

## **Penerapan Metode *Hierarchical Clustering* untuk Klusterisasi Provinsi di Indonesia berdasarkan Indikator Status Gizi Anak Baduta (Bawah Dua Tahun) Tahun 2023**

**Syfriza Davies Raihannabil**<sup>1</sup>

<sup>1</sup> Politeknik Statistika STIS, Jl. Otto Iskandardinata No.64C 1, RT.1/RW.4, Bidara Cina, Kecamatan Jatinegara, Kota Jakarta Timur, Daerah Khusus Ibukota Jakarta 13330, Indonesia

\*Corresponding author : [212212893@stis.ac.id](mailto:212212893@stis.ac.id)



**P-ISSN:** 2986-4178  
**E-ISSN:** 2988-4004

### **Riwayat Artikel**

Dikirim: 25 Juli 2024  
Direvisi: 26 Oktober 2024  
Diterima: 27 Oktober 2024

### **ABSTRAK**

Indonesia sebagai negara dengan geografis yang sangat luas dan kondisi sosial-ekonomi yang beragam menyebabkan ketimpangan status gizi antardaerah. Pada masa baduta (bawah dua tahun), terjadi perbaikan gizi anak yang sangat kritis sehingga intervensi gizi setelah masa ini berdampak kurang signifikan pada perkembangan fisik dan kognitif anak. Penelitian ini menggunakan metode *hierarchical clustering* untuk klusterisasi provinsi di Indonesia berdasarkan indikator status gizi anak baduta tahun 2023. Analisis menghasilkan jumlah kluster optimum sebanyak 2 kluster dengan algoritma terbaik adalah AGNES. Kluster 1 beranggotakan 16 provinsi dengan permasalahan kekurangan gizi yang tinggi sehingga status gizi pada kluster ini dapat dikategorikan buruk. Sementara itu, kluster 2 beranggotakan 22 provinsi dengan permasalahan kekurangan gizi yang rendah sehingga status gizi pada kluster ini dapat dikategorikan baik. Dengan demikian, pemerintah dapat melakukan evaluasi kebijakan terkait intervensi gizi khususnya pada daerah yang tergolong status gizi buruk, seperti menyediakan program pemberian makanan tambahan (PMT) pada baduta, meningkatkan peran posyandu, dan melakukan pemantauan secara berkala terhadap kondisi status gizi baduta.

**Kata Kunci:** Status gizi, baduta, *hierarchical clustering*, AGNES, DIANA.

## **ABSTRACT**

*Indonesia, as a country with a very wide geographical area and diverse socio-economic conditions, causes disparities in nutritional status between regions. In the period under two years, there is a very critical improvement in children's nutrition so nutritional interventions after this period have a less significant impact on children's physical and cognitive development. This research uses a hierarchical clustering method to cluster provinces in Indonesia based on indicators of the nutritional status of under-two children in 2023. The analysis produces an optimum number of clusters of 2 clusters with the best algorithm being AGNES. Cluster 1 consists of 16 provinces with high malnutrition problems so the nutritional status in this cluster can be categorized as poor. Meanwhile, cluster 2 consists of 22 provinces with low malnutrition problems so the nutritional status in this cluster can be categorized as good. In this way, the government can evaluate policies related to nutritional interventions, especially in areas classified as having poor nutritional status, such as providing supplementary feeding programs (PMT) for under-two children, increasing the role of posyandu, and carrying out regular monitoring of the nutritional status of under-two children.*

**Keywords:** *Nutritional status, under-two children, hierarchical clustering, AGNES, DIANA.*

## **1. Pendahuluan**

Gizi merupakan masalah yang krusial bagi seluruh kelompok usia, baik anak, remaja, maupun dewasa. Permasalahan gizi masih menjadi topik utama yang harus diselesaikan oleh pemerintah di berbagai negara di dunia, termasuk Indonesia. Gizi dapat diartikan sebagai elemen dalam makanan yang dibutuhkan oleh tubuh untuk pertumbuhan, pemeliharaan, dan perbaikan jaringan tubuh [1]. Dalam tujuan pembangunan berkelanjutan kedua (*zero hunger*), pemerintah menargetkan untuk mengakhiri segala kondisi kekurangan gizi pada tahun 2030 [2]. Tidak adanya permasalahan gizi dapat menunjang pembangunan manusia, baik dari dimensi pendidikan, kesehatan, maupun standar hidup yang layak. Peningkatan kualitas gizi akan dibarengi dengan peningkatan kapasitas belajar sehingga kualitas akademik menjadi baik dan akses terhadap pendidikan makin luas [3]. Kualitas gizi yang bermasalah terbukti signifikan berpengaruh terhadap imunitas anak sehingga jumlah penyakit menular makin meningkat yang menyebabkan tingginya jumlah kematian [4]. Selain itu, gizi yang buruk dapat menyebabkan penurunan produktivitas sehingga pendapatan yang dihasilkan menurun dan angka kemiskinan makin tinggi sehingga masyarakat tidak mampu mencapai standar hidup yang layak [5].

