

## 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\*

Perbandingan Metode Hierarchical Clustering, K-Means, K-Medoids, dan Fuzzy C-Means dalam Pengelompokan Provinsi di Indonesia Menurut Indeks Khusus Penanganan Stunting

Ghina Rofifa Suraya<sup>1</sup>, Arie Wahyu Wijayanto<sup>2‡</sup>

<sup>1,2</sup>Politeknik Statistika STIS, Indonesia

<sup>‡</sup>corresponding author: [ariewahyu@stis.ac.id](mailto:ariewahyu@stis.ac.id)

Copyright © 2022 Siswanto, Edy Saputra R, Nurtiti Sunusi and Nirwan Ilyas. 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 has been widely known as the highest case of malnutrition suffered by toddlers in the world and has a bad impact on children's future. In 2018, Indonesia was ranked the 31st highest stunting in the world and ranked 4th in Southeast Asia. About 30.8% (roughly 3 out of 10) of children under 5 years suffer from stunting in Indonesia. To support the government policy making in handling stunting, it is undoubtedly necessary to classify the levels of stunting handling in regions in Indonesia. In this work, the hierarchical agglomerative and non-hierarchical clustering is compared and evaluated to perform clustering on stunting data. The agglomerative hierarchical cluster uses Single Linkage, Average Linkage, Complete Linkage, and Ward Method, while the non-hierarchical cluster uses K-Means, K-Medoids (PAM) Clustering, and Fuzzy C-Means. This study uses data from 12 IKPS indicators in 34 provinces in Indonesia in 2018. Based on the results of the evaluation using the Connectivity Coefficient, Dunn Index, Silhouette Coefficient, Davies Bouldin Index, Xie & Beni Index, and Calinski-Harabasz Index, the results show that the Average Linkage is the best cluster method with the optimal number of clusters is four clusters. The first cluster is a cluster with a good level of stunting management which consists of 28 provinces. The second cluster consists of only one province, DI Yogyakarta with a very good level of stunting handling. The third cluster consists of four provinces with poor stunting handling rates. Finally, the last cluster consisting of one province, Papua, has a very poor level of stunting handling.

**Keywords:** Stunting, agglomerative hierarchical, K-Means, K-Medoids and Fuzzy C-Means

\* Received: Jan 2021; Reviewed: Jun 2022; Published: Aug 2022

## 1. Pendahuluan

Malnutrisi merupakan suatu ancaman bagi kelangsungan hidup seseorang dan pembangunan suatu negara. Hal ini perlu menjadi perhatian penting yang harus ditangani oleh setiap negara. Jumlah kasus malnutrisi secara global masih tergolong tinggi dan Asia merupakan benua yang memiliki jumlah kasus malnutrisi tertinggi di dunia. Malnutrisi memiliki dampak yang buruk bagi kehidupan seseorang dan bahkan menjadi salah satu kontributor terbesar penyebab kematian balita yakni sebesar 45% .

*Stunting* merupakan salah satu kasus malnutrisi tertinggi yang dialami oleh anak di dunia. *Stunting* didefinisikan sebagai anak-anak yang memiliki tinggi di bawah -2 standar deviasi (pendek) dan -3 standar deviasi (sangat pendek) jika dibandingkan dengan standar baku pertumbuhan anak WHO. Pada tahun 2017, sebanyak 55% anak berusia di bawah 5 tahun (balita) mengalami *stunting*, 69% mengalami *wasting*, dan 46% balita mengalami kelebihan berat badan. Kemudian, pada tahun 2018 sekitar 150,8 juta (22,2%) balita di dunia mengalami *stunting*, 50,5 juta (7,5%) balita mengalami *wasting*, dan 38,3 juta (5,6%) balita mengalami kelebihan berat badan (WHO, 2018). *Stunting* memiliki dampak jangka panjang dan jangka pendek bagi pertumbuhan dan perkembangan anak. Menurut WHO, dampak jangka pendek *stunting* dapat berupa peningkatan kejadian kesakitan, kematian, tidak optimalnya perkembangan kognitif, motorik, dan verbal serta terjadinya peningkatan biaya kesehatan (Kemenkes, 2018). Sedangkan dampak jangka panjangnya dapat berupa tubuh yang lebih pendek dari orang-orang seumurnya, meningkatnya resiko obesitas, kesehatan reproduksi menurun, kurang optimalnya performa dan kapasitas belajar saat sekolah, serta menurunnya produktivitas kerja. Selain itu, jika yang mengalami *stunting* adalah seorang wanita maka resiko untuk melahirkan anak dengan berat badan lahir rendah (BBLR) lebih besar (WHO, 2019). Hal ini dapat mengakibatkan terjadinya kasus malnutrisi secara terus menerus pada generasi selanjutnya karena anak dengan BBLR cenderung untuk tumbuh lebih pendek saat dewasa.

Menurut Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K), *stunting* juga berpengaruh terhadap Produk Domestik Bruto (PDB). Pada tahun 2017 *stunting* menyebabkan hilangnya PDB hingga 11%, mengurangi pendapatan seseorang hingga 20%, memperbesar terjadinya kesenjangan, dan menyebabkan terjadinya kemiskinan antar generasi (TNP2K, 2017). Kemudian, *stunting* juga dapat menurunkan kualitas sumber daya manusia (SDM) di suatu wilayah yang pada akhirnya berdampak pada terhambatnya pertumbuhan ekonomi dan terjadinya kemiskinan.

Di Indonesia, pada tahun 2018 sekitar 3 dari 10 balita mengalami *stunting* (UNICEF, 2019). Jika dilihat berdasarkan peringkat *stunting* pada balita, pada tahun 2018 Indonesia menduduki peringkat ke-31 *stunting* tertinggi di dunia dan peringkat ke-4 se-Asia Tenggara setelah Timor Leste, India dan Nepal (WHO, 2021). Walaupun Indonesia tergolong memiliki tingkat *stunting* yang tinggi, sebenarnya Indonesia sudah mengalami penurunan angka *stunting* di tahun 2018. Berdasarkan data Riset Kesehatan Dasar (Riskesdas) prevalensi *stunting* balita di Indonesia sudah mengalami penurunan dari 37,2% di tahun 2013 menjadi 30,8% di tahun 2018. Namun, angka ini masih tergolong sangat tinggi jika mengacu pada standar WHO. WHO membagi prevalensi *stunting* menjadi lima bagian yaitu sangat rendah (< 2,5%), rendah (2,5 sampai < 10%), sedang (10 sampai < 20%), tinggi (20 sampai < 30%), dan sangat tinggi ( $\geq 30\%$ ) (WHO, 2019).

Jumlah kasus *stunting* yang masih tinggi dan memiliki dampak yang buruk bagi anak serta negara, perlu menjadi perhatian khusus bagi pemerintah dan lembaga terkait agar dapat memperbaiki gizi anak secara efektif dan berkelanjutan dalam jangka waktu yang panjang.

Pemerintah telah berupaya untuk menangani kasus *stunting* pada balita salah satunya dengan memasukkan *stunting* sebagai target yang harus dicapai dalam *The Sustainable Development Goals* (SDGs). SDGs merupakan suatu agenda pembangunan pemerintah yang disepakati oleh negara-negara di dunia dalam rangka meningkatkan kualitas hidup manusia. SDGs bertujuan untuk menjaga peningkatan kesejahteraan ekonomi masyarakat secara berkesinambungan, menjaga keberlanjutan kehidupan sosial masyarakat, menjaga kualitas lingkungan hidup serta pembangunan yang menjamin keadilan dan terlaksananya tata kelola yang mampu menjaga peningkatan kualitas hidup dari satu generasi ke generasi berikutnya (Bappenas, 2020). SDGs memiliki 17 tujuan pembangunan yang terbagi menjadi 169 target yang ingin dicapai pada tahun 2030. Salah satu tujuan yang ingin dicapai yaitu tujuan kedua yang diharapkan dapat mengakhiri kelaparan, mencapai ketahanan pangan dan nutrisi yang lebih baik serta mendukung pertanian berkelanjutan pada tahun 2030. Pada tujuan kedua ini terdapat target yaitu ingin mengakhiri kelaparan, malnutrisi dan gizi buruk pada seluruh rakyat termasuk balita. Kemudian, *stunting* juga menjadi target dalam *The Global Nutrition* sebagai target pertama yang ingin mengurangi 40% jumlah penderita *stunting* pada balita pada tahun 2025 (WHO, 2014). Selain itu, dalam Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2020-2024 diharapkan prevalensi balita yang mengalami *stunting* menurun menjadi 19% pada tahun 2024 (Bappenas, 2019).

