

PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI PAPUA BERDASARKAN FAKTOR PENYEBAB STUNTING

Elva Grace Fernanda Gamay¹, Rita Raya², Nicea Roona Paranoan³

^{1,2,3} Universitas Cenderawasih

e-mail: nicearoonal2@gmail.com

ABSTRAK

Stunting merupakan masalah gizi akibat kekurangan asupan gizi yang berlangsung lama terutama pada 1000 hari pertama kehidupan. Berdasarkan Survei Status Gizi Indonesia, pada tahun 2022 angka prevalensi *stunting* pada balita di Provinsi Papua mencapai 34,6%. Angka prevalensi ini masih tergolong tinggi sehingga perlu dilakukan upaya agar dapat mengatasi masalah *stunting* di Provinsi Papua. Salah satu upaya yang dapat dilakukan adalah memahami karakteristik dari masing-masing Kabupaten/Kota di Provinsi Papua melalui analisis pengelompokan. Penelitian ini bertujuan untuk melakukan pengelompokan Kabupaten/Kota di Provinsi Papua berdasarkan faktor penyebab *stunting*. Terdapat dua metode dalam analisis pengelompokan yaitu hierarki dan non hierarki. Metode hierarki memiliki lima metode yaitu Pautan tunggal, Pautan lengkap, Pautan rata-rata, Metode Ward dan Metode *Centroid*. Pemilihan metode yang terbaik pada metode hierarki, digunakan uji korelasi *Cophenetic*. Metode pengelompokan yang digunakan pada penelitian ini adalah metode *Centroid* dan *K-Means*. Uji validasi yang digunakan untuk membandingkan metode *Centroid* dan *K-Means* berdasarkan faktor penyebab *stunting* di Provinsi Papua yaitu uji *Silhouette Coefficient Index*. Berdasarkan uji *Silhouette Coefficient Index*, diperoleh bahwa pada metode *Centroid* dan *K-Means* jumlah kelompok optimal yang terbentuk adalah dua kelompok. Kelompok pertama beranggotakan 21 Kabupaten/Kota dan Kelompok kedua beranggotakan 3 Kabupaten/Kota. Berdasarkan uji *Silhouette Coefficient Index* juga diperoleh bahwa metode yang lebih baik pada penelitian ini adalah metode *Centroid* karena nilai *Silhouette Coefficient Index* yang diperoleh lebih tinggi dari metode *K-Means*.

Kata Kunci : *Centroid, K-Means, Silhouette Coefficient Index, Stunting*.

1. PENDAHULUAN

Kejadian balita pendek biasa disebut dengan *stunting* merupakan salah satu masalah gizi yang dihadapi oleh anak-anak di seluruh dunia (Kementrian Kesehatan RI, 2018). Pada tahun 2020 terdapat sekitar 149,2 juta anak di bawah usia 5 tahun di seluruh dunia yang mengalami *stunting* (Salsabila, dkk, 2021). *Stunting* merupakan suatu kondisi tumbuh kembang anak terhambat yang disebabkan kekurangan gizi yang berkepanjangan. Kondisi ini mempengaruhi pertumbuhan tubuh dan otak, sehingga balita yang mengalami *stunting* memiliki tubuh lebih pendek dibandingkan teman sebayanya. Selain itu, *stunting* juga berhubungan dengan keterlambatan perkembangan kognitif pada balita yang mengakibatkan mereka mengalami kesulitan dalam berpikir dan belajar (Kementrian Kesehatan RI, 2018).

Pada tahun 2017, sekitar 150,8 juta balita di seluruh dunia mengalami *stunting*. Lebih dari separuh anak di bawah usia lima tahun tersebut, yaitu sekitar 55% berasal dari wilayah Asia, sementara sekitar 39% tinggal di benua Afrika. Dari total 83,6 juta balita *stunting* di Asia, sebagian besar berasal dari wilayah Asia Selatan sebesar 58,7%, wilayah Asia Tenggara sebesar 14,9%, wilayah Asia Timur sebesar 4,8%, wilayah Asia Barat sebesar 4,2%. Sedangkan proporsi yang paling sedikit berasal dari Asia Tengah sebesar 0,9%. Pada

tahun 2005 hingga 2017, Indonesia termasuk dalam negara ketiga dengan prevalensi *stunting* tertinggi di wilayah Asia Tenggara, yaitu sebesar 36,4% (Kementerian Kesehatan RI, 2018).

Berdasarkan hasil Survei Status Gizi Indonesia (SSGI), persentase anak balita yang memiliki tinggi di bawah tinggi rata-rata (prevalensi) pada tahun 2019 mencapai 27,7%. Sedangkan pada tahun 2021 dan 2022 prevalensi *stunting* mengalami penurunan dengan prevalensi masing-masing sebesar 24,4% dan 21,6%. Meskipun terjadi penurunan angka prevalensi *stunting* pada tahun 2021 dan 2022, angka tersebut masih relatif tinggi daripada standar yang ditetapkan oleh WHO, yaitu 20%. Penurunan angka *stunting* juga termasuk dalam rencana misi Presiden tahun 2020-2024 yang dijelaskan oleh Kementerian Kesehatan Indonesia. Presiden Joko Widodo memiliki target untuk menurunkan angka prevalensi *stunting* di Indonesia menjadi 14% pada tahun 2024. Menurut hasil survei tersebut, Provinsi Papua menduduki peringkat ketiga dengan tingkat prevalensi *stunting* sebesar 34,6% yang melebihi prevalensi *stunting* nasional sebesar 21,6% (Kementerian Kesehatan RI, 2023).

Beberapa faktor yang menyebabkan kasus *stunting* pada balita antara lain antara lain tidak lengkapnya imunisasi dasar, riwayat penyakit seperti diare dan ISPA pada balita, kurangnya akses sanitasi yang layak, kurangnya akses air bersih, tingginya prevalensi gizi buruk dan kekurangan gizi, serta kurangnya pelayanan kesehatan yang memadai (Bappenas, 2019). Faktor penyebab *stunting* dapat bervariasi antara daerah satu dengan daerah lain. Agar pemerintah dapat mengatasi masalah *stunting* dengan lebih efektif pada Kabupaten/Kota di Provinsi Papua, penting untuk memahami karakteristik dari masing-masing Kabupaten/Kota tersebut sehingga diharapkan dapat memberikan penanganan yang sesuai. Salah satu metode yang umum digunakan untuk pengelompokan berdasarkan kesamaan karakteristik adalah analisis kluster.

Analisis kluster membantu dalam mengelompokkan data ke dalam dua kelompok atau lebih berdasarkan kesamaan karakteristik yang dimiliki. Analisis kluster dibagi dua metode yaitu metode hierarki dan non hierarki. Ghaisani, dkk (2019) mengelompokkan provinsi di Indonesia berdasarkan indikator demokