

## Grouping Provinces in Indonesia Based on the Causes of Stunting Variables using Hierarchical Clustering Analysis\*

Pengelompokan Provinsi di Indonesia Berdasarkan Peubah Penyebab Stunting Menggunakan Analisis Cluster Hierarki

Detia Meilani<sup>1</sup>, Mohammad Masjkur<sup>2‡</sup>, and Farit M. Afendi<sup>3</sup>

<sup>1,2,3</sup>Department of Statistics, IPB University, Indonesia

<sup>‡</sup>corresponding author: [masjkur@apps.ipb.ac.id](mailto:masjkur@apps.ipb.ac.id)

Copyright © 2023 Detia Meilani, Mohammad Masjkur, and Farit M. Afendi. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

Stunting is a condition due to chronic malnutrition that causes children to be shorter in height compared to their age. The prevalence of stunting in Indonesia still exceeds the standards set by WHO. This study aims to classify provinces in Indonesia based on the characteristics of the causes of stunting. Cluster analysis is a statistical method used to group objects with similar characteristics. Province grouping is done using hierarchical cluster analysis consisting of Single Linkage, Complete Linkage, Average Linkage, Ward's method, and Centroid method. The Cophenetic correlation coefficient was used to determine the best cluster method and the optimal number of clusters using the Silhouette coefficient. The results show that the centroid method has the highest Cophenetic correlation coefficient with four clusters. The first cluster consists of 1 province with low stunting characteristics, the second cluster consists of 3 provinces with high stunting characteristics, the third cluster consists of 22 provinces with very high stunting characteristics, and the fourth cluster consists of 8 provinces with moderate stunting characteristics.

**Keywords:** cophenetic correlation coefficient, hierarchical clustering analysis, silhouette coefficient, stunting.

---

\* Received: Nov 2022; Reviewed: Feb 2023; Published: Oct 2023

## 1. Pendahuluan

Masalah kekurangan gizi merupakan salah satu target pembangunan yang tertulis dalam dokumen *Sustainable Development Goals* (SDGs) dengan tujuan menghilangkan kelaparan, mencapai ketahanan pangan dan gizi baik, serta meningkatkan pertanian berkelanjutan. *Stunting* merupakan suatu kondisi akibat kekurangan gizi kronis yang menyebabkan tinggi badan anak lebih pendek jika dibandingkan dengan anak seusianya. *Stunting* dapat menyebabkan penurunan pada fungsi kekebalan tubuh, hambatan perkembangan, menurunnya fungsi kognitif, dan gangguan sistem pembakaran. Selain itu, *stunting* dapat meningkatkan risiko penyakit jantung, obesitas, hipertensi, dan osteoporosis. *Stunting* dapat disebabkan oleh pendidikan orang tua, tinggi badan ibu, kategori IMT (indeks masa tubuh) ayah, tempat persalinan, tempat tinggal, akses sanitasi, akses air bersih, berat badan lahir, dan kelahiran prematur (Aditianti et al., 2020; Widari et al., 2021).

Hasil Riset Kesehatan Dasar (Riskesdas) 2018 menunjukkan prevalensi *stunting* Indonesia sebesar 30,8% (Kemenkes, 2019). Prevalensi *stunting* Indonesia masih cukup tinggi karena melebihi standar prevalensi *stunting* nasional suatu negara yang ditetapkan oleh WHO, yaitu sebesar 20%. Saat ini *stunting* menjadi salah satu masalah kesehatan yang diprioritaskan oleh Kemenkes RI.

Setiap provinsi di Indonesia memiliki faktor penyebab *stunting* yang berbeda sehingga penanganan dan kebijakan yang berbeda antar provinsi sangatlah diperlukan (Satriawan & Styawan, 2021). Pengelompokan akan memudahkan pemerintah melihat karakteristik penyebab *stunting* di setiap provinsi sehingga kebijakan yang berbeda dapat diterapkan dengan tepat. Provinsi-provinsi di Indonesia akan dikelompokkan berdasarkan kemiripan karakteristik penyebab *stunting* menggunakan analisis *cluster*. Analisis *cluster* merupakan suatu metode statistika yang mengelompokkan objek-objek yang memiliki kemiripan ke dalam kelompok yang sama (Dani et al., 2019). Provinsi-provinsi yang memiliki kemiripan karakteristik penyebab *stunting* akan berada pada kelompok yang sama. Penelitian ini menggunakan analisis *cluster* hierarki dengan metode *single linkage*, *complete linkage*, *average linkage*, metode *centroid*, dan metode *Ward's*.

## 2. Metodologi

### 2.1 Bahan dan Data

Data yang digunakan pada penelitian ini merupakan data Riset Kesehatan Dasar (Riskesdas) tahun 2018 yang diperoleh dari lembaga Pusat Penelitian dan Pengembangan Kementerian Kesehatan Republik Indonesia (Puslitbang Kemenkes RI). Adapun peubah yang digunakan adalah sebagai berikut:

Tabel 1: Peubah yang digunakan

Peubah	Keterangan	Rujukan
X1	Persentase kelahiran prematur	(Widari et al., 2021)
X2	Persentase tinggi badan ibu kurang dari 150 cm	(Widari et al., 2021)
X3	Persentase pendidikan ayah tidak tamat SD	(Aditianti et al., 2020)
X4	Persentase pendidikan ibu tidak tamat SD	(Aditianti et al., 2020)
X5	Persentase ayah dengan kategori IMT <i>underweight</i>	(Aditianti et al., 2020)
X6	Persentase tempat tinggal di perdesaan	(Widari et al., 2021)
X7	Persentase akses air bersih kurang	(Widari et al., 2021)
X8	Persentase tempat penampungan air limbah dapur terbuka	(Widari et al., 2021)
X9	Persentase tempat penampungan air limbah kamar mandi/tempat cuci terbuka	(Widari et al., 2021)
X10	Persentase tempat persalinan di rumah	(Aditianti et al., 2020)
X11	Persentase berat badan lahir balita kurang dari 2500 gram	(Widari et al., 2021)

## 2.2 Metode Penelitian

Prosedur penelitian:

### 1) Praproses data

Data awal berupa individu yang berusia 0-59 bulan di Indonesia, yaitu sebanyak 93.596 amatan. Data masih memiliki duplikat dan *missing value* sehingga perlu penanganan sebelum dianalisis. Selain itu, data masih terdiri dari amatan dengan status gizi normal sehingga perlu pelabelan status gizi untuk memisahkan setiap amatan.

#### a. Menghilangkan duplikat

Duplikat data yang dimaksud merupakan amatan-amatan yang memiliki nilai yang sama pada semua peubah. Terdapat 13 duplikat yang dihapus pada data.

#### b. Mengatasi *missing value*

Pada proses ini, peubah X11 (persentase berat badan lahir balita kurang dari 2500 gram) dihapus karena memiliki *missing value* lebih dari 50%. Penanganan *missing value* lainnya dilakukan dengan cara imputasi modus untuk peubah kategorik dan median untuk peubah numerik.

#### c. Perhitungan status gizi *stunting*

Status gizi *stunting* dihitung berdasarkan standar antropometri yang terdapat pada Peraturan Menteri Kesehatan RI No.2 Tahun 2020 tentang Standar Antropometri Anak.

### 2) Eksplorasi data untuk mengetahui gambaran umum data.

### 3) Agregasi data untuk mengubah amatan anggota rumah tangga menjadi provinsi.

### 4) Uji asumsi analisis *cluster*. Menurut Hair et al. (2014), terdapat dua asumsi yang harus dipenuhi pada analisis *cluster*, yaitu:

#### a. Sampel yang mewakili (*Representative*)

Uji *Kaiser Meyer Olkin* (KMO) dapat digunakan untuk melihat kecukupan suatu sampel dengan rumus (Shrestha, 2021):

$$KMO = \frac{\sum_{j=1}^p \sum_{k=1, k \neq j}^p r_{x_j x_k}^2}{\sum_{j=1}^p \sum_{k=1, k \neq j}^p r_{x_j x_k}^2 + \sum_{j=1}^p \sum_{k=1, k \neq j}^p \rho_{x_j x_k, x_l}^2}$$

Keterangan:

$p$  = Banyaknya peubah

$r_{x_j x_k}^2$  = Korelasi antar peubah  $X_j$  dan  $X_k$

$\rho_{x_j x_k, x_l}^2$  = Korelasi parsial antara peubah  $X_j$  dan  $X_k$  dengan  $X_l$

Sampel dapat dikatakan representatif apabila memiliki nilai  $KMO > 0,5$ .

b. Tidak ada Multikolinearitas

Multikolinearitas terjadi ketika terdapat hubungan linear antara beberapa atau semua peubah. Salah satu cara yang dapat dilakukan untuk mendeteksi multikolinearitas yaitu dengan menghitung nilai *Variance Inflation Factor* (VIF). Multikolinearitas terindikasi apabila nilai  $VIF > 10$  (Nurhasanah et al., 2021).

5) Pengelompokkan provinsi dengan analisis *cluster* hierarki, yaitu (Rachmatin, 2014):

a. *Single Linkage*

Mengelompokkan dua objek yang memiliki jarak terdekat (minimum) atau aturan tetangga dekat (*nearest neighbour rule*).

$$d_{(i,j)k} = \min \{d_{ik}, d_{jk}\}$$

b. *Complete Linkage*

Mengelompokkan dua objek yang memiliki jarak terjauh (maksimum) atau *farthest-neighbour*.

$$d_{(i,j)k} = \max \{d_{ik}, d_{jk}\}$$

c. *Average Linkage*

Jarak antara dua *cluster* dianggap sebagai jarak rata-rata antara semua objek dalam satu *cluster* dengan *cluster* lainnya.

$$d_{(i,j)k} = \frac{\sum_a \sum_b d_{ab}}{N_{ij} N_k}$$

Dengan  $d_{ab}$  merupakan jarak antara objek  $a$  pada *cluster* ( $IJ$ ) dan objek  $b$  pada *cluster*  $K$ ,  $N_{ij}$  merupakan jumlah item pada *cluster* ( $IJ$ ), dan  $N_k$  merupakan jumlah item pada *cluster* ( $IJ$ ) dan *cluster*  $K$ .

d. *Ward's Method*

Metode *Ward's* bertujuan untuk memperoleh *cluster* dengan varians internal *cluster* kecil. Dua *cluster* digabungkan dengan meminimumkan total jumlah kuadrat galat dalam kelompok. Persamaan SSE adalah sebagai berikut (Muthahharah & Juhari, 2021):

$$SSE = \sum_{j=1}^p \left( \sum_{i=1}^n x_{ij}^2 - \left( \frac{1}{n} \sum_{i=1}^n x_{ij} \right)^2 \right)$$

dengan  $x_{ij}$  adalah nilai untuk objek ke-  $i$  pada *cluster* ke-  $j$ ,  $p$  adalah banyaknya peubah, dan  $n$  adalah banyaknya objek dalam *cluster* yang terbentuk.

e. *Centroid Method*

Metode *centroid* memerhatikan rata-rata dari semua objek dalam *cluster*. *Centroid* dihitung setiap kali objek digabungkan sehingga ketika anggota

*cluster* bertambah maka *centroid* akan berubah.

6) Memilih metode *cluster* terbaik

Koefisien korelasi *Cophenetic* merupakan salah satu ukuran untuk menguji validitas hasil analisis *cluster* dan hanya dapat digunakan untuk analisis *cluster* hierarki (Alfina et al., 2012). Koefisien korelasi *Cophenetic* dapat dihitung dengan rumus:

$$r_{coph} = \frac{\sum_{i < j}^n (d_{ij} - \bar{d})(d_{coph \sim ij} - \bar{d}_{coph})}{\sqrt{\left[ \sum_{i < j}^n (d_{ij} - \bar{d})^2 \right] \left[ \sum_{i < j}^n (d_{coph \sim ij} - \bar{d}_{coph})^2 \right]}}$$

Keterangan:

$r_{coph}$  = koefisien korelasi *Cophenetic*

$d_{ij}$  = jarak asli antara objek ke- $i$  dan ke- $j$

$\bar{d}$  = rata-rata  $d_{ij}$

$d_{coph \sim ij}$  = jarak *Cophenetic* objek ke- $i$  dan ke- $j$

$\bar{d}_{coph}$  = rata-rata  $d_{coph \sim ij}$

Proses *clustering* dikatakan baik jika nilai  $r_{coph}$  mendekati 1 (Dani et al., 2019).

7) Menentukan jumlah *cluster* optimum

Koefisien *Silhouette* dapat digunakan untuk menentukan jumlah *cluster* optimum dan evaluasi untuk mengetahui seberapa baik pengelompokan (Dani et al., 2019). Jumlah *cluster* optimum adalah jumlah *cluster* yang memiliki nilai rata-rata koefisien *Silhouette* paling besar atau mendekati 1 (Simanjuntak & Khaira, 2021). Koefisien *Silhouette* dapat dihitung dengan rumus (Dani et al., 2019):

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)}$$

Keterangan

$a_i$  = rata-rata jarak objek ke- $i$  dengan semua objek pada kelompok yang sama

$b_i$  = rata-rata jarak objek ke- $i$  dengan semua objek pada kelompok yang berbeda

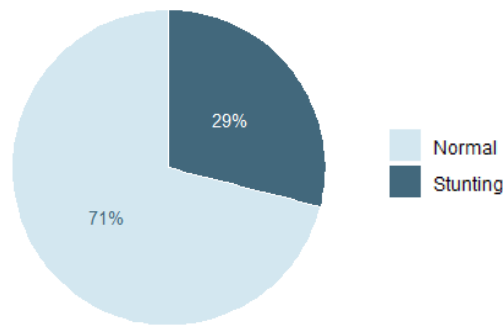
$S_i$  = nilai koefisien *Silhouette*

8) Identifikasi karakteristik *cluster*.