Dalam rangka mencapai target yang harus dipenuhi baik target SDGs, *The Global Nutrition* maupun RPJMN, maka masalah *stunting* ini tidak cukup jika hanya dilihat dari segi asupan gizi yang diterima oleh anak saja. *Stunting* merupakan suatu permasalahan kesehatan yang dipengaruhi oleh berbagai aspek. Oleh karena itu, untuk menangani kasus *stunting*, perlu dipertimbangkan aspek-aspek yang mungkin memengaruhi terjadinya *stunting* pada balita. Badan Pusat Statistik (BPS) telah melakukan penghitungan Indeks Khusus Penanganan *Stunting* (IKPS) guna mengetahui tingkat penanganan *stunting* yang telah dilakukan di Indonesia. Penghitungan indeks ini telah dilakukan BPS pada tahun 2018 yang menghitung IKPS pada tingkat nasional, provinsi maupun kabupaten/kota. Dalam penghitungan IKPS ini, BPS menggunakan 6 dimensi dan 12 indikator yang diharapkan mampu menggambarkan tingkat penanganan *stunting* di Indonesia.

Kemudian, agar dapat menentukan tingkatan prioritas dalam penanganan *stunting* di Indonesia, maka perlu dilakukan pengelompokan terhadap provinsi-provinsi di Indonesia. Pengelompokan ini dilakukan berdasarkan kemiripan karakteristik wilayahnya. Terdapat beberapa penelitian serupa terkait pengelompokan wilayah berdasarkan indikator tertentu. Thamrin dan Wijayanto (2021) melakukan penelitian untuk menentukan metode *clustering* terbaik dalam mengelompokkan kabupaten/kota di Pulau Jawa menggunakan metode *hard clustering* dan *soft clustering* berdasarkan tingkat kesejahteraan. Berdasarkan penelitian tersebut diperoleh hasil bahwa metode terbaik dalam pengelompokan kabupaten/kota di Pulau Jawa berdasarkan tingkat kesejahteraan yaitu menggunakan *Agglomerative Ward Linkage* dengan tiga *cluster*. Selanjutnya, Luthfi dan Wijayanto (2021) juga melakukan penelitian yang serupa.

Penelitian tersebut mengelompokkan kabupaten/kota di Indonesia menggunakan metode *Hierarchical*, *K-Means*, dan *K-Medoids Clustering* berdasarkan Indeks Pembangunan Manusia (IPM). Hasil penelitian menunjukkan bahwa metode terbaik dalam pengelompokan kabupaten/kota berdasarkan IPM yaitu menggunakan *K-Medoids* dengan lima *cluster*.

Penelitian kali ini melakukan pengelompokan pada 34 provinsi di Indonesia menggunakan metode *Hierarchical Clustering*, *K-Means*, *K-Medoids*, dan *Fuzzy C-Means* berdasarkan indikator IKPS. Kemudian, dari beberapa metode tersebut dipilih metode *clustering* terbaik dalam mengelompokkan provinsi di Indonesia berdasarkan indikator IKPS. Dengan proses *clustering* ini diharapkan pemerintah dapat menentukan prioritas dan mengambil kebijakan yang tepat mengenai provinsi mana yang perlu ditindak lanjuti terlebih dahulu.

## 2. Metodologi

### 2.1 Bahan dan Data

Penelitian ini menggunakan data sekunder yang diperoleh dari publikasi dan tabel dinamis BPS, publikasi Kementerian Kesehatan, Kementerian Pemberdayaan Perempuan dan Perlindungan Anak (KPPPA), serta Kementerian Pertanian pada tahun 2018. Terdapat 34 provinsi di Indonesia yang dianalisis dalam penelitian ini menggunakan 12 indikator IKPS yang disajikan pada Tabel 1.

Tabel 1: Indikator IKPS dan sumbernya

Dimensi	Indikator	Keterangan	Sumber Data
Kesehatan	Imunisasi	Persentase Anak Usia 12-23 bulan yang menerima imunisasi lengkap Tahun 2018	Profil Kesehatan Ibu dan Anak (2020)
	Penolong Persalinan Oleh Tenaga Kesehatan	Persentase Perempuan Pernah Kawin Berusia 15-49 Tahun Yang Proses Kelahiran Terakhirnya Ditolong Oleh Tenaga Kesehatan Terlatih	Tabel Dinamis BPS
	Keluarga Berencana (KB) Modern	Persentase Pasangan Usia Subur (PUS) Umur 15-49 Tahun yang Sedang Menggunakan Alat KB Modern untuk Menunda atau Mencegah Kehamilan	Profil Kesehatan Ibu dan Anak (2020)
Gizi	ASI Eksklusif	Persentase Bayi Usia Kurang Dari 6 Bulan Yang Mendapatkan Asi Eksklusif	Tabel Dinamis BPS
	Makanan Pendamping (MP) ASI	Persentase Penduduk Umur 0-23 Bulan (Baduta) yang Mendapatkan Makanan/Cairan Tambahan dalam 24 Jam Terakhir	Statistik Kesejahteraan Rakyat (2018) (2019)
Perumahan	Air Minum	Persentase Rumah Tangga	Tabel dinamis

	Layak	menurut Provinsi, Tipe Daerah dan Sumber Air Minum Layak	BPS
	Sanitasi Layak	Persentase Rumah Tangga menurut Provinsi, Tipe Daerah dan Sanitasi Layak	Tabel dinamis BPS
Pangan	Mengalami Kerawanan Pangan	Prevelensi penduduk yang mengalami kerawanan pangan	Statistik Ketahanan Pangan (2020)
	Ketidacukupan Konsumsi Pangan	Prevalensi penduduk yang tidak mengalami kecukupan konsumsi pangan ( <i>Prevalence of Undernourishment (PoU)</i> )	Statistik Ketahanan Pangan (2020)
Pendidikan	Pendidikan Anak Usia Dini (PAUD)	Angka partisipasi Anak Usia 3-6 Tahun yang Sedang Mengikuti PAUD	Profil Anak Indonesia (2019)
Perlindungan Sosial	Pemanfaatan Jaminan Kesehatan	Persentase Penduduk yang Memiliki Jaminan Kesehatan Nasional (JKN)/Jamkesda dan Pernah Memanfaatkannya untuk Pemeriksaan Kesehatan dalam Setahun Terakhir	Profil Kesehatan Ibu dan Anak (2020)
	Penerima KPS/KKS	Persentase Rumah Tangga yang Menerima Kartu Perlindungan Sosial (KPS)/Kartu Keluarga Sejahtera (KKS)	Statistik Kesejahteraan Rakyat (2018)

## 2.2 Metode Penelitian

Metode analisis yang digunakan dalam penelitian ini yaitu *Principal Component Analysis* (PCA) dan analisis *cluster* dengan metode *Hierarchical* dan *Non-Hierarchical Clustering*. Metode *hierarchical clustering* yang digunakan yaitu metode *Single Linkage*, *Average Linkage*, *Complete Linkage* dan metode *Ward*. Sementara itu, metode *non-hierarchical clustering* yang digunakan yaitu *K-Means*, *K-Medoids* (PAM) dan *Fuzzy C-Means*. Analisis dalam penelitian ini juga digunakan untuk melihat metode *clustering* terbaik dengan membandingkan tingkat akurasi dari setiap metode dalam mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator IKPS.

### a. *Principal Component Analysis* (PCA)

Sebelum melakukan analisis *cluster*, perlu dilakukan pemeriksaan terhadap asumsi multikolinearitas pada setiap indikator atau variabel dalam data. Pengecekan multikolinearitas dapat dilakukan dengan melihat nilai korelasi pada setiap variabel. Jika terdapat multikolinearitas dalam data maka perlu diatasi. PCA merupakan salah satu metode yang dapat digunakan untuk mengatasi terjadinya multikolinearitas. PCA adalah teknik interdependensi yang tidak membedakan antara variabel dependen atau independen (Pramana, Yuniarto, Mariyah, Santoso, & Nooraeni, 2018). PCA berkaitan dengan menjelaskan struktur varians-kovarians dari satu set variabel melalui beberapa kombinasi linear dari variabel (Johnson & Wichern, 2007). Metode PCA ini bertujuan

untuk mengatasi masalah dimensi data yang kompleks dan adanya multikolinearitas dalam data dengan cara mereduksi dimensi data yang kompleks menjadi lebih sederhana. Proses pereduksian ini dilakukan dengan membentuk komponen baru dengan mempertahankan total variasi variabel asal sehingga meminimalkan kehilangan informasi awal. Hubungan antar komponen baru yang terbentuk saling bebas dan setiap komponen tersebut dapat menggantikan atau memiliki informasi mengenai variabel-variabel dalam data.

Penentuan jumlah komponen utama yang terbentuk dapat didasarkan pada beberapa ketentuan yaitu (Pramana, Yuniarto, Mariyah, Santoso, & Nooraeni, 2018):

- 1) **Scree plot**; banyaknya komponen utama yang digunakan yaitu ketika garis kurva *scree plot* mulai melandai.
- 2) **Kumulatif proporsi varians total**; batasan minimal kumulatif proporsi varians dianggap cukup menjelaskan total keragaman variabel asal yaitu sebesar 80%.
- 3) **Eigenvalue**; jika nilai *eigenvalue* > 1 maka komponen tersebut akan dipilih sebagai komponen utama.

