

Analisis Kluster K-Means Dan Agglomerative Nesting Pada Indikator Stunting Balita Di Indonesia

(K-Means Cluster Analysis and Agglomerative Nesting Analysis on Toddler Stunting Indicators in Indonesia)

Rosi Anisya Faujia^{1*}, Eni Sawitri Setianingsih¹, Hasih Pratiwi¹

¹Program Studi Statistika, Fakultas MIPA, Universitas Sebelas Maret

E-mail: rosianisya@student.uns.ac.id

ABSTRAK

Target kedua dari tujuan kedua SDGs, yaitu menghapus segala bentuk kekurangan gizi, salah satunya indikator prevalensi stunting balita. Penelitian ini bertujuan untuk mengelompokkan dan mengidentifikasi karakteristik provinsi di Indonesia berdasarkan indikator stunting pada balita. Metode yang digunakan dalam penelitian yaitu algoritma k-means dan *agglomerative nesting* (AGNES) *clustering*. Dengan membandingkan nilai rata-rata *average silhouette* dapat diketahui bahwa metode *hierarchical clustering* dari algoritma AGNES dengan *single linkage* memiliki nilai *average silhouette* tertinggi sebesar 0,67 yang merupakan kluster kuat. Berdasarkan hasil analisis diperoleh 2 kluster optimum. Karakteristik kluster 2 yaitu Provinsi Papua termasuk indikator kejadian stunting yang tinggi, karena pada kluster ini imunisasi, akses sanitasi, akses fasilitas kesehatan, tingkat pendidikan SMA rendah dan BBLR <2.500 gram. Sedangkan untuk 33 provinsi lainnya masuk kluster 1 dengan indikator imunisasi, akses sanitasi, akses fasilitas kesehatan, pendidikan SLTA dan BBLR <2.500 gram sehingga pada kluster ini termasuk dalam indikator rendah kejadian stunting.

Kata kunci: *clustering*; stunting; *k-means*; AGNES.

ABSTRACT

The second target of the second goal of the SDGs is to eliminate all forms of malnutrition, one of which is an indicator of the prevalence of stunting in children under five. This study aims to classify and identify the characteristics of provinces in Indonesia based on stunting indicators in children under five. The method used is the k-means algorithm and *agglomerative nesting* (AGNES) *clustering*. By comparing the average silhouette value, it can be seen that the hierarchical clustering method of the AGNES algorithm with single linkage has the highest average silhouette value of 0,67 which is a strong cluster. Based on the results of the analysis obtained 2 optimum clusters. The characteristics of cluster 2, namely Papua Province, are indicators of a high incidence of stunting, because in this cluster immunization, access to sanitation, access to health facilities, high school education levels are low and LBW <2.500 grams. Meanwhile, the other 33 provinces are in cluster 1 with indicators of immunization, access to sanitation, access to health facilities, high school education and LBW <2.500 grams so that this cluster is included in the low incidence of stunting indicators.

Keywords: *clustering*; stunting; *k-means*; AGNES

PENDAHULUAN

Stunting adalah kondisi kekurangan gizi pada balita yang menyebabkan anak tumbuh terlalu pendek pada usianya (Yulisa dkk., 2021). Stunting merupakan salah satu masalah gizi kronis yang dialami oleh balita di dunia saat ini. Secara global pada tahun 2020, anak di bawah usia 5 tahun mengalami stunting sebanyak 149,2 juta, anak bertubuh kurus 45,4 juta, dan anak kelebihan berat badan sebanyak 38,9 juta (WHO, 2021). Menurut data yang dikumpulkan *World Health Organization*, prevalensi stunting anak tahun 2020 di Asia Tenggara termasuk kategori tertinggi nomor dua setelah Afrika (31,7%) yaitu sebesar 30,1%. Batas maksimal angka prevalensi stunting sebesar 20%. Sedangkan di tahun 2021 Indonesia memiliki angka prevalensi stunting yang masih cukup tinggi yaitu sebesar 24,4% (Kemdikbud, 2022).

Pada tanggal 25 September 2015, Agenda Tujuan Pembangunan Berkelanjutan (*Sustainable Development Goals*) disahkan sebagai kesepakatan pembangunan global oleh para pemimpin dunia. Agenda SDGs ini mempunyai 17 tujuan dan 169 target. Tujuan kedua SDGs yaitu mengakhiri kelaparan, mencapai ketahanan pangan dan peningkatan gizi, serta mencanangkan pertanian berkelanjutan. Sesuai target tahun 2025 yang sudah disepakati dalam menghilangkan segala bentuk kekurangan gizi, indikator yang perlu diperhatikan yaitu prevalensi anak di bawah usia lima tahun (balita) yang mengalami stunting (Bappenas, 2017).

Masalah stunting merupakan masalah gizi akut yang sebenarnya bisa dicegah. Beberapa faktor penyebab tingginya angka stunting di antaranya aspek pola makan, yaitu kurangnya akses terhadap pangan dalam hal kuantitas dan kualitas gizi. Aspek kedua yaitu perilaku pola asuh yang kurang baik dalam hal pemberian makan

bagi bayi maupun balita, dan ketiga yaitu aspek sanitasi dan akses air bersih terhadap pelayanan kesehatan, termasuk di dalamnya risiko anak terhadap ancaman penyakit infeksi (Kemenkes, 2018).

Penelitian mengenai kejadian stunting sudah banyak dilakukan. Menurut penelitian Uliyanti (2017) kejadian stunting dipengaruhi oleh dua hal, yaitu faktor langsung dan tidak langsung. Faktor langsung dipengaruhi oleh variabel asupan gizi, riwayat infeksi, pengetahuan gizi ibu dan kadar gizi, sedangkan faktor tidak langsung dipengaruhi oleh Pola Hidup Bersih dan Sehat (PHBS) melalui riwayat penyakit infeksi yang diderita. Hasil penelitian lain membuktikan bahwa asupan energi, protein, pengetahuan ibu, dan pemberian ASI eksklusif berpengaruh terhadap kejadian balita stunting (Bertalina & P. R. Amelia, 2018). Faktor-faktor lain yang mempengaruhi kejadian stunting pada balita antara lain asupan energi, berat badan lahir, tingkat pendidikan ibu, tingkat pendapatan keluarga, pola asuh dan keragaman pangan (Nugroho *et al.*, 2021).

Berbagai pendekatan terus dilakukan untuk menurunkan angka stunting di Indonesia. Masing-masing provinsi di Indonesia memiliki karakteristik masing-masing yang tidak bisa disamakan antardaerah, sehingga perlunya analisis pengelompokan balita stunting. Adanya pengelompokan ini, mempermudah dalam melihat karakteristik faktor penyebab masalah stunting, terutama di Indonesia. Dalam penelitian Satriawan & Styawan (2021), pengelompokan faktor penyebab stunting menggunakan metode *hierarchical clustering ward* dan menghasilkan 4 kluster. Analisis kluster adalah proses pengelompokan sekumpulan objek ke dalam beberapa kluster sehingga objek di dalam suatu kluster bersifat homogen sedangkan objek antar-kluster bersifat heterogen (Pratiwi *et al.*, 2021).

Berdasarkan studi literatur dan permasalahan yang ada, dilakukan pengelompokan provinsi di Indonesia terkait terjadinya stunting dengan metode *k-means* dan *hierarchical clustering* algoritma *Agglomerative Nesting* (AGNES) untuk mengetahui karakteristik dari setiap kluster. Dengan adanya pengelompokan ini, diharapkan pemerintah lebih mudah dalam memahami karakteristik masing-masing kelompok sehingga dalam upaya pencegahan dapat dilakukan secara tepat.

METODE

Metode yang digunakan dalam penelitian yaitu metode *k-means* yang mana termasuk metode *partitioning clustering* dan metode *agglomerative nesting* bagian dari *non-hierarchical clustering*. Penulis menggunakan algoritma dari *package-package* yang sudah tersedia di *software R-studio*.

Data dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder dari laman BPS dan publikasi dari Kementerian Kesehatan RI “Buku Saku Hasil Studi Kasus Gizi Tingkat Nasional, Provinsi, dan Kabupaten/kota tahun 2021”. Data terdiri dari 34 provinsi di Indonesia. Variabel yang terdapat pada penelitian ini yaitu proporsi anak 12-23 bulan yang mendapatkan imunisasi dasar lengkap sebagai X_1 , proporsi rumah tangga dengan akses terhadap layanan sanitasi dasar sebagai X_2 , proporsi rumah tangga dengan akses terhadap layanan dasar indeks pada fasilitas kesehatan sebagai X_3 , tingkat penyelesaian pendidikan menurut jenjang pendidikan SMA sebagai X_4 , dan proporsi balita dengan berat badan Lahir < 2.500 gram sebagai X_5 .

Stunting

Stunting yaitu gangguan pertumbuhan fisik yang menurunkan laju pertumbuhan karena ketidakseimbangan nutrisi (Losong & Adriani M., 2017). Salah satu tanda stunting adalah memiliki nilai *z-score* tinggi badan menurut umur (TB/U) kurang dari -2 standar deviasi (Loya & Nuryanto, 2017).

Klastering

Klastering merupakan teknik dalam mengelompokkan sebuah data ke dalam kelompok tertentu yang memiliki kedekatan pada masing-masing objek. Tujuan utama dari klastering adalah mengelompokkan sebuah dataset ke dalam kluster yang memiliki karakteristik yang hampir sama dan antar kluster mempunyai karakteristik yang berbeda-beda (Ngurah *et al.*, 2017).

Pada analisis klastering terdapat asumsi yang harus dipenuhi sebagai berikut:

1. Sampel yang Mewakili

Sampel yang mewakili merupakan sampel yang dipilih yang dapat merepresentasikan populasi dari penelitian. Dalam hal ini, Uji *Kaiser-Meyer-Olkin* (KMO) digunakan untuk mengukur kecukupan sampling baik secara menyeluruh maupun untuk setiap indikator. Hasil uji KMO menunjukkan jika

nilai KMO berkisar 0,5 – 1 maka sampel dapat dikatakan mewakili populasi atau representatif (J.F. Hair *et al.*, 2006).

2. Non Multikolinearitas

Multikolinearitas merupakan hubungan antar variabel-variabel independen baik berhubungan positif maupun negatif. Untuk menguji ada tidaknya multikolinearitas dapat dilakukan dengan cara melihat koefisien korelasi antar variabel - variabel independen. Jika nilai koefisien korelasi > 0,8 maka antar satu atau lebih variabel independen terjadi masalah multikolinearitas (Gujarati, 1995).

K-Means

K-Means klastering merupakan salah satu metode klastering non hirarki yang membagi data yang ke dalam satu atau lebih kelompok berdasarkan *k* kelompok yang ditentukan. Metode *K-Means* klastering pengelompokannya berdasarkan nilai awal titik pusat (*centroid*) di mana nilai *centroid* awal mempengaruhi nilai *centroid* berikutnya dan penentuan nilai klaster berikutnya, apabila klaster sebelumnya memiliki pola yang sama dengan pola klaster selanjutnya maka perhitungan dihentikan (Johnson & Wichern, 2007).

Hierarchical Clustering

Hierarchical clustering yaitu metode pengelompokan objek data ke dalam sebuah kelompok hirarki. *Hierarchical clustering* dapat dilakukan menggunakan dua cara, yaitu teknik *bottom up* yang biasa disebut *agglomerative nesting* (AGNES) dan *top down* yang bisa disebut *divisive analysis* (DIANA) (Witten *et al.*, 2011). Pada analisis kluster, langkah awal sebelum pengelompokan yaitu penentuan ukuran jarak. Ukuran jarak bisa menggunakan jarak *euclidean*, *manhattan*, dan *mahalanobis*.

Silhouette Coefficient

Silhouette coefficient merupakan sebuah metode interpretasi dan validasi data kluster yang diperkenalkan oleh Rousseeuw pada tahun 1987. *Silhouette* menunjukkan seberapa baik objek terletak dalam suatu kluster dibandingkan berada pada kluster lain (Rousseeuw & Leroy, 1987). Metode ini mengukur validasi kebaikan sebuah data, kluster tunggal atau keseluruhan kluster. Pengelompokan terbaik dicapai apabila nilai *silhouette coefficient* maksimal, artinya meminimalkan jarak dalam kluster sekaligus memaksimalkan jarak antarkluster. Kategori nilai *silhouette coefficient* dan interpretasinya menurut Kaufman dan Rousseeuw (2005) ditunjukkan pada Tabel 1 (Prasetyo, 2012)

Rumus untuk menghitung nilai *silhouette coefficient* adalah

$$s(i) = (b_i - a_i) / \max(b_i, a_i) \dots\dots\dots(1)$$

dimana a_i = nilai rata-rata jarak objek i ke semua objek dalam kluster sendiri
 b_i = nilai rata-rata jarak objek i ke semua objek dalam kluster lain

Tabel 1. Kategori nilai *Silhouette Coefficient* (SC).

SC	Interpretasi
$0,7 < SC < 0,1$	Terdapat ikatan yang sangat kuat antara objek dan kluster yang terbentuk
$0,5 < SC \leq 0,7$	Terdapat ikatan yang cukup kuat antara objek dan kluster yang terbentuk
$0,25 < SC \leq 0,5$	Terdapat ikatan yang lemah antara objek dan kluster yang terbentuk
$SC \leq 0,25$	Tidak terdapat ikatan antara objek dan kluster yang terbentuk

HASIL DAN PEMBAHASAN

Analisis Deskriptif

Langkah awal yang dilakukan yaitu analisis deskriptif dari data indikator stunting yang digunakan. Analisis deskriptif digunakan untuk mengetahui sebaran dari sebuah data. Sebaran data tersebut meliputi nilai minimum, median, nilai maksimum, dan rata-rata yang terdapat pada Tabel 2.