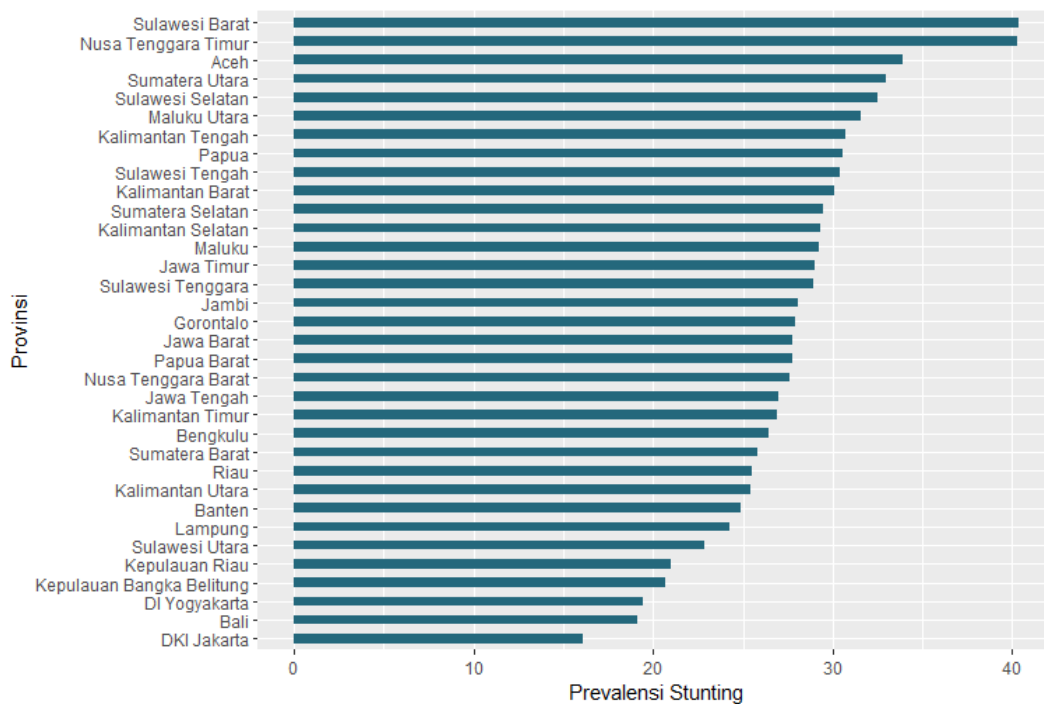
### 3. Hasil dan Pembahasan

#### 3.1 Eksplorasi Data

Data yang digunakan merupakan data Riset Kesehatan Dasar (Riskesdas) tahun 2018 yang terdiri dari balita berusia 0-59 bulan di Indonesia. Status gizi *stunting* balita dapat dilihat dari nilai *z-score* berdasarkan indikator TB/U. Berdasarkan Gambar 1, pada tahun 2018 prevalensi *stunting* di Indonesia sebesar 29%. Gambar 2 menunjukkan bahwa mayoritas provinsi di Indonesia memiliki prevalensi *stunting* yang tinggi. Hanya tiga provinsi yang memiliki prevalensi *stunting* di bawah 20%, yaitu DI Yogyakarta, Bali, dan DKI Jakarta dengan prevalensi *stunting* setiap provinsi sebesar 19,42%; 19,12%; dan 16,06%. Sulawesi Barat merupakan provinsi yang memiliki prevalensi *stunting* tertinggi, yaitu sebesar 40.37%.



Gambar 1 Prevalensi *stunting* Indonesia tahun 2018



Gambar 2: Prevalensi *stunting* Indonesia tahun 2018 berdasarkan provinsi

### 3.2 Pengujian Asumsi

Uji *Kaiser Meyer Olkin* (KMO) digunakan untuk melihat kecukupan sampel. Nilai KMO dari data adalah 0,5974 yang lebih besar dari 0,5. Hal ini menunjukkan bahwa asumsi kecukupan sampel atau sampel yang representatif terpenuhi.

Multikolinearitas dideteksi dengan nilai VIF. Nilai VIF lebih dari 10 menunjukkan bahwa terdapat multikolinearitas pada data. Berdasarkan Tabel 2, terdapat dua peubah yang memiliki nilai VIF lebih dari 10, yaitu peubah X8 (persentase tempat penampungan air limbah dapur terbuka) dan X9 (persentase tempat penampungan air limbah kamar mandi/tempat cuci terbuka). Hal ini menunjukkan adanya multikolinearitas pada data sehingga asumsi tentang multikolinearitas tidak terpenuhi. Masalah multikolinearitas dapat diatasi dengan analisis komponen utama (Prabowo et al., 2020).

Tabel 2: Nilai VIF setiap peubah

X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
3,440	2,653	9,427	9,453	1,791	2,754	3,507	11,113	12,749	3,016

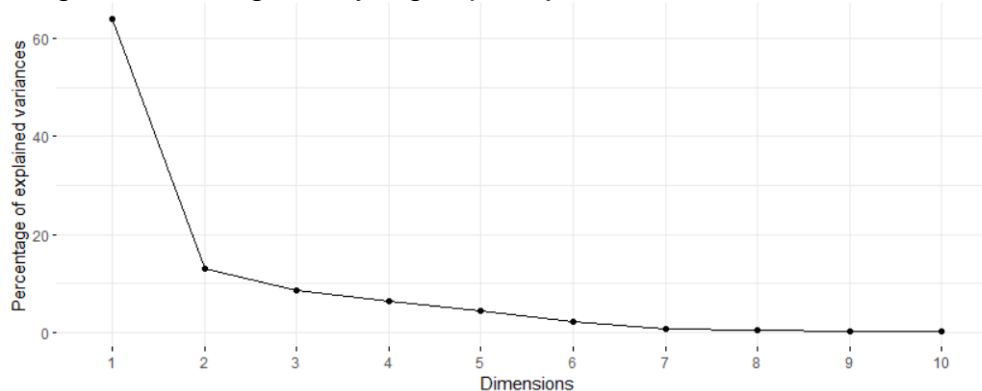
### 3.3 Analisis Komponen Utama

Analisis komponen utama merupakan analisis multivariat untuk mereduksi data menjadi kombinasi linear dari peubah-peubah yang saling bebas dengan tetap mempertahankan informasi yang terkandung di dalamnya (Prabowo et al., 2020). Tabel 3 menunjukkan proporsi keragaman yang dapat dijelaskan oleh setiap komponen utama.

