

Perbandingan *K-Means* dan *Hierarchical Clustering* dalam Pengelompokan Daerah Beresiko Stunting

Indra¹, Nahya Nur², Muh. Iqram³, Nurul Inayah⁴

^{1,2,3,4}Universitas Sulawesi Barat, Talumung, Majene, Sulawesi Barat, Indonesia

Email: indra@unsulbar.ac.id¹, nahya.nur@unsulbar.ac.id², muhammadiqram298@gmail.com³, naya.mmj@gmail.com⁴

Abstract – Currently, Indonesia is one of the countries with a fairly high stunting rate in the world, where the prevalence of stunting is still in the range of 21.6%, while the minimum standard for stunting prevalence set by WHO is 20%. Stunting is a condition of failure to thrive that occurs early in life, usually in children aged 0-5 years. To overcome this problem, the government and related parties have carried out various efforts and intervention programs, one of which is determining priority areas for handling stunting by clustering. In this research, we will cluster stunting areas based on provinces in Indonesia by referring to several parameters, namely the percentage of immunization, proportion of stunting, coverage of exclusive breastfeeding, coverage of vitamins and blood supplement tablets, as well as access to proper sanitation and drinking water. This research will compare clusters formed using Hierarchical Clustering and K Means. The results of the comparison between the K-Means and Hierarchical Clustering methods show that K-Means produces better cluster grouping in terms of the Silhouette Coefficient value of 0.48 and the Calinski-Harabasz index of 10.49 with the number of clusters formed being 2 clusters. In the Hierarchical Clustering Algorithm, the resulting Silhouette Coefficient value is 0.47 and the Calinski-Harabasz index is 9.54. The greater the value of the Silhouette Coefficient and Calinski-Harabasz index, the better the cluster that is formed.

Keywords - stunting, K-Means, Hierarchical Clustering

Intisari - Saat ini Indonesia merupakan salah satu negara dengan tingkat stunting yang cukup tinggi di dunia, dimana prevalensi stunting masih berada pada kisaran 21,6% sedangkan standar minimal prevalensi stunting yang telah ditetapkan oleh WHO adalah 20%. Stunting merupakan kondisi gagal tumbuh yang terjadi pada masa awal kehidupan, biasanya pada anak usia 0-5 tahun. Untuk mengatasi permasalahan ini, pemerintah dan pihak terkait telah melakukan berbagai upaya dan program intervensi, salah satunya adalah menentukan daerah yang menjadi prioritas penanganan stunting dengan melakukan klusterisasi. Dalam penelitian ini akan melakukan klusterisasi daerah stunting berdasarkan provinsi yang ada di Indonesia dengan mengacu pada beberapa parameter yaitu persentase imunisasi, proporsi stunting, cakupan pemberian ASI eksklusif, cakupan pemberian vitamin dan tablet tambah darah, serta akses terhadap sanitasi dan air minum yang layak. Penelitian ini akan membandingkan antara kluster yang terbentuk menggunakan *Hierarchical Clustering* dan K Means. Hasil perbandingan antara metode *K-Means* dan *Hierarchical Clustering* memperlihatkan bahwa *K-Means* menghasilkan pengelompokan klaster yang lebih baik ditinjau dari nilai *Silhouette Coefficient* sebesar 0.48 dan *Calinski-Harabasz index* sebesar 10.49 dengan jumlah klaster yang terbentuk sebanyak 2 klaster. Pada Algoritma Hierarchical Clustering, nilai *Silhouette Coefficient* yang dihasilkan adalah 0.47 dan *Calinski-Harabasz index* sebesar 9.54. Semakin besar nilai *Silhouette Coefficient* dan *Calinski-Harabasz index*, maka semakin baik klaster yang terbentuk.

Kata Kunci - stunting, K-Means, Hierarchical Clustering

I. PENDAHULUAN

Stunting adalah kondisi gagal tumbuh atau pertumbuhan terhambat yang terjadi pada masa awal kehidupan, biasanya pada anak usia 0-5 tahun. Stunting terjadi ketika anak tidak mencapai tinggi badan yang seharusnya pada usia tertentu, karena kurangnya asupan gizi

yang cukup dan berkualitas, serta paparan infeksi yang berulang-ulang. Stunting sangat berdampak pada kesehatan dan perkembangan anak, serta berpotensi menyebabkan masalah kesehatan yang serius di masa dewasa, seperti keterlambatan mental dan fisik, rendahnya produktivitas, dan masalah kesehatan kronis seperti penyakit jantung dan diabetes. Stunting juga dapat menyebabkan kecacatan fisik dan perkembangan otak yang tidak normal [1]. Stunting sangat erat kaitannya dengan pertumbuhan fisik pada balita. Akan tetapi, tidak semua anak yang pendek dapat dikategorikan sebagai stunting. Tinggi badan anak dipengaruhi oleh berbagai faktor, termasuk faktor genetik, lingkungan, gizi, dan kesehatan secara umum. Selain pertumbuhan fisik, stunting juga dipengaruhi oleh kondisi lainnya seperti perkembangan otak dan sebagainya [2].