Tabel 2. Statistik deskriptif.

Variabel	Minimum	Median	Maksimum	Rata-rata
Imunisasi	27,40	68,95	94,90	67,28
Sanitasi	40,81	80,67	97,12	80,97
Faskes	36,97	80,06	91,30	78,74
SMA	32,95	66,02	90,12	65,51
BBLR	4,50	7,10	10,30	6,868

Pada Tabel 2 diketahui rata-rata proporsi anak 12-23 bulan yang mendapatkan imunisasi dasar lengkap sebesar 67,28. Nilai rata-rata proporsi rumah tangga dengan akses terhadap layanan sanitasi dasar sebesar 80,97. Nilai rata-rata proporsi rumah tangga dengan akses terhadap layanan dasar indeks pada fasilitas kesehatan sebesar 78,74. Nilai rata-rata tingkat penyelesaian pendidikan menurut jenjang pendidikan SMA, dan nilai rata-rata proporsi anak di bawah usia lima tahun dengan berat badan lahir < 2.500 gram sebesar 6,868.

Langkah selanjutnya setelah analisis deskriptif yaitu uji asumsi *Kaiser-Meyer-Olkin* (KMO). Uji ini digunakan untuk mengukur kecukupan sampel untuk setiap variabel. Menurut Hair (2006) apabila nilai KMO berkisar 0,5 – 1 maka uji asumsi terpenuhi yang artinya sampel dapat mewakili populasi. Hasil uji KMO dapat dilihat pada tabel 3.

Tabel 3. Uji KMO.

Variabel	Nilai KMO
Keseluruhan	0,69
Imunisasi	0,68
Sanitasi	0,65
Faskes	0,83
SMA	0,65
BBLR	0,61

Tabel 3 menunjukkan nilai KMO pada keseluruhan dan masing-masing variabel. Dari Tabel 3 dapat diketahui baik untuk keseluruhan maupun masing-masing variabelnya, nilai KMO $> 0,5$. Hal ini menunjukkan bahwa sampel sudah mewakili populasi atau sampel representatif. Sehingga analisis bisa dilanjutkan.

Dalam penelitian ini, asumsi non multikolinearitas harus dipenuhi dalam analisis kluster. Asumsi ini digunakan untuk mengetahui korelasi parsial antar variabel atau bisa disebut juga hubungan linier dari dua variabel yang terkait. Suatu variabel dikatakan terjadi multikolinearitas apabila nilai korelasinya $> 0,8$. Hasil korelasi parsial terdapat pada Tabel 4.

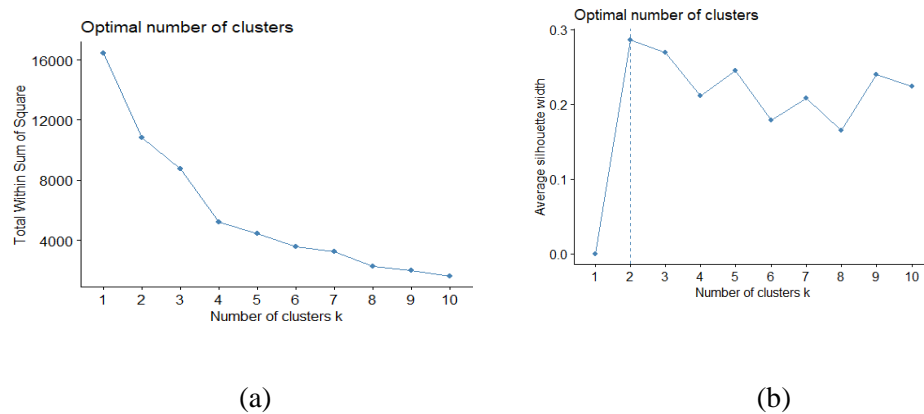
Tabel 4. Korelasi parsial.

Variabel	Imunisasi	Sanitasi	Faskes	SMA	BBLR
Imunisasi	1	0,437	0,105	-0,264	-0,034
Sanitasi	0,437	1	0,449	0,623	0,091
Faskes	0,105	0,449	1	0,097	0,086
SMA	-0,264	0,623	0,097	1	-0,406
BBLR	-0,034	0,091	0,086	-0,406	1

Tabel 4 menunjukkan korelasi parsial antar variabel. Nilai korelasi parsial antar variabel tidak ada yang lebih dari 0,80 maka asumsi non multikolinearitas terpenuhi.