Tabel 3: Nilai akar ciri dan proporsi keragaman yang dapat dijelaskan oleh setiap komponen utama

Komponen Utama	Nilai Akar Ciri	Proporsi (%)	Proporsi Kumulatif (%)
PC1	0,0709	63,99	63,99
PC2	0,0144	13	76,99
PC3	0,0095	8,57	85,56
PC4	0,0071	6,41	91,97
PC5	0,0049	4,42	96,39
PC6	0,0023	2,08	98,47
PC7	0,0008	0,72	99,19
PC8	0,0006	0,54	99,73
PC9	0,0002	0,18	99,91
PC10	0,0001	0,09	100,00

Banyaknya komponen utama yang digunakan ditentukan berdasarkan *scree plot* pada Gambar 3. Garis membentuk siku saat berada di komponen utama kedua. Maka, dua komponen utama yaitu PC1 dan PC2 akan digunakan untuk melakukan analisis *cluster* dengan total keragaman yang dapat dijelaskan adalah sebesar 76,95%.

Gambar 3: *Scree plot* banyaknya komponen utama yang digunakan

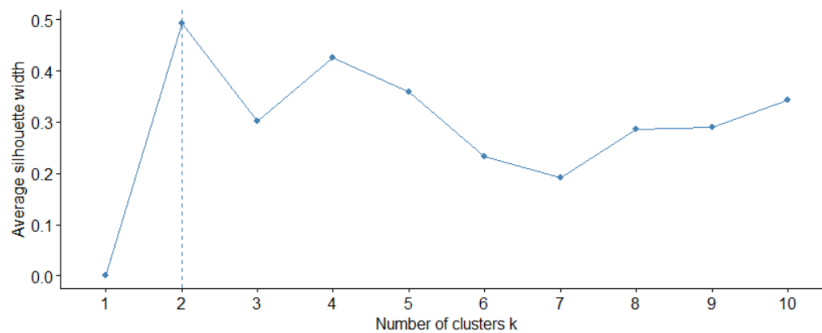
### 3.4 Analisis Cluster Hierarki

Pengelompokkan dilakukan dengan analisis *cluster* hierarki menggunakan jarak *euclidean* dan dipilih metode terbaik untuk mengelompokkan provinsi. Pemilihan metode terbaik dilakukan dengan membandingkan nilai koefisien korelasi *Cophenetic* setiap metode. Metode dengan nilai koefisien korelasi *Cophenetic* tertinggi merupakan metode terbaik yang akan dipilih untuk mengelompokkan provinsi. Tabel 4 menunjukkan bahwa metode *centroid* memiliki nilai koefisien korelasi *Cophenetic* tertinggi. Maka, pengelompokkan provinsi akan dilakukan dengan metode *centroid*.

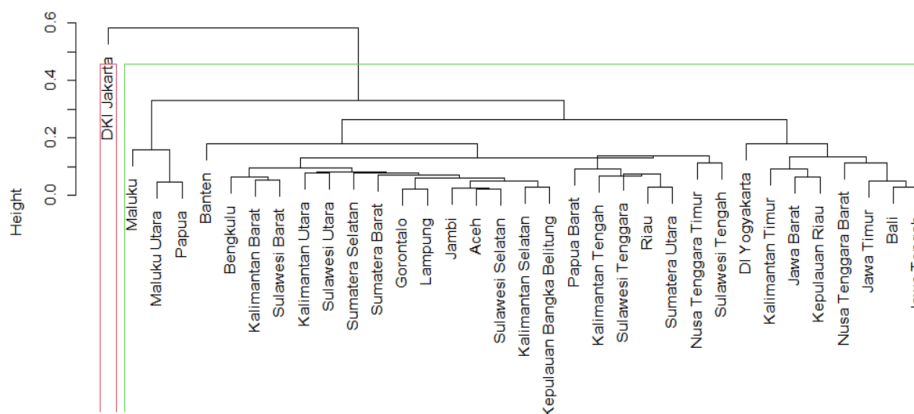
Tabel 4: Nilai koefisien korelasi *Cophenetic*

Metode Cluster	Koefisien Korelasi <i>Cophenetic</i>
<i>Single Linkage</i>	0,7097
<i>Average Linkage</i>	0,7412
<i>Complete Linkage</i>	0,7259
<b>Metode <i>Centroid</i></b>	<b>0,7797</b>
Metode <i>Ward's</i>	0,6839

Banyaknya *cluster* akan ditentukan menggunakan nilai koefisien *Silhouette*. Berdasarkan Gambar 4, jumlah *cluster* optimum adalah dua dengan nilai *Silhouette* sebesar 0,5601 sehingga provinsi akan dikelompokkan ke dalam dua *cluster*. Hasil analisis *cluster* hierarki akan membentuk suatu dendrogram seperti Gambar 5.



