



## Segmentasi Wilayah Provinsi di Indonesia Berdasarkan Indeks Penanganan Stunting Menggunakan PCA dan Partition Clustering

Farhan Syah Putra Wiyono<sup>1</sup>, Laisal Kaffi<sup>2</sup>, Mohammad Hikmal Maulana<sup>3</sup>, Aviolla

Terza Damaliana<sup>4</sup>, Shindi Sella May Wara<sup>5</sup>

<sup>1,2,3,4,5</sup>Afiliasi pertama (Sains Data, UPN “Veteran” Jawa Timur)

<sup>1</sup>[23083010088@student.upnjatim.ac.id](mailto:23083010088@student.upnjatim.ac.id)

<sup>2</sup>[23083010058@student.upnjatim.ac.id](mailto:23083010058@student.upnjatim.ac.id)

<sup>3</sup>[23083010053@student.upnjatim.ac.id](mailto:23083010053@student.upnjatim.ac.id)

<sup>4</sup>[aviolla.terza.sada@upnjatim.ac.id](mailto:aviolla.terza.sada@upnjatim.ac.id)

<sup>5</sup>[shindi.shella.fasilkom@upnjatim.ac.id](mailto:shindi.shella.fasilkom@upnjatim.ac.id)

**Abstract:** Stunting is a public health issue that has long-term impacts on children's cognitive development, productivity, and quality of life. This study aims to cluster provinces in Indonesia based on stunting-related indicators using two partition clustering methods: K-Means and K-Medoids. The analysis involved nine indicators, including immunization coverage, delivery assistance by health professionals, use of modern contraception, exclusive breastfeeding, access to safe drinking water, proper sanitation, food insecurity rate, early childhood education (PAUD) participation, and the percentage of young children from poor families. Dimensionality reduction was performed using Principal Component Analysis (PCA) to improve clustering efficiency and visualization. The clustering performance was evaluated using four internal metrics: Silhouette Coefficient, Davies-Bouldin Index, Calinski-Harabasz Index, and Dunn Index. The results showed that K-Means produced more optimal segmentation than K-Medoids, forming two main clusters. The first cluster includes provinces with strong performance across the indicators, while the second cluster represents provinces with lower performance and higher stunting vulnerability. These findings can serve as a basis for the government to design more targeted stunting interventions tailored to the specific characteristics of each region.

**Keywords:** Stunting, Clustering, PCA, K-Means, K-Medoids

**Abstrak:** Stunting merupakan masalah kesehatan masyarakat yang berdampak jangka panjang terhadap perkembangan kognitif, produktivitas, dan kualitas hidup anak. Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator penanganan stunting menggunakan metode partition clustering K-Means dan K-Medoids. Sembilan indikator digunakan dalam analisis, meliputi cakupan imunisasi dasar lengkap, persalinan oleh tenaga kesehatan, penggunaan kontrasepsi modern, cakupan ASI eksklusif, akses air minum layak, sanitasi layak, tingkat ketidakcukupan konsumsi pangan, partisipasi anak usia dini dalam PAUD, serta persentase anak usia dini dari keluarga miskin. Reduksi dimensi dilakukan menggunakan Principal Component Analysis (PCA) untuk meningkatkan efisiensi dan visualisasi kluster. Evaluasi performa kluster dilakukan dengan empat metrik internal: Silhouette Coefficient, Davies-Bouldin Index, Calinski-Harabasz Index, dan Dunn Index. Hasil analisis menunjukkan bahwa metode K-Means menghasilkan segmentasi yang lebih optimal dibandingkan K-Medoids, dengan pembentukan dua kluster utama. Kluster pertama terdiri dari wilayah-wilayah dengan capaian indikator yang baik, sementara kluster kedua mencerminkan wilayah dengan capaian rendah dan tingkat kerentanan stunting yang lebih tinggi. Hasil ini dapat menjadi dasar bagi pemerintah dalam merancang intervensi stunting yang lebih terarah dan sesuai dengan karakteristik masing-masing wilayah.

**Kata kunci:** Stunting, Klusterisasi, PCA, K-Means, K-Medoids

### I. PENDAHULUAN

Peningkatan kualitas sumber daya manusia merupakan fondasi utama dalam membangun bangsa yang berdaya saing tinggi. Namun, masih banyak tantangan dalam upaya ini, salah satunya adalah persoalan kesehatan dan gizi anak sejak usia dini. Di antara berbagai masalah yang ada, stunting menjadi perhatian utama karena dampaknya yang signifikan terhadap pertumbuhan dan perkembangan anak secara jangka panjang. Stunting, atau yang dikenal sebagai sindrom gagal tumbuh merupakan kondisi ketika pertumbuhan fisik anak terutama tinggi badan berada di bawah rata-rata anak seusianya [1]. Masalah ini berkembang akibat kekurangan gizi kronis yang dialami sejak masa kehamilan hingga anak berusia dua tahun, serta dipengaruhi oleh berbagai faktor seperti kesehatan ibu, pola asuh, dan lingkungan yang tidak mendukung [2]. Selain itu, stunting juga menjadi indikator penting dari masalah



gizi dan kesehatan masyarakat, karena dapat berdampak jangka panjang terhadap perkembangan kognitif, produktivitas, dan kualitas hidup anak di masa depan.

Pada tahun 2024, prevalensi stunting nasional tercatat sebesar 19,8% [3], sebagaimana diumumkan secara resmi oleh Kementerian Kesehatan RI pada 26 Mei 2025. Angka ini menunjukkan penurunan dibandingkan tahun sebelumnya (21,5% pada 2023) dan berhasil melampaui target tahunan sebesar 20,1%. Namun demikian, angka tersebut masih tergolong tinggi apabila dibandingkan dengan target jangka panjang pemerintah, yaitu 14,2% pada tahun 2029 dan 5% pada tahun 2045 sebagaimana tercantum dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN). Capaian ini mengindikasikan adanya kemajuan, tetapi sekaligus menegaskan pentingnya strategi intervensi yang lebih terarah dan berbasis data.

Untuk memahami faktor-faktor yang memengaruhi prevalensi stunting, penelitian ini menggunakan sembilan indikator utama yang mencerminkan kondisi kesehatan ibu dan anak, lingkungan tempat tinggal, serta status sosial ekonomi. Indikator tersebut meliputi cakupan imunisasi dasar lengkap, persentase persalinan oleh tenaga kesehatan, penggunaan kontrasepsi modern, cakupan ASI eksklusif, akses terhadap air minum dan sanitasi layak, tingkat ketidacukupan pangan, partisipasi PAUD bagi anak usia dini, serta persentase anak usia dini dari keluarga miskin. Keragaman nilai dari indikator-indikator ini antarprovinsi mencerminkan adanya ketimpangan wilayah dalam penanganan stunting.

Dalam konteks tersebut, pendekatan klusterisasi menjadi metode yang relevan untuk mengelompokkan wilayah berdasarkan kemiripan karakteristik indikator. Dengan melakukan klustering, dapat diidentifikasi pola-pola regional yang khas, sehingga intervensi kebijakan dapat disesuaikan dengan kebutuhan spesifik tiap daerah. Tujuan utama penelitian ini adalah untuk mengelompokkan 34 provinsi di Indonesia berdasarkan indikator penanganan stunting, guna mendukung penyusunan strategi penurunan stunting yang lebih adil dan efektif. Penelitian ini juga bertujuan mengevaluasi metode klusterisasi mana yang paling optimal dalam menghasilkan segmentasi wilayah yang bermakna dan dapat dijadikan dasar perumusan kebijakan.

Penelitian ini merujuk pada studi sebelumnya [4] yang membandingkan empat metode klusterisasi—Hierarchical Clustering, K-Means, K-Medoids, dan Fuzzy C-Means—dalam pengelompokan provinsi di Indonesia berdasarkan 12 indikator IKPS tahun 2018. Hasilnya menunjukkan bahwa metode K-Means memberikan performa terbaik dibandingkan metode lainnya. Oleh karena itu, penelitian ini juga mempertimbangkan penggunaan K-Means dalam mengelompokkan provinsi berdasarkan indikator terkini tahun 2023, sekaligus mengkaji keunggulannya dibandingkan metode lainnya dalam konteks data terbaru.

Berbeda dengan penelitian [5] yang menggunakan data tahun 2018 dan membandingkan berbagai metode klusterisasi, penelitian ini memfokuskan pada dua metode *partition clustering*, yaitu K-Means dan K-Medoids, serta menggunakan data IKPS tahun 2023 yang terbaru. Selain itu, penelitian ini menyesuaikan indikator yang digunakan dengan membaginya berdasarkan dimensi, yang terdiri dari dimensi kesehatan (cakupan imunisasi dasar lengkap, persalinan oleh tenaga kesehatan, dan keluarga berencana), dimensi gizi (cakupan ASI eksklusif), dimensi perumahan (akses air minum layak dan sanitasi layak), dimensi pangan (tingkat ketidacukupan konsumsi pangan), dan dimensi pendidikan (pendidikan anak usia dini). Penelitian ini juga menambahkan variabel baru berupa persentase anak usia dini dari keluarga miskin untuk menggantikan dimensi perlindungan sosial, sehingga hasil klusterisasi diharapkan dapat merepresentasikan kondisi nyata masing-masing wilayah dan memberikan dasar yang kuat dalam menetapkan prioritas intervensi serta arah kebijakan percepatan penanganan stunting secara lebih tepat sasaran.

## II. METODE PENELITIAN

### 2.1 Principal Component Analysis (PCA)

PCA membentuk sejumlah komponen utama (*principal components*) yang merupakan kombinasi linier dari variabel-variabel awal dan saling bebas (tidak berkorelasi). Dengan kata lain, PCA mengubah



sistem koordinat awal ke sistem baru yang lebih ringkas namun tetap mewakili informasi data asli secara maksimal. Metode ini banyak diterapkan dalam pengolahan data spasial dan kesehatan masyarakat karena kehandalannya dalam menangani korelasi antar variabel dan menyederhanakan struktur data yang kompleks.