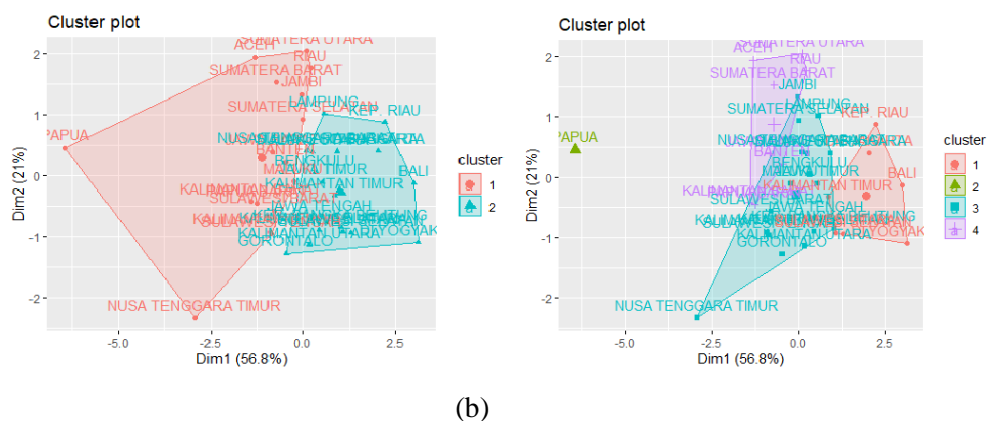
K-Means Clustering

Pada analisis *k-means* langkah awal yang dilakukan yaitu menetapkan banyaknya *k* kluster optimal. Untuk menentukan *k* kluster optimal digunakan metode *Elbow* dan *Silhouette*. Berikut hasil dari kedua metode dalam penentuan banyaknya *k* kluster optimal.



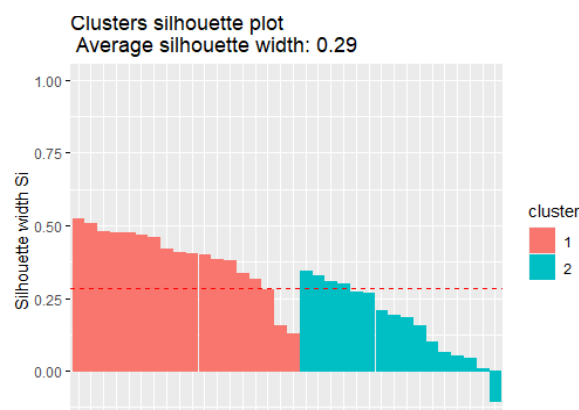
Gambar 1. Penentuan k kluster optimal metode K -Means (a) metode Elbow; (b) metode $Silhouette$.

Pada Gambar 1(a) penentuan k optimal menggunakan metode $Elbow$, diketahui banyaknya kluster ditentukan pada titik k di mana garis tersebut terdapat patahan yang berbentuk siku. Gambar 1 bagian (a) terlihat bahwa garis yang mendekati siku yaitu berada pada $k = 4$. Oleh sebab itu, berdasarkan metode $Elbow$, jumlah kluster yang maksimum yaitu sebanyak empat kluster, sedangkan pada bagian (b) penentuan k optimal pada metode $silhouette$ dari nilai rata-rata koefisien $silhouette$ yang paling besar. Dari Gambar 1(b) terlihat bahwa kluster optimal ketika $k = 2$.



Gambar 2. Hasil kluster metode K -Means (a) $k = 2$, (b) $k = 4$.

Dari Gambar 2 terlihat bahwa untuk $k = 4$ terdapat *overlapping* antar anggota kluster. Jadi, untuk metode k -means kluster yang akan digunakan yaitu untuk $k = 2$. Untuk melihat ikatan antara objek dari kluster yang terbentuk digunakan nilai dari *average silhouette width*. Gambar 3 menunjukkan nilai *average silhouette width* menggunakan metode K -means clustering yaitu sebesar 0,29. Menurut Kaufman dan Rousseeuw (2005), nilai koefisien *silhouette* sebesar 0,29 menunjukkan bahwa ikatan antara objek dan kluster yang terbentuk tergolong lemah (*weak cluster*).



