



Pengelompokan Data *Stunting* di Indonesia Menggunakan Metode *X-Means* dan *Agglomerative Hierarchical Clustering*

Nur Dhea Wahab¹, Salmun K. Nasib^{1*}, Nurwan², Djihad Wungguli², Nisky Imansyah yahya²

¹Program Studi Statistika, Universitas Negeri Gorontalo, Bone Bolango 96554, Indonesia

¹Program Studi Matematika, Universitas Negeri Gorontalo, Bone Bolango 96554, Indonesia

Info Artikel

*Penulis Korespondensi.
Email: salmun@ung.ac.id

Diterima: 10 November 2024
Direvisi: 12 Februari 2025
Disetujui: 15 Februari 2025



Under the licence
CC BY-NC-SA 4.0

Diterbitkan oleh:



Copyright ©2025 by Author(s)

Abstrak

Stunting merupakan salah satu permasalahan serius yang mengancam kualitas sumber daya manusia di Indonesia. Penelitian ini bertujuan untuk menganalisis pola dan karakteristik stunting di Indonesia dengan menerapkan metode clustering *X-Means* dan *Agglomerative Hierarchical Clustering* (AHC). Metode *X-Means* digunakan untuk menentukan jumlah cluster optimal secara otomatis dengan memanfaatkan kriteria Bayesian Information Criterion (BIC), sedangkan AHC membentuk dendrogram untuk memahami struktur bertingkat dari cluster yang terbentuk. Berdasarkan hasil analisis, metode *X-Means* menghasilkan tiga cluster optimal dengan nilai BIC terkecil sebesar 651.9475, dimana cluster 1 terdiri dari 17 provinsi, cluster 2 mencakup 12 provinsi, dan cluster 3 meliputi 5 provinsi. Metode AHC dengan pendekatan Single Linkage juga menghasilkan tiga cluster optimal, dengan cluster 1 mencakup 32 provinsi, cluster 2 terdiri dari 1 provinsi (Nusa Tenggara Barat), dan cluster 3 mencakup 1 provinsi (Nusa Tenggara Timur), serta nilai Silhouette Index tertinggi sebesar 0.28. Hasil penelitian menunjukkan bahwa kedua metode memberikan gambaran yang komprehensif tentang pola stunting di Indonesia, yang dapat digunakan sebagai dasar dalam merancang program intervensi yang lebih terarah sesuai dengan karakteristik masing-masing cluster. Strategi berbasis data ini diharapkan mampu meningkatkan efektivitas kebijakan dalam menurunkan angka stunting di Indonesia.

Kata Kunci: *X-Means*; *Agglomerative Hierarchical Clustering*; *Stunting*

Abstract

Stunting is one of the serious problems that threaten the quality of human resources in Indonesia. This study aims to analyze the patterns and characteristics of stunting in Indonesia by applying the X-Means clustering method and Agglomerative Hierarchical Clustering (AHC). The X-Means method is used to determine the optimal number of clusters automatically by utilizing the Bayesian Information Criterion (BIC), while AHC forms a dendrogram to understand the multilevel structure of the clusters formed. Based on the analysis, the X-Means method produces three optimal clusters with the smallest BIC value of 651.9475, where cluster 1 consists of 17 provinces, cluster 2 includes 12 provinces, and cluster 3 includes 5 provinces. The AHC method with the Single Linkage approach also produced three optimal clusters, with cluster 1 covering 32 provinces, cluster 2 consisting of 1 province (West Nusa Tenggara), and cluster 3 covering 1 province (East Nusa Tenggara), as well as the highest Silhouette Index value of 0.28. The results show that both methods provide a comprehensive picture of stunting patterns in Indonesia, which can be used as a basis for designing more targeted intervention programs according to the characteristics of each cluster. This data-driven strategy is expected to increase policy effectiveness in reducing stunting in Indonesia.

Keywords: *X-Means*; *Agglomerative Hierarchical Clustering*; *Stunting*

1. Pendahuluan

Stunting merupakan salah satu permasalahan serius yang mengancam kualitas sumber daya manusia di Indonesia. Kondisi ini tidak hanya berdampak pada pertumbuhan fisik anak tetapi juga berpengaruh terhadap perkembangan otak dan kecerdasan, yang pada akhirnya dapat menurunkan produktivitas di masa depan. Stunting terjadi ketika balita mengalami hambatan pertumbuhan akibat kekurangan gizi dalam jangka panjang, terutama selama 1000 hari pertama kehidupannya. Kondisi ini disebabkan oleh asupan gizi yang tidak mencukupi serta pola asuh yang kurang optimal pada masa awal pertumbuhan [1]. Berdasarkan hasil Studi Status Gizi Indonesia (SSGI) tahun 2021, angka stunting di Indonesia mencapai 24,4%, lebih tinggi dari standar yang ditetapkan oleh World Health Organization (WHO) sebesar 20%. Beberapa provinsi menunjukkan angka stunting yang cukup tinggi, seperti Nusa Tenggara Timur yang mencapai 37,8%, sementara provinsi dengan angka terendah adalah Jawa Tengah dengan 8,9% [2]. Tingginya angka stunting ini menunjukkan perlunya analisis lebih lanjut untuk memahami pola dan karakteristik stunting di berbagai wilayah di Indonesia.

Clustering merupakan teknik penelitian yang digunakan untuk menemukan kelompok observasi serupa dalam suatu kumpulan data. Analisis cluster juga digunakan untuk mengidentifikasi kesamaan dan mengembangkan kelompok objek serupa, serta dapat berguna dalam menemukan kelompok sebenarnya yang diyakini ada dan juga berguna secara informatif. Hal ini melibatkan identifikasi perbedaan, persamaan, dan perubahan dalam konteks yang berbeda [3]. Analisis *clustering* mencakup dua pendekatan yang berbeda: metodologi non-hirarki dan hirarki. Metode non-hierarki adalah teknik yang digunakan untuk pengelompokan data, di mana jumlah cluster dapat diatur [4]. Metode non hierarki merupakan metode yang mengelompokkan data ke dalam beberapa cluster berdasarkan kriteria tertentu, seringkali menggunakan konsep jarak atau kesamaan, dan tidak melibatkan pendekatan *bottom-up* atau *top-down* [5]. Metode non-hierarki yang populer adalah metode *K-Means*, yang mengelompokkan data ke dalam sejumlah cluster yang telah ditentukan. Algoritma ini bekerja dengan meminimalkan jarak antara titik data dengan pusat cluster (centroid). Meskipun K-means sederhana dan cepat, kelemahan utamanya adalah persyaratan untuk memasukkan jumlah cluster di awal [6]. Kelemahan pada metode *K-Means* menyebabkan terdapat pengembangan metode yang digunakan untuk menyelesaikan kelemahan dari K-means yaitu metode *X-Means*.

X-means adalah ekstensi dari K-means yang secara otomatis menentukan jumlah cluster yang optimal. Algoritma ini melakukan "penyusunan ulang" atau "pembagian" pada cluster yang ada, untuk menemukan struktur yang paling sesuai dengan data, sehingga X-means lebih mudah menemukan jumlah cluster yang paling akurat [7]. Algoritma X-means merupakan pengembangan dari algoritma K-means yang digunakan untuk menyelesaikan salah satu kelemahan utama dari algoritma K-means, yaitu kebutuhan untuk memiliki pengetahuan sebelumnya tentang jumlah cluster (K) [8]. X-means dapat menentukan jumlah cluster yang optimal secara otomatis, dengan menggunakan kriteria statistik seperti Bayesian Information Criterion (BIC) [9]. Hal ini didukung oleh penelitian yang dilakukan [10], dimana pada penelitiannya tentang pengelompokan penjualan barang dari tahun 2019 sampai 2020 menggunakan K-means yang menghasilkan 3 cluster yang diinterpretasikan sangat diminati, diminati dan kurang diminati. Namun usulan pada penelitian tersebut yaitu *clustering* menggunakan metode X-means. Dengan adanya usulan tersebut, pada penelitian [11] digunakan metode X-means untuk menganalisis pengelompokan penjualan barang dari tahun 2019 sampai 2020 yang menghasilkan dua cluster utama, dengan cluster 0 berisi 271 item dan cluster 1 berisi 62 item. Evaluasi menunjukkan bahwa nilai k optimal adalah 2, dan penggunaan operator normalisasi mempengaruhi indeks Davies Bouldin, dengan nilai indeks yang lebih baik sebesar 0.574 dicapai tanpa normalisasi. Penelitian ini juga menunjukkan efektivitas algoritma X-means dalam analisis pengelompokan data. Dengan algoritma X-means kontribusinya terhadap perusahaan yaitu dapat mengetahui pembagian cluster berdasarkan produk yang banyak laku dan kurang laku . oleh sebab itu X-means sangatlah cocok dan lebih efektif digunakan.

Selain metode X-means terdapat metode hirarki yang sering digunakan dalam analisis cluster yaitu metode Agglomerative Hierarchical Clustering (AHC). Agglomerative Hierarchical Clustering

(AHC) adalah teknik pengelompokan hirarki yang membangun dendrogram yang memungkinkan kita melihat struktur bertingkat dari klusterisasi. Teknik ini membentuk struktur bertingkat yang memungkinkan pengguna untuk memilih jumlah cluster yang diinginkan dengan memotong dendrogram pada level tertentu [12]. Ada beberapa penelitian metode AHC yang dilakukan sebelumnya, antara lain penelitian oleh [13] diperoleh bahwa Metode Agglomerative Hierarchical Clustering (AHC) dalam sistem yang telah dibuat dapat mengatasi kerumitan dalam menghitung jarak terjauh, namun belum bisa menampilkan hasil pohon atau grafiknya yang menentukan sehat atau tidak sehat dalam hasil akhirnya. Penelitian lainnya dilakukan oleh [14] tentang Klasifikasi Sistemik Lingkungan Perumahan Pop-Up (PUH) menggunakan metode AHC. Dari penelitian ini diperoleh sekitar 75% objek yang diteliti menggunakan bahan konstruksi ringan dan kaku, dengan dua pertiga dari struktur yang dianalisis terdiri dari satu unit hunian dengan satu lantai. 60% dari hunian pop-up yang didokumentasikan mengandung unit yang dapat digunakan kembali atau didaur ulang, yang menunjukkan konstruksi modular yang sederhana, fondasi yang dapat dilepas, dan bahan yang dapat didaur ulang. Sekitar 37% objek yang didokumentasikan menggunakan kayu sebagai bahan bangunan utama dan isolasi termal dengan insulator konvensional. Penelitian ini juga menekankan aplikasi efektif dari algoritma pengelompokan hierarkis untuk menganalisis dan mengklasifikasikan data yang beragam terkait solusi perumahan sementara.

Penelitian ini dilakukan dengan tujuan untuk menganalisis pola dan karakteristik stunting di Indonesia menggunakan metode *X-Means* dan *Agglomerative Hierarchical Clustering*. Selain itu, penelitian ini juga berupaya menentukan jumlah *cluster* optimal untuk mengelompokkan provinsi di Indonesia berdasarkan tingkat stunting yang terjadi. Evaluasi terhadap efektivitas metode *X-Means* dan *AHC* dalam mengelompokkan data stunting juga menjadi salah satu fokus utama dalam penelitian ini. Dengan adanya analisis *clustering* ini, diharapkan dapat diperoleh pemetaan yang lebih jelas mengenai wilayah dengan tingkat stunting tinggi dan rendah, serta faktor-faktor yang berkontribusi terhadap fenomena tersebut. Hasil penelitian ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai pola stunting di Indonesia serta menjadi dasar dalam merancang strategi intervensi yang lebih efektif sesuai dengan karakteristik masing-masing kelompok, sehingga dapat membantu dalam upaya menurunkan angka stunting secara lebih terarah dan efisien.

2. Metode Penelitian

2.1 Sumber Data

Metode Penelitian ini menggunakan data sekunder yaitu data faktor-faktor penyebab stunting. Data diperoleh dari hasil [15], Publikasi Kementerian Kesehatan Indonesia, serta website Kemenkes Indonesia. Teknik penarikan sampel yang digunakan adalah *sampling jenuh*, dimana semua populasi yang ada dijadikan sampel.

2.2 Tahapan Penelitian

Penelitian ini dilakukan berdasarkan prosedur sebagai berikut :

1. Melakukan penginputan data
2. Pra-procesing
3. Melakukan analisis klaster menggunakan metode *X-Means*
 - a. Analisis data menggunakan *K-means* untuk menghasilkan *cluster* awal dan centroid
 - b. Evaluasi hasil dari metode *K-Means* menggunakan BIC
 - c. Melakukan pembagian *cluster* berdasarkan nilai BIC terkecil
 - d. Melakukan pengelompokan dengan *X-Means* berdasarkan BIC terbaik
4. Melakukan analisis klaster menggunakan metode AHC (*Agglomerative Hierarchical Clustering*)
 - a. Melakukan pengelompokan untuk single linkage, complete linkage, dan average linkage.
 - b. Evaluasi hasil pengelompokan single linkage, complete linkage, dan average linkage menggunakan silhouette coefficient.

- c. Melakukan pengelompokkan menggunakan metode ahc terbaik berdasarkan nilai silhouette coefficient terbaik.
5. Interpretasi hasil pengelompokan stunting berdasarkan faktor-faktor yang mempengaruhi dengan menggunakan metode X-Means dan AHC

3. Hasil dan Pembahasan

3.1 Statistik Deskriptif Data Penelitian

Data yang digunakan dalam penelitian ini, merupakan data faktor penyebab stunting yang terdiri dari 7 variabel yaitu Balita yang mendapatkan ASI Eksklusif (X1), Balita yang diberikan MPASI (X2), Berat Badan Lahir Rendah (X3), Rumah Tangga yang memiliki Akses Sanitasi lengkap (X4), Rumah Tangga yang Memiliki Akses Air Minum Layak (X5), Rata-rata pengeluaran Perkapita Sebulan (X6), dan Rata-rata Konsumsi Kalori Perkapita Sehari (X7). Adapun statistika deskriptif disajikan pada Tabel 1.

Tabel 1. Statistik Deskriptif

	X1	X2	X3	X4	X5	X6	X7
Count	34	34	34	34	34	34	34
Min	36.20	26.90	4.500	57.00	42.40	840359	1862
Q1	46.10	36.02	5.925	80.35	60.52	1112810	2021
Median	51.15	44.40	7.100	85.05	69.60	1188904	2112
Mean	52.85	44.44	6.868	84.06	69.36	1288077	2105
Q3	59.12	51.20	7.700	90.95	76.45	1412359	2166
Max	75.90	64.30	10.300	96.80	94.50	2336429	2543

Berdasarkan Tabel 1, terlihat bahwa adanya perbedaan rentang nilai yang cukup jauh antara variabel 6 dengan yang lainnya. Hal ini dipandang perlu dilakukannya proses standarisasi data terlebih dahulu dengan tujuan untuk menstabilkan varian antara variabel sehingga mencegah adanya data bias yang mungkin timbul akibat skala yang berbeda.

3.2 Standarisasi Data

Dalam penelitian ini, proses standarisasi data dilakukan melalui perhitungan Z Score. Berikut adalah hasil standarisasi data yang diolah menggunakan RStudio, disajikan pada Tabel 2.

Tabel 2. Standarisasi Data

Provinsi	X1	X2	X3	X4	...	X7
Aceh	-1.1682	-1.7354	-0.0506	-0.3265	...	0.4916
Bali	0.0246	0.4510	-1.2481	1.4059	...	0.6814
Banten	-1.3260	-1.2506	0.6229	-0.2161	...	1.1415
Bengkulu	0.4682	0.0750	-0.7242	-1.1099	...	0.1477
:	:	:	:	:	...	:
Sumatera Utara	-1.1978	-1.2704	-1.6971	-0.9775	...	0.3298

3.3 Uji Multikolinearitas

Uji ini dilakukan untuk mendeteksi keberadaan multikolinearitas antara variabel bebas yang dilakukan dengan menghitung nilai koefisien korelasi. Hasil uji multikolinearitas disajikan pada Tabel 3.

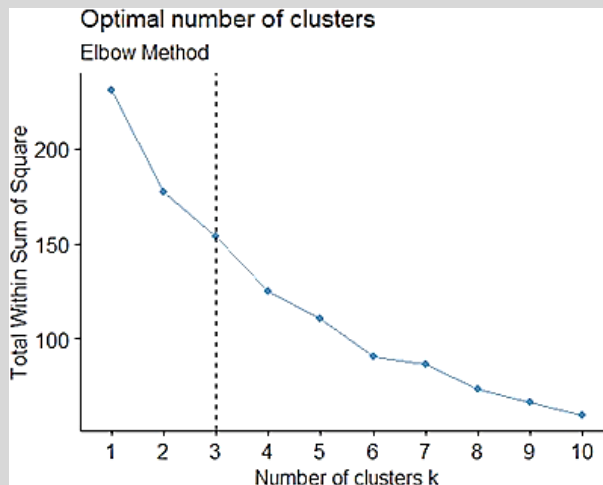
Tabel 3. Uji Multikolinearitas

Variabel	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
X ₁	0	3.0848	1.0069	1.0739	1.1457	1.0274	1.0129
X ₂	3.0848	0	1.0001	1.0098	1.0510	1.0001	1.0830
X ₃	1.0069	1.0001	0	1.0489	1.0026	1.0093	1.0915
X ₄	1.0739	1.0098	1.0489	0	1.0705	1.0940	1.0192
X ₅	1.1457	1.0510	1.0026	1.0705	0	1.7848	1.0135
X ₆	1.0274	1.0001	1.0093	1.0940	1.7848	0	1.0302
X ₇	1.0129	1.0830	1.0915	1.0192	1.0135	1.0302	0

Berdasarkan Tabel 3 dapat dilihat bahwa nilai $VIF \leq 10$ maka dapat disimpulkan bahwa variabel penelitian tidak terindikasi multikolinearitas.