Gizi dalam makanan yang dikonsumsi oleh manusia dapat menentukan kondisi tubuhnya. Kondisi tubuh tersebut dinamakan status gizi. Status gizi merupakan kondisi tubuh manusia yang disebabkan oleh proses konsumsi makanan [6]. Maka dari itu, seseorang yang mengalami kekurangan gizi akan memiliki kondisi tubuh yang kurang baik karena status gizi yang buruk. Kementerian Kesehatan RI mengelompokkan status gizi menjadi status gizi pada baduta, balita, anak dan remaja, serta dewasa [7]. Masa baduta berkisar dari pembuahan hingga anak berusia dua tahun yang merupakan tahapan penting bagi perkembangan intelektual dan kesehatan anak untuk masa yang akan datang [8]. Pada

masa tersebut terjadi perbaikan gizi anak yang kritis sehingga intervensi gizi setelah masa baduta akan berdampak kurang signifikan untuk perkembangan fisik anak [9]. Selain itu, nutrisi yang optimal dalam dua tahun pertama kehidupan sangat penting bagi pertumbuhan, kesehatan, dan perkembangan anak. Malagizi selama periode ini sering kali bersifat permanen dan dapat memiliki konsekuensi jangka panjang pada perkembangan fisik dan kognitif anak [10]. Oleh karena itu, peneliti memilih untuk menganalisis status gizi pada anak baduta. Pengukuran status gizi baduta dilakukan berdasarkan umur, panjang badan (PB), dan berat badan (BB). Terdapat tiga indikator pengukuran status gizi pada baduta, yaitu berat badan terhadap umur (BB/U), panjang badan terhadap umur (PB/U), dan berat badan terhadap panjang badan (BB/PB). Berdasarkan indikator tersebut, didapatkan *z-score* sehingga status gizi pada baduta dapat diklasifikasikan pada **tabel 1** berikut [7]:

**Tabel 1** Klasifikasi Status Gizi pada Baduta

Indikator	Status Gizi	Z-score
BB/U	<i>Severely Underweight</i>	$Z\text{-score} < -3$
	<i>Underweight</i>	$-3 \leq Z\text{-score} < -2$
	<i>Normal</i>	$-2 \leq Z\text{-score} < 2$
	<i>Overweight and Obesse</i>	$Z\text{-score} > 2$
PB/U	<i>Severely Stunting</i>	$Z\text{-score} < -3$
	<i>Stunting</i>	$-3 \leq Z\text{-score} < -2$
	<i>Normal</i>	$Z\text{-score} \geq -2$
BB/PB	<i>Severely Wasting</i>	$Z\text{-score} < -3$
	<i>Wasting</i>	$-3 \leq Z\text{-score} < -2$
	<i>Normal</i>	$-2 \leq Z\text{-score} < 2$
	<i>Overweight and Obesse</i>	$Z\text{-score} > 2$

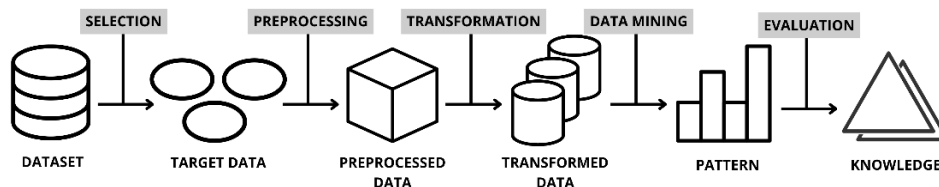
Indonesia merupakan negara dengan geografis yang sangat luas dengan kondisi sosial-ekonomi yang beragam. Kondisi sosial-ekonomi merupakan faktor penentu status gizi. Kondisi seperti tingginya kemiskinan, rendahnya tingkat pendidikan, dan kurangnya akses terhadap layanan kesehatan yang memadai dapat mengakibatkan malagizi [11]. Luas geografis yang besar tentu menyebabkan ketimpangan status gizi antardaerah di Indonesia yang menggambarkan perbedaan kondisi ekonomi, pendidikan, dan kesehatan di tiap-tiap daerah. Dengan demikian, klasterisasi provinsi-provinsi di Indonesia berdasarkan indikator status gizi pada baduta dapat membantu mengidentifikasi wilayah-wilayah dengan permasalahan gizi yang berbeda-beda sehingga dapat diketahui wilayah mana yang memerlukan perhatian khusus terkait masalah gizi pada baduta. Oleh karena itu, peneliti bertujuan untuk melakukan pengklasteran provinsi-provinsi di Indonesia berdasarkan indikator status gizi pada baduta tahun 2023 yang diharapkan dapat membantu pemegang kebijakan untuk merencanakan program strategis terhadap wilayah-wilayah yang memerlukan perhatian khusus terkait masalah gizi sehingga program menjadi lebih tepat sasaran.

Beberapa penelitian sebelumnya terkait metode *clustering* telah diidentifikasi. Penelitian yang dilakukan oleh Jannah dan Rosha pada tahun 2024, yaitu mengelompokkan provinsi berdasarkan indikator status gizi menggunakan *Cluster-Based Similarity Partitioning Algorithm* (CSPA), yaitu gabungan antara metode hierarki dan non-hierarki dan menghasilkan 2 klaster di mana klaster pertama terdiri atas provinsi-provinsi dengan masalah *stunting*, *wasting*, dan *underweight* yang tinggi, sedangkan klaster kedua dengan masalah *overweight* yang tinggi [12]. Selain itu, Andriany, Kusnandar, dan Perdana pada tahun 2021 juga melakukan penelitian untuk klasterisasi permasalahan gizi tiap-tiap provinsi di Indonesia dengan *fuzzy subtractive clustering*. Berdasarkan penelitian tersebut diperoleh 3 klaster dengan jumlah anggota klaster masing-masing adalah 9, 12, dan 13 provinsi dengan prevalensi *stunting*, *wasting*, dan *underweight* tertinggi pada klaster kedua [13]. Teknik *clustering* menggunakan metode hierarki juga telah banyak dilakukan pada penelitian sebelumnya. Pada tahun 2024, Wulandari melakukan *clustering* provinsi di