Langkah-Langkah PCA dalam Penelitian Ini:

1. Normalisasi data dilakukan menggunakan teknik Min-Max Scaling untuk menyamakan skala antar indikator dan mencegah dominasi satu variabel terhadap yang lain. Rumus Min-Max Scaling untuk setiap nilai variabel  $x$  :

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

$x$  = nilai asli dari variabel

$x'$  = nilai yang telah dinormalisasi

$x_{min}$  = nilai minimum dari variabel tersebut

$x_{max}$  = nilai maksimum dari variabel tersebut

2. Dibentuk matriks kovarians antar variabel untuk mengukur hubungan linier.
3. Dilakukan dekomposisi nilai eigen dari matriks kovarians untuk memperoleh eigenvalue dan eigenvector, yang menjadi dasar pembentukan komponen utama.

Rumus Matriks Kovarians

$$\Sigma = \frac{1}{n - 1} (X^T X)$$

$X$  = matriks data terstandarisasi

$\Sigma$  = matriks kovarians ukuran  $p \times p$

4. Komponen utama disusun berdasarkan nilai eigen terbesar, karena nilai tersebut mencerminkan seberapa besar variasi yang dijelaskan oleh setiap komponen.

Sebagaimana dicontohkan oleh Prayoga dan Pramana, penggunaan PCA dalam tahap awal segmentasi wilayah memungkinkan model clustering bekerja lebih optimal dan mengurangi bias akibat redundansi antar indikator [6].

## 2.2. K-Means Clustering

Setelah dilakukan reduksi dimensi menggunakan PCA, tahap selanjutnya dalam penelitian ini adalah melakukan pengelompokan (clustering) terhadap provinsi-provinsi di Indonesia berdasarkan indeks penanganan stunting menggunakan algoritma K-Means. K-Means merupakan salah satu metode partition clustering yang paling populer dan banyak digunakan karena kesederhanaan dan efisiensinya. Algoritma ini bertujuan untuk membagi sekumpulan objek ke dalam  $k$  cluster berdasarkan kedekatan objek dengan pusat cluster (centroid), menggunakan jarak Euclidean sebagai ukuran kemiripan. Hasilnya adalah sekumpulan cluster yang memiliki internal homogenitas tinggi dan antar cluster yang heterogen [6].

Langkah-Langkah Algoritma K-Means dalam Penelitian ini:

1. Menentukan jumlah cluster ( $k$ ) awal.
2. Menginisialisasi centroid awal secara acak dari data.
3. Menghitung jarak setiap objek ke centroid menggunakan rumus Euclidean:



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

$d(x, c)$  = jarak antara objek data  $x$  dan centroid  $c$

$x_i$  = nilai atribut ke  $-i$  objek data

$c_i$  = nilai atribut ke  $-i$  centroid

$n$  = jumlah atribut (dimensi data)

4. Menempatkan objek ke cluster dengan jarak terdekat.
5. Menghitung ulang centroid berdasarkan rata-rata semua anggota dalam cluster.
6. Mengulangi proses iteratif (langkah 3–5) hingga posisi centroid tidak berubah atau perubahan sangat kecil (konvergen).

Dalam penelitian ini, K-Means digunakan untuk mengelompokkan provinsi berdasarkan skor komponen utama hasil PCA, sehingga menghasilkan segmentasi wilayah penanganan stunting yang berbasis data.

### 2.3 K-Medoids Clustering

Setelah dilakukan segmentasi data menggunakan metode K-Means, penelitian ini juga menerapkan algoritma K-Medoids Clustering sebagai metode pembandingan. K-Medoids merupakan algoritma partition-based clustering. Algoritma K-Medoids cenderung menghasilkan kluster yang lebih stabil dan akurat pada data spasial dengan distribusi tidak simetris [7]. Langkah-Langkah Algoritma K-Medoids dalam Penelitian ini:

1. Menentukan jumlah cluster ( $k$ ) yang diinginkan.
2. Memilih  $k$  objek secara acak dari dataset sebagai medoid awal.
3. Mengelompokkan objek ke medoid terdekat berdasarkan jarak menggunakan rumus Manhattan Distance:

$$d(x, m) = \sum_{i=1}^n |x_i - m_i|$$

$d(x, m)$  = jarak antara objek  $x$  dan medoid  $m$

$x_i, m_i$  = nilai atribut ke  $-i$  dari objek dan medoid

$n$  = jumlah atribut

4. Menghitung total biaya (total dissimilarity) untuk seluruh cluster.
5. Untuk setiap cluster, lakukan pertukaran medoid dengan objek non-medoid dalam cluster yang sama, lalu hitung total biaya baru.
6. Jika total biaya berkurang, maka pertukaran medoid diterima.
7. Iterasi berlanjut hingga tidak ada perubahan medoid atau hingga jumlah iterasi maksimum tercapai.

### 2.4 Evaluasi dan Validasi Hasil Clustering

Untuk menilai kualitas hasil pengelompokan yang dilakukan oleh algoritma K-Means dan K-Medoids, penelitian ini menggunakan tiga indeks evaluasi internal, yaitu: Silhouette Coefficient, Davies-Bouldin Index (DBI), dan Calinski-Harabasz Index (CHI). Ketiga indeks ini digunakan untuk menilai seberapa baik struktur cluster yang terbentuk berdasarkan hasil PCA dan algoritma clustering yang digunakan. Menurut Dzikrillah et al [8], kombinasi tiga metrik ini memberikan “gambaran kuantitatif mengenai kepadatan dalam cluster dan keterpisahan antar cluster, sehingga mendukung keputusan jumlah cluster yang optimal”.