## b. Analisis Cluster

Analisis *cluster* merupakan salah satu contoh penerapan dari *unsupervised learning* yang tidak bergantung pada kelas yang telah ditentukan. Oleh karena itu, pengelompokan data pada analisis *cluster* merupakan suatu bentuk proses belajar berdasarkan data observasi. *Cluster* merupakan objek-objek dari suatu data yang memiliki kemiripan satu sama lain pada *cluster* yang sama dan tidak mirip dengan objek yang berada pada *cluster* lainnya (Han & Kamber, 2006). Proses mengelompokkan objek-objek yang memiliki karakteristik yang serupa ini ke dalam kelas yang sama disebut dengan *clustering*. Proses *cluster* disebut juga dengan proses segmentasi data karena mempartisi sekumpulan data yang besar ke dalam kelompok-kelompok data lebih kecil yang memiliki kemiripan karakteristik. *Cluster* memiliki prinsip bahwa objek-objek yang terdapat dalam satu kelas memiliki kemiripan karakteristik yang tinggi (homogen) dan antar kelas memiliki karakteristik yang berbeda (heterogen). Untuk mengukur kedekatan/kemiripan antara satu objek dengan objek yang lainnya, dilakukan perhitungan kemiripan (*similarity*) menggunakan "Jarak Euclidean" dengan rumus sebagai berikut (Han & Kamber, 2006):

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2} \quad (1)$$

dimana  $d(i, j)$  adalah jarak antara objek  $i$  dengan objek  $j$ ,  $x_{in}$  adalah data dari subjek ke- $i$  pada variabel ke- $n$ , dan  $x_{jn}$  adalah data dari subjek ke- $j$  pada variabel ke- $n$ .

Terdapat dua metode dalam analisis *cluster* yaitu metode *hierarchical* dan *non-hierarchical clustering*. Metode *hierarchical clustering* dirancang untuk mengubah bentuk data menjadi bentuk yang lebih sederhana (dekomposisi) dari kumpulan data yang ada (Han & Kamber, 2006). Metode *hierarchical clustering* terbagi menjadi *agglomerative* dan *divisive hierarchical*. *Agglomerative hierarchical* menggunakan pendekatan *bottom-up*, yang awalnya menganggap setiap objek memiliki *cluster* tersendiri. Kemudian dilakukan penggabungan objek secara bertahap, mulai dari objek-objek yang paling mirip hingga semua kelompok bergabung menjadi satu *cluster* atau hingga proses penggabungan tidak bisa dilakukan lagi. Sementara itu, *divisive hierarchical* menggunakan pendekatan *top-up*, yang pada awalnya semua objek akan

berada dalam satu *cluster* yang sama. Lalu, pada setiap tahapannya *cluster-cluster* tersebut dipecah menjadi *cluster* yang lebih kecil hingga setiap objek berada dalam *cluster*-nya sendiri.

Metode *non-hierarchical clustering* dirancang untuk mengelompokkan  $i$  ke dalam sejumlah  $K$  *cluster* (Johnson & Wichern, 2007). Penentuan jumlah  $K$  *cluster* dapat ditentukan sebelumnya atau berdasarkan prosedur *clustering*. Metode ini diawali dengan mempartisi sekumpulan objek ke dalam kelompok atau ketika sekumpulan titik membentuk inti *cluster*. Pada penelitian ini, metode *hierarchical clustering* yang digunakan yaitu *agglomerative hierarchical*, di antaranya *single linkage*, *average linkage*, *complete linkage* dan metode *ward*. Sementara itu, pada *non-hierarchical clustering* analisis dilakukan dengan menggunakan *k-means*, *adaptive k-means* dan *fuzzy c-means*.

### Metode Agglomerative Hierarchical

#### Single Linkage

Metode ini menggunakan jarak terdekat antara sebuah objek dari suatu *cluster* dengan sebuah objek dari *cluster* lainnya (Johnson & Wichern, 2007). Objek-objek dengan jarak terdekat dikelompokkan menjadi satu *cluster*.

$$d_{(UV)W} = \min\{d_{UW}, d_{VW}\} \quad (2)$$

dimana  $d_{(UV)W}$  adalah jarak *cluster* (UV) dengan *cluster* W;  $d_{UW}$  adalah jarak tetangga terdekat antara U dan W; dan  $d_{VW}$  adalah jarak tetangga terdekat antara V dan W.

#### Complete Linkage

Metode ini menggunakan jarak terjauh antara sebuah objek dari suatu *cluster* dengan sebuah objek dari *cluster* lainnya (Johnson & Wichern, 2007). Objek-objek dengan jarak terjauh dikelompokkan menjadi satu *cluster*.

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \quad (3)$$

dimana  $d_{(UV)W}$  adalah jarak *cluster* (UV) dengan *cluster* W;  $d_{UW}$  adalah jarak tetangga terjauh antara U dan W; dan  $d_{VW}$  adalah jarak tetangga terjauh antara V dan W.

#### Average Linkage

Metode ini menggunakan jarak rata-rata antara objek-objek dari suatu *cluster* dengan objek-objek dari *cluster* lainnya (Johnson & Wichern, 2007). Pengelompokan objek-objek ini didasarkan pada pasangan objek yang memiliki jarak terdekat dengan jarak rata-rata.

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{UV} N_W} \quad (4)$$

dimana  $d_{(UV)W}$  adalah jarak *cluster* (UV) dengan *cluster* W;  $d_{ik}$  adalah jarak antara objek  $i$  pada *cluster* (UV) dan objek  $k$  pada *cluster* W;  $N_{UV}$  adalah jumlah objek pada *cluster* (UV); dan  $N_W$  adalah jumlah objek pada *cluster* W.

#### Ward

Pada metode ini jarak antara dua *cluster* dihitung menggunakan jumlah kuadrat pada setiap variabelnya (Johnson & Wichern, 2007). Metode ini

mempertimbangkan keragaman dari *cluster* dengan meminimumkan nilai *Sum of Squared Error (SSE)*, dimana dua *cluster* yang memiliki peningkatan nilai SSE terkecil akan digabungkan.

$$SSE = \sum_{j=1}^N (x_j - \bar{x})'(x_j - \bar{x}) \quad (5)$$

dimana  $x_j$  adalah data dari objek ke- $j$  dan  $\bar{x}$  adalah rata-rata dari semua objek dalam *cluster*.

## Metode Non-Hierarchical

### **K-Means Clustering**

*K-means clustering* merupakan metode yang digunakan untuk mengelompokkan objek ke dalam  $K$  *cluster* (Pramana, Yuniarto, Mariyah, Santoso, & Nooraeni, 2018). Setiap objek akan dikelompokkan pada suatu *cluster* berdasarkan titik pusat (*centroid*) *cluster* yang terdekat dengan objek tersebut. Setiap *cluster* yang terbentuk akan diwakili oleh nilai rata-rata dari objek yang terdapat di dalam *cluster* (Han & Kamber, 2006). Tahapan pembentukan *cluster* menggunakan *k-means clustering* adalah sebagai berikut (Han & Kamber, 2006):

1. Menginisiasi atau menentukan jumlah  $k$  *cluster* awal yang diinginkan.
2. Setiap objek akan dimasukkan ke dalam *cluster* tertentu berdasarkan rata-rata terdekat. Penghitungan jarak biasanya menggunakan jarak *Euclidean*.
3. Lakukan perhitungan rata-rata kembali untuk *cluster* yang mendapatkan objek baru dan *cluster* yang kehilangan objek.
4. Ulangi langkah kedua sampai tidak ada lagi objek antar *cluster* yang berpindah.

*K-means clustering* merupakan algoritma yang sensitif terhadap *outlier* atau atau jarak data yang terlalu timpang dan *noise*. Ini disebabkan terdapat jarak observasi yang terlalu jauh dari sekumpulan data yang dapat merubah rata-rata dari *cluster*. Hal ini dapat memengaruhi hasil *clustering* yang dapat menyebabkan pengelompokan yang tidak tepat.

### **K-Medoids Clustering**

*K-Medoids* merupakan pengembangan dari metode *k-means clustering*. Jika *k-means* merupakan algoritma yang sensitif terhadap *outlier* atau jarak data yang terlalu timpang dan *noise* maka metode *k-medoids* dapat mengatasi hal tersebut. Hal ini dikarenakan *k-medoids* menggunakan nilai median dari data yang tidak terpengaruh walaupun jika ada *outlier* dan *noise* dalam data sehingga metode ini lebih *robust* dibandingkan *k-means* (Han & Kamber, 2006). Setiap *cluster* yang terbentuk akan diwakili oleh salah satu objek yang terletak di dekat pusat *cluster*. Pada metode ini, partisi dilakukan dengan prinsip meminimalkan jumlah ketidaksamaan antara setiap objek  $p$  dan objek  $c_i$  perwakilan yang sesuai. Tahapan pembentukan *cluster* menggunakan *k-medoids clustering* adalah sebagai berikut (Han & Kamber, 2006):

1. Menginisiasi atau menentukan jumlah  $k$  *cluster* awal yang diinginkan.

2. Bagi setiap objek ke dalam *cluster* berdasarkan objek representatif yang memiliki jarak terdekat.
3. Memilih secara acak objek baru yang berfungsi sebagai *medoid*.
4. Gunakan set *medoids* baru untuk menghitung ulang biaya.
5. Jika biaya < 0 ganti *medoid* ke-*j* dengan memilih secara *medoid* acak dari dataset.
6. Ulangi langkah kedua hingga kelima sampai tidak ada perubahan dalam *medoid* (tidak ada objek yang berpindah).

### Fuzzy C-Means (FCM)

Menurut J. C. Bezdek (1981) algoritma *fuzzy c-means clustering* merupakan ekuivalen *fuzzy* dari algoritma *hard clustering* (Xie & Beni, 1991). Algoritma ini berguna untuk meminimalkan fungsi tujuan sehubungan dengan *fuzzy membership*  $\mu_{ij}$  dan *cluster centroid*  $V_i$ . Eksponen  $m$  dalam *fuzzy c-means* mewakili parameter *fuzziness* yang bertujuan untuk mengontrol bobot dalam fungsi jarak  $d(X_j, V_i)$  dan mendefinisikan seberapa janggal atau tidak jelasnya hasil yang diperoleh. Ketika jumlah  $m = 1$  berarti hasil yang diperoleh jelas, namun ketika  $m$  mendekati tak hingga maka terdapat ketidakjelasan yang sempurna. Tidak ada bukti teoritis dan matematis dalam pemilihan jumlah  $m$  terbaik, namun secara umum biasanya menggunakan nilai antara 1 dan 3 untuk mendapatkan hasil yang lebih tepat (Grekousis & Thomas, 2012). Sementara itu, untuk klasifikasi kebanyakan peneliti menggunakan nilai  $m$  antara 1,5 sampai 3 untuk mendapatkan hasil yang bagus.

$$J_m = \sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m d(X_j, V_i)^2 \quad (6)$$

dimana  $\mu_{ij}$  adalah nilai keanggotaan objek ke-*j* *cluster* ke-*i*;  $d(X_j, V_i)$  adalah jarak titik objek ke-*j* dengan *cluster centroid* ke-*i*;  $V_i$  adalah *cluster centroid* ke-*i*;  $X_j$  adalah objek ke-*j*;  $n$  adalah jumlah data; dan  $m$  adalah parameter *fuzziness* atau parameter ukuran kesamaran hasil *clustering* ( $m > 1$ ).

Adapun tahapan algoritma FCM adalah sebagai berikut:

- a. Menginisialisasi *membership*  $\mu_{ij}$  dari objek  $X_j$  pada *cluster* ke-*i*.

$$\sum_{i=1}^c \mu_{ij} = 1 \quad (7)$$

- b. Menghitung *fuzzy centroid* ( $V_i$ ).

$$V_i = \frac{\sum_{j=1}^n (\mu_{ij})^m X_j}{\sum_{j=1}^n (\mu_{ij})^m} \quad (8)$$

- c. Melakukan *updating fuzzy membership*.

$$\mu_{ij} = \frac{\left( \frac{1}{d^2(X_j, V_i)} \right)^{\frac{1}{m-1}}}{\sum_{i=1}^c \left( \frac{1}{d^2(X_j, V_i)} \right)^{\frac{1}{m-1}}} \quad (9)$$

d. Ulangi tahap 2 dan 3 sampai tidak ada nilai  $J_m$  yang menurun.

**Validasi Cluster.**

**a. Cophenetic Correlation Coefficient**

Koefisien korelasi *cophenetic* adalah korelasi antara elemen asli matriks *dissimilarity* dan elemen-elemen dari matriks yang dihasilkan dendrogram (Silva & Dias, 2013). Koefisien korelasi *cophenetic* dapat dihitung menggunakan rumus berikut:

$$r_{coph} = \frac{\sum_{i=1}^{n-1} \sum_{j>i}^n (c_{ij} - \bar{c}) (d_{ij} - \bar{d})}{\left(\sum_{i=1}^{n-1} \sum_{j>i}^n (c_{ij} - \bar{c})^2\right)^{\frac{1}{2}} \left(\sum_{i=1}^{n-1} \sum_{j>i}^n (d_{ij} - \bar{d})^2\right)^{\frac{1}{2}}} \quad (10)$$

dengan

$$\bar{c} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j>i}^n c_{ij} \quad (11)$$

$$\bar{d} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j>i}^n d_{ij} \quad (12)$$

dimana  $r_{coph}$  adalah koefisien korelasi *cophenetic*;  $n$  adalah jumlah observasi;  $d_{ij}$  adalah jarak (jarak Euclidean) antara objek  $i$  dan  $j$ ;  $\bar{d}$  adalah rata-rata jarak (jarak Euclidean) objek;  $c_{ij}$  adalah jarak *cophenetic* antara objek  $i$  dan  $j$ ; dan  $\bar{c}$  adalah rata-rata jarak *cophenetic* objek.

**b. Agglomerative Coefficient**

Koefisien *agglomerative* menggambarkan kekuatan struktur pengelompokan yang bernilai antara 0 sampai 1 (Kaufman & Rousseeuw, 1990). Semakin mendekati 1 struktur pengelompokan yang diperoleh semakin jelas. Namun, ketika jumlah data ( $n$ ) meningkat koefisien *agglomerative* cenderung menjadi lebih besar, sehingga metode ini tidak baik digunakan untuk membandingkan kumpulan data dengan ukuran yang sangat berbeda. Adapun rumus yang digunakan untuk menghitung koefisien *agglomerative* yaitu (Kaufman & Rousseeuw, 1990):

$$AC = \frac{1}{n} \sum_{i=1}^n l(i) \quad (13)$$

dimana AC adalah koefisien *agglomerative*;  $n$  adalah jumlah data atau observasi; dan  $l(i)$  adalah panjang garis label untuk setiap observasi.

**c. Connectivity Coefficient**

Nilai koefisien *connectivity* berada pada rentang nol (0) sampai tak hingga ( $\infty$ ) (Brock, Pihur, Datta, & Datta, 2008). Semakin rendah nilai koefisien *connectivity* maka *cluster* yang terbentuk semakin baik. Adapun rumus yang digunakan untuk menghitung koefisien *connectivity* yaitu:

$$Conn(C) = \sum_{i=1}^N \sum_{j=1}^L x_{i,nn_{i(j)}} \quad (14)$$

dimana  $Conn(C)$  adalah koefisien *agglomerative*;  $x_{i,nn_{i(j)}}$  adalah nol jika  $i$  dan  $j$  berada dalam *cluster* yang sama dan  $1/j$  jika sebaliknya;  $nn_{i(j)}$  adalah tetangga terdekat objek ke- $j$  dari objek ke- $i$ ;  $N$  adalah jumlah data atau observasi; dan  $L$  adalah parameter yang menentukan jumlah tetangga terdekat untuk digunakan.

#### d. Silhouette Coefficient

Koefisien *silhouette* berguna untuk mengukur tingkat kepercayaan dalam proses pengelompokan (*cluster*) suatu observasi (Brock, Pihur, Datta, & Datta, 2008). Nilai koefisien *silhouette* berada pada rentang antara -1 sampai 1. *Cluster* dengan nilai koefisien *silhouette* mendekati 1 dikategorikan baik dan jika mendekati -1 suatu *cluster* akan dikategorikan buruk (tidak baik). Rumus yang digunakan untuk menghitung koefisien *silhouette* yaitu (Brock, Pihur, Datta, & Datta, 2008):

$$S(i) = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (15)$$

$$a_i = \frac{1}{n(C(i))} \sum_{j \in C(i)} dist(i, j) \quad (16)$$

$$b_i = \min_{C_k \in C} \sum_{j \in C_k} \frac{dist(i, j)}{n(C_k)} \quad (17)$$

dimana  $a_i$  adalah jarak rata-rata antara objek  $i$  dengan semua objek dalam *cluster* yang sama;  $b_i$  adalah jarak rata-rata antara objek  $i$  dengan objek pada *cluster* tetangga terdekat;  $C(i)$  adalah *cluster* dengan observasi ke- $i$ ;  $dist(i, j)$  adalah jarak *dissimilarity* (Euclidean, Manhattan, dsb) antara objek  $i$  dan  $j$ ; dan  $n(C)$  adalah kardinalitas *cluster*  $C$ .

#### e. Dunn Index

Indeks *dunn* merupakan rasio jarak terkecil antara observasi yang pada *cluster* yang berbeda dengan jarak intra *cluster* terbesar (Brock, Pihur, Datta, & Datta, 2008). Indeks *dunn* berada pada rentang antara 0 sampai tak hingga ( $\infty$ ). Semakin besar nilai indeks *dunn* pada suatu *cluster*, semakin baik hasil *clustering*-nya. Rumus yang digunakan untuk menghitung indeks *dunn* yaitu (Brock, Pihur, Datta, & Datta, 2008):

$$D(C) = \frac{\min_{C_k, C_l \in C, C_k \neq C_l} \left( \min_{i \in C_k, j \in C_l} dist(i, j) \right)}{\max_{C_m \in C} (diam(C_m))} \quad (18)$$

dimana  $D(C)$  adalah indeks *dunn*; dan  $diam(C_m)$  adalah jarak maksimum antar observasi dalam *cluster*  $C_m$ .

#### f. Davies Bouldin Index

Indeks *davies bouldin* adalah suatu ukuran yang menunjukkan kesamaan *cluster* yang kerapatan datanya merupakan fungsi penurunan jarak dari karakteristik vektor *cluster* (Davies & Bouldin, 1979). Semakin kecil nilai indeks *davies bouldin* pada suatu *cluster*, semakin baik hasil *clustering*-nya. Dengan meminimalkan

indeks ini terdapat perbedaan yang jelas antara satu *cluster* dengan *cluster* lainnya sehingga dapat mencapai partisi terbaik. Rumus yang digunakan untuk menghitung indeks *davies bouldin* yaitu:

$$DB = \frac{1}{k} \sum_{p=1}^k R_p \quad (19)$$

dimana DB adalah indeks *davies bouldin*; dan  $R_p$  adalah ukuran kesamaan *cluster* (maksimum).

#### g. *Calinski-Harabasz Index*

Indeks *calinski-harabasz* merupakan rasio antara nilai *Sum of Square between cluster* (SSB) dan nilai *Sum of Square within-cluster* (SSW) yang dikalikan dengan faktor normalisasi (Baarsch & Celebi, 2012). Faktor normalisasi ini akan mengurangi nilai indeks ketika jumlah *cluster* ( $k$ ) bertambah. Semakin besar nilai indeks *calinski-harabasz* pada suatu *cluster*, semakin baik hasil *clustering*-nya. Rumus yang digunakan untuk menghitung indeks *calinski-harabasz* yaitu (Baarsch & Celebi, 2012):

$$CH = \frac{\text{trace}(SSB)}{\text{trace}(SSW)} \times \frac{N - k}{k - 1} \quad (20)$$

dimana CH adalah indeks *calinski-harabasz*; N adalah jumlah observasi; dan K adalah jumlah *cluster*.

#### h. *Xie and Beni Index*

Indeks *xie and beni* mengukur rasio total *varians within cluster* dan *separation* pada *cluster* (Xie & Beni, 1991). Indeks *xie and beni* juga didefinisikan sebagai hasil bagi antara kesalahan kuadrat rata-rata dan jarak kuadrat minimal antara titik-titik dalam *cluster* (Desgraupes, 2017). Semakin kecil nilai indeks *xie and beni* semakin baik hasil *clustering*-nya. Rumus yang digunakan untuk menghitung indeks *xie and beni* yaitu (Xie & Beni, 1991):

$$XB = \frac{\sum_{i=1}^m \sum_{k=1}^N \mu_{ik}^C |x_k - u_i|^2}{N \min_{i,k} |x_k - u_i|^2} \quad (21)$$

### 3. Hasil dan Pembahasan

#### 3.1 Pengecekan Asumsi

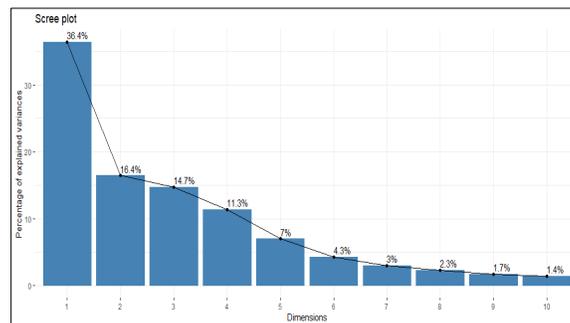
Pada analisis *cluster* terdapat asumsi yang harus dipenuhi terlebih dahulu yaitu asumsi non-multikolinearitas. Berdasarkan pengujian korelasi pada setiap indikator diketahui bahwa terdapat indikator yang memiliki nilai korelasi lebih dari 0,85 yaitu ketidakcukupan konsumsi pangan (PoU) dengan indikator penolong kesehatan ditolong tenaga kesehatan (0,86) dan antara indikator PoU dengan kerawanan pangan (0,87). Hal ini menunjukkan bahwa terdapat multikolinearitas pada data, sehingga hal ini perlu untuk ditangani agar analisis *cluster* dapat dilakukan.

Selanjutnya akan dilakukan analisis menggunakan PCA. Tahapan pertama yaitu pembentukan komponen utama PCA yang terdiri dari kombinasi linear variabel dalam data. Penentuan jumlah komponen utama didasarkan pada *eigenvalue*, nilai proporsi

kumulatif varians dan *scree plot*. Berdasarkan Tabel 2 terlihat bahwa terdapat empat komponen yang memiliki nilai eigen > 1.

Tabel 2: Hasil analisis komponen PCA

Komponen Utama	Eigen Value	% Variance	% Cumulative Variance
Komponen 1	4,37	36,39	36,39
Komponen 2	1,97	16,44	52,84
Komponen 3	1,77	14,72	67,56
Komponen 4	1,36	11,34	78,89
Komponen 5	0,84	6,99	85,89
Komponen 6	0,52	4,31	90,19
Komponen 7	0,36	3,02	93,21
Komponen 8	0,28	2,34	95,56
Komponen 9	0,20	1,71	97,26
Komponen 10	0,17	1,43	98,69
Komponen 11	0,09	0,78	99,47
Komponen 12	0,06	0,53	100

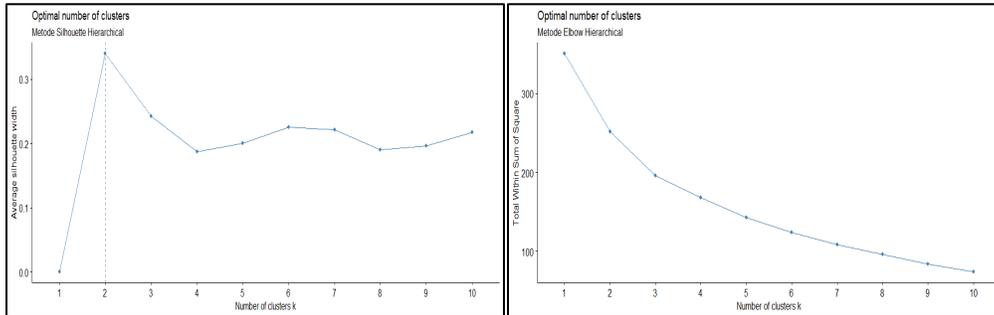


Gambar 1: Scree plot komponen PCA

Namun, jika dilihat dari nilai proporsi kumulatif varians keempat komponen tersebut nilainya belum mencapai 80%, yaitu sebesar 78,89%, sehingga perlu ditambahkan satu atau beberapa komponen lagi. Berdasarkan tabel terlihat bahwa komponen utama sudah mencapai nilai proporsi kumulatif varians 80% yaitu saat menggunakan lima komponen utama yaitu sebesar 85,89%. Hal ini berarti bahwa keragaman data dapat dijelaskan sebesar 85,89% sehingga cukup untuk menggambarkan struktur data. Kemudian, jika dilihat menggunakan *scree plot*, posisi titik kurva sudah mulai melandai saat lima komponen utama. Dengan demikian, pada penelitian ini digunakan lima kompoen utama untuk analisis selanjutnya.

### 3.2 Agglomerative Hierarchical

Selanjutnya dilakukan analisis *hierarchical clustering* menggunakan metode *single linkage*, *average linkage*, *complete linkage* dan metode *ward*. Penentuan jumlah *cluster* optimal dilakukan menggunakan metode *elbow*, *silhouette* dan *package Nbclust*. Berdasarkan plot dari metode *silhouette* terlihat bahwa jumlah *cluster* optimal yang terbentuk yaitu sebanyak dua *cluster*. Kemudian, pada metode *elbow* jumlah *cluster* optimal yang terbentuk yaitu sebanyak empat *cluster* yang dilihat berdasarkan pada patahan *gradient* terbesarnya.



Gambar 2: Plot penentuan jumlah *cluster* optimal metode *hierarchical clustering*

Berdasarkan *package* Nbclust jumlah *cluster* optimal yang terbentuk untuk setiap metode yaitu dua *cluster* dengan rekomendasi terbanyak dan empat *cluster* dengan rekomendasi kedua terbanyak. Namun, jika menggunakan dua *cluster* sebaran data pada setiap *cluster* terlalu timpang. Oleh karena itu, pada metode *hierarchical* jumlah *cluster* optimal yang dipilih untuk analisis selanjutnya yaitu menggunakan empat *cluster*.