3.4 Analisis cluster X-Means

Pada penelitian ini, penentuan jumlah K cluster optimal yaitu menggunakan metode Elbow. Untuk menentukan cluster yang optimal dapat dilihat pada titik yang membentuk siku pada Gambar 1.

**Gambar 1.** Grafik Metode *Elbow*

Dari Gambar 1 dapat dilihat bahwa titik yang membentuk siku yaitu cluster 3. Maksud dari titik yang membentuk siku adalah penurunan nilai yang besar antara dua titik yang kemudian diikuti oleh penurunan nilai yang kecil. Sehingga, jumlah cluster optimal untuk pengelompokan menggunakan metode X-Means clustering dan AHC yang terbentuk adalah 3 cluster karena terdapat penurunan nilai yang besar pada cluster 3 dan diikuti oleh penurunan nilai yang rendah pada cluster selanjutnya.

Setelah menentukan jumlah cluster, selanjutnya menetapkan nilai *Centroid* awal. Dari hasil pengelompokan menggunakan *K-Means*, diperoleh nilai centroid pada Tabel 4.

Tabel 4. Centroid Awal K-Means Clustering

Variabel	C1	C2	C3
X1	-0.6358962	0.2198525	0.8092478
X2	-0.7687725	0.6746110	0.8080065
X3	0.2267277	-0.3200602	-0.1878391
X4	-0.1343658	0.8409871	-0.1600597
X5	0.2674236	1.5092473	-1.0077031
X6	-0.07672144	1.72191907	-0.60877757
X7	-0.28718162	0.07626635	0.37506298

Tabel 4. menunjukkan nilai centroid awal untuk tiga cluster yang diperoleh dari hasil pengelompokan *K-Means clustering*. Selanjutnya ditampilkan hasil pengelompokan *K-Means clustering* dengan nilai centroid dari hasil pengelompokan yang diolah menggunakan software R, pada Tabel 5.

Tabel 5. Hasil *K-Means Clustering*

Cluster	Jumlah	Anggota
1	17	Aceh, Banten, Gorontalo, Jawa Barat, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kep. Bangka Belitung, Maluku, Maluku Utara, Papua, Papua Barat, Riau, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Utara
2	5	Bali, DKI Jakarta, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau.
3	12	Bengkulu, DI Yogyakarta, Jambi, Jawa Tengah, Jawa Timur, Lampung, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Barat, Sulawesi Selatan, Sumatera Baratm Sumatera Selatan.

Selanjutnya evaluasi hasil pengelompokan *K-Means clustering* dengan BIC. Penggunaan BIC pada algoritma K-means penting untuk menentukan jumlah kluster yang optimal dengan pendekatan yang lebih objektif. K-means membutuhkan input jumlah kluster (K) yang ditentukan sebelumnya, tetapi secara intuitif sulit untuk memilih nilai K tanpa panduan. BIC menawarkan solusi dengan mengevaluasi keseimbangan antara kecocokan model (seberapa baik data diwakili oleh kluster) dan kompleksitas model (jumlah kluster yang digunakan). Semakin banyak kluster, kecocokan model dapat meningkat, tetapi pada saat yang sama, kompleksitas model bertambah, yang bisa menyebabkan overfitting. BIC menghitung penalti untuk kompleksitas ini, sehingga memilih jumlah kluster yang memberikan representasi terbaik dengan model yang sederhana. Selain itu, BIC memberikan dasar statistik yang lebih kuat, karena mempertimbangkan log-likelihood dan ukuran dataset secara kuantitatif. Oleh karena itu, BIC sangat berguna untuk memastikan bahwa hasil klusterisasi K-means tidak hanya efektif dalam mendeteksi pola data, tetapi juga efisien dalam representasi modelnya. Berikut adalah nilai log-likelihood dengan perhitungan BIC.

Tabel 6. Log-likelihood

Variabel	Log-Likelihood
K3	-57.54869
K4	-56.97076
K5	-59.12484
K6	-58.87639
K7	-56.41534
K8	-57.22027
K9	-59.05738

Berdasarkan Tabel 6, nilai log-likelihood, model dengan jumlah kluster K=7 memiliki nilai log-likelihood terbaik (-56.41534), yang menunjukkan bahwa model ini memiliki kecocokan yang paling baik dalam merepresentasikan data dibandingkan jumlah kluster lainnya. Nilai log-likelihood yang lebih tinggi (atau nilai negatif yang lebih kecil) menunjukkan bahwa model lebih mampu menjelaskan variasi dalam data dengan lebih baik. Sebaliknya, jumlah kluster seperti K=5 dan K=9 menunjukkan log-likelihood yang lebih rendah (-59.12484 dan -59.05738), menandakan bahwa model dengan jumlah kluster tersebut kurang optimal dalam mengelompokkan data. Oleh karena itu, model dengan K=7 dapat dianggap sebagai kandidat terbaik untuk hasil klusterisasi, meskipun perlu dilakukan

evaluasi tambahan menggunakan kriteria seperti BIC untuk mempertimbangkan kompleksitas model dan memastikan hasil yang optimal.

Setelah mendapatkan nilai log-likelihood, selanjutnya didapatkan nilai BIC pada Tabel 7.