Indonesia berdasarkan prevalensi balita *stunting* menggunakan *agglomerative nesting* (AGNES) dengan membandingkan hasil kluster antara algoritma *single* dan *complete linkage*, kemudian didapatkan hasil kluster pertama beranggotakan 6 provinsi, sedangkan kluster kedua beranggotakan 24 provinsi [14]. Penelitian lainnya dilakukan oleh Satriawan dan Styawan tahun 2021 yang mengelompokkan provinsi di Indonesia berdasarkan determinan *stunting* pada balita menggunakan analisis kluster hierarki dengan metode *Ward*. Penelitian tersebut menghasilkan 4 kluster dengan jumlah anggota dan karakteristik kluster secara berturut-turut adalah 16 provinsi dengan faktor *stunting* tinggi, 8 provinsi sedang, 6 provinsi rendah, dan 4 provinsi sangat tinggi [15].

## 2. Metodologi Penelitian

Metode penelitian yang digunakan adalah *Knowledge Discovery in Database* (KDD). Peneliti membandingkan hasil kluster dari dua algoritma *hierarchical clustering*, yaitu *agglomerative nesting* (AGNES) dan *divisive analysis* (DIANA) untuk klusterisasi provinsi-provinsi di Indonesia berdasarkan indikator status gizi anak baduta tahun 2023. Langkah-langkah penelitian menggunakan metode KDD dapat diilustrasikan melalui pada **gambar 1** berikut:



**Gambar 1.** Langkah-Langkah dalam KDD

Langkah-langkah penelitian yang dilakukan dengan metode KDD dapat dijabarkan sebagai berikut:

1. **Selection.** Data yang digunakan merupakan data sekunder dari Survei Kesehatan Indonesia (SKI) 2023 Kementerian Kesehatan RI yang terdiri atas 3 data, yaitu prevalensi status gizi indikator BB/U, PB/U, dan BB/PB pada baduta.
2. **Preprocessing.** Menggabungkan (*merge*) ketiga data menjadi satu kemudian memilih atribut yang diperlukan saja, yaitu prevalensi (%) status gizi pada baduta untuk tiap-tiap indikator dan membuang atribut 95% *confidence interval*, prevalensi (%) baduta normal, dan N tertimbang. Kemudian, melakukan analisis deskriptif, seperti nilai minimum, maksimum, dan rata-rata dari tiap-tiap variabel, serta menghitung statistik *Hopkins* untuk menentukan apakah dari sekumpulan data yang digunakan terdapat tendensi pengelompokan.
3. **Transformation.** Melakukan pengujian asumsi pada analisis kluster, yaitu ketercukupan sampel menggunakan uji *Kaiser-Meyer-Olkin* (KMO) dan non-multikolinieritas menggunakan *variance inflation factor* (VIF). Jika tidak memenuhi asumsi, dapat dilakukan penanganan, seperti transformasi *log* dan *ln* atau mereduksi variabel dengan KMO rendah dan variabel yang mengalami multikolinieritas. Jika asumsi pada analisis kluster telah terpenuhi, maka selanjutnya dilakukan standarisasi data.
4. **Data Mining.** Melakukan analisis kluster hierarki dengan algoritma AGNES dan DIANA yang diawali dengan penentuan jumlah kluster optimum menggunakan metode *silhouette*.
5. **Evaluation.** Mengevaluasi hasil kluster dengan membandingkan *Average Silhouette Width* (ASW), *Davies-Bouldin Index* (DBI), *Calinski-Harabasz Index*, dan *Dunn*

Index dari algoritma AGNES dan DIANA. Kemudian, memilih algoritma analisis kluster terbaik berdasarkan evaluasi hasil kluster.

6. **Interpretation.** Menginterpretasikan hasil kluster.

### 2.1. Data dan Sumber Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang bersumber dari Survei Kesehatan Indonesia (SKI) tahun 2023. Data yang digunakan mencakup 38 provinsi di Indonesia. Variabel-variabel yang digunakan disajikan pada **tabel 2** berikut:

**Tabel 2** Deskripsi Variabel Penelitian

Variabel	Keterangan	Penghitungan	Satuan
<i>Underweight</i>	Prevalensi baduta <i>underweight</i>	Prevalensi baduta <i>severely underweight</i> + Prevalensi baduta <i>underweight</i>	Persen
<i>Stunting</i>	Prevalensi baduta <i>stunting</i>	Prevalensi baduta <i>severely stunting</i> + Prevalensi baduta <i>stunting</i>	Persen
<i>Wasting</i>	Prevalensi baduta <i>wasting</i>	Prevalensi baduta <i>severely wasting</i> + Prevalensi baduta <i>wasting</i>	Persen

### 2.2. Statistik Hopkins

Statistik *Hopkins* merupakan suatu ukuran yang digunakan untuk mengidentifikasi tendensi pengelompokan dari sekumpulan data sebelum dilakukan analisis kluster [16]. Rumus untuk menghitung statistik *Hopkins* adalah sebagai berikut:

$$H = \frac{\sum_{i=1}^n w_i}{\sum_{i=1}^n (w_i + u_i)} \quad (1)$$

di mana:

$n$  = jumlah sampel,

$w_i$  = jarak terkecil dari titik acak ke titik terdekat dalam kumpulan data asli,

$u_i$  = jarak terkecil dari titik acak ke titik terdekat dalam kumpulan data acak.