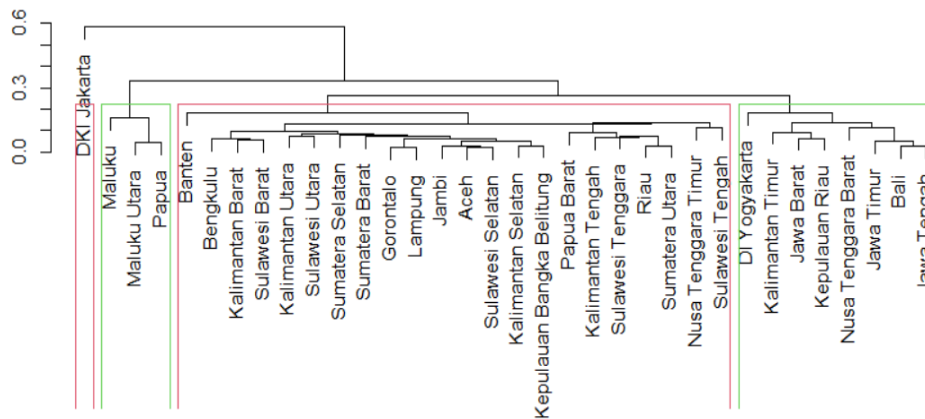
Gambar 4: Jumlah *cluster* optimal dengan koefisien *Silhouette*



Gambar 5: Dendrogram hasil analisis *cluster* dengan metode *centroid*



DKI Jakarta merupakan satu-satunya anggota pada *cluster* 1. Dengan kata lain, DKI Jakarta memiliki karakteristik penyebab *stunting* yang paling berbeda dengan provinsi-provinsi lain di Indonesia. Namun, hasil *cluster* belum cukup informatif sehingga jumlah *cluster* yang berbeda dipilih untuk mengelompokkan provinsi. Jumlah *cluster* ditentukan berdasarkan nilai koefisien *Silhouette* tertinggi kedua, yaitu sebanyak empat *cluster*. Hasil analisis *cluster* yang baru dapat dilihat pada Gambar 6. Hasil analisis *cluster* juga dapat dilihat pada Tabel 5.



Gambar 6: Dendrogram hasil analisis *cluster*

Tabel 5: Hasil analisis *cluster*

Provinsi	<i>Cluster</i>
DKI Jakarta	1
Maluku, Maluku Utara, Papua	2
Aceh, Banten, Bengkulu, Gorontalo, Jambi, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kalimantan Utara, Kepulauan Bangka Belitung, Lampung, Nusa Tenggara Timur, Papua Barat, Riau, Sulawesi Barat, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Utara, Sumatera Barat, Sumatera Selatan, Sumatera Utara	3
Bali, DI Yogyakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Timur, Kepulauan Riau, Nusa Tenggara Barat	4

### 3.5 Karakteristik *Cluster*

Tabel 6 menunjukkan rata-rata peubah untuk mengetahui karakteristik yang ada pada setiap *cluster*. Rata-rata nasional digunakan sebagai acuan untuk menggolongkan setiap peubah pada *cluster* memiliki nilai yang baik atau buruk.

*Cluster* 1 dan *cluster* 4 memiliki karakteristik ayah dengan kategori IMT *underweight* yang relatif tinggi. Karakteristik lainnya, yaitu kelahiran prematur, tinggi badan ibu kurang, pendidikan orang tua tidak tamat SD, akses air bersih, tempat penampungan limbah dapur dan kamar mandi/tempat cuci terbuka, serta tempat persalinan di rumah

relatif rendah. Ketika dilihat lebih lanjut, beberapa peubah pada *cluster* 4 memiliki nilai yang lebih tinggi jika dibandingkan dengan *cluster* 1. Hal ini menandakan bahwa provinsi-provinsi pada *cluster* 1 memiliki karakteristik penyebab *stunting* yang rendah dan *cluster* 4 memiliki karakteristik penyebab *stunting* yang sedang.

Tabel 6: Rata-rata peubah pada setiap *cluster*

Peubah	Cluster				Nasional*
	1	2	3	4	
Kelahiran prematur	24,12	<b>54,58</b>	<b>36,46</b>	21,32	34,13
Tinggi badan ibu kurang dari 150 cm	34,17	32,72	<b>44,98</b>	42,58	43,01
Pendidikan ayah tidak tamat SD	3,52	<b>13,58</b>	<b>11,82</b>	8,31	10,91
Pendidikan ibu tidak tamat SD	5,53	<b>19,21</b>	<b>12,84</b>	8,52	12,17
Ayah dengan kategori IMT <i>underweight</i>	<b>7,54</b>	4,02	<b>6,68</b>	<b>6,63</b>	6,46
Tempat tinggal di perdesaan	0	<b>79,77</b>	<b>71,96</b>	46,56	64,56
Akses air bersih kurang	3,02	<b>12,77</b>	<b>7,25</b>	4,05	6,86
Tempat penampungan air limbah dapur terbuka	89,45	<b>95,49</b>	<b>92,03</b>	82,98	90,13
Tempat penampungan air limbah kamar mandi/tempat cuci terbuka	83,92	<b>91,52</b>	<b>88,74</b>	74,99	85,61
Tempat persalinan di rumah	2,01	<b>65,78</b>	<b>33,02</b>	9,39	29,44