Tabel 7. BIC

Variabel	BIC Values
BIC ₃	651.9475
BIC ₄	655.4738
BIC ₅	659.0002
BIC ₆	662.5266
BIC ₇	666.0529
BIC ₈	669.5793
BIC ₉	673.1056

Berdasarkan nilai BIC pada Tabel 7, hasil clustering menggunakan metode X-Means menunjukkan bahwa jumlah kluster optimal adalah K=3. Hal ini didasarkan pada nilai BIC terkecil, yaitu 651.9475, yang mengindikasikan bahwa model dengan K=3 memberikan keseimbangan terbaik antara kecocokan data dan kompleksitas model. Nilai BIC yang lebih besar pada jumlah kluster lainnya, seperti pada K=4 hingga K=9, menunjukkan bahwa penambahan kluster tidak memberikan peningkatan signifikan dalam representasi data, melainkan hanya menambah kompleksitas model. Dengan demikian, K=3 adalah hasil clustering yang paling efisien dan optimal dalam menggambarkan struktur data yang ada.

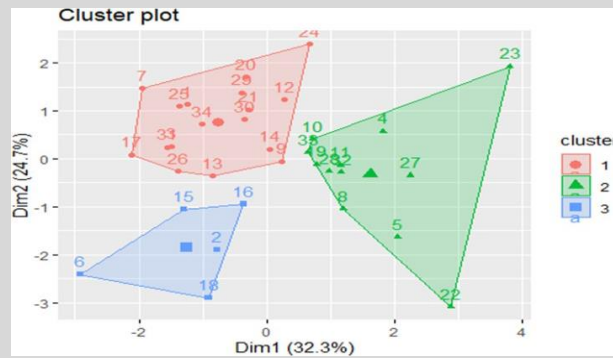
Tampilan hasil pengelompokan X-Means clustering dari hasil pengelompokan yang diolah menggunakan software R, disajikan pada Tabel 8.

Tabel 8. Hasil X-Means Clustering

Cluster	Jumlah	Anggota
1	17	Aceh, Banten, Gorontalo, Jawa Barat, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kep. Bangka Belitung, Maluku, Maluku Utara, Papua, Papua Barat, Riau, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Utara.
2	12	Bengkulu, DI Yogyakarta, Jambi, Jawa Tengah, Jawa Timur, Lampung, Nusa Tenggara Barat, Nusa Tenggara Timur, Sulawesi Barat, Sulawesi Selatan, Sumatera Baratm Sumatera Selatan.
3	5	Bali, DKI Jakarta, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau.

Tabel 8 menunjukkan cluster yang terbentuk serta jumlah anggota masing-masing cluster setelah dikelompokkan dengan *X-Means clustering*. Anggota Provinsi yang masuk ke dalam satu cluster merupakan wilayah yang memiliki sifat-sifat yang homogen berdasarkan data yang ada. Cluster 1 terdiri atas 17 anggota Provinsi, Cluster 2 terdiri atas 12 anggota Provinsi, dan Cluster 3 terdiri atas 5 anggota Provinsi.

Hasil pengelompokan dengan visualisasi menggunakan *X- Means Clustering* disajikan pada Gambar 2.

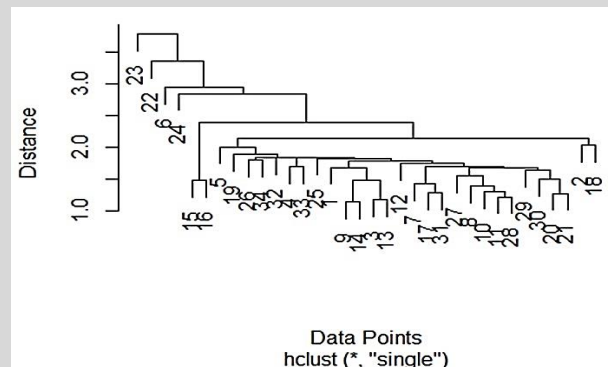


Gambar 2. X-Means Visualisasi

3.5 Analisis cluster X-Means

Agglomerative Hierarchical Clustering (AHC) mencakup beberapa tahap utama, dimulai dari penentuan metode penggabungan kluster seperti Single Linkage, Complete Linkage, yang dipilih berdasarkan karakteristik data. Hasil clustering divisualisasikan menggunakan dendrogram untuk melihat struktur hierarkis pengelompokan data secara intuitif. Selanjutnya, jumlah kluster optimal ditentukan dengan memotong dendrogram pada ketinggian tertentu atau menggunakan evaluasi seperti Silhouette Score untuk memastikan hasil yang terbaik. Terakhir, evaluasi kluster dilakukan untuk mengukur kualitas kluster yang terbentuk, memastikan data dalam satu kluster homogen dan antar kluster terpisah dengan baik, sehingga hasil klusterisasi sesuai dengan tujuan analisis.

Metode pertama yang diuji dalam AHC yaitu *Single linkage* yang menggabungkan kluster berdasarkan jarak terdekat antara dua titik data dari kluster yang berbeda. Dalam pendekatan ini, jarak antar kluster dihitung sebagai jarak minimum antara pasangan titik-titik dari dua kluster tersebut. Single Linkage sering disebut juga sebagai minimum linkage [16] Di bawah ini merupakan penentuan pusat kluster pada single linkage menggunakan dendrogram.

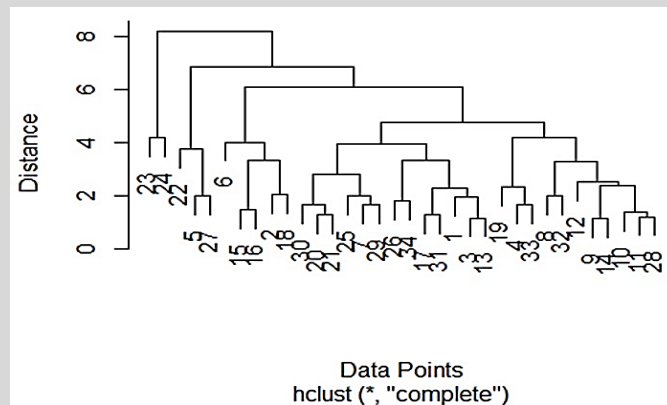


Gambar 3. Single Linkages

Berdasarkan dendrogram yang menggunakan metode Single Linkage, jumlah kluster optimal dapat ditentukan dengan memotong dendrogram pada ketinggian tertentu yang menunjukkan pemisahan paling jelas antara kelompok data. Pada ketinggian sekitar 2.0 pada sumbu jarak (Distance), dendrogram menunjukkan pemisahan menjadi 3 kluster utama, ditandai dengan adanya tiga cabang besar yang belum saling bergabung. Pemilihan ini didasarkan pada jarak antar kluster yang cukup signifikan sebelum penggabungan lebih lanjut terjadi. Dengan demikian, jumlah kluster paling optimal adalah 3 kluster.