Jika  $H > 0,5$  atau makin mendekati 1, maka terdapat tendensi pengelompokan yang kuat dan adanya kluster dalam kumpulan data [17].

### 2.3. Asumsi pada Analisis Kluster

#### 2.3.1. Uji Kaiser-Meyer-Olkin (KMO)

Uji *Kaiser-Meyer-Olkin* (KMO) merupakan metode pengujian untuk mengukur ketercukupan sampel. Nilai KMO  $> 0,5$  atau mendekati 1 menunjukkan bahwa sampel yang digunakan sudah mewakili populasi [18].

#### 2.3.2. Variance Inflation Factor (VIF)

Pemeriksaan terjadinya multikolinieritas antarvariabel dalam data dilakukan dengan VIF [19]. Multikolinieritas terjadi ketika nilai VIF  $> 10$ .

### 2.4. Metode Silhouette

Metode *silhouette* menilai sejauh mana objek lebih cocok dengan kluster sendiri daripada kluster lain berdasarkan *silhouette score* [20]. *Silhouette score* dapat dihitung dengan rumus sebagai berikut:

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

dimana:

$a(i)$  = jarak rata-rata antara titik  $i$  dan titik lain dalam kluster sendiri,

$b(i)$  = jarak rata-rata antara titik  $i$  dan titik lain dalam kluster yang lainnya.

Penentuan jumlah kluster optimum didasarkan pada jumlah kluster yang dapat memaksimalkan *average silhouette width* (ASW) [21]. Rumus untuk menghitung ASW adalah sebagai berikut:

$$ASW = \frac{\sum_{i=1}^n s(i)}{n} \quad (3)$$

## 2.5. Hierarchical Clustering

Metode *hierarchical clustering* merupakan metode pengelompokan (*clustering*) dengan membentuk kluster secara bertahap melalui pendekatan *bottom-up* (*agglomerative*) atau *top-down* (*divisive*) yang digambarkan dalam bentuk *dendogram* [16].

### 2.5.1. Agglomerative Nesting (AGNES)

AGNES menganggap suatu objek sebagai kluster terpisah kemudian menggabungkan kluster lain yang mempunyai kemiripan secara bertahap berdasarkan metrik jarak hingga terbentuk suatu kluster besar [21]. Teknik yang digunakan pada penelitian ini adalah *Ward's method*. Metode ini menggabungkan antarkluster yang menyebabkan penambahan variasi yang paling kecil dan meminimalkan total variasi dalam kluster sehingga dalam kluster yang sama memiliki karakteristik yang lebih homogen [21]. Rumus untuk menghitung jarak antardua kluster yang berbeda ( $C_i$  dan  $C_j$ ) adalah sebagai berikut:

$$d^2(C_i, C_j) = \frac{n_i \cdot n_j}{n_i + n_j} \|\bar{x}_i - \bar{x}_j\|^2 \quad (4)$$

di mana:

$n_i$  dan  $n_j$  = jumlah objek pada kluster  $C_i$  dan  $C_j$ ,

$\bar{x}_i$  dan  $\bar{x}_j$  = *centroid* (*mean*) dari kluster  $C_i$  dan  $C_j$ ,

$\|\bar{x}_i - \bar{x}_j\|^2$  = kuadrat jarak *euclidean* antara *centroid* dari kluster  $C_i$  dan  $C_j$ .

### 2.5.2. Divisive Analysis (DIANA)

DIANA membagi satu kluster besar yang berisi seluruh objek secara rekursif menjadi kluster-kluster yang lebih kecil dan homogen [21]. DIANA menggunakan ukuran heterogenitas, yaitu *within-cluster sum of squares* (WCSS) untuk mengevaluasi kluster pada saat pembagian. Rumus untuk menghitung WCSS adalah sebagai berikut:

$$WCSS = \sum_{k=1}^K \sum_{i \in C_k} \|x_i - \bar{x}_k\|^2 \quad (5)$$

di mana:

$K$  = jumlah kluster,

$C_k$  = kluster ke- $k$ ,

$x_i$  = objek ke-  $i$ ,

$\bar{x}_k$  = *centroid* (*mean*) dari kluster ke- $k$ ,

$\|x_i - \bar{x}_k\|^2$  = kuadrat jarak *euclidean* antara objek  $x_i$  dan *centroid*  $\bar{x}_k$ .

## 2.6. Evaluasi Hasil Kluster

Penelitian ini mengevaluasi hasil kluster berdasarkan beberapa ukuran, yaitu *Average Silhouette Width* (ASW), *Davies-Bouldin Index* (DBI), *Calinski-Harabasz Index*, dan *Dunn Index*.

### 2.6.1. *Davies-Bouldin Index* (DBI)

DBI mengukur rata-rata rasio kemiripan setiap kluster dengan kluster lain yang memiliki kemiripan tertinggi dengan kluster tersebut [22]. DBI yang makin kecil menunjukkan hasil kluster yang makin baik. Rumus untuk menghitung DBI adalah sebagai berikut:

$$DBI = \frac{1}{K} \sum_{k=1}^K \max_{j \neq k} \left( \frac{s_k + s_j}{d_{kj}} \right) \quad (6)$$

di mana:

$s_k$  dan  $s_j$  = *intra-cluster distance* (jarak dalam kluster),

$d_{kj}$  = *inter-cluster distance* (jarak antarkluster).