Indonesia masih menjadi salah satu negara dengan tingkat stunting yang cukup tinggi di dunia. Menurut data dari Kementerian Kesehatan Indonesia pada tahun 2022, angka stunting di Indonesia mencapai 21,6%. Meskipun prevalensi stunting menurun dimana pada tahun 2021 angka stunting masih mencapai 24,4%, tetapi masih belum memenuhi standar WHO yang menekankan bahwa prevalensi stunting harus kurang dari 20% [3].

Untuk mengatasi masalah stunting di Indonesia, pemerintah dan berbagai pihak telah melakukan berbagai upaya dan program intervensi, antara lain dengan meningkatkan pengetahuan dan kesadaran masyarakat tentang gizi sehat dan praktik gizi yang baik, meningkatkan akses pada makanan yang sehat dan bergizi, mengembangkan program-program gizi khusus bagi kelompok rentan, dan meningkatkan kebersihan lingkungan dalam mencegah penyakit infeksi. Untuk menentukan target daerah yang menjadi prioritas penanganan stunting menjadi hal yang perlu untuk ketahui, salah satunya dengan melakukan klusterisasi. Penting untuk melakukan klusterisasi daerah stunting karena dengan melakukan klusterisasi, kita dapat membagi daerah-daerah yang mengalami stunting berdasarkan karakteristik tertentu seperti tingkat kemiskinan, status sosial-ekonomi, tingkat pendidikan, akses sanitasi, dan faktor lingkungan lainnya. Dengan memahami karakteristik setiap kluster atau kelompok daerah, kita dapat merancang program intervensi yang lebih sesuai dan efektif dalam menangani kasus stunting di daerah masing-masing. Selain itu, klusterisasi juga dapat membantu dalam pemetaan sumber daya dan alokasi anggaran yang lebih tepat dan efisien. Dengan cara ini, upaya penurunan angka stunting di Indonesia dapat dilakukan secara terarah dan terfokus sehingga memberikan manfaat yang lebih besar pada anak-anak yang rentan mengalami stunting.

Dalam penelitian ini akan melakukan klusterisasi daerah stunting berdasarkan provinsi yang ada di Indonesia dengan mengacu pada beberapa parameter seperti persentase anak usia 12-23 bulan yang mendapatkan imunisasi dasar lengkap, proporsi stunting (TB/U) pada balita, persentase rerata balita umur 6-59 bulan ditimbang perbulan, cakupan imunisasi dasar lengkap, cakupan bayi baru lahir yang mendapat inisiasi menyusui dini, cakupan pemberian kapsul vitamin A, cakupan pemberian tablet tambah darah, persentase rumah tangga dengan akses terhadap sanitasi yang layak, serta persentase rumah tangga dengan akses terhadap air minum yang layak. Pengelompokan daerah stunting akan dilakukan dengan menggunakan metode *Hierarchical Clustering* dan *K-Means*, untuk selanjutnya dilakukan perbandingan dari hasil klusterisasi kedua metode tersebut. Hasil identifikasi pengempokan daerah dapat dijadikan sebagai rujukan dalam menyusun target daerah yang menjadi prioritas penanganan stunting.

II. SIGNIFIKANSI STUDI

A. Studi Literatur

1. Hierarchical Clustering

Hierarchical Clustering adalah sebuah metode clustering atau pengelompokan data yang berbasis pada struktur hierarki. Hierarki dalam hal ini berarti bahwa data akan dielompokkan

ke dalam subkelompok yang semakin spesifik dan terorganisir hingga level terkecil yang disebut leaf atau daun. *Hierarchical Clustering* dapat dilakukan dengan dua cara yaitu *agglomerative* dan *divisive* [4].

Kelebihan dari *Hierarchical Clustering* adalah kemampuannya dalam menganalisis struktur hierarki dari data. Selain itu, dengan *Hierarchical Clustering* kita dapat melihat gambaran besar dari data dengan mudah. Namun, kelemahan *Hierarchical Clustering* adalah waktu komputasi yang memakan waktu cukup lama pada dataset yang besar serta membutuhkan pemilihan parameter atau pengelompokan yang tepat untuk menghasilkan hasil clustering yang lebih akurat.

Berikut adalah langkah-langkah dalam algoritma Hierarchical Clustering:

- a. Menentukan tipe Hierarchical Clustering
Pada penelitian ini akan menggunakan *agglomerative Clustering*. Pada metode kluster tersebut, awalnya setiap data akan dianggap sebagai sebuah kelompok terpisah. Kemudian, kelompok-kelompok tersebut akan digabungkan berdasarkan jarak antar kelompok terdekat. Pada setiap iterasi, kelompok yang memiliki jarak terdekat akan digabungkan hingga hanya tersisa satu kelompok yang memuat seluruh data.
- b. Menentukan Metode Pengukuran Similaritas
Metode pengukuran similaritas atau jarak antara data sangat penting dalam *Hierarchical Clustering*. Beberapa metode pengukuran similaritas yang umum digunakan termasuk *euclidean distance*, *manhattan distance*, dan *cosine similarity*.
- c. Membentuk Matriks Similaritas
Langkah berikutnya adalah membentuk matriks similaritas yang menunjukkan jarak antara setiap pasangan data. Matriks similaritas ini akan digunakan untuk membentuk dendrogram, yaitu representasi visual dari struktur hierarki dataset.
- d. Menggabungkan Cluster Berdasarkan Similaritas Terdekat
Pada *agglomerative Clustering*, setiap data awalnya dianggap sebagai kelompok terpisah. Selanjutnya, secara iteratif, dua kelompok yang memiliki jarak terdekat akan digabungkan satu sama lain.
- e. Menghitung Jarak Baru Antara Kluster yang Dibentuk
Setelah dua cluster digabungkan, jarak baru antara cluster tersebut harus dihitung. Ada beberapa metode untuk menghitung jarak baru, seperti *single linkage*, *complete linkage*, *average linkage*, dan *ward's method*.
- f. Membentuk Dendrogram
Setiap kali sebuah cluster baru terbentuk, dendrogram perlu dibentuk untuk merepresentasikan struktur hierarki dataset. Dendrogram ini akan memudahkan interpretasi dari hasil clustering. Setelah dendrogram terbentuk, tahapan selanjutnya adalah memotong dendrogram agar terbentuk kelompok-kelompok cluster yang diinginkan. Pemotongan dilakukan dengan cara memilih tinggi potongan tertentu pada dendrogram.

Setelah langkah-langkah di atas dilakukan, kita akan mendapatkan hasil kluster dari dataset yang telah diproses. Namun, perlu diketahui bahwa penentuan jumlah kluster pada *Hierarchical Clustering* bisa cukup subjektif dan memakai metode lain misalnya *elbow method* untuk membantu penentuan k optimal.

Cophenetic correlation coefficient merupakan sebuah metode yang digunakan untuk melakukan validasi terhadap metode *Hierarchical Clustering* dengan menghitung seberapa mirip kluster yang digabungkan pada metode tersebut. *Cophenetic Correlation Coefficient* mengukur seberapa baik representasi jarak asli antara data pada kluster-kluster yang digabungkan dalam hasil dendrogram yang dihasilkan oleh algoritma *Hierarchical Clustering*.

Metode ini membandingkan jarak asli antara semua pasang data dalam dataset dengan jarak yang diukur dari dendrogram [5].

Cophenet Correlation Coefficient dapat dihitung menggunakan persamaan (1):

$$c = \frac{\sum_{i < j} (x(i,j) - \bar{x})(t(i,j) - \bar{t})}{\sqrt{(\sum_{i < j} (x(i,j) - \bar{x})^2)(\sum_{i < j} (t(i,j) - \bar{t})^2)}} \quad (1)$$

Korelasi yang dihasilkan memiliki rentang dari 0 hingga 1. Nilai yang lebih mendekati 1 menunjukkan bahwa dendrogram dengan baik merepresentasikan jarak asli antara data, sehingga penggabungan kluster-kluster dalam algoritma *Hierarchical Clustering* dianggap efisien. Sebaliknya, nilai yang lebih rendah menunjukkan bahwa dendrogram tidak efisien dalam merepresentasikan jarak asli antara data.

2. *K Means Clustering*

K-Means adalah salah satu algoritma clustering yang populer dalam analisis data. Algoritma ini digunakan untuk membagi data menjadi beberapa kelompok atau kluster yang memiliki karakteristik yang sama [4]. Tujuannya adalah untuk memaksimalkan kesamaan antarobservasi dalam satu kelompok dan meminimalkan kesamaan antarkelompok.

K-Means bekerja dengan cara memilih K titik pusat secara acak (*centroid*) di dalam data dan menghitung jarak antara setiap observasi dengan setiap pusat. Setelah itu, setiap observasi akan diatribusikan ke kelompok dengan pusat terdekat. Kemudian, rata-rata dari setiap kelompok akan dihitung dan digunakan sebagai pusat baru untuk setiap kelompok[6]. Langkah ini diulang selama nilai pusat tidak berubah atau telah mencapai batas iterasi.

Berikut adalah langkah-langkah *K-Means* dalam proses clustering data:

- a. Menentukan jumlah kluster
Langkah pertama dalam *K-Means* adalah menentukan jumlah kluster atau kelompok yang akan dibuat. Jumlah kluster dipilih berdasarkan pada tujuan dari clustering dan karakteristik data yang akan dianalisis.
- b. Memilih centroid awal
Setelah jumlah kluster ditentukan, langkah selanjutnya adalah memilih titik centroid awal secara acak dari seluruh data. Titik centroid inilah yang akan menjadi pusat dari setiap kelompok yang dibuat.
- c. Menghitung jarak
Setelah titik centroid awal dipilih, langkah selanjutnya adalah menghitung jarak antara setiap observasi dengan titik centroid. Jarak ini biasanya dihitung dengan menggunakan metrik *Euclidean* yang ditunjukkan pada persamaan (2).

$$dist = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

- d. Mengelompokkan data
Setelah jarak dihitung, langkah selanjutnya adalah mengelompokkan setiap data ke dalam kelompok yang paling dekat dengan titik centroid. Tiap data hanya bisa masuk ke dalam satu kelompok saja.
- e. Menghitung titik centroid baru
Setelah semua data dikategorikan, langkah selanjutnya adalah menghitung ulang titik centroid baru dari masing-masing kelompok berdasarkan rata-rata dari setiap data pada kelompok tersebut.

Setelah titik *centroid* baru dihitung, langkah selanjutnya adalah mengulangi proses dari langkah 3-5 dengan menggunakan titik centroid yang baru sebagai pusat dari kelompok-

kelompok tersebut. Prosedur ini diulang terus-menerus hingga titik centroid tidak berubah atau jumlah iterasi yang telah ditentukan telah tercapai.

Terdapat beberapa kelebihan dan kekurangan dari K-Means. Kelebihannya adalah algoritma ini cukup sederhana dan efisien untuk ukuran data yang besar. Selain itu, data yang terbagi ke dalam kelompok yang kompak dan homogen dapat membantu mempercepat analisis data selanjutnya. Namun, kelemahannya adalah sensitif terhadap posisi awal pusat yang dipilih serta pada data yang mengandung pencilan. Oleh karena itu, pengguna harus berhati-hati dalam memilih jumlah k dan memahami karakteristik data yang akan dianalisis sebelum menggunakan algoritma ini.

3. *Silhouette Coefficient*

Silhouette Coefficient adalah salah satu teknik evaluasi kualitas kluster dalam data mining dan machine learning. Metrik ini mengukur seberapa dekat setiap observasi dalam kluster dengan observasi dalam kluster lain dan seberapa terpisah antara kluster [7]. Tujuan dari *Silhouette Coefficient* adalah untuk mengevaluasi apakah kluster yang dibentuk dari data adalah kluster yang baik atau tidak.

Proses dalam menghitung koefisien Silhouette dimulai dengan mencari jarak rata-rata data ke-i dengan semua data dalam kluster yang sama, dengan asumsi bahwa data ke-i berada dalam kluster A. Untuk menghitung $a(i)$ dapat dituliskan dalam persamaan (3).

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (3)$$

Dimana A adalah banyaknya data di kluster A

Langkah berikutnya adalah menghitung nilai $b(i)$, yang merupakan jarak rata-rata minimum antara data ke-i dengan semua data yang berada pada kluster yang berbeda. Misalkan kluster A dan kluster C adalah kluster yang berbeda, sehingga untuk menghitung jarak rata-rata antara data ke-i dengan semua data di kluster C dapat ditunjukkan pada persamaan (4):

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (4)$$

Dimana C adalah banyaknya data di kluster C.

Setelah diperoleh nilai $d(i, C)$ pada semua kluster $C \neq A$, langkah berikutnya adalah mencari nilai jarak paling rendah sebagai nilai $b(i)$.

$$b(i) = \min_{C \neq A} d(i, C) \quad (5)$$

Jika kluster B memiliki nilai jarak yang paling kecil, maka $d(i, B) = b(i)$ dapat dikatakan sebagai tetangga dari data ke-i dan merupakan kluster terbaik kedua untuk data ke-i setelah kluster A. Setelah nilai $a(i)$ dan $b(i)$ diketahui, langkah terakhir adalah menghitung *Silhouette Coefficient*. *Silhouette Coefficient* dapat dinyatakan dalam persamaan (6) berikut:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (6)$$

Rentang nilai *Silhouette Coefficient* mulai dari -1 hingga 1, dimana nilai yang baik adalah antara 0 dan 1, sementara nilai negatif menunjukkan hasil clustering yang buruk. Ketika *Silhouette Coefficient* dihitung untuk setiap klaster, nilai-nilai ini dapat digunakan untuk mengevaluasi kualitas klaster terkait seberapa baik pengelompokan tersebut. Misalnya, dalam kasus di mana ditemukan satu klaster yang memiliki nilai *Silhouette Coefficient* yang buruk pada pengukuran awal, maka kita mungkin akan mengubah beberapa parameter dalam clustering dengan tujuan menghasilkan hasil clustering yang lebih baik [8].

4. *Calinski-Harabasz Index*

Calinski-Harabasz index merupakan salah satu metrik evaluasi yang biasanya digunakan dalam melakukan evaluasi klaster dimana nilai yang lebih tinggi mengindikasikan bahwa model kluster yang terbentuk lebih jelas. *Calinski-Harabasz Indeks* didefinisikan sebagai rasio antara total dispersi antar klaster dan dispersi dalam klaster untuk semua klaster. Untuk kumpulan data X berukuran n yang telah dikelompokkan menjadi k klaster, *Calinski-Harabasz indeks* dinyatakan menurut persamaan berikut

$$s(k) = \frac{tr(B_k)m - k}{tr(w_k)k - 1} \tag{7}$$

$$tr(B_k) = \sum_{q=1}^k \sum_{x \in C_q} (x - c_q)(x - c_q)^T \tag{8}$$

$$tr(W_k) = \sum_{q=1}^k n_q(c_q - c_E)(c_q - c_E)^T \tag{9}$$

dimana $tr(B_k)$ merupakan trace dari matriks disperse antar klaster, $tr(w_k)$ adalah trace matriks disperse dalam klaster, k adalah jumlah klaster, C_q adalah himpunan titik yang berada pada klaster q dengan c_q merupakan pusat dari kluster q, c_E adalah pusat dari himpunan data E, dan n_q menunjukkan banyaknya titik yang berada pada klaster q[9].