Selanjutnya dilakukan uji metode *Complete linkage* yang menggabungkan kluster berdasarkan jarak terjauh antara dua titik data dari kluster yang berbeda. Dalam pendekatan ini, jarak antar kluster dihitung sebagai jarak maksimum antara pasangan titik-titik dari dua kluster tersebut. Metode ini cenderung menghasilkan kluster yang lebih kompak dan homogen dibandingkan dengan Single Linkage, karena jarak maksimum memastikan bahwa semua titik dalam kluster memiliki

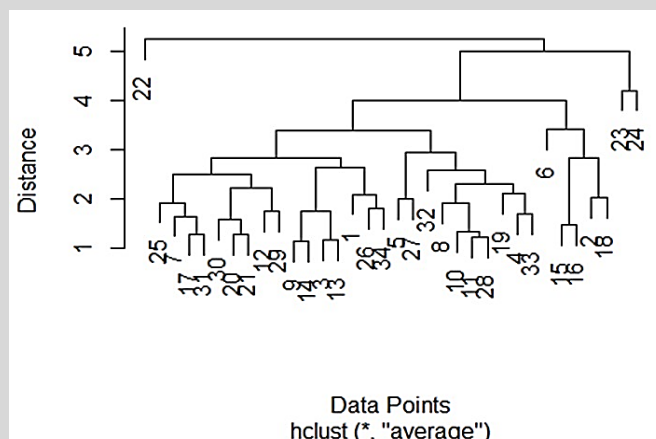
keterhubungan yang lebih ketat [16]. Di bawah ini merupakan penentuan pusat kluster pada Complete Linkage menggunakan dendrogram, di mana pemisahan antar kluster terlihat lebih jelas dengan cabang-cabang yang lebih seragam.



Gambar 4. Complete Linkage

Jumlah kluster optimal berdasarkan dendrogram tersebut adalah 4 kluster, karena pada titik ini struktur kluster terlihat lebih jelas dan pemisahan antar kluster cukup signifikan.

Selanjutnya dilakukan uji metode *Average Linkage* yang menggabungkan kluster rata-rata jarak antara semua pasangan titik dari dua kluster yang berbeda. Metode ini memberikan pendekatan yang seimbang antara Single Linkage dan Complete Linkage, sehingga menghasilkan kluster yang lebih stabil dan merata. Average Linkage mampu menghindari efek chaining pada Single Linkage dan menghasilkan pemisahan yang tidak terlalu ketat seperti Complete Linkage. Pemisahan kluster dalam metode ini divisualisasikan melalui dendrogram, di mana cabang-cabangnya menunjukkan pembentukan kluster yang lebih seimbang dan natural [16].



Gambar 5. Average Linkage

Berdasarkan dendrogram yang menggunakan metode Average Linkage, jumlah kluster optimal dapat ditentukan dengan melihat ketinggian tertentu pada sumbu Distance (sumbu Y) di mana pemisahan antar kluster terlihat jelas. Pada ketinggian sekitar 3.0, dendrogram menunjukkan adanya 4 kluster utama, ditandai dengan cabang-cabang besar yang terpisah sebelum penggabungan lebih lanjut. Pemilihan ketinggian ini memastikan bahwa data dalam satu kluster memiliki keterhubungan yang lebih erat dengan jarak rata-rata antar titik yang dihitung lebih seimbang. Dengan demikian, jumlah kluster optimal pada dendrogram ini adalah 4 kluster, di mana struktur pengelompokan terlihat lebih stabil dan natural.

Selanjutnya akan dilakukan evaluasi AHC menggunakan *silhouette score*. Uji validasi AHC dilakukan menggunakan *Silhouette Index* (SI). SI yang memiliki nilai mendekati 1 mempunyai arti

bahwa objek semakin cocok dalam cluster. SI yang memiliki nilai negatif menandakan objek tersebut tidak cocok berada dalam cluster. SI yang memiliki nilai 0 atau mendekati nol menandakan objek terletak pada batas antara dua cluster [17]. Dibawah ini merupakan hasil evaluasi indeks pada metode Single Linkage, Complete Linkage dan Average Linkage yang disajikan pada Tabel 9.

Tabel 9. Nilai Indeks Validitas Cluster *Silhouette Index* (SI) Metode AHC

<i>Metode</i>	<i>Nilai SI</i>
<i>Single Linkage</i>	0,28
<i>Complete Linkage</i>	0,24
<i>Average Linkage</i>	0,27

Dari Tabel 9, terlihat bahwa pengelompokan dengan menggunakan *Single Linkage* memberikan nilai lebih besar untuk SI dibandingkan pengelompokan yang menggunakan *Complete Linkage* dan *Average Linkage*. Hal ini membuktikan metode *Single Linkage* dengan memberikan hasil dari pengelompokan lebih optimal dari metode *Complete Linkage* dan *Average Linkage*.