## 1. Silhouette Coefficient (SC)

Silhouette Coefficient mengukur seberapa mirip suatu objek dengan anggota cluster-nya sendiri dibandingkan dengan cluster lain. Nilai mendekati 1 menandakan bahwa objek ditempatkan dalam cluster yang tepat, sedangkan nilai mendekati -1 menunjukkan potensi salah pengelompokan. Rumus Silhouette untuk objek  $i$  adalah:

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

$a(i)$  = rata-rata jarak objek  $i$  ke seluruh objek dalam cluster yang sama

$b(i)$  = rata-rata jarak objek  $i$  ke objek-objek dalam cluster terdekat berikutnya

Tyas *et al.* [9] menyatakan bahwa nilai Silhouette di atas 0,5 sudah menunjukkan pemisahan klaster yang baik.

## 2. Davies-Bouldin Index (DBI)

Davies-Bouldin Index mengukur kemiripan antar cluster dengan mempertimbangkan jarak antara pusat cluster dan sebaran internalnya. Nilai DBI yang lebih rendah menandakan struktur cluster yang lebih baik dan pemisahan yang lebih jelas antar kelompok. Suraya dan Wijayanto [10] menyebut bahwa DBI merupakan indikator penting dalam mengevaluasi efektivitas pemisahan antar klaster spasial. Nilai DBI yang rendah menunjukkan bahwa klaster terbentuk secara kompak dan terpisah dengan baik, sedangkan nilai yang tinggi mengindikasikan adanya tumpang tindih atau ketidakteraturan dalam pembentukan klaster.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{s_i + s_j}{d_{ij}} \right)$$

$k$  = jumlah klaster

$s_i$  = rata-rata jarak antara setiap titik dalam klaster  $i$  dengan centroid-nya

$d_{ij}$  = jarak antara centroid klaster  $i$  dan  $j$

$\frac{s_i + s_j}{d_{ij}}$  = rasio kesamaan antar klaster  $i$  dan  $j$

## 3. Calinski-Harabasz Index (CHI)

Calinski-Harabasz Index, dikenal sebagai Variance Ratio Criterion, membandingkan varians antar cluster dengan varians dalam cluster. Semakin tinggi nilai menunjukkan pengelompokan yang lebih baik. Indra *et al.* [11], dalam studi tentang pengelompokan daerah risiko stunting, menyimpulkan bahwa nilai CHI yang tinggi menunjukkan stabilitas antar kelompok serta adanya variasi data yang terpisah dengan baik.

## 4. Dunn Index

Dunn Index merupakan indeks yang berfokus pada rasio antara jarak minimum antar cluster (inter-cluster distance) dan jarak maksimum dalam cluster (intra-cluster distance). Tujuan utama dari metrik ini adalah untuk memaksimalkan pemisahan antar cluster dan meminimalkan penyebaran dalam cluster, sehingga semakin besar nilai Dunn Index, maka semakin baik kualitas struktur cluster yang terbentuk. Menurut Rudianto dan Wijayanto, “Dunn Index efektif digunakan



sebagai pembanding antar algoritma clustering karena ia mempertimbangkan kedekatan dan kepadatan cluster secara bersamaan” [12]

## 2.5. Data dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Republik Indonesia. Publikasi resmi melalui Laporan Indeks Khusus Penanganan Stunting (IKPS) 2022–2023 dan Profil Anak Usia Dini 2023, yang dapat diakses di situs web [13], [14]. Data ini untuk tahun 2022 mencakup 34 provinsi Indonesia. Faktor-faktor yang digunakan dalam publikasi ini termasuk imunisasi, penolong persalinan oleh tenaga kesehatan, penggunaan kontrasepsi modern, cakupan ASI eksklusif, akses air minum dan sanitasi yang layak, ketidakcukupan konsumsi pangan, tingkat partisipasi kasar dalam PAUD pada usia 3–6 tahun, dan anak-anak usia dini yang berasal dari keluarga miskin. Semua variabel ditunjukkan dalam bentuk persentase, dan nilai kuantitatifnya diberikan dalam bentuk angka.

## III. HASIL DAN PEMBAHASAN

### 3.1 Statistika Deskriptif

Pada tabel 1 berikut ringkasan statistik deskriptif dari sembilan variabel yang digunakan dalam analisis. Faktor-faktor berikut termasuk aspek kesehatan anak dan ibu, sanitasi, dan gizi, serta akses ke layanan dasar lainnya yang berkaitan dengan penanganan stunting.