5. *Davies-Bouldin index*

Davies-Bouldin index (DBI) mengevaluasi hasil dari model clustering. DBI menerapkan proses pengukuran suatu hasil pengelompokan berdasar pada nilai kohesi dan separasi. Hasil pengelompokan kohesi didefinisikan sebagai jumlah dari kedekatan data terhadap suatu titik pusat (*centroid*) dari suatu hasil pengelompokan sedangkan hasil pengelompokan separasi dihitung berdasarkan jarak antar *centroid* dari klasternya. Hasil nilai DBI yang paling optimum adalah nilai DBI yang terkecil [10]. Nilai *Davies-Bouldin Index* (DBI) dari nilai Rasio yang diperoleh sebelumnya dengan menggunakan persamaan berikut.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \tag{10}$$

dimana

- DBI : Davies-Bouldin Index
- k : jumlah cluster
- R_{ij} : Rasio antara cluster i dan j

B. Dataset

Data yang digunakan dalam penelitian ini terdiri atas 34 record dimana masing-masing record merupakan data dari setiap provinsi di Indonesia. Adapun total atribut yang digunakan sebanyak 9 dengan mengacu pada penelitian sebelumnya [10] dan menambahkan parameter eksternal terkait akses sanitasi dan air minum yang layak mengingat bahwa kedua parameter tersebut juga dapat mempengaruhi tumbuh kembang anak [11]. Atribut yang digunakan dapat dilihat pada Tabel I.

TABEL I
ATRIBUT DATA

No	Kode	Atribut
1	X ₁	Persentase anak usia 12-23 Bulan yang mendapatkan imunisasi dasar lengkap
2	X ₂	Proporsi stunting (TB/U) pada balita
3	X ₃	Persentase rerata balita umur 6-59 bulan ditimbang perbulan
4	X ₄	Cakupan imunisasi dasar lengkap
5	X ₅	Cakupan bayi baru lahir yang mendapat inisiasi menyusui dini
6	X ₆	Cakupan pemberian kapsul vitamin A
7	X ₇	Cakupan pemberian tablet tambah darah
8	X ₈	Persentase rumah tangga dengan akses terhadap sanitasi yang layak
9	X ₉	Persentase rumah tangga dengan akses terhadap air minum yang layak

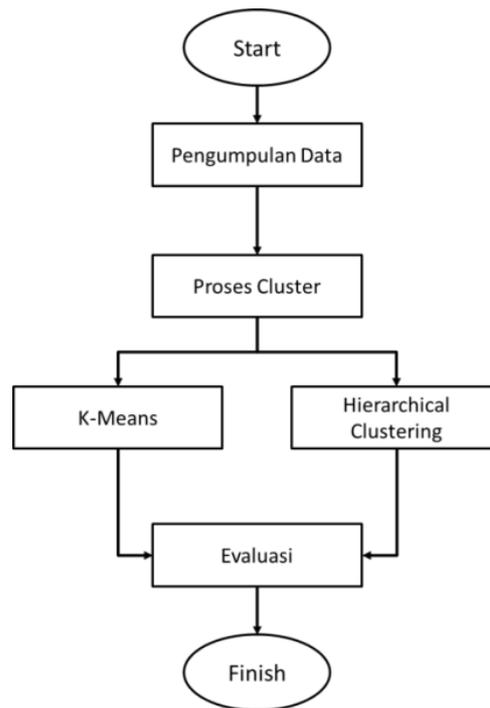
Berikut adalah sample data yang digunakan dalam penelitian ini.

TABEL II
SAMPEL DATA

Provinsi	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
Aceh	55.4	85.4	33.2	42.7	76	94.3	21.2	81	88.79
Sumatera utara	42.1	76.7	25.8	80.9	67.2	91.1	31.4	85.3	90.89
DKI Jakarta	68.6	40.7	16.8	63.3	98.5	69.4	7.6	93.5	99.86
Jawa barat	59.4	73.6	24.5	89.8	87.9	96.6	21.8	85.9	93.24
Kalimantan Barat	52.1	48.2	29.8	73.8	71.3	77.4	35	73.6	78.76
Kalimantan Tengah	44.7	48	27.4	84.9	88.6	79.3	18.9	73.9	77.05
Sulawesi barat	45.8	58.3	33.8	76.4	91.2	78.9	3.9	86	78.35
NTB	82.4	82.6	31.4	95.5	87.3	96.8	35.7	90.8	94.6
Maluku Utara	55.9	82.4	27.5	81	87.8	84.2	2.1	78.2	88.66
Papua	13	21.7	29.5	53.5	84.4	22	6.6	56.5	64.92

