. Pola pengelompokan ini memudahkan identifikasi desa dengan karakteristik serupa, seperti kluster dengan prevalensi tinggi (poin di atas sumbu Y) atau jumlah balita besar (poin di kanan sumbu X), untuk intervensi berbasis risiko.



Gambar 7. Silhouette Score

3.2.6 Hasil Clustering dengan K-Medoids

Hasil clustering membagi desa-desa menjadi 3 kelompok berdasarkan karakteristik prevalensi stunting dan jumlah balita. Kluster 0 mencakup desa seperti Pangkalan Siata (0.7%) dan Tanjung Mulia (4.4%), menunjukkan variasi prevalensi dengan jumlah balita menengah. Kluster 1 berisi desa seperti Paya Tusan (5.5%) dan Besilam (2.6%), sementara Kluster 2 didominasi desa dengan jumlah balita besar seperti Air Hitam (0.4%) dan Pulau Kampai (5.6%), mengindikasikan kebutuhan strategi berbeda untuk tiap kelompok:

Contoh Desa dari Setiap Kluster:

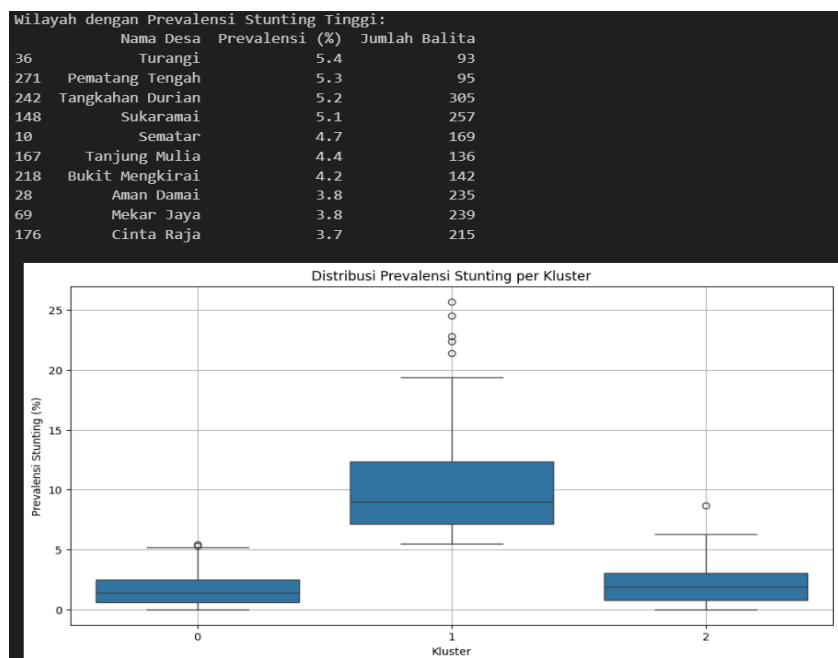
Kluster 0:			
	Nama Desa	Prevalensi (%)	Jumlah Balita
258	Pangkalan Siata	0.7	287
167	Tanjung Mulia	4.4	136
271	Pematang Tengah	5.3	95

Kluster 1:			
	Nama Desa	Prevalensi (%)	Jumlah Balita
129	Paya Tusan	5.5	165
...			
152	Besilam	2.6	390
210	Air Hitam	0.4	489
268	Pulau Kampai	5.6	412

Gambar 8. Hasil Clustering dengan K-Medoids

3.2.7 Analisis Faktor Dominan Berdasarkan Kluster

Data ini menampilkan 10 desa dengan prevalensi stunting tertinggi (3,7%-5,4%) di Kabupaten Langkat, di mana Turangi (5,4%) dan Pematang Tengah (5,3%) merupakan wilayah prioritas utama yang memerlukan intervensi segera. Terdapat variasi signifikan dalam jumlah balita (93-305) pada desa-desa tersebut, menunjukkan bahwa tingkat keparahan stunting tidak selalu berkorelasi dengan besarnya populasi, seperti terlihat pada Tangkahan Durian (305 balita) dan Sukaramai (257 balita) yang membutuhkan pendekatan khusus. Analisis kluster yang disajikan dapat mengidentifikasi pola pengelompokan desa berdasarkan karakteristik prevalensi dan jumlah balita, memungkinkan intervensi yang lebih terarah dan efektif. Hasil ini memberikan dasar penting bagi penyusunan strategi penanganan stunting yang tepat sasaran, khususnya untuk desa-desa dengan beban kasus tinggi dan populasi balita besar:



Gambar 9. Analisis Faktor Dominan Berdasarkan Kluster

3.3 Pembahasan

3.3.1 Interpretasi Hasil Clustering K-Medoids

Hasil clustering dengan metode K-Medoids berhasil mengelompokkan desa-desa di Kabupaten Langkat menjadi 3 kluster berdasarkan prevalensi stunting dan jumlah balita. Kluster 0 terdiri dari desa dengan prevalensi rendah hingga menengah dan jumlah balita menengah, seperti Pangkalan Siata (0.7%) dan Tanjung Mulia (4.4%). Kluster 1 mencakup desa dengan prevalensi menengah seperti Paya Tusan (5.5%), sedangkan Kluster 2 didominasi oleh desa dengan jumlah balita besar

seperti Air Hitam (0.4%). Pengelompokan ini menunjukkan bahwa faktor jumlah balita tidak selalu berkorelasi langsung dengan tingkat stunting, sehingga intervensi harus disesuaikan dengan karakteristik masing-masing kluster.

3.3.2 Implikasi Kebijakan Berdasarkan Analisis Kluster

Temuan ini memberikan dasar untuk kebijakan yang lebih terarah. Desa dalam Kluster 2, meskipun memiliki jumlah balita besar, tidak selalu memiliki prevalensi stunting tinggi. Hal ini mengindikasikan bahwa faktor lain seperti akses layanan kesehatan atau kesadaran gizi mungkin berperan. Sementara itu, desa dalam Kluster 1 dengan prevalensi menengah memerlukan program pencegahan seperti edukasi gizi dan pemeriksaan kesehatan rutin. Untuk desa dengan prevalensi tinggi (beberapa di Kluster 0), intervensi intensif seperti suplementasi gizi dan pelatihan kader kesehatan menjadi prioritas.

3.3.3 Keterbatasan dan Rekomendasi Untuk Penelitian Lanjutan

Penelitian ini memiliki keterbatasan, seperti hanya menggunakan dua variabel (prevalensi stunting dan jumlah balita) untuk clustering. Variabel tambahan seperti kondisi ekonomi, geografis, atau ketersediaan fasilitas kesehatan dapat memperkaya analisis. Selain itu, metode K-Medoids dipilih karena ketahanannya terhadap outlier, tetapi algoritma lain seperti DBSCAN dapat diuji untuk membandingkan hasil. Rekomendasi untuk penelitian selanjutnya mencakup eksplorasi faktor-faktor dominan di setiap kluster secara kualitatif serta uji coba intervensi berbasis kluster untuk mengukur efektivitasnya.

4. KESIMPULAN

Penelitian ini menggunakan algoritma K-Medoids untuk mengelompokkan 277 desa di Kabupaten Langkat menjadi 3 kluster berdasarkan prevalensi stunting dan jumlah balita. Hasilnya menunjukkan bahwa jumlah balita tidak selalu berkorelasi langsung dengan tingkat stunting, dan faktor lain seperti akses layanan kesehatan serta kesadaran gizi turut berperan. Temuan ini dapat menjadi dasar bagi pemerintah untuk menyusun intervensi yang lebih terarah, meskipun penelitian ini memiliki keterbatasan dalam cakupan variabel yang dianalisis.