**Tabel 1.** Statistika Deskriptif

Variabel	Mean	Standar Deviasi	Minimum	Maksimum
Imunisasi	27.5	14.47	27.5	93.9
Penolong persalinan	88.42	10.87	52.6	100
Keluarga Berencana (KB) modern	62.82	14.52	18.8	83.2
ASI eksklusif	87.91	9.44	68.9	100
Air minum layak	88.18	7.75	66.5	99.4
Sanitasi layak	82.57	9.45	43	96.4
Ketidakcukupan konsumsi pangan	80.85	13.02	40.6	96.4
Pendidikan Anak Usia Dini (PAUD)	37.85	11.12	13.3	70.1
Anak Usia Dini dari Keluarga Miskin	13.81	6.57	5.7	33.07

Cakupan imunisasi pada layanan dasar kesehatan rata-rata terendah sebesar 27.5%, menunjukkan bahwa masih ada provinsi dengan cakupan yang sangat rendah. Sebaliknya, metrik seperti penolong persalinan dan akses air minum layak mencapai titik tertinggi mendekati 100%, yang menunjukkan bahwa beberapa provinsi telah menerima layanan penuh. Cakupan ASI eksklusif juga



cukup tinggi, dengan rata-rata 87.91%, tetapi ada perbedaan dari minimumnya, yaitu 68.9%. Namun, ketidakcukupan konsumsi pangan dan akses sanitasi layak menunjukkan keragaman yang signifikan di antara provinsi, yang menunjukkan perbedaan dalam tingkat kerentanan terhadap risiko stunting. PAUD dan jumlah anak dari keluarga miskin masih menunjukkan nilai minimum yang cukup rendah dalam hal pendidikan dan kesejahteraan anak.

### 3.2 Pengujian Asumsi

Uji multikolinieritas dan kecukupan sampel adalah bagian dari uji asumsi clustering. Nilai KMO Kaiser-Meyer-Olkin (KMO) sebesar 0.707 dihasilkan dari uji kecukupan sampel. Nilai yang lebih tinggi dari 0.5 menunjukkan bahwa sampel yang digunakan cukup untuk menggambarkan populasi saat ini. Selain itu, hasil uji Bartlett’s Test of Sphericity menghasilkan nilai Chi-Square sebesar 127.129 dengan signifikansi p-value 0.0004 ( $< 0.05$ ), yang menandakan bahwa terdapat korelasi yang cukup kuat antar variabel sehingga data layak untuk direduksi melalui analisis faktor seperti PCA. Selanjutnya, uji multikolinieritas dilakukan untuk mengevaluasi korelasi antara masing-masing variabel. Nilai korelasi maksimum untuk tiap variabel adalah 0.702 dan masih di bawah ambang 0.8, yang biasanya dianggap sebagai indikasi awal multikolinieritas tinggi. Oleh karena itu, data yang dievaluasi tidak menunjukkan masalah multikolinieritas yang signifikan.

### 3.3 Reduksi Dimensi

Pada uji asumsi seperti uji Bartlett, KMO, dan nilai MSA yang sudah dilakukan untuk memastikan kelayakan analisis sebelum penerapan PCA. Semua uji ini menunjukkan bahwa data layak untuk direduksi secara statistik. Tabel berikut menunjukkan hasil PCA.

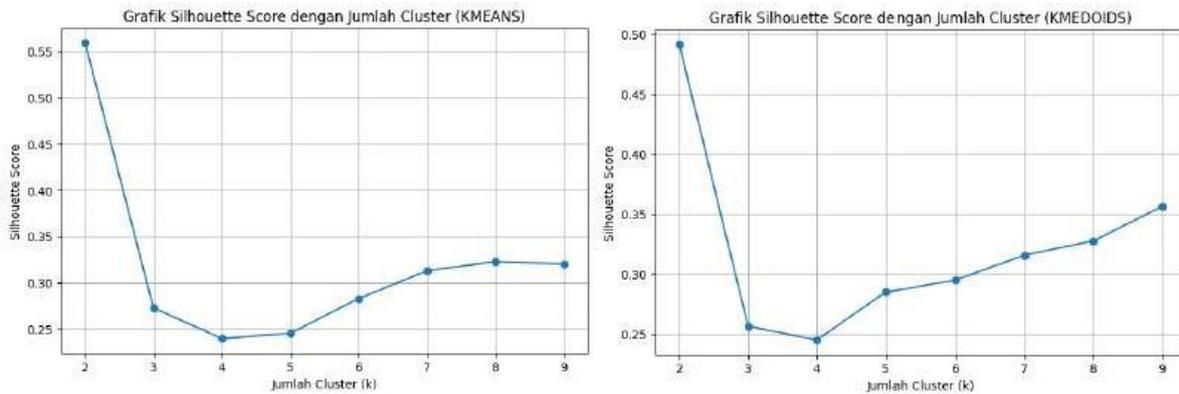
**Tabel 2.** Proporsi variansi setiap variabel

Komponen PC	Variance Kumulatif
Pc 1	0.45
Pc 2	0.59
Pc 3	0.71
Pc 4	0.81

Hasil PCA menunjukkan bahwa tiga komponen utama juga dipilih dengan varians kumulatif sebesar 71%. Meskipun empat komponen dapat bertanggung jawab atas 81% varians total, pilihan untuk menggunakan hanya tiga komponen dianggap sebagai kompromi antara efisiensi pengurangan dimensi dan retensi data. Dengan tiga bagian, sebagian besar informasi penting dari sembilan variabel asli telah diwakili dengan baik tanpa menghilangkan makna substantifnya. Selain itu, mempertahankan lebih sedikit komponen mempermudah interpretasi struktur data dan mencegah overfitting saat digunakan dalam analisis kluster. Pilihan di rentang tujuh puluh persen juga didasarkan pada praktik umum dalam analisis faktor eksploratori, di mana nilai kumulatif varians sebesar lebih dari  $\geq 70\%$  dianggap cukup untuk menunjukkan struktur data asli dengan cukup. Akibatnya, tiga komponen ini dianggap paling cocok untuk menjelaskan hubungan antarprovinsi dalam penanganan stunting.