C. Metode Penelitian

Tahapan utama dalam penelitian ini terdiri atas tiga tahapan yang dimulai dengan pengumpulan data, pembentukan kluster menggunakan *K-Means* dan *Hierarchical Clustering*, dan tahapan terakhir adalah evaluasi. Alur penelitian ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

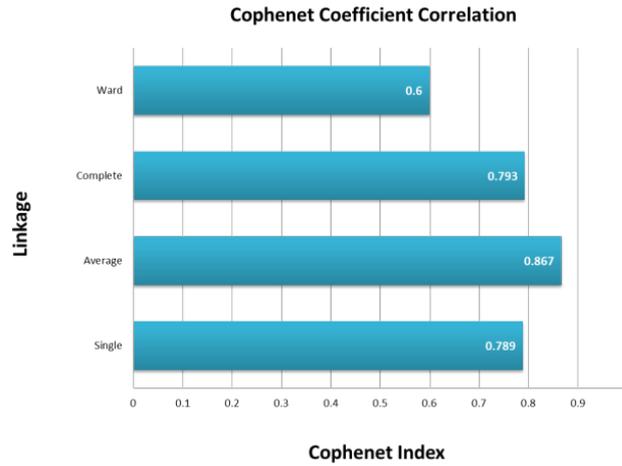
Tahapan pertama dalam penelitian ini dimulai dengan melakukan pengumpulan data. Data yang digunakan merupakan hasil survey status gizi Indonesia tahun 2021 dengan mempertimbangkan beberapa parameter untuk mengelompokkan daerah beresiko stunting, diantaranya persentase anak usia 12-23 bulan yang mendapatkan imunisasi dasar lengkap, proporsi stunting (TB/U) pada balita, persentase rerata balita umur 6-59 bulan ditimbang perbulan, cakupan imunisasi dasar lengkap, cakupan bayi baru lahir yang mendapat inisiasi menyusui dini, cakupan pemberian kapsul vitamin A, cakupan pemberian tablet tambah darah, persentase rumah tangga dengan akses terhadap sanitasi yang layak, serta persentase rumah tangga dengan akses terhadap air minum yang layak. Setelah itu dilakukan pembentukan kluster menggunakan dua metode yaitu *K-Means* dan *Hierarchical Clustering*. Hasil kluster yang terbentuk akan dievaluasi dengan menggunakan tiga metric yaitu *Silhouette Coefficient*, *Calinski-Harabasz index*, dan *Davies-Bouldin index*.

III. HASIL DAN PEMBAHASAN

Data pada penelitian ini diperoleh dari hasil survey status gizi Indonesia (SSGI) tahun 2021[12]. Data tersebut akan diolah dengan menggunakan dua pendekatan, yaitu *Hierarchical Clustering* dan *k means* untuk melihat hasil pengelompokan daerah beresiko stunting berdasarkan beberapa parameter. Berikut adalah hasil klusterisasi yang diperoleh

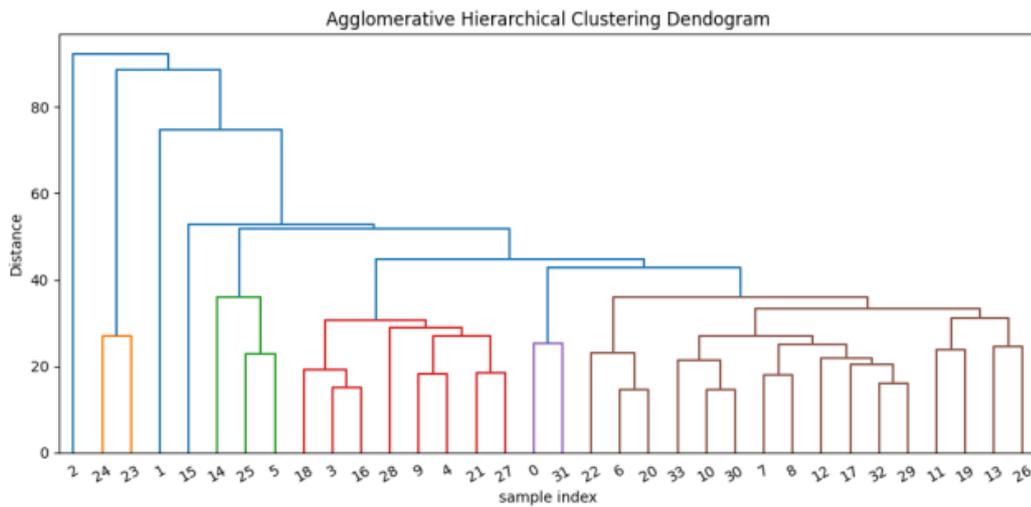
A. Hasil Hierarchical Clustering

Pada penelitian ini, pengelompokan daerah dengan hierarchical clustering menggunakan *agglomerative Clustering* dengan terlebih dahulu mencari jenis *linkage* terbaik yang membentuk dendrogram. Berikut adalah hasil perhitungan *cophenet coefficient correlation* dari masing-masing jenis *linkage*, yaitu *simple*, *average*, *complete*, dan *ward*.