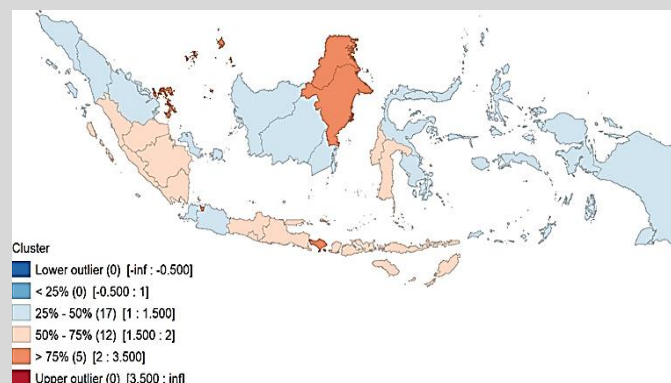
Selanjutnya dilakukan pembentukan *cluster* menggunakan *single linkage* dengan nilai *Silhouette Coefficient* terbaik dari hasil pengelompokan yang diolah menggunakan *software R*.

Tabel 10. Hasil *Single Linkage Clustering*

Cluster	Jumlah	Anggota
1	32	Aceh, Banten, Gorontalo, Jawa Barat, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kep. Bangka Belitung, Maluku, Maluku Utara, Papua, Papua Barat, Riau, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Utara, Bali, DKI Jakarta, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau, Bengkulu, DI Yogyakarta, Jambi, Jawa Tengah, Jawa Timur, Lampung, Sulawesi Barat, Sulawesi Selatan, Sumatera Barat, Sumatera Selatan.
2	1	Nusa Tenggara Barat
3	1	Nusa Tenggara Timur

3.6 Interpretasi Hasil

Hasil pengelompokan stunting dengan metode X-Means Clustering membentuk 3 cluster. Cluster 1 terdiri atas 17 provinsi, cluster 2 terdiri atas 12 provinsi dan *cluster* 3 terdiri atas 5 provinsi. Visualisasi menggunakan *X-Means* dengan 34 provinsi di Indonesia dibesarkan pada Gambar 6.



Gambar 6. Pemetaan Stunting di 34 Provinsi

Dari hasil pengelompokkan yang diperoleh, kemudian dihitung nilai rata-rata tiap variabel pada cluster-cluster yang ada sebagai dasar untuk menginterpretasikan tiap cluster. Hasil perhitungannya menggunakan software R, disajikan pada Tabel 11.

Tabel 11. Nilai Rata-Rata Variabel Tiap Cluster

Variabel	Cluster 1	Cluster 2	Cluster 3
Balita yang Mendapatkan ASI Eksklusif	-0.6358962	0.8092478	0.2198525
Balita yang diberika MPASI	-0.7687725	0.8080065	0.6746110
Berat Badan Lahir Rendah (BBLR)	0.2267277	-0.1878391	-0.3200602
Rumah Tangga yang Mmeiliki Akses Sanitasi Lengkap	-0.1343658	-0.1600597	0.8409871
Rumah Tangga yang Memiliki Akses Air Minum Layak	0.2674236	-1.0077031	1.5092473
Rata-rata Pengeluaran Perkapita Sebulan	-0.07672144	-0.60877757	1.72191907
Rata-rata Konsumsi Kalori Perkapita Sehari	-0.28718162	0.37506298	0.07626635

Berdasarkan nilai rata-rata variabel tiap klaster, terdapat perbedaan signifikan dalam karakteristik setiap klaster. Pada Cluster 1, terdapat kekurangan yang cukup mencolok dalam beberapa aspek seperti pemberian ASI eksklusif (-0.6358962) dan pemberian MPASI (-0.7687725), yang menunjukkan bahwa praktik ini perlu ditingkatkan secara signifikan. Selain itu, kasus BBLR (Berat Badan Lahir Rendah) di Cluster 1 memiliki nilai positif (0.2267277), mengindikasikan tingginya prevalensi kasus BBLR yang perlu diatasi melalui peningkatan gizi ibu hamil. Di sisi lain, variabel akses sanitasi lengkap (-0.1343658) dan akses air minum layak (0.2674236) masih memerlukan perbaikan meskipun sedikit lebih baik. Cluster 1 juga mengalami masalah pada konsumsi kalori perkapita (-0.28718162), yang berada di bawah rata-rata, serta pengeluaran perkapita (-0.07672144) yang menunjukkan kondisi ekonomi masih cukup lemah.

Sementara itu, Cluster 2 menghadapi masalah yang lebih serius, khususnya pada variabel akses air minum layak (-1.0077031) yang menjadi perhatian utama. Nilai ini menunjukkan keterbatasan signifikan dalam akses air bersih yang layak, yang berpengaruh terhadap kesehatan masyarakat. Selain itu, pengeluaran perkapita (-0.60877757) dan akses sanitasi lengkap (-0.1600597) juga rendah, menandakan adanya keterbatasan ekonomi dan infrastruktur. Namun, Cluster 2 memiliki kondisi yang lebih baik dalam pemberian ASI eksklusif (0.8092478) dan MPASI (0.8080065), serta konsumsi kalori (0.37506298) yang cukup memadai. Di sisi lain, Cluster 3 menunjukkan performa yang lebih baik secara keseluruhan, terutama dalam akses air minum layak (1.5092473), akses sanitasi lengkap (0.8409871), dan pengeluaran perkapita (1.72191907). Namun, Cluster 3 masih perlu meningkatkan pemberian ASI eksklusif (0.2198525) dan konsumsi kalori per kapita (0.07626635). Dengan demikian, Cluster 1 dan Cluster 2 menjadi fokus utama intervensi, terutama dalam peningkatan akses infrastruktur, ekonomi, serta pemenuhan gizi bagi balita dan rumah tangga.

Selanjutnya Hasil pengelompokkan stunting dengan metode AHC membentuk 3 cluster. Cluster 1 terdiri atas 32 provinsi, cluster 2 terdiri atas 1 provinsi dan cluster 3 terdiri atas 1 provinsi. Tampilan hasil pengelompokkan Single Linkage dengan nilai Silhouette Coefficient terbaik dari hasil pengelompokkan yang diolah menggunakan software R, disajikan pada Tabel 12.