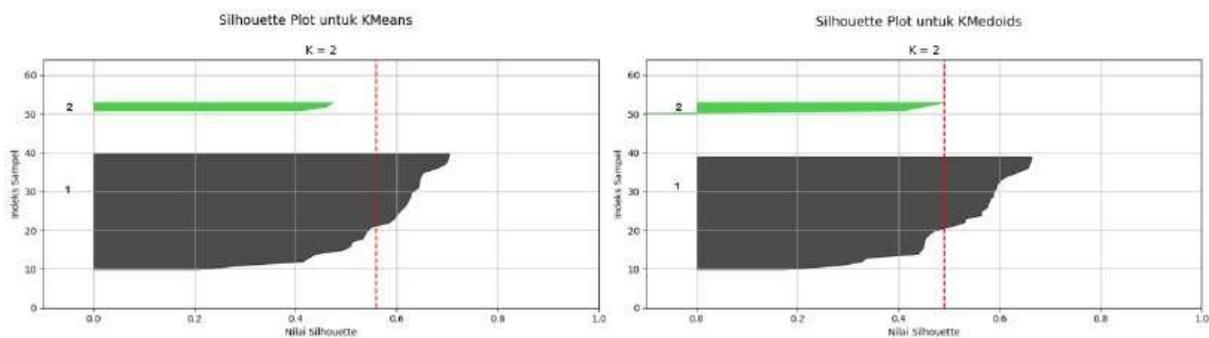
### 3.4 Proses Clustering

Dalam proses penentuan jumlah kluster optimal Penelitian ini tidak menggunakan metode Elbow karena titik tekuknya sering tidak jelas, terutama saat struktur kluster kurang kontras. Sebagai gantinya, digunakan Silhouette Score Plot yang lebih representatif dan kuantitatif dalam mengevaluasi kualitas kluster berdasarkan keserupaan dalam kluster dan perbedaannya antar kluster.



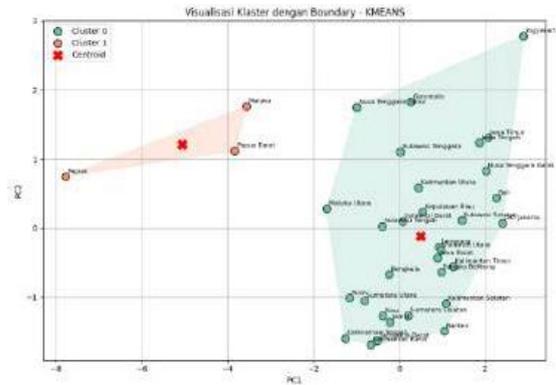
**Gambar 1.** Silhouette Score Plot dengan penerapan PCA

Pada gambar 1 untuk kedua metode K-Means dan K-Medoids, grafik Silhouette Score menunjukkan bahwa pembentukan dua kluster memberikan pemisahan yang paling jelas dan kompak antar objek dalam kluster, dengan nilai 0.56 untuk K-Means dan 0.49 untuk K-Medoids masing-masing. Nilai Silhouette Score turun ketika jumlah kluster ( $k$ ) lebih besar dari 2 yang menunjukkan bahwa pembentukan lebih dari dua kluster memiliki efek yang sama. Oleh karena itu, jumlah kluster yang ideal untuk penelitian ini adalah dua kluster yang memiliki skor Silhouette tertinggi.

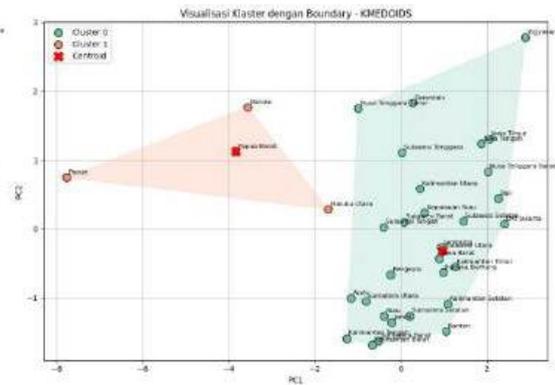


**Gambar 2.** Silhouette Coefficient Score Plot

Pada gambar 2 nilai silhouette untuk masing-masing anggota kluster ditunjukkan pada gambar plot silhouette untuk K-Means dan K-Medoids dengan jumlah kluster ( $k$ ) = 2 dengan masing-masing menunjukkan nilai silhouette rata-rata di atas 0.5, yang menunjukkan pemisahan kluster yang cukup baik. Namun, K-Medoids menunjukkan beberapa nilai negatif, menunjukkan bahwa sebagian data lebih mirip dengan kluster lain yang berdampak pada kualitas pengelompokan, sedangkan K-Means menunjukkan nilai silhouette yang lebih konsisten dan stabil tanpa nilai negatif.