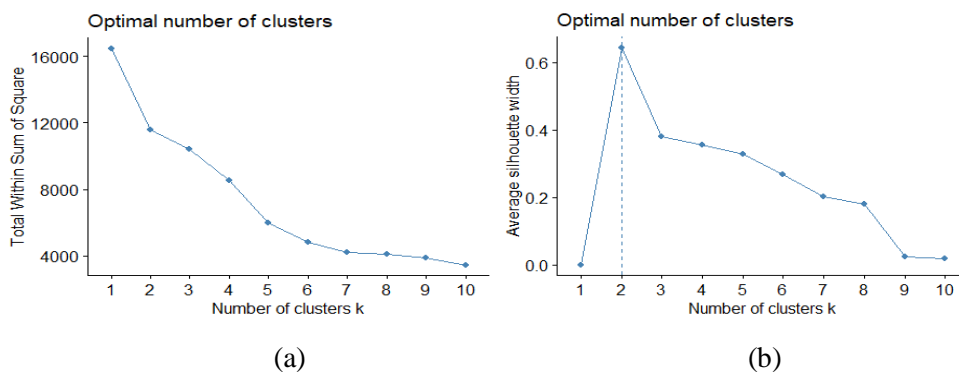
Agglomerative Nesting (AGNES)

Algoritma *agglomerative nesting* (AGNES) termasuk ke dalam *hierarchical clustering*. Terdapat empat metode dalam pembentukan kluster ini yaitu, *ward linkage*, *average linkage*, *complete linkage*, dan *single linkage*. Untuk membandingkan setiap linkage yang optimal digunakan koefisien korelasi *cophenetic*. Nilai koefisien korelasi *cophenetic* berkisar antara -1 sampai 1, apabila nilai koefisien korelasi *cophenetic* mendekati 1 berarti semakin baik *cluster* yang dihasilkan. Dengan menggunakan bantuan *software* R Studio diperoleh Tabel 5 yang menunjukkan perbandingan koefisien korelasi *cophenetic* dari empat metode pembentukan kluster.

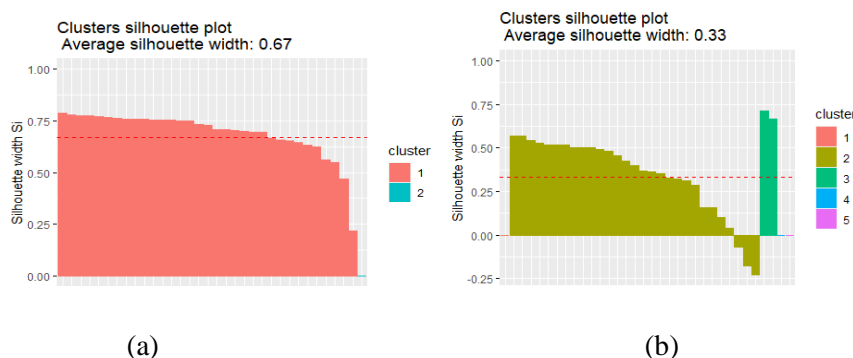
Tabel 5. Koefisien korelasi *Cophenetic*.

Metode	Koefisien Korelasi <i>Cophenetic</i>
<i>Single</i>	0,8680676
<i>Complete</i>	0,8081806
<i>Average</i>	0,8674761
<i>Ward</i>	0,4377979

Tabel 5 menunjukkan bahwa nilai koefisien *cophenetic* yang dihasilkan metode *single linkage* mempunyai nilai koefisien korelasi *cophenetic* paling tinggi diantara metode pengelompokan lainnya yaitu sebesar 0,8680676. Penelitian ini menggunakan algoritma *agglomerative nesting* dengan *single linkage* dan jarak *manhattan*. Dalam penentuan banyaknya *k* kluster optimal dalam metode AGNES sama dengan metode sebelumnya, yaitu menggunakan metode *elbow* dan *silhouette*. Berikut hasil dari kedua metode dalam penentuan banyaknya *k* kluster optimal.

Gambar 4. Penentuan *k* kluster optimal algoritma AGNES (a) metode Elbow; (b) metode *Silhouette*.

Pada Gambar 4 (a), banyaknya kluster pada metode *elbow* ditentukan pada titik *k* di mana garis tersebut terdapat patahan yang berbentuk siku. Pada Gambar 4 (a) terlihat bahwa garis yang mendekati siku yaitu berada pada $k = 5$. Oleh karena itu berdasarkan metode ini, jumlah kluster yang maksimum yaitu sebanyak 5 kluster. Pemilihan jumlah kluster menggunakan metode *silhouette* dilihat dari nilai rata-rata koefisien *silhouette* yang paling tinggi. Dari Gambar 4 (b) terlihat bahwa kluster optimal untuk metode *silhouette* ketika $k = 2$.

Gambar 5. Hasil kluster algoritma AGNES (a) $k = 2$, (b) $k = 5$.