Tabel 3: Penentuan jumlah *cluster* optimal menggunakan *package* Nbclust

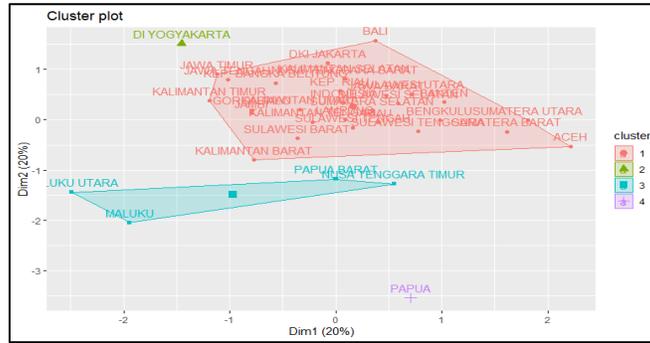
Jumlah <i>Cluster</i>	Jumlah Rekomendasi Menggunakan Nbclust					
	Single	Complete	Average	Ward	K-Means	K-Medoids
2	10	7	10	11	10	14
3	4	7	2	7	11	8
4	2	7	9	4	-	1
5	7	2	3	2	3	1

Setelah mendapatkan jumlah *cluster* optimal sebanyak empat *cluster*, dilakukan uji validasi pada setiap metode untuk menentukan *cluster* terbaik dari metode *hierarchical clustering*. Uji validasi yang digunakan yaitu *cophenetic correlation coefficient*, *agglomerative coefficient* (AC), *connectivity coefficient* (CC), *dunn index* (DI), dan *silhouette coefficient* (SC).

Tabel 4: Uji validasi metode *hierarchical clustering*

Metode	Cophenetic	AC	CC	DI	SC
Single	0,80	0,63	8,99	0,37	0,20
Complete	0,66	0,81	30,10	0,19	0,15
Average	0,83	0,73	15,41	0,39	0,24
Ward	0,65	0,85	29,43	0,21	0,21

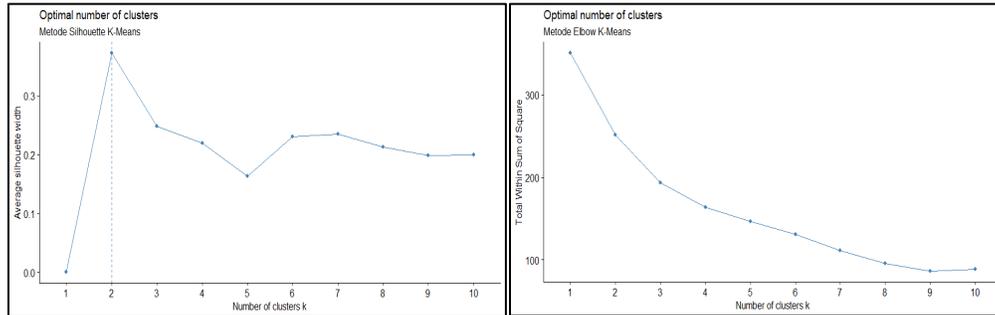
Berdasarkan Tabel 4, diketahui bahwa metode *average linkage* memiliki nilai koefisien korelasi *cophenetic*, *dunn index* dan koefisien *silhouette* terbaik dibandingkan metode lainnya. Oleh karena itu, metode *average linkage* merupakan *hierarchical clustering* terbaik dan digunakan untuk analisis selanjutnya. Berikut visualisasi *cluster* menggunakan metode *average linkage* dengan empat *cluster* yang disajikan pada Gambar 3.



Gambar 3: Visualisasi *average linkage* dengan empat *cluster*

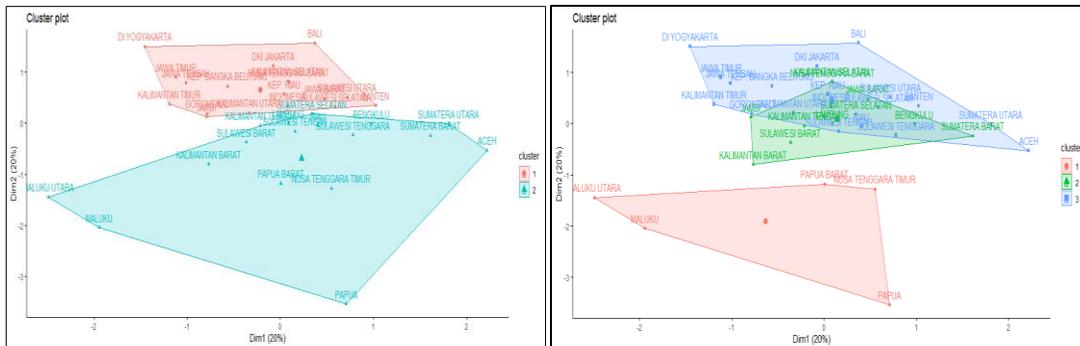
### 3.3 K-Means Clustering

Pada metode *k-means clustering* setiap objek dikelompokkan berdasarkan titik pusat (*centroid*) *cluster* yang terdekat dengan objek tersebut. Tahapan pertama yang dilakukan yaitu penentuan jumlah *cluster* optimum menggunakan metode *elbow*, *silhouette* dan *package Nbclust*.



Gambar 4: Plot penentuan jumlah *cluster* optimal metode *k-means*

Berdasarkan plot dari metode *silhouette* terlihat bahwa jumlah *cluster* optimal yang terbentuk yaitu sebanyak dua *cluster*, sedangkan pada metode *elbow* jumlah *cluster* optimal yang terbentuk yaitu sebanyak tiga *cluster*. Berdasarkan *Nbclust* jumlah *cluster* optimal yang terbentuk yaitu tiga *cluster* dengan rekomendasi terbanyak dan dua *cluster* dengan rekomendasi kedua terbanyak.

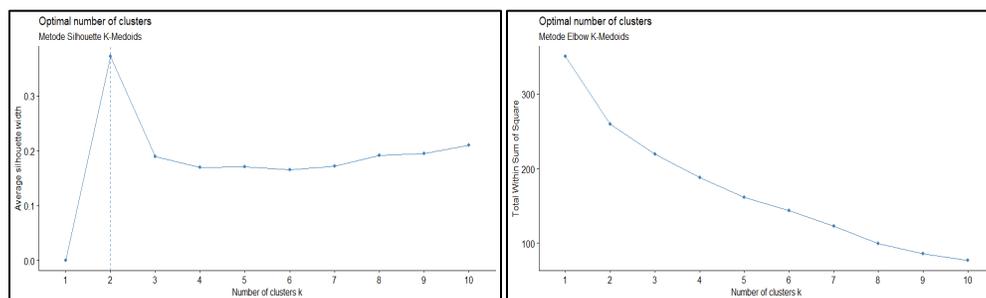


Gambar 5: Visualisasi *k-means* dengan dua *cluster* dan tiga *cluster*

Namun, ketika diplot hasil pengelompokan dengan tiga *cluster* terdapat provinsi yang saling tumpang tindih antar *cluster*-nya. Sementara itu, jika diplot menggunakan dua *cluster* hasil pengelompokannya jauh lebih baik dan tidak terdapat provinsi yang tumpang tindih. Dengan demikian, jumlah *cluster* optimal yang dipilih untuk analisis selanjutnya menggunakan metode *k-means* yaitu dua *cluster*.

### 3.4 K-Medoids Clustering (PAM)

Pada *k-medoids clustering* setiap *cluster* yang terbentuk diwakili oleh salah satu objek yang terletak di dekat pusat *cluster*. Sama seperti metode sebelumnya, tahapan pertama yang dilakukan yaitu penentuan jumlah *cluster* optimum menggunakan metode *elbow*, *silhouette* dan *package Nbclust*. Berdasarkan plot dari metode *silhouette* dan *elbow* terlihat bahwa jumlah *cluster* optimal yang terbentuk yaitu sebanyak dua *cluster*.

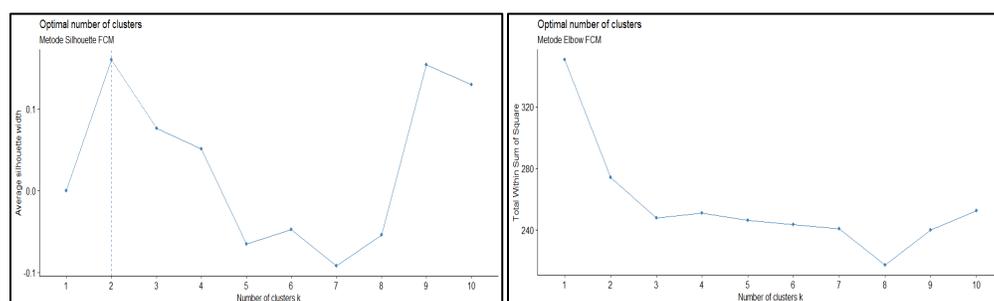


Gambar 6: Plot penentuan jumlah *cluster* optimal metode *k-medoids*

Berdasarkan *Nbclust* jumlah *cluster* optimal yang terbentuk untuk setiap metode yaitu dua *cluster* dengan rekomendasi terbanyak dan tiga *cluster* dengan rekomendasi kedua terbanyak. Oleh karena itu, jumlah *cluster* optimal yang dipilih untuk analisis selanjutnya dengan metode *k-medoids* yaitu dua *cluster*.