Gambar 3a. Kluster K-Means



Gambar 3b. Kluster K-Medoids

Pada visualisasi pemetaan kluster menggunakan algoritma K-Means dan K-Medoids, terlihat bahwa kedua metode membagi provinsi-provinsi di Indonesia menjadi dua kluster utama. Pola pemetaan kluster dari kedua algoritma K-Means dan K-Medoids berbeda, pada K-Means posisi geometris centroid yang dihitung berdasarkan rata-rata mempengaruhi pembagian kluster sedangkan K-Medoids menggunakan medoid, sehingga pembagian kluster sedikit berbeda terutama berlaku untuk Provinsi Maluku Utara.

### 3.5 Validasi Kluster

Setelah dilakukan pengelompokkan dengan kedua metode, kemudian dilakukan pemilihan metode terbaik melalui validasi kluster menggunakan empat ukuran yaitu Davies-Bouldin Index (DBI), Calinski-Harabasz Index (CHI), Dunn Index, dan Silhouette Score.

Tabel 3. Hasil uji Kluster.

Metode	Jumlah kluster optimal	Ukuran Validasi			
		DBI	CHI	Dunn	Silhouette
K-Means	2	0.6637	21.15	0.3823	0.55
K-Medoids	2	0.7682	22.73	0.2422	0.49

Menurut hasil Tabel 3, algoritma K-Means memiliki performa terbaik dalam validasi kluster. Dengan nilai DBI terendah, nilai indeks Dunn dan nilai Silhouette Score tertinggi, sementara K-Medoids menghasilkan kluster yang lebih kompak, konsisten, dan terpisah dengan nilai CHI K-Medoids sedikit lebih tinggi tetapi peningkatan ini tidak cukup untuk mengungguli performa K-Means pada metrik. Selain itu, algoritma K-Means dinilai lebih sesuai dibandingkan K-Medoids, yang biasanya lebih kuat terhadap pencilon dan lebih cocok untuk dataset berskala kecil.



### 3.6 Interpretasi

Untuk data Indikator penanganan stunting Indonesia tahun 2023, hasil validasi kluster menunjukkan bahwa algoritma k-Means memberikan hasil pengelompokan terbaik. Dua kluster terbentuk. Rincian wilayah kluster yang terbentuk dapat dilihat pada peta gambar 4.

Peta Hasil Klastering Provinsi Indonesia



Gambar 4. Peta wilayah kluster

Pada gambar peta menunjukkan bagaimana wilayah Indonesia dibagi menurut kemiripan dalam sembilan indikator penanganan stunting. Hasil klasifikasi membagi provinsi menjadi dua kluster utama. Kluster 1 terdiri dari provinsi di wilayah barat dan tengah Indonesia, seperti Jawa, Sumatera, dan Kalimantan, yang umumnya memiliki kinerja lebih baik dalam penanganan stunting. Kluster 2 terdiri dari provinsi di wilayah timur Indonesia, seperti Papua dan Maluku, yang memiliki kinerja lebih rendah dalam indikator-indikator penting tentang penanganan stunting.

Tabel 4. Perbandingan rata-rata indikator kluster

Variabel	Klaster 1	Klaster 2
Imunisasi	70.1	60.7
Penolong Persalinan	90.6	65.8
KB Modern	65.8	31.9
ASI Eksklusif	89.2	75
Air Minum Layak	88.9	80.4
Sanitasi Layak	84.2	65.8
Ketidakcukupan Pangan	83.8	50
PAUD	39.1	25.5
Anak Usia Dini Keluarga Miskin	12.5	27.5

Tabel 4 menunjukkan bahwa Klaster 1 memiliki kinerja yang lebih baik dalam semua indikator penanganan stunting dibandingkan Klaster 2. Indikator-indikator tersebut termasuk imunisasi, persalinan oleh tenaga kesehatan, penggunaan kontrasepsi modern, ASI eksklusif, akses yang lebih baik ke air minum dan sanitasi, dan tingkat partisipasi PAUD yang lebih tinggi. Di sisi lain, Klaster 2 memiliki tingkat ketidakcukupan pangan dan proporsi anak usia dini berasal dari keluarga miskin yang lebih tinggi. Perbedaan ini menunjukkan perbedaan besar antara kluster.



**Tabel 5.** Hasil klusterisasi

<b>Nama Klaster</b>	<b>Nama Provinsi</b>
Wilayah Berkinerja Baik dalam Penanganan Stunting	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara.
Wilayah Berkinerja Rendah dalam Penanganan Stunting	Maluku, Papua Barat, Papua

Hasil dari pembagian wilayah berdasarkan indeks penanganan stunting menunjukkan bahwa ada dua klaster utama. Klaster pertama dianggap berkinerja baik dalam penanganan stunting mencakup 31 provinsi, termasuk Jawa Barat, Sumatera Selatan, Kalimantan Timur, dan Sulawesi Selatan. Provinsi-provinsi ini menunjukkan capaian yang relatif tinggi pada metrik penting seperti cakupan imunisasi, persalinan oleh tenaga kesehatan, akses air minum dan sanitasi yang layak, dan cakupan ASI eksklusif. Klaster kedua, yang terdiri dari Maluku, Papua Barat, dan Papua, memerlukan perhatian lebih lanjut karena menunjukkan capaian yang lebih rendah dalam sebagian besar metrik, termasuk partisipasi pendidikan anak usia dini dan akses layanan dasar. Segmentasi ini sangat penting untuk membantu pemerintah membuat kebijakan yang lebih terfokus dan sesuai dengan kebutuhan wilayah.