### 2.6.2. *Calinski-Harabasz Index*

*Calinski-Harabasz Index* mengukur variasi antarkluster terhadap variasi dalam kluster [22]. Nilai yang makin tinggi menunjukkan kualitas hasil kluster yang makin baik. Rumus untuk menghitung *Calinski-Harabasz Index* adalah sebagai berikut:

$$CH = \frac{\left[ \sum_{k=1}^K n_k \|c_k - \bar{c}\|^2 / (K - 1) \right]}{\left[ \sum_{k=1}^K \sum_{i=1}^{n_k} \|d_i - c_k\|^2 / (N - K) \right]} \quad (7)$$

di mana:

$N$  = jumlah total sampel,

$n_k$  = jumlah sampel pada kluster ke-  $k$ ,

$\bar{x}_k$  = *centroid (mean)* dari kluster ke- $k$ ,

$\bar{x}$  = *centroid* keseluruhan,

$x_i$  = objek ke-  $i$ .

### 2.6.3. *Dunn Index*

*Dunn Index* mengukur rasio antara jarak minimum antarkluster dan jarak maksimum dalam kluster [22]. Nilai yang makin tinggi menunjukkan hasil kluster yang makin baik. Rumus untuk menghitung *Dunn Index* adalah sebagai berikut:

$$DI = \frac{\min_{1 \leq k \leq j \leq K} d(C_k, C_j)}{\max_{1 \leq k \leq K} \delta(C_k)} \quad (8)$$

di mana:

$d(C_k, C_j)$  = jarak minimum antara titik dalam kluster  $k$  dan titik dalam kluster  $j$ ,

$\delta(C_k)$  = jarak maksimum antardua titik dalam kluster  $k$ ,

### 3. Hasil dan Pembahasan

#### 3.1. Analisis Deskriptif

Analisis deskriptif meliputi nilai minimum, maksimum, dan rata-rata dari prevalensi baduta *underweight*, *stunting*, dan *wasting* yang disajikan pada **tabel 3**.

**Tabel 3** Statistik Deskriptif Indikator Status Gizi Baduta

Variabel	Minimum	Maksimum	Mean
<i>Underweight</i>	4,60	24,80	15,56
<i>Stunting</i>	7,00	36,60	20,80
<i>Wasting</i>	3,70	20,40	11,31

Berdasarkan **tabel 3**, prevalensi baduta *underweight* berada pada rentang 4,60% hingga 24,80% dengan rata-rata prevalensi sebesar 15,56%. Kemudian, prevalensi baduta *stunting* berada pada rentang 7,00% hingga 36,60% dengan rata-rata prevalensi sebesar 20,80%. Sementara itu, prevalensi baduta *wasting* berada pada rentang 3,70% hingga 20,40% dengan rata-rata prevalensi sebesar 11,31%.

#### 3.2. Statistik Hopkins

Penghitungan statistik *Hopkins* diperlukan untuk mengetahui kecenderungan pengelompokan dari suatu kumpulan data sehingga dapat diketahui kumpulan data tersebut cocok untuk dilakukan analisis kluster atau tidak. Hasil penghitungan statistik *Hopkins* ditunjukkan pada **tabel 4**.

**Tabel 4** Statistik Hopkins

$H$
0,754

Berdasarkan **tabel 4**, statistik *Hopkins* yang dihasilkan adalah sebesar 0,754. Nilai tersebut lebih dari 0,5 yang menunjukkan bahwa terdapat tendensi pengelompokan sehingga dapat dilanjutkan untuk analisis kluster.

#### 3.3. Asumsi pada Analisis Kluster

##### 3.3.1. Ketercukupan Sampel

Ketercukupan sampel adalah asumsi penting dalam analisis kluster yang menentukan data yang digunakan sudah representative atau belum. Pengujian asumsi ini dilakukan menggunakan Uji *Kaiser-Meyer-Olkin* (KMO) yang hasilnya ditunjukkan pada **tabel 5**.

**Tabel 5** Hasil Uji *Kaiser-Meyer-Olkin* (KMO)

Variabel	KMO
<i>Underweight</i>	0,65
<i>Stunting</i>	0,86
<i>Wasting</i>	0,68
Keseluruhan	0,71

Berdasarkan **tabel 5**, nilai KMO yang dihasilkan, baik untuk tiap-tiap variabel, maupun keseluruhan, bernilai lebih dari 0,5 yang berarti sampel sudah mewakili populasi sehingga asumsi ketercukupan sampel terpenuhi.

##### 3.3.2. Non-Multikolinieritas

Asumsi selanjutnya yang harus terpenuhi adalah non-multikolinieritas yang dapat diketahui dari *variance inflation factor* (VIF). Nilai VIF untuk tiap-tiap variabel ditunjukkan pada **tabel 6**.