Gambar 2. Hasil perbandingan metode linkage

Gambar 2 menunjukkan hasil perbandingan metode linkage yang digunakan dalam Hierarchical Clustering. Dalam hal ini, terdapat empat jenis metode yaitu, *ward method*, *Complete linkage*, *average linkage*, dan *single linkage*. Nilai *Cophenet coefficient* yang dihasilkan menunjukkan bahwa metode *average linkage* menghasilkan dendrogram terbaik dalam proses klasterisasi dengan nilai sebesar 0.867.



Gambar 3. Dendrogram Hierarchical Clustering

Gambar 3 merupakan visualisasi hasil dendrogram yang diperoleh dengan menggunakan *Hierarchical Clustering* dengan metode *average linkage*.

TABEL III
EVALUASI *HIERARCHICAL CLUSTERING*

Jumlah Kluster	<i>Silhouette Coefficient</i>	<i>Calinski-Harabasz</i>	<i>Davies-Bouldin</i>
2	0.46	4.88	0.39
3	0.47	9.54	0.46
4	0.26	8.95	0.44
5	0.2	7.52	0.49

Berdasarkan tabel 3 jumlah kluster optimum yang diperoleh adalah 3 kluster jika ditinjau dari nilai *Silhouette Coefficient* dan *Calinski-Harabasz index*, dimana untuk *Silhouette Coefficient* dan *Calinski-Harabasz index* semakin besar nilainya makanya semakin baik kluster yang terbentuk. Adapun pembagian klasternya adalah sebagai berikut.

a. Kluster 1

Terdapat 31 provinsi yang menjadi anggota pada Kluster 1, yaitu : 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, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, dan Maluku Utara.

b. Kluster 2

Anggota dari kluster 2 hanya terdiri dari satu Provinsi yaitu Provinsi Banten.

c. Kluster 3

Pada Kluster 3, terdapat 2 Provinsi yang menjadi anggota kluster tersebut yaitu Papua dan Papua Barat.

B. Hasil K-Means Clustering

TABEL IV
EVALUASI K-MEANS

Jumlah Kluster	<i>Silhouette Coefficient</i>	<i>Calinski-Harabasz</i>	<i>Davies-Bouldin</i>
2	0.48	10.49	0.54
3	0.27	12.66	1.07
4	0.29	13.95	0.82
5	0.24	14.52	1.02

Berdasarkan tabel 4 kluster optimum yang diperoleh menggunakan *K-Means* adalah 2 kluster jika ditinjau dari *Silhouette Coefficient* dan *Davies-Bouldin index*, dimana pada *Silhouette Coefficient* semakin besar nilainya maka kluster yang terbentuk semakin baik, sedangkan nilai *Davies-Bouldin index* semakin kecil nilainya, maka semakin baik kluster yang terbentuk. Berikut adalah pembagian dari masing-masing kluster.

a. Kluster 1

Daerah-daerah yang termasuk dalam kluster 1 terdiri dari 32 Provinsi yaitu : Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, Jawa Barat, Banten, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, NTB, NTT, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, dan Maluku Utara.

b. Kluster 2

Daerah-daerah yang termasuk dalam kluster 2 adalah Provinsi Papua dan Papua Barat.

C. Evaluasi

Evaluasi pembentukan kluster yang dilakukan dengan menghitung *Silhouette Coefficient*. Pada penelitian ini akan dibandingkan hasil klusterisasi menggunakan dua metode, yaitu *hierarchical clustering* dan *k means clustering*. Berikut adalah hasil perbandingan nilai

Silhouette Coefficient dari dua algoritma yang digunakan.

TABEL V
HASIL PERBANDINGAN ALGORITMA

No	Algoritma	<i>Silhouette Coefficient</i>	<i>Calinski-Harabasz</i>	<i>Davies-Bouldin</i>	Jumlah Klaster
1	<i>Hierarchical Clustering</i>	0.47	9.54	0.46	3
2	<i>K-Means Clustering</i>	0.48	10.49	0.54	2

Berdasarkan hasil perbandingan nilai *Silhouette Coefficient* dan *Calinski-Harabasz index* dari dua metode yang digunakan, diperoleh bahwa *K-Means* menghasilkan klaster yang lebih baik. Akan tetapi, jika dilihat dari nilai *Davies-Bouldin Index*, maka *Hierarchical Clustering* menghasilkan pembagian klaster yang lebih baik. Hasil pengelompokan daerah menggunakan *Hierarchical Clustering* dan *K-Means Clustering* hampir sama, dimana anggota klaster 3 pada *Hierarchical Clustering* dan klaster 2 pada *K-Means* anggotanya adalah Papua dan Papua Barat. Pada klaster tersebut memiliki karakteristik data yang lebih kecil dibandingkan dengan data pada provinsi lainnya jika ditinjau dari beberapa parameter, seperti persentase anak usia 12-23 bulan yang mendapatkan imunisasi dasar lengkap, proporsi stunting (TB/U) pada balita, cakupan imunisasi dasar lengkap, cakupan pemberian kapsul vitamin A, cakupan pemberian tablet tambah darah, persentase rumah tangga dengan akses terhadap sanitasi yang layak, serta persentase rumah tangga dengan akses terhadap air minum yang layak. Sedangkan anggota klaster 2 pada *Hierarchical Clustering* hanya beranggotakan 1 Provinsi yaitu Banten, dimana Banten pada *K-Means* merupakan klaster 1.