\* balita *stunting*

*Cluster* 2 memiliki karakteristik tinggi badan ibu kurang dan ayah dengan kategori IMT *underweight* yang relatif rendah menandakan bahwa kesehatan orang tua pada *cluster* 2 sudah cukup baik. Kategori lainnya seperti kelahiran prematur, pendidikan orang tua tidak tamat SD, tempat tinggal di perdesaan, akses air bersih, penampungan air limbah dapur dan kamar mandi/tempat cuci, serta tempat persalinan di rumah memiliki nilai yang relatif tinggi. Hal ini menandakan bahwa provinsi-provinsi pada *cluster* 2 memiliki karakteristik penyebab *stunting* yang tinggi.

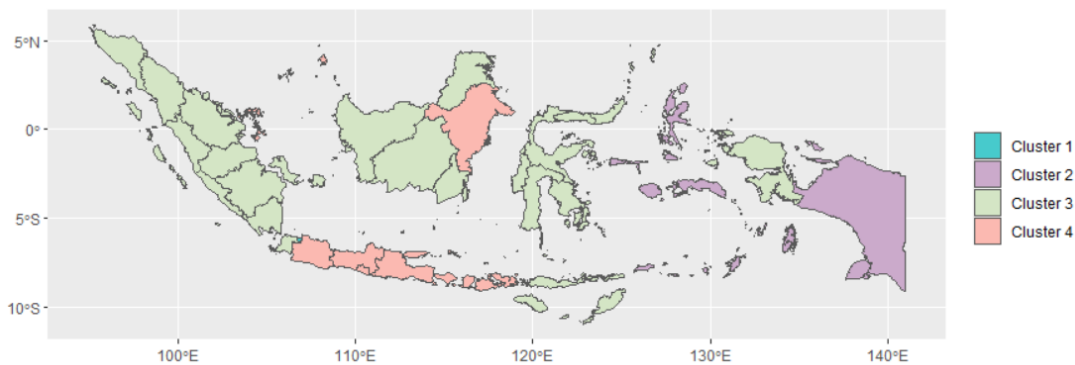
Karakteristik-karakteristik pada *cluster* 3 memiliki nilai yang relatif tinggi. Kelahiran prematur, pendidikan orang tua tidak tamat SD, tempat tinggal di perdesaan, akses air bersih, penampungan air limbah dapur dan kamar mandi/tempat cuci terbuka, dan persalinan di rumah pada *cluster* 3 merupakan karakteristik yang memiliki nilai rata-rata tertinggi di antara *cluster* lainnya. Sebagian besar anggota pada *cluster* 3 merupakan provinsi-provinsi yang memiliki prevalensi *stunting* tinggi. Maka *cluster* 3 merupakan provinsi-provinsi yang memiliki karakteristik penyebab *stunting* sangat tinggi.

Terdapat peubah yang secara nasional memiliki persentase yang cukup tinggi, yaitu tempat tinggal di perdesaan, tempat penampungan air limbah dapur terbuka, dan tempat penampungan air limbah kamar mandi/tempat cuci terbuka. Tempat penampungan air limbah terbuka baik untuk dapur ataupun kamar mandi/tempat cuci memiliki nilai yang cukup tinggi dibandingkan dengan karakteristik lainnya. Hal ini menunjukkan akses sanitasi balita *stunting* yang buruk. Tempat penampungan air limbah yang benar adalah dengan penampungan tertutup (Hidayangsih et al., 2021).

Semua balita *stunting* pada *cluster* 1 tidak ada yang berasal dari perdesaan. Hal ini

karena semua wilayah di *cluster 1* atau DKI Jakarta diklasifikasikan sebagai perkotaan menurut Peraturan Kepala Badan Pusat Statistik Nomor 37 Tahun 2010. Salah satu indikator dalam penetapan wilayah adalah fasilitas perkotaan. Wilayah yang diklasifikasikan sebagai perkotaan menandakan wilayah tersebut memiliki akses yang mudah terhadap fasilitas perkotaan. Akses rumah sakit merupakan salah satu kategori dalam fasilitas perkotaan. Semakin dekat jarak tempat tinggal dengan rumah sakit maka akses kesehatan semakin mudah didapatkan anak.

Gambar 7 menunjukkan peta provinsi di Indonesia berdasarkan *cluster*. Provinsi-provinsi dengan karakteristik penyebab *stunting* rendah dan sedang didominasi oleh provinsi-provinsi di Pulau Jawa. Provinsi dengan karakteristik penyebab *stunting* yang tinggi berada di wilayah Indonesia Timur. Provinsi-provinsi dengan karakteristik penyebab *stunting* sangat tinggi tersebar di beberapa pulau di Indonesia. Banten merupakan satu-satunya provinsi pada Pulau Jawa yang memiliki karakteristik penyebab *stunting* sangat tinggi.



Gambar 7: Peta provinsi berdasarkan *cluster*

#### 4. Simpulan

Provinsi di Indonesia dikelompokkan ke dalam empat *cluster* berdasarkan karakteristik penyebab *stunting*. Pengelompokkan dilakukan menggunakan metode *centroid* yang memiliki nilai koefisien korelasi *Cophenetic* tertinggi. *Cluster* pertama terdiri dari 1 provinsi dengan karakteristik penyebab *stunting* rendah, *cluster* kedua terdiri dari 3 provinsi dengan karakteristik penyebab *stunting* tinggi, *cluster* ketiga terdiri dari 22 provinsi dengan karakteristik penyebab *stunting* sangat tinggi, dan *cluster* keempat terdiri dari 8 provinsi dengan karakteristik penyebab *stunting* sedang. Balita *stunting* mayoritas bertempat tinggal di perdesaan. Tempat penampungan air limbah dapur dan kamar mandi/tempat cuci terbuka memiliki nilai yang sangat tinggi yang menandakan bahwa balita *stunting* di Indonesia memiliki akses sanitasi yang buruk.

#### Daftar Pustaka

- Aditianti, Sudikno, Raswanti I, Izwardy D, Irianto SE. 2020. Prevalensi dan faktor risiko *stunting* pada balita 24-59 bulan di Indonesia: analisis data riset kesehatan dasar 2018. *Nutr Food Res.* 43(2):51–64.
- Alfina T, Santosa B, Barakbah AR. 2012. Analisa perbandingan metode hierarchical clustering, k-means dan gabungan keduanya dalam membentuk cluster data (studi kasus: problem kerja praktek Jurusan Teknik Industri ITS). *J Tek POMITS.* 1(1):1–5.

- Dani ATR, Wahyuningsih S, Rizki NA. 2019. Penerapan hierarchical clustering metode agglomerative pada data runtun waktu. *Jambura J Math.* 1(2):64–78. doi:<http://dx.doi.org/10.34312%2Fjjom.v1i2.2354>.
- Hair JF, Black WC, Babin BJ, Anderson RE. 2014. *Multivariate Data Analysis*. Ed ke-12. Harlow: Pearson.
- Hidayangsih PS, Dharmayanti I, Tjandrarini DH, Anwar A, Azhar K, Paramita A, Kusriani I, Lestari H, Setiadji B. 2021. *Indeks Wash (Water Sanitation Hygiene) Indonesia*. Jakarta: Nas Media Pustaka.
- [Kemenkes] Kementerian Kesehatan. 2019. *Laporan Nasional Riskesdas 2018*. Jakarta: Badan Penelitian dan Pengembangan Kesehatan.
- Muthahharah I, Juhari A. 2021. Cluster analysis with complete linkage and ward's method for health service data in Makassar City. *J Varian.* 4(2):109–116. doi:[10.30812/varian.v4i2.883](https://doi.org/10.30812/varian.v4i2.883).
- Nurhasanah N, Salwa N, Ornita L, Hasan A, Mardhani M. 2021. Classifying regencies and cities on human development index dimensions: application of k-means cluster analysis. *J Sains Sosio Hum.* 5(2):913–918. doi:[10.22437/jssh.v5i2.15801](https://doi.org/10.22437/jssh.v5i2.15801).
- Prabowo RA, Nisa K, Faisol A, Setiawan E. 2020. Simulasi pemilihan metode analisis cluster hirarki agglomerative terbaik antara average linkage dan ward pada data yang mengandung masalah multikolinearitas. *J Siger Mat.* 1(2):49–55. doi:[10.23960/jsm.v1i2.2497](https://doi.org/10.23960/jsm.v1i2.2497).
- Rachmatin D. 2014. Aplikasi metode-metode agglomerative dalam analisis kluster pada data tingkat polusi udara. *Infin J.* 3(2):133–149. doi:[10.22460/infinity.v3i2.59](https://doi.org/10.22460/infinity.v3i2.59).
- Satriawan D, Styawan DA. 2021. Pengelompokan provinsi di Indonesia berdasarkan faktor penyebab balita stunting menggunakan analisis cluster hierarki. *J Stat dan Apl.* 5(1):61–70. doi:[10.21009/jsa.05106](https://doi.org/10.21009/jsa.05106).
- Shrestha N. 2021. Factor analysis as a tool for survey analysis. *Am J Appl Math Stat.* 9(1):4–11. doi:[10.12691/ajams-9-1-2](https://doi.org/10.12691/ajams-9-1-2).
- Simanjuntak KP, Khaira U. 2021. Pengelompokan titik api di Provinsi Jambi dengan algoritma agglomerative hierarchical clustering. *MALCOM Indones J Mach Learn Comput Sci.* 1(1):7–16.
- Widari S, Bachtiar N, Primayesa E. 2021. Faktor penentu stunting: analisis komparasi masa Millenium Development Goals (MDGs) dan Sustainable Development Goals (SDGs) di Indonesia. *J Ilm Univ Batanghari Jambi.* 21(3):1338–1346. doi:[10.33087/jjubj.v21i3.1726](https://doi.org/10.33087/jjubj.v21i3.1726).