**Tabel 6** Hasil Pemeriksaan Non-Multikolinieritas

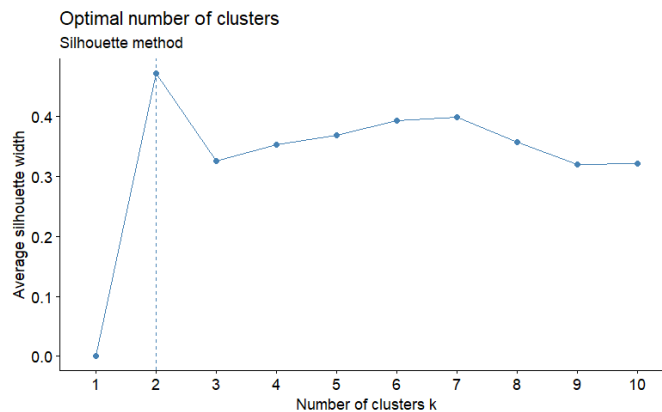


Variabel	VIF
<i>Underweight</i>	4,306
<i>Stunting</i>	2,034
<i>Wasting</i>	3,833

Berdasarkan **tabel 6**, nilai VIF dari variabel *underweight*, *stunting*, dan *wasting* bernilai kurang dari 10 sehingga asumsi untuk tidak terjadinya multikolinieritas terpenuhi.

### 3.4. Jumlah Kluster Optimum

Dalam menentukan jumlah kluster optimum, penelitian ini menggunakan metode *silhouette* yang hasilnya dapat dilihat pada **gambar 2**.

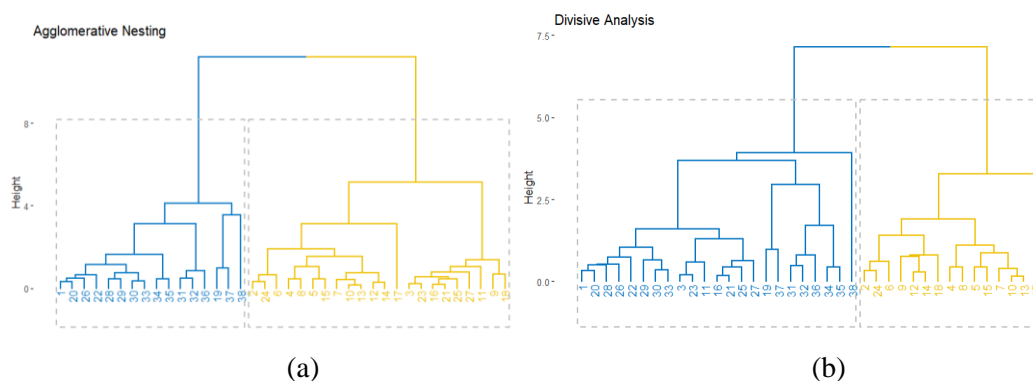


**Gambar 2.** Penentuan Jumlah Kluster Optimum Menggunakan Metode *Silhouette*

Berdasarkan **gambar 2**, jumlah kluster optimum yang diperoleh dari metode *silhouette* adalah sebanyak 2 kluster karena jumlah kluster  $k = 2$  dapat memaksimumkan *average silhouette width* (ASW).

### 3.5. Hierarchical Clustering

Metode *hierarchical clustering* menggunakan algoritma AGNES dan DIANA dengan  $k = 2$ . *Dendrogram* yang dihasilkan dapat dilihat pada **gambar 3**.



**Gambar 3.** Dendrogram (a) AGNES; (b) DIANA

Berdasarkan **gambar 3**, algoritma AGNES menghasilkan sebagian besar provinsi berada pada kluster 2, sedangkan algoritma DIANA menghasilkan sebagian besar provinsi berada pada kluster 1. Untuk menentukan hasil kluster terbaik, dilakukan evaluasi terhadap kedua algoritma tersebut.

### 3.6. Evaluasi Hasil Klaster

Dalam menentukan metode terbaik, perlu dilakukan evaluasi hasil klaster berdasarkan beberapa ukuran, di antaranya *average silhouette width* (ASW), *Davies-Bouldin Index* (DBI), *Calinski-Harabasz* (CH) *Index*, dan *Dunn Index* untuk algoritma AGNES dan DIANA yang hasilnya dapat dilihat pada **tabel 7**.

**Tabel 7** Evaluasi Hasil Klaster

Algoritma	ASW	DBI	CH	Dunn
AGNES	0,472	0,872	46,91	0,151
DIANA	0,463	0,821	46,44	0,141

**Tabel 7** menunjukkan ASW untuk AGNES lebih besar daripada DIANA. Kemudian, DIANA menghasilkan DBI yang lebih kecil daripada AGNES. Akan tetapi, CH *Index* dan *Dunn Index* yang dihasilkan dari AGNES lebih besar daripada DIANA. Dengan demikian, algoritma terbaik yang dipilih adalah *agglomerative nesting* (AGNES).

### 3.7. Interpretasi Hasil Klaster

Hasil pengklasteran provinsi-provinsi di Indonesia berdasarkan indikator status gizi baduta menggunakan metode *hierarchical clustering* dengan algoritma AGNES dapat ditunjukkan pada **tabel 8**.