REFERENCES

- [1] Herlena, A. C. P. (2023). Implementasi Data Mining Untuk Klasifikasi Stunting Gizi Pada Balita di Surabaya Menggunakan Metode K-Medoids. *Jurnal Publikasi Teknik Informatika*, 2(1), 61-67.
- [2] عبدالمنعم احمد جاسم الجنابي. "نظام تقويمي لمستوى القدرات الحركية لتلاميذ الصفوف (1، 2، 3) الابتدائي بطبني التعلّم." *الثقافة والتعليم* and حسين فهمي سليمان. *الرياضية* 15.1 (2024): 86-72.
- [3] Pohan, H., Zarlis, M., Irawan, E., Okprana, H., & Pranayama, Y. (2021). Penerapan Algoritma K-Medoids dalam Pengelompokan Balita Stunting di Indonesia. *JUKI: Jurnal Komputer dan Informatika*, 3(2), 97-104.
- [4] Apriyani, P., Dikananda, A. R., & Ali, I. (2023). Penerapan Algoritma K-Means dalam Klusterisasi Kasus Stunting Balita Desa Tegalwangi. *Hello World Jurnal Ilmu Komputer*, 2(1), 20–33. <https://doi.org/10.56211/helloworld.v2i1.230>.
- [5] Hidayat, F. P., Putra, R. P., Alfitrah, M. D., & Widodo, E. (2023). Implementasi Clustering K-Medoids dalam Pengelompokan Kabupaten di Provinsi Aceh Berdasarkan Faktor yang Mempengaruhi Kemiskinan. *Indonesian Journal of Applied Statistics*, 5(2), 121. <https://doi.org/10.13057/ijas.v5i2.55080>.
- [6] Puspita, I., Hayati, M. N., & Nohe, D. A. (2023). Pengelompokan Puskesmas Berdasarkan Kasus Balita Stunting di Kabupaten Paser Menggunakan Metode K-Medoids. *EKSPONENSIAL*, 14(1), 1-10.
- [7] Ranjawali, R., Carmen Talakua, A., Thimotius Abineno, R., Wira Wacana Sumba JIRSuprato No, K., Waingapu, K., Sumba Timur, K., & Tenggara Timu, N. (2023). *Universitas Kristen Wira Wacana Sumba Fakultas Sains dan Teknologi SATI: Sustainable Agricultural Technology Innovation CLUSTERING STUNTING PADA BALITA DENGAN METODE K-MEANS DI PUSKESMAS KANATANG*.
- [8] A. R. Cahyani, E. Riga Puspita, L. A. Pratiwi, and M. Muhajir, "Pengelompokan Stunting Menggunakan Metode K-Medoids Di Daerah Istimewa Yogyakarta (Diy)," 2024.
- [9] Samudi, S., Widodo, S., & Brawijaya, H. (2020). The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic. *Sinkron*, 5(1), 116. <https://doi.org/10.33395/sinkron.v5i1.10649>.
- [10] Hermanto, E. M. P., Rochmanto, H. B., & Agustini, R. (2023). Pemetaan Program Indonesia Sehat dengan Pendekatan Keluarga (PIS PK) di Kabupaten Bondowoso dengan K-Medoids. *Jurnal Statistika Dan Komputasi*, 2(2), 83–92. <https://doi.org/10.32665/statkom.v2i2.2307>.
- [11] Maulana, P. R., Hananto, A. L., Hananto, A., & Priyatna, B. (2024). Klusterisasi Tingkat Penjualan Kedai Kopi Hallo Burjois Menggunakan Algoritma K-Medoids Sebagai Evaluasi. *JURNAL FASILKOM*, 14(1), 226-233.
- [12] Meiriza, A., Lestari, E., Putra, P., Oktadini, N. R., & Audya, M. (2022). Pemanfaatan Data Mining Dalam Penentuan Penyuluhan Penyakit Stunting Menggunakan Partitioning Around Medoids (PAM). *TeKa*, 12(02), 89-96.
- [13] E. Engineering, "Penerapan Algoritma Clustering K-Medoids Untuk Menentukan Status Gizi Balita Application of the K-Medoids Clustering Algorithm to Determine the Nutritional Status of Toddlers," vol. 7, pp. 94–99, 2025.
- [14] Nikmah, A., Nisa, C., & Riefky, M. (2025). Penerapan Algoritma K-Medoids untuk Pengelompokan Provinsi di Indonesia Berdasarkan Status Gizi Anak Balita. *Emerging Statistics and Data Science Journal*, 3(1).
- [15] Sari, D. P. (2023). Implementasi Algoritma K-Medoids dan Aplikasi RapidMiner dalam Pengelompokan Kasus BALITA

- Stunting. *Brahmana: Jurnal Penerapan Kecerdasan Buatan*, 5(1), 1-8..
- [16] Taufik Hidayat, Mohamad Jajuli, & Susilawati. (2023). Clustering daerah rawan stunting di Jawa Barat menggunakan algoritma K-Means. *INFOTECH: Jurnal Informatika & Teknologi*, 4(2), 137–146. <https://doi.org/10.37373/infotech.v4i2.642>.
- [17] Schubert, E., & Rousseeuw, P. J. (2021). Fast and eager k-medoids clustering: O(k) runtime improvement of the PAM, CLARA, and CLARANS algorithms. *Information Systems*, 101. <https://doi.org/10.1016/j.is.2021.101804>.
- [18] Tiwari, M., Kang, R., Lee, D., Thrun, S., Piech, C., Shomorony, I., & Zhang, M. J. (2023). BanditPAM++: Faster k-Medoids Clustering. *arXiv preprint arXiv:2310.18844*. <https://doi.org/10.48550/arXiv.2310.18844>.
- [19] de Mathelin, A., Cecchi, N. E., Deheeger, F., Mougéot, M., & Vayatis, N. (2025). *OneBatchPAM: A Fast and Frugal K-Medoids Algorithm*. <http://arxiv.org/abs/2501.19285>.