IV. KESIMPULAN

Berdasarkan hasil perbandingan metode *K-Means* dan *Hierarchical Clustering*, diperoleh bahwa *K-Means* menghasilkan pengelompokan klaster yang lebih baik ditinjau dari nilai *Silhouette Coefficient* sebesar 0.48 dan *Calinski-Harabasz index* sebesar 10.49 dengan jumlah klaster yang terbentuk sebanyak 2 klaster. Pada Algoritma *Hierarchical Clustering*, nilai *Silhouette Coefficient* yang dihasilkan adalah 0.47 dan *Calinski-Harabasz index* sebesar 9.54. Setelah dilakukan analisis terhadap hasil klaster yang terbentuk, daerah-daerah yang berada pada klaster kedua yaitu Provinsi Papua dan Papua Barat merupakan daerah dengan resiko stunting yang tinggi sehingga dapat dilakukan pendampingan serta upaya intervensi pencegahan dan penanganan stunting di daerah tersebut seperti pemberian imunisasi, perbaikan sanitasi serta akses air minum yang layak, dan sebagainya.

Penelitian ini masih dapat dikembangkan dengan menggunakan data dengan cakupan yang lebih spesifik seperti data per kabupaten untuk masing-masing provinsi di Indonesia serta menambahkan parameter-parameter lain yang berkaitan dengan stunting.

REFERENSI

- [1] A. D. N. Yadika, K. N. Berawi, and S. H. Nasution, "Pengaruh stunting terhadap perkembangan kognitif dan prestasi belajar," *J. Major.*, vol. 8, no. 2, pp. 273–282, 2019.
- [2] F. Wajidi and N. Nur, "Sistem Pakar Diagnosa Stunting Pada Balita Menggunakan Metode Forward Chaining," *J. Inform. Univ. Pamulang*, vol. 6, no. 2, pp. 401–407, 2021.
- [3] Kemenkes, "Prevalensi Stunting di Indonesia Turun ke 21,6% dari 24,4%," 2023. <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20230125/3142280/prevalensi-stunting-di-indonesia-turun-ke-216-dari-244/> (accessed Jul. 17, 2023).

- [4] A. Fadliana and F. Rozi, “Penerapan Metode Agglomerative Hierarchical Clustering untuk Klasifikasi Kabupaten/Kota di Propinsi Jawa Timur Berdasarkan Kualitas Pelayanan Keluarga Berencana,” *CAUCHY J. Mat. Murni dan Apl.*, vol. 4, no. 1, pp. 35–40, 2015, doi: 10.18860/ca.v4i1.3172.
- [5] G. R. Suraya and A. W. Wijayanto, “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,” *Indones. J. Stat. Its Appl.*, vol. 6, no. 2, pp. 180–201, Aug. 2022, doi: 10.29244/ijsa.v6i2p180-201.
- [6] M. Benri, H. Metisen, and S. Latipa, “Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila,” *J. Media Infotama*, vol. 11, no. 2, pp. 110–118, 2015, [Online]. Available: <https://core.ac.uk/download/pdf/287160954.pdf>
- [7] R. A. Faujia, E. S. Setianingsih, and H. Pratiwi, “Analisis Klaster K-Means Dan Agglomerative Nesting Pada Indikator Stunting Balita Di Indonesia,” *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 1249–1258, 2022, doi: 10.34123/semnasoffstat.v2022i1.1511.
- [8] N. S. Fatonah and T. K. Pancarani, “Analisa Perbandingan Algoritma Clustering untuk Pemetaan Status Gizi Balita di Puskesmas Pasir Jaya,” *KONVERGENSI*, vol. 18, no. 1, 2022.
- [9] A. M. Sikana and A. W. Wijayanto, “Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering,” *J. Ilmu Komput.*, vol. 14, no. 2, p. 66, 2021, doi: 10.24843/jik.2021.v14.i02.p01.
- [10] A. Fadilah, M. N. Pangestu, S. Lumbanbatu, and S. Defiyanti, “Pengelompokan Kabupaten/Kota Di Indonesia Berdasarkan Faktor Penyebab Stunting Pada Balita Menggunakan Algoritma K-Means,” *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, p. 223, 2022, doi: 10.26798/jiko.v6i2.581.
- [11] B. P. P. Nasional, “Air Minum dan Sanitasi.” <https://cegahstunting.id/intervensi/intervensi-sensitif/air-minum-dan-sanitasi/>
- [12] Kemenkes, *Buku Saku Hasil Studi Status Gizi Indonesia (SSGI) Tingkat Nasional, Provinsi, dan Kabupaten/ Kota Tahun 2021*. 2021.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada DIPA Universitas Sulawesi Barat yang telah mendanai penelitian ini.