**Tabel 8** Anggota Klaster

Klaster	Jumlah	Anggota
1	16	Aceh, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Selatan, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah, Papua Pegunungan
2	22	Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan

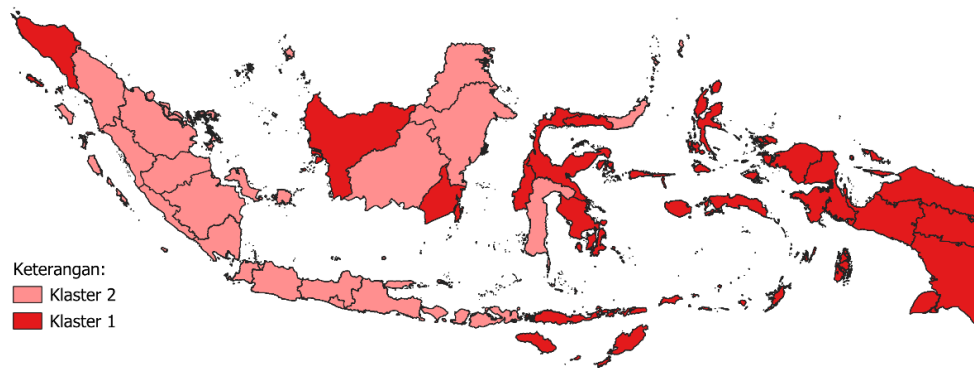
Berdasarkan **tabel 8**, dapat dilihat bahwa terdapat 16 provinsi yang tergabung dalam klaster 1, sedangkan 22 provinsi lainnya tergabung dalam klaster 2. Untuk melihat karakteristik klaster masing-masing, dapat dilakukan profilisasi dengan melihat rerata klaster.

**Tabel 9** Rerata Klaster

Klaster	<i>Underweight</i>	<i>Stunting</i>	<i>Wasting</i>
1	19,89	25,48	14,73
2	12,41	17,40	8,82

Berdasarkan **tabel 9**, karakteristik klaster masing-masing dapat diinterpretasikan sebagai berikut:

1. Klaster 1  
Klaster 1 memiliki karakteristik prevalensi baduta *underweight*, *stunting*, dan *wasting* yang lebih tinggi daripada klaster 2. Dengan demikian, klaster 1 beranggotakan provinsi-provinsi dengan permasalahan kekurangan gizi yang tinggi sehingga status gizi pada klaster ini dapat dikategorikan buruk.
2. Klaster 2  
Klaster 2 memiliki karakteristik prevalensi baduta *underweight*, *stunting*, dan *wasting* yang lebih rendah daripada klaster 1. Dengan demikian, klaster 2 beranggotakan provinsi-provinsi dengan permasalahan kekurangan gizi yang rendah sehingga status gizi pada klaster ini dapat dikategorikan baik.



**Gambar 4.** Peta Tematik Persebaran Status Gizi Anak Baduta di Indonesia

Persebaran status gizi anak baduta ditunjukkan oleh **gambar 4**. Provinsi dengan permasalahan kekurangan gizi yang rendah atau status gizi yang dikategorikan baik ditandai dengan warna merah muda (klaster 2). Sementara itu, provinsi dengan permasalahan kekurangan gizi yang tinggi atau status gizi yang dikategorikan buruk ditandai dengan warna merah (klaster 1). Provinsi Aceh menjadi satu-satunya daerah yang tergabung dalam klaster 1 di Pulau Sumatera. Di Pulau Kalimantan, masih terdapat 2 provinsi yang tergolong status gizi buruk, yaitu Kalimantan Barat dan Kalimantan Selatan. Hanya terdapat 2 provinsi di Pulau Sulawesi yang tergolong status gizi baik, yaitu Sulawesi Utara dan Sulawesi Selatan. Di bagian selatan, terdapat Provinsi NTT yang tergabung dalam klaster 1. Sementara itu, seluruh provinsi di Pulau Maluku dan Papua masih tergabung dalam klaster 1, yaitu daerah dengan permasalahan kekurangan gizi yang tinggi atau status gizi yang dikategorikan buruk.

#### 4. Kesimpulan

Hasil analisis menunjukkan bahwa metode *hierarchical clustering* yang terbaik untuk klasterisasi provinsi di Indonesia berdasarkan indikator status gizi anak baduta tahun 2023 adalah menggunakan algoritma *agglomerative nesting* (AGNES). Jumlah klaster optimum yang dihasilkan adalah sebanyak 2 klaster. Klaster 1 beranggotakan 16 provinsi dengan permasalahan kekurangan gizi, yaitu *underweight*, *stunting*, dan *wasting* yang tinggi sehingga status gizi pada klaster ini dapat dikategorikan buruk. Sementara itu, klaster 2 beranggotakan 22 provinsi dengan permasalahan kekurangan gizi, yaitu *underweight*, *stunting*, dan *wasting* yang rendah sehingga status gizi pada klaster ini dapat dikategorikan baik. Dengan demikian, pemerintah dapat melakukan evaluasi kebijakan terkait intervensi gizi khususnya pada daerah yang tergolong status gizi buruk, seperti menyediakan program pemberian makanan tambahan (PMT) pada baduta, meningkatkan peran posyandu, dan melakukan pemantauan secara berkala terhadap kondisi status gizi baduta.

#### 5. Daftar Pustaka

- [1] T. Soenardi, *Teori Dasar Kuliner: Teori, Dasar Memasak untuk Siswa, Peminat, dan Calon Profesional*, 1st ed. Jakarta: PT. Gramedia Pustaka Utama, 2013.
- [2] Bappenas, "Tanpa Kelaparan: Menghilangkan Kelaparan, Mencapai Ketahanan Pangan dan Gizi yang Baik, serta Meningkatkan Pertanian Berkelanjutan," Kementerian Perencanaan Pembangunan Nasional/Badan Perencanaan Pembangunan Nasional Republik Indonesia.