Pada Gambar 5 menunjukkan nilai *average silhouette width* dari hasil kluster menggunakan metode AGNES dengan *single linkage*. Untuk $k = 2$, nilai *average silhouette width* lebih tinggi dibandingkan untuk $k = 5$. Nilai sebesar 0,67 tersebut menunjukkan bahwa ikatan antara objek dan kluster yang terbentuk kuat (*strong cluster*).

Pemilihan Metode Terbaik

Nilai *average silhouette width* digunakan untuk menentukan metode terbaik. Perbandingan nilai *average silhouette width* terdapat pada Tabel 6.

Tabel 6. Perbandingan metode pembentukan kluster.

Metode	Jumlah Kluster	Average Silhouette Width
K-Means	2	0,29
	4	0,27
AGNES	2	0,67
	5	0,33

Tabel 6 menunjukkan hasil perbandingan nilai dari *average silhouette width* pada beberapa metode. Metode terbaik dilihat dari nilai *average silhouette width*, semakin tinggi nilai yang dihasilkan maka ikatan objek dengan kluster yang terbentuk semakin baik. Dari tabel perbandingan metode pembentukan kluster terlihat bahwa metode *hierarchical clustering* algoritma *Agglomerative Nesting* dengan *single linkage* memiliki nilai *average silhouette width* paling tinggi yaitu sebesar 0,67, yang mana hasil ini termasuk *strong cluster*. Berikut merupakan anggota-anggota dari masing-masing kluster yang dihasilkan dari algoritma *agglomerative nesting* dengan *single linkage*.

Berdasarkan Tabel 7 diketahui kluster 1 memiliki 33 anggota yang meliputi Provinsi Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, dan Papua Barat. Sedangkan pada kluster 2 memiliki satu anggota kluster yaitu Provinsi Papua.

Tabel 7. Anggota hasil kluster.

Kluster	Anggota Kluster	Jumlah Anggota Kluster
1	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Jawa Tengah, Di Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Dan Papua Barat.	33
2	Papua	1

Karakteristik dari masing-masing kluster dapat dilihat pada Tabel 8, yang mana pada kluster 1, rata-rata anak 12-23 bulan yang mendapatkan imunisasi dasar lengkap sebagai variabel X_1 , rumah tangga dengan akses terhadap layanan sanitasi dasar sebagai variabel X_2 dan indeks pada fasilitas kesehatan sebagai variabel X_3 , serta tingkat penyelesaian pendidikan SMA sebagai variabel X_4 kategori tinggi. Sedangkan proporsi balita dengan berat badan lahir < 2500 gram sebagai variabel X_5 kategori rendah. Untuk kluster 2, rata-rata dari proporsi anak 12-23 bulan yang mendapatkan imunisasi dasar lengkap, rumah tangga dengan akses terhadap sanitasi dasar maupun indeks fasilitas kesehatan, serta tingkat penyelesaian pendidikan SMA yang rendah, sedangkan proporsi balita lahir < 2500 gram tinggi.

Tabel 8. Rerata klaster.

Variabel	Klaster 1	Klaster 2
Imunisasi	67,78788	50,5
Sanitasi	82,18424	40,81
Faskes	80,00485	36,97
SMA	66,49727	32,95
BBLR	6,839394	7,8

KESIMPULAN

Berdasarkan hasil analisis dan pembahasan, dapat disimpulkan bahwa metode yang baik digunakan yaitu metode AGNES dengan single linkage dimana nilai average silhouette width yang dihasilkan sebesar 0,67. Hasil klastering didapatkan 2 klaster, dimana pada klaster 2 yaitu Provinsi Papua termasuk indikator kejadian stunting yang tinggi, karena pada klaster ini imunisasi, akses terhadap sanitasi, akses terhadap fasilitas kesehatan, tingkat pendidikan jenjang SMA yang rendah dan bayi lahir <2500 gram tinggi. Untuk 33 provinsi lainnya termasuk ke dalam klaster 1 dengan indikator imunisasi, akses terhadap sanitasi, akses terhadap fasilitas kesehatan, tingkat pendidikan jenjang SMA yang tinggi dan bayi lahir <2500 gram rendah. Jadi, klaster ini termasuk dalam indikator kejadian stunting yang rendah. Dari hasil analisis klastering ini, Pemerintah diharapkan dapat membuat kebijakan yang tepat dalam menanggulangi masalah kejadian stunting di Indonesia, terutama pada provinsi Papua dengan kejadian stunting yang tergolong tinggi.