Tabel 12. Hasil *single Linkage Clustering*

Cluster	Jumlah	Anggota
1	32	Aceh, Banten, Gorontalo, Jawa Barat, Kalimantan Barat, Kalimantan Selatan, Kalimantan Tengah, Kep. Bangka Belitung, Maluku, Maluku Utara, Papua, Papua Barat, Riau, Sulawesi Tengah, Sulawesi Tenggara, Sulawesi Utara, Sumatera Utara, Bali, DKI Jakarta, Kalimantan Timur, Kalimantan Utara, Kepulauan Riau, Bengkulu, DI Yogyakarta, Jambi, Jawa Tengah, Jawa Timur, Lampung, Sulawesi Barat, Sulawesi Selatan, Sumatera Baratm Sumatera Selatan.
2	1	Nusa Tenggara Barat
3	1	Nusa Tenggara Timur

4. Kesimpulan

Berdasarkan analisis menggunakan metode X-Means dan Agglomerative Hierarchical Clustering (AHC) dengan pendekatan Single Linkage, diperoleh tiga kluster optimal yang mencerminkan pola dan karakteristik stunting di Indonesia. Metode X-Means menunjukkan klusterisasi yang efisien dengan keseimbangan terbaik antara kecocokan data dan kompleksitas model (BIC = 651.9475), sementara metode AHC menghasilkan klusterisasi dengan nilai Silhouette Index (SI) tertinggi sebesar 0.28. Hasil pengelompokan ini memberikan gambaran yang jelas untuk menyusun strategi intervensi stunting yang lebih terarah sesuai dengan karakteristik masing-masing kluster.

Referensi

- [1] S. Ningsih, M. R. Madonsa, S. L. Mahmud, I. Djakaria, and S. K. Nasib, "Implementasi Regresi Logistik Biner Stratifikasi Pada Pemodelan Stunting Untuk Anak Balita Di Kabupaten Gorontalo," *Jambura J. Probab. Stat.*, vol. 5, no. 1, pp. 19–23, 2024, doi: 10.37905/jjps.v5i1.19793.
- [2] Kemenkes, "hasil studi status gizi indonesia (SSGI) tingkat nasional, provinsi dan kabupaten/kota tahun 2021," *Buana Ilmu*, vol. 2, no. 1, 2017, doi: 10.36805/bi.v2i1.301.
- [3] N. Moroke, "Faculty of Economic and Management Sciences ' On Multivariate Analysis of High-dimensional data ,'" 2019.
- [4] N. A. Kilo, M. R. Katili, and I. K. Hasan, "Research in the Mathematical and Natural Sciences Perbandingan Metode K-Means dan K-Medoids Dengan Validitas Davies-Bouldin Indeks , Dunn Indeks dan Indeks Connectivity Pada Pengelompokan Masyarakat Penerima Bantuan Langsung Tunai," vol. 4, no. 1, pp. 8–15, 2025, doi: 10.55657/rmns.v4i1.190.
- [5] G. D. Rembulan, T. Wijaya, D. Palullungan, K. N. Alfina, and M. Qurthuby, "Kebijakan Pemerintah Mengenai Coronavirus Disease (COVID-19) di Setiap Provinsi di Indonesia Berdasarkan Analisis Kluster," *JIEMS (Journal Ind. Eng. Manag. Syst.*, vol. 13, no. 2, 2020, doi: 10.30813/jiems.v13i2.2280.
- [6] Y. T. Chen and D. M. Witten, "Selective inference for k-means clustering," vol. 24, pp. 1–20, 2022, [Online]. Available: <http://arxiv.org/abs/2203.15267>
- [7] D. Aulia *et al.*, "Penerapan Algoritma K-Means dalam Proses Clustering Penilaian Kinerja ASN," *J. Ris. Sist. Inf. Dan Tek. Inform.*, vol. 6, p. 47, 2021, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jurasik>
- [8] P. H. Putra, A. Hasibuan, and E. A. Marpaung, "Analisis Klasifikasi Metode X-Means Pada Minat dan Bakat Anak Dimasa Pandemi," *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 19, no. 2, pp. 424–429, 2022, [Online]. Available: <https://ejournal.uin-suska.ac.id/index.php/sitekin/article/download/17889/7834>
- [9] C. Ramdani, "ELANG: Journal of Interdisciplinary Research Komparasi Algoritme X-Means

- dan K-Medoids pada Klasterisasi Data Akademik Mahasiswa,” 2023.
- [10] dan S. Putra, Y. D., Sudarma, M., “Clustering history data penjualan menggunakan algoritma k-means,” vol. 20(2):195, 2021.
 - [11] M. R. Sulistio, N. Suarna, and O. Nurdiawan, “Analisa Penerapan Metode Clustering X-Means Dalam Pengelompokan Penjualan Barang,” *J. Teknol. Ilmu Komput.*, vol. 1, no. 2, pp. 37–42, 2023, doi: 10.56854/jtik.v1i2.49.
 - [12] L. Zahrotun *et al.*, “Analisis Persebaran UMKM Bidang Jasa Menggunakan Metode AHC Complete Linkage,” vol. 4, no. 2, pp. 255–265, 2023.
 - [13] S. Suhirman and H. Wintolo, “System for Determining Public Health Level Using the Agglomerative Hierarchical Clustering Method,” *Compiler*, vol. 8, no. 1, p. 95, 2019, doi: 10.28989/compiler.v8i1.425.
 - [14] T. Märzinger, J. Kotík, and C. Pfeifer, “Application of hierarchical agglomerative clustering (Hac) for systemic classification of pop-up housing (puh) environments,” *Appl. Sci.*, vol. 11, no. 23, 2021, doi: 10.3390/app112311122.
 - [15] Kemenkes, “Hasil Survei Status Gizi Indonesia (SSGI) 2022. Kementerian Kesehatan Indonesia,” pp. 1–7, 2023.
 - [16] H. A. dan N. Abbass, *Data Mining: A Heuristic Approach*. 2003. doi: 10.1108/14684520310502315.
 - [17] S. Prasetyadi, *Buku Ajar Data Maining*. 2024.