#### IV. KESIMPULAN

Berdasarkan hasil pembahasan sebelumnya, algoritma K-Means dengan dua klaster adalah algoritma partition clustering terbaik untuk segmentasi wilayah berdasarkan indeks penanganan stunting. Secara umum kedua klaster memiliki fitur yang cukup berbeda. Klaster pertama terdiri dari sebagian besar provinsi yang menunjukkan kinerja yang baik dalam hal layanan kesehatan ibu dan anak, gizi, sanitasi, dan partisipasi pendidikan anak usia dini. Sementara itu, klaster kedua terdiri dari provinsi Papua Barat, Maluku, dan Papua, yang menunjukkan kinerja yang lebih rendah dalam semua indikator tersebut. Perbedaan ini menunjukkan bahwa ada perbedaan dalam perawatan stunting di antara daerah. Akibatnya, untuk mencapai pembangunan yang lebih baik, strategi kebijakan yang fleksibel dan fokus harus digunakan untuk meningkatkan capaian indikator di provinsi-provinsi dalam klaster kedua guna mencapai kemajuan yang lebih merata dan inklusif.



## REFERENSI

- [1] P. Apriyani, A. R. Dikananda, and I. Ali, “Penerapan Algoritma K-Means dalam Klasterisasi Kasus Stunting Balita Desa Tegalwangi,” *Hello World Jurnal Ilmu Komputer*, vol. 2, no. 1, pp. 20–33, Mar. 2023, doi: 10.56211/helloworld.v2i1.230.
- [2] S. Gerakan Cegah Stunting Menggunakan Metode Penyuluhan Di Desa Lubuk Belimbing *et al.*, “Empowerment) STRATEGI GERAKAN CEGAH STUNTING MENGGUNAKAN METODE PENYULUHAN DI DESA LUBUK BELIMBING 1 BENGKULU MOVEMENT STRATEGY TO PREVENT STUNTING USING EXPLANATION METHODS IN LUBUK BELIMBING 1 VILLAGE BENGKULU,” *JOURNAL OF EMPOWERMENT*, vol. 3, no. 1, 2022, [Online]. Available: <https://jurnal.unsur.ac.id/index.php/JE>
- [3] Kementerian Kesehatan Republik Indonesia, “‘SSGI 2024: Prevalensi Stunting Nasional Turun Menjadi 19,8%,’ Sehat Negeriku.”
- [4] G. R. Suraya and A. W. Wijayanto, “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*, vol. 6, no. 2, pp. 180–201, Aug. 2022, doi: 10.29244/ijsa.v6i2p180-201.
- [5] S. W. Prayoga and S. Pramana, “Pemetaan Kejadian Balita Stunting Melalui Integrasi Citra Satelit Multisumber dan Official Statistics di Provinsi Nusa Tenggara Timur Menggunakan Machine Learning,” *Seminar Nasional Sains Data*, vol. 2024.
- [6] E. Saputra, H. Oktavianto, and L. A. Muharom, “SEGMENTASI DAERAH RENTAN STUNTING DI INDONESIA MENGGUNAKAN METODE K-MEANS.”
- [7] R. H. Sachrial and A. Iskandar, “Analisa Perbandingan Complete Linkage AHC dan K-Medoids Dalam Pengelompokan Data Kemiskinan di Indonesia,” *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 2, Sep. 2023, doi: 10.47065/bits.v5i2.4310.
- [8] M. Dzikrillah, N. Althaf, S. P. Utari, S. N. Hariono, and M. Ula, “SENASTIKA Universitas Malikussaleh SEGMENTASI RESIKO OBESITAS DAN PENYAKIT JANTUNG DENGAN ALGORITMA K-MEANS.”
- [9] L. Barokah, F. O. Lorenza, and Fitri Ayuning Tyas, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Brebes,” *Jurnal PROCESSOR*, vol. 20, no. 1, May 2025, doi: 10.33998/processor.2025.20.1.2050.
- [10] A. Astri Az-Zahra *et al.*, “Penerapan Algoritma K-Modes Clustering dengan Validasi Davies Bouldin Index pada Pengelompokan Tingkat Minat Belanja Online di Provinsi Daerah Istimewa Yogyakarta.”
- [11] N. Nur, M. Iqram, and N. Inayah, “Perbandingan K-Means dan Hierarchical Clustering dalam Pengelompokan Daerah Beresiko Stunting,” vol. 8, no. 2, p. 2023.
- [12] R. D. Rudianto and A. W. Wijayanto, “Analisis Perbandingan K-Means dan K-Medoids dalam Pengelompokan Provinsi Berdasarkan Indeks Demokrasi Indonesia 2021,” *Komputika : Jurnal Sistem Komputer*, vol. 13, no. 1, pp. 19–26, Mar. 2024, doi: 10.34010/komputika.v13i1.10812.
- [13] Badan Pusat Statistik, “Indeks Khusus Penanganan Stunting (IKPS) Tahun 2022-2023.”
- [14] Badan Pusat Statistik, “Profil Anak Usia Dini Tahun 2023.”