DAFTAR PUSTAKA

- Bappenas. (2017). Ringkasan Metadata Indikator Tujuan Pembangunan Berkelanjutan (TPB)/Sustainable Development Goals (SDGs).
- Bertalina, & P. R. Amelia. (2018). Hubungan Asupan Gizi, Pemberian Asi Eksklusif, dan Pengetahuan Ibu dengan Status Gizi (Tb/U) Balita 6-59 Bulan. In *Jurnal Kesehatan* (Vol. 9, Issue 1). Online. <http://ejurnal.poltekkes-tjk.ac.id/index.php/JK>
- Gujarati, D. N. (1995). *Basic Econometrics* (4th ed.). United State Military Academy.
- Indrastuty, D., & Pujiyanto. (2014). Determinan Sosial Ekonomi Rumah Tangga dari Balita Stunting : Analisis Data Indonesia Family Life Survey (IFLS). *Jurnal Ekonomi Kesehatan Indonesia*, 3(2), 68–70.
- J.F. Hair, Black, W. C., Babin, W. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* 6th Edition. Pearson Prentice Hall.
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). Pearson, Prentice Hall.
- Kemdikbud. (2022). Prevalensi Stunting Indonesia 2022 Masih di atas WHO, 37% Pasangan Usia Subur Alami Anemia. <https://pauddpedia.kemdikbud.go.id/berita/prevalensi-stunting-indonesia-2022-masih-diatas-standar-who-37-pasangan-usia-subur-alami-anemia?id=812&ix=11> [7 Juli 2022]
- Kemkes. (2018). Cegah Stunting dengan Perbaikan Pola Makan, Pola Asuh dan Sanitasi. <http://p2ptm.kemkes.go.id/kegiatan-p2ptm/subdit-penyakit-diabetes-melitus-dan-gangguan-metabolik/cegah-stunting-dengan-perbaikan-pola-makan-pola-asuh-dan-sanitasi> [1 Juli 2022]
- Losong, N. H. F., & Adriani Merryana. (2017). Perbedaan Kadar Hemoglobin , Asupan Zat Besi , dan Zinc pada Balita Stunting dan Non Stunting. 117–123. <https://doi.org/10.2473/amnt.v1i2.2017.117-123>
- Loya, R. R. P., & Nuryanto. (2017). POLA ASUH PEMBERIAN MAKAN PADA BALITA STUNTING USIA 6-12 BULAN DI KABUPATEN SUMBA TENGAH NUSA TENGGARA TIMUR. *Journal of Nutrition College*, 6(2), 84–95. <http://ejournal-s1.undip.ac.id/index.php/jnc>
- Ngurah, G., Paramartha, W., Ratnawati, D. E., & Widodo, A. W. (2017). Analisis Perbandingan Metode K-Means Dengan Improved Semi-Supervised K-Means Pada Data Indeks Pembangunan Manusia (IPM) (Vol. 1, Issue 9). <http://j-ptiik.ub.ac.id>

- Nugroho, M. R., Sasongko, R. N., & Kristiawan, M. (2021). Faktor-faktor yang Mempengaruhi Kejadian Stunting pada Anak Usia Dini di Indonesia. *Jurnal Obsesi : Jurnal Pendidikan Anak Usia Dini*, 5(2), 2269–2276. <https://doi.org/10.31004/obsesi.v5i2.1169>
- Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasi Menggunakan Matlab*. ANDI OFFSET.
- Pratiwi, H., Handajani, S. S., & Respatiwan. (2021). *Clustering dan Penerapannya dalam Bidang Seismologi*. Yogyakarta: Deepublish
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. Wiley Series in Probability and Mathematical Statistics, . John Wiley, New.
- Satriawan, D., & Styawan, D. A. (2021). Pengelompokan Provinsi di Indonesia Berdasarkan Faktor Penyebab Balita Stunting Menggunakan Analisis Cluster Hierarki. *Jurnal Statistika Dan Aplikasinya*, 5(1), 61–70.
- Uliyanti, G. Tamtono, D., & Anantanyu, S. (2017). Faktor yang Berhubungan Dengan Kejadian Stunting pada Balita Usia 24-59 Bulan. *Jurnal Vokasi Kesehatan*.
- WHO. (2021). Child Malnutrition. <http://www.who.int/gho/child-malnutrition/en/> [5 Juli 2022]
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques* (3rd ed.). Morgan Kaufmann Publishers.
- Yulisa, N., Saragih, R., & Ambarita, I. (2021, June). Data Mining Pengelompokan Anak Stunting Berdasarkan Usia, Penyebab dan Pekerjaan Orang Tua Dengan Menggunakan Metode Clustering (Studi Kasus: Dinas Kesehatan Kabupaten Langkat). In *Seminar Nasional Informatika (SENATIKA)* (pp. 295-306).