### 3.5 Fuzzy C-Means (FCM)

Penentuan jumlah *cluster* optimal pada metode FCM dilakukan menggunakan metode *elbow*, *silhouette* dan ukuran evaluasi. Ukuran evaluasi yang digunakan di antaranya yaitu *davies bouldin index* (DBI), *silhouette index* (SI), *calinski-harabasz index* (CHI), *xie and beni index* (XBI), dan *dunn index* (DI). Berdasarkan plot dari metode *silhouette* dan *elbow* terlihat bahwa jumlah *cluster* optimal yang terbentuk yaitu sebanyak dua *cluster*.



Gambar 7. Plot penentuan jumlah *cluster* optimal metode FCM

Jika menggunakan ukuran evaluasi seperti yang disajikan pada Tabel 5 diketahui bahwa jumlah *cluster* optimal yang terbentuk pada metode FCM yaitu sebanyak tiga *cluster*. Ditinjau kembali pada metode *Silhouette* diketahui bahwa jumlah *cluster* tertinggi kedua yaitu saat tiga *cluster*. Oleh karena itu, jumlah *cluster* optimal yang dipilih untuk analisis selanjutnya dengan metode FCM yaitu sebanyak tiga *cluster*.

Tabel 5: Uji validasi metode FCM

Metode	Jumlah Cluster (k)			
	2	3	4	5
DBI	1,69	1,39	1,30	1,39
SI	0,17	0,20	0,18	0,14
CHI	9,90	12,11	11,37	9,60
XBI	4,49	1,54	3,15	2,90
DI	0,16	0,30	0,19	0,21

### 3.6 Evaluasi Metode Terbaik

Setelah melakukan analisis pada setiap metode *cluster*, selanjutnya dilakukan evaluasi terhadap semua metode *hierarchical* dan *non-hierarchical clustering*. Hal ini bertujuan untuk menentukan metode *cluster* terbaik dalam mengelompokkan kasus penanganan *stunting* pada provinsi-provinsi di Indonesia. Ukuran evaluasi yang digunakan yaitu *connectivity coefficient* (CC), *dunn index* (DI), *silhouette coefficient* (SC), *davies bouldin index* (DBI), *silhouette index* (SI), *calinski-harabasz index* (CHI), dan *xie and beni index* (XBI) yang dapat dilihat pada Tabel 6.

Berdasarkan evaluasi yang telah dilakukan, dapat disimpulkan bahwa metode terbaik dalam mengelompokkan tingkat penanganan *stunting* pada provinsi-provinsi di Indonesia pada tahun 2018 yaitu menggunakan *average linkage* dengan empat *cluster*.

Tabel 6: Evaluasi metode terbaik

Metode	Jumlah Cluster (k)	CC	DI	SC	DBI	CHI	XBI
<i>Average Linkage</i>	4	15,4056	0,3850	0,2387	0,8840	6,8872	0,9261
<i>K-Means</i>	2	10,1933	0,3567	0,3722	1,1358	13,0889	1,1062
<i>K-Medoids</i>	2	8,0976	0,3328	0,3716	1,0256	11,5720	1,1426
<i>Fuzzy C-Means (FCM)</i>	3	33,5369	0,1485	0,0763	1,3880	12,1100	1,5376

Hal ini dapat dilihat bahwa metode *average linkage* memiliki tiga indeks dengan kriteria terbaik dibandingkan metode lainnya. Indeks tersebut terdiri dari *dunn index* dengan nilai tertinggi, sedangkan nilai *davies bouldin index* dan *xie and beni index* merupakan nilai terkecil di antara metode lainnya. Kemudian, metode terbaik berdasarkan *silhouette coefficient*, *calinski-harabasz index connectivity* yaitu metode *k-means* dengan dua *cluster* dan untuk *connectivity coefficient* metode terbaiknya yaitu *k-medoids* dengan dua *cluster*.

### 3.7 Pengelompokan Provinsi dan *Profiling Hasil Analisis Cluster*

Berikut merupakan pengelompokan provinsi di Indonesia berdasarkan indikator-indikator IKPS tahun 2018 yang disajikan pada Tabel 7. Pengelompokan dilakukan menggunakan metode *average linkage* dengan 4 *cluster*. Berdasarkan pengelompokan tersebut diketahui bahwa *cluster* satu terdiri dari 28 provinsi, satu provinsi di *cluster* dua, empat provinsi di *cluster* tiga, dan satu provinsi di *cluster* empat.

Tabel 7: Hasil pengelompokan menggunakan metode *average linkage*

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

Untuk mengetahui kategori dari setiap *cluster* digunakan nilai rata-rata pada setiap indikator IKPS. Nilai rata-rata pada setiap indikator ini dijadikan sebagai dasar dalam pengkategorian *cluster*. Berikut disajikan nilai rata-rata setiap indikator IKPS di masing-masing *cluster* yang dapat dilihat pada Tabel 8. Berdasarkan nilai rata-rata setiap indikator, diperoleh hasil bahwa *cluster* kedua merupakan *cluster* yang memiliki tingkat penanganan *stunting* sangat baik dibandingkan *cluster* lainnya. Hal ini dilihat dari banyaknya indikator yang memiliki nilai rata-rata tertinggi di *cluster* kedua. Misalnya indikator imunisasi, penolong persalinan oleh tenaga kesehatan, ASI eksklusif, air minum layak, sanitasi layak, APK, PAUD, dan penerima KPS/KKS. Kemudian, nilai rata-rata pada indikator ketidakcukupan konsumsi pangan *cluster* dua memiliki nilai yang rendah dibandingkan *cluster* lainnya. Hal ini menggambarkan bahwa penanganan masalah konsumsi pada provinsi di *cluster* dua sudah baik. Namun, *cluster* dua juga memiliki satu indikator dengan kriteria masih sangat buruk yaitu indikator MPASI. Hal ini perlu diatasi dan ditingkatkan lagi agar kasus *stunting* dapat menurun.

Kemudian, *cluster* pertama merupakan *cluster* yang memiliki tingkat penanganan *stunting* yang baik dengan nilai rata-rata terbaik kedua setelah *cluster* dua. *Cluster* satu memiliki dua indikator dengan kriteria sangat baik (indikator penggunaan KB modern dan mengalami kerawanan pangan) dan lima indikator dengan kriteria baik (Imunisasi, penolong persalinan oleh tenaga kesehatan, sanitasi layak, ketidakcukupan konsumsi pangan, dan APK PAUD). *Cluster* satu terdiri dari 28 provinsi yang dapat dilihat pada Tabel 7. Hal ini menunjukkan bahwa tingkat penanganan *stunting* di Indonesia sudah dalam kategori baik.

*Cluster* tiga merupakan *cluster* dengan kategori tingkat penanganan *stunting* buruk. *Cluster* tiga memiliki delapan indikator dengan kriteria buruk yaitu indikator imunisasi, penolong persalinan oleh tenaga kesehatan, penggunaan KB modern, sanitasi layak, mengalami kerawanan pangan, ketidakcukupan konsumsi pangan, APK

PAUD, dan jaminan kesehatan. Terdapat 4 provinsi dalam *cluster* tiga yaitu Nusa Tenggara Timur (NTT), Maluku, Maluku Utara, dan Papua Barat.

Tabel 8: Rata-rata indikator IKPS menggunakan metode *average linkage*

Cluster	1	2	3	4
Imunisasi	58.60	87.13	42.04	30.45
Nakes	93.81	100.00	76.08	61.55
KB Modern	56.31	52.30	40.94	17.84
ASI	44.55	55.70	49.46	43.48
MPASI	86.38	84.74	87.08	86.70
Air Minum	73.10	80.62	73.79	58.35
Sanitasi	69.30	88.92	65.17	33.75
PoU	7.83	8.74	24.99	38.49
Rawan_Pangan	31.61	18.09	51.89	56.27
APK_PAUD	35.97	72.00	30.97	13.01
JKN_Jamkesda	26.82	34.50	31.32	43.26
KPS_KKS	12.87	21.36	15.60	15.52

Keterangan :

- : Indikator dengan kriteria sangat baik
- : Indikator dengan kriteria baik
- : Indikator dengan kriteria buruk
- : Indikator dengan kriteria sangat buruk

Sementara itu, *cluster* yang memiliki tingkat penanganan *stunting* yang sangat buruk terdapat pada *cluster* empat. Pada *cluster* empat terlihat bahwa nilai rata-rata dari indikator imunisasi, penolong persalinan oleh tenaga kesehatan, penggunaan KB Modern, ASI eksklusif, air minum layak, sanitasi layak, dan APK PAUD memiliki nilai rata-rata terendah dibandingkan *cluster* lainnya. Sementara itu, indikator ketidakcukupan konsumsi pangan dan kerawanan pangan pada *cluster* empat memiliki nilai yang tertinggi dibandingkan *cluster* lainnya. Hal ini menunjukkan bahwa tingkat penanganan *stunting* di Provinsi Papua perlu menjadi prioritas utama bagi pemerintah dalam upaya penurunan kasus *stunting* di Indonesia.