- [3] L. H. Jomaa, E. McDonnell, and C. Probart, "School Feeding Programs in Developing Countries: Impacts on Children's Health and Educational Outcomes," *Nutr Rev*, vol. 69, no. 2, pp. 83–98, Feb. 2011, doi: 10.1111/j.1753-4887.2010.00369.x.
- [4] WHO, "Nutrition." Accessed: Jul. 20, 2024. [Online]. Available: [https://www.who.int/health-topics/nutrition#tab=tab\\_1](https://www.who.int/health-topics/nutrition#tab=tab_1)
- [5] A. Ayala and B. M. Meier, "A Human Rights Approach to The Health Implications of Food and Nutrition Insecurity," *Public Health Rev*, vol. 38, no. 1, p. 10, Dec. 2017, doi: 10.1186/s40985-017-0056-5.
- [6] Kemenkes, "Seperti Apa Masalah Status Gizi pada Balita?," Jakarta, Jan. 2024. Accessed: Jul. 21, 2024. [Online]. Available: [https://yankes.kemkes.go.id/view\\_artikel/3145/seperti-apa-masalah-status-gizi-pada-balita](https://yankes.kemkes.go.id/view_artikel/3145/seperti-apa-masalah-status-gizi-pada-balita)
- [7] Kemenkes, *Survei Kesehatan Indonesia (SKI) 2023 Dalam Angka: Data Akurat Kebijakan Tepat*. Jakarta: Kementerian Kesehatan Republik Indonesia, 2023. Accessed: Jul. 21, 2024. [Online]. Available: <https://www.badankebijakan.kemkes.go.id/ski-2023-dalam-angka/>
- [8] C. G. Victora et al., "Maternal and Child Undernutrition: Consequences for Adult Health and Human Capital," *The Lancet*, vol. 371, no. 9609, pp. 340–357, Jan. 2008, doi: 10.1016/S0140-6736(07)61692-4.
- [9] R. Shrimpton, C. G. Victora, M. de Onis, R. C. Lima, M. Blössner, and G. Clugston, "Worldwide Timing of Growth Faltering: Implications for Nutritional Interventions," *Pediatrics*, vol. 107, no. 5, pp. e75–e75, May 2001, doi: 10.1542/peds.107.5.e75.
- [10] WHO, *Infant and Young Child Feeding: Model Chapter for TextBooks for Medical Students and Allied Health Professionals*. Geneva: WHO Press, World Health Organization, 2009. Accessed: Jul. 22, 2024. [Online]. Available: <https://www.who.int/publications/i/item/9789241597494>
- [11] WHO, "Social Determinants of Health," World Health Organization. Accessed: Jul. 21, 2024. [Online]. Available: <https://www.who.int/teams/social-determinants-of-health>
- [12] R. Jannah and M. Rosha, "Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Status Gizi Balita dengan Analisis Cluster Ensemble," *Jurnal Pendidikan Tambusai*, vol. 8, no. 2, pp. 22741–22751, 2024, Accessed: Jul. 21, 2024. [Online]. Available: <https://jptam.org/index.php/jptam/article/view/16982>
- [13] S. Andriany, D. Kusnandar, and H. Perdana, "Pengelompokan Provinsi di Indonesia Berdasarkan Permasalahan Gizi Balita Menggunakan Metode Fuzzy Subtractive Clustering," *Bimaster : Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 10, no. 3, pp. 351–360, 2021, doi: <https://dx.doi.org/10.26418/bbimst.v10i3.48340>.
- [14] S. Wulandari, "Analisis Clustering Provinsi Berdasarkan Prevalensi Stunting Balita Menggunakan Algoritma Single dan Complete Linkage," *Semnas Ristek (Seminar Nasional Riset dan Inovasi Teknologi)*, vol. 8, no. 01, pp. 26–33, Jan. 2024, doi: 10.30998/semnasristek.v8i01.7128.
- [15] D. Satriawan and D. A. Styawan, "Pengelompokan Provinsi di Indonesia Berdasarkan Faktor Penyebab Balita Stunting Menggunakan Analisis Cluster Hierarki," *Jurnal Statistika dan Aplikasinya*, vol. 5, no. 1, pp. 61–70, Jun. 2021, doi: 10.21009/JSA.05106.
- [16] C. M. Bishop, *Pattern Recognition and Machine Learning*, 1st ed. New York: Springer, 2006. Accessed: Jun. 14, 2024. [Online]. Available: <https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf>
- [17] A. Banerjee and R. N. Dave, "Validating clusters using the Hopkins statistic," in 2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542), IEEE, 2004, pp. 149–153. doi: 10.1109/FUZZY.2004.1375706.
- [18] J. F. Hair, W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis*, 7th ed. New York: Pearson, 2009.

- [19] J. Stock and M. Watson, Introduction to Econometrics, 4th ed. London: Pearson Education, 2018.
- [20] P.-N. Tan, M. Steinbach, and V. Kumar, Introduction to Data Mining, 1st ed. Pearson Addison Wesley, 2006.
- [21] L. Kaufman and P. J. Rousseeuw, Finding Groups in Data: An Introduction to Cluster Analysis, 1st ed. Wiley, 1990. doi: 10.1002/9780470316801.
- [22] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Elsevier, 2012. doi: 10.1016/C2009-0-61819-5.