#### 4. Simpulan

Berdasarkan analisis yang telah dilakukan diperoleh bahwa metode terbaik dalam pengelompokan provinsi di Indonesia berdasarkan indikator-indikator IKPS tahun 2018 yaitu menggunakan metode *average linkage* dengan empat *cluster*. Metode *average linkage* terpilih sebagai metode terbaik setelah dilakukan beberapa pengujian atau evaluasi validasi. Berdasarkan hasil evaluasi validasi tersebut diketahui bahwa metode *average linkage* dengan empat *cluster* memiliki nilai evaluasi yang lebih baik dibandingkan metode *cluster* lainnya. Terdapat empat *cluster* yang terbentuk berdasarkan hasil pengelompokan tersebut. *Cluster* 1 merupakan *cluster* dengan

kategori tingkat penanganan *stunting* baik yang terdiri dari 28 provinsi. *Cluster 1* memiliki kriteria terbaik kedua setelah *cluster* dua yaitu memiliki dua indikator dengan kriteria sangat baik dan lima indikator dengan kriteria baik. Tingkat penanganan *stunting* pada provinsi-provinsi di *cluster* satu perlu dipertahankan dan lebih ditingkatkan lagi. *Cluster 2* merupakan *cluster* dengan kategori tingkat penanganan *stunting* sangat baik yang terdiri dari satu provinsi yaitu DI Yogyakarta. Terdapat delapan indikator dengan kriteria terbaik dari 12 indikator IKPS 2018 pada *cluster 2*. Tingkat penanganan *stunting* di Provinsi DI Yogyakarta sudah baik, namun perlu dipertahankan dan lebih ditingkatkan lagi agar tidak terjadi penurunan. *Cluster 3* merupakan *cluster* dengan kategori tingkat penanganan *stunting* buruk yang terdiri dari empat provinsi. Terdapat delapan indikator dengan kriteria buruk di *cluster 3* dibandingkan *cluster* lainnya. Hal ini menunjukkan bahwa tingkat penanganan *stunting* pada keempat provinsi tersebut perlu diperhatikan dan diatasi dengan segera oleh pemerintah agar terjadi penurunan kasus *stunting* di Indonesia. *Cluster 4* merupakan *cluster* dengan kategori tingkat penanganan *stunting* sangat buruk yang terdiri dari satu provinsi yaitu Provinsi Papua. Terdapat sembilan indikator dengan kriteria terburuk dari 12 indikator IKPS 2018 pada *cluster 4*. Hal ini menunjukkan bahwa Tingkat penanganan *stunting* pada Provinsi Papua harus diperhatikan dan perlu menjadi prioritas utama bagi pemerintah dalam upaya penurunan kasus *stunting* di Indonesia.

## Daftar Pustaka

- Baarsch, J., & Celebi, M. E. (2012). Investigation of Internal Validity Measures for K-Means Clustering.
- Badan Ketahanan Pangan Kementerian Pertanian. (2020). *Statistik Ketahanan Pangan 2019*. Jakarta: Kementerian Pertanian.
- Bappenas. (2017). *Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2015-2019*. Jakarta: Bappenas.
- Bappenas. (2019). *Rancangan Teknokratik : Rencana Pembangunan Jangka Menengah Nasional 2020-2024*. Jakarta: Bappenas.
- Bappenas. (2020). *Pedoman Teknis Penyusunan Rencana Aksi Tujuan Pembangunan Berkelanjutan (TPB)/ Sustainable Development Goals (SDGs)*. Jakarta: Bappenas.
- BPS. (2018). *Statistik Kesejahteraan Rakyat*. Jakarta: BPS.
- BPS. (2019). *Statistik Kesejahteraan Rakyat*. Jakarta: BPS.
- BPS. (2020). *Profil Kesehatan Ibu dan Anak*. Jakarta: BPS.
- Brock, G., Pihur, V., Datta, S., & Datta, S. (2008, March). clValid: An R Package for Cluster Validation. *Journal of Statistical Software*, 25(4), 1–22.

- Chowdhury, T. R., Chakrabarty, S., Rakib, M., Afrin, S., Saltmarsh, S., & Winn, S. (2020, September). Factors associated with stunting and wasting in children under 2 years in Bangladesh. *Heliyon*, 6(9), e04849.
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 224-227.
- Desgraupes, B. (2017). Clustering Indices. 1-34.
- Grekousis, G., & Thomas, H. (2012). Comparison of two fuzzy algorithms in geodemographic segmentation analysis: The Fuzzy C-Means and Gustafsson-Kessel methods. *Applied Geography*, 125-136.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques, Second Edition*. San Francisco: Elsevier, Inc.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis 6th Edition*. United States of America: Pearson Education, Inc.
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data : An Introduction to Cluster Analysis*. United States: John Wiley & Sons.
- Kemenkes. (2018). *Situasi Balita Pendek (Stunting) di Indonesia*. Jakarta: Kementerian Kesehatan.
- Kemenkes. (2019). *Laporan Nasional Riskesdas 2018*. Jakarta: Badan Penelitian dan Pengembangan Kesehatan.
- KPPPA dan BPS. (2019). *Profil Anak Indonesia*. Jakarta: KPPPA.
- Luthfi, E., & Wijayanto, A. W. (2021). Analisis perbandingan metode hirarchical, k-means, dan k-medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia. *Inovasi*, 761-773.
- Oot, L., Sethuraman, K., Ross, J., & Sommerfelt, A. E. (2016, February). *The Effect of Chronic Malnutrition (Stunting) on Learning Ability, a Measure of Human Capital: A Model in PROFILES for Country-Level Advocacy*. Retrieved 11 26, 2021, from Food and Nutrition Technical Assistance III Project: <https://www.fantaproject.org/sites/default/files/resources/PROFILES-brief-stunting-learning-Feb2016.pdf>
- Pramana, S., Yuniarto, B., Mariyah, S., Santoso, I., & Nooraeni, R. (2018). *Data Mining dengan R : Konsep serta Implementasi*. Bogor: In Media.
- Silva, A. R., & Dias, C. T. (2013). A cophenetic correlation coefficient for Tocher's method. *Pesq. agropec. bras., Brasília*, 589-596.
- Thamrin, N., & Wijayanto, A. W. (2021). Comparison of Soft and Hard Clustering: A Case Study on Welfare Level in Cities on Java Island. *Indonesian Journal of Statistics and Its Applications*, 141-160.

- TNP2K. (2017). *100 Kabupaten/Kota Prioritas Untuk Intervensi Anak Kerdil (Stunting)*. Jakarta: Sekretariat Wakil Presiden Republik Indonesia.
- UNICEF. (2019). *Gizi di Indonesia*. Retrieved 11 26, 2021, from Status Anak Dunia 2019: <https://www.unicef.org/indonesia/id/status-anak-dunia-2019>
- WHO. (2014). *Global Nutrition Targets 2025 Stunting Policy Brief*. Geneva: WHO.
- WHO. (2014, December 30). *Global nutrition targets 2025: stunting policy brief*. Retrieved November 22, 2021, from WHO/NMH/NHD/14.3: <https://www.who.int/publications/i/item/WHO-NMH-NHD-14.3>
- WHO. (2018). *Level and Trends in Child Malnutrition*. Retrieved 11 12, 2021, from <https://www.who.int/nutgrowthdb/2018-jme-brochure.pdf>
- WHO. (2019). *Nutrition Landscape Information System (NLIS) Country Profile Indicators: Interpretation Guide, Second Edition*. Geneva: World Health Organization (WHO).
- WHO. (2021, June 09). *Malnutrition*. Retrieved 11 23, 2021, from WHO: <https://www.who.int/news-room/fact-sheets/detail/malnutrition>
- WHO. (2021). *Stunting prevalence among children under 5 years of age (% height-for-age <-2 SD) (JME country)*. Retrieved 11 30, 2021, from The Global Health Observatory: [https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-jme-country-children-aged-5-years-stunted-\(-height-for-age--2-sd\)](https://www.who.int/data/gho/data/indicators/indicator-details/GHO/gho-jme-country-children-aged-5-years-stunted-(-height-for-age--2-sd))
- WHO. (2021). *The Global Health Observatory*. Retrieved 11 30, 2021, from WHO: <https://www.who.int/data/gho/data/countries/country-details/GHO/indonesia?countryProfileId=3584815c-0c4d-4f7b-b7c6-11487adf5df0>
- Xie, X. L., & Beni, G. (1991). A Validity Measure for Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8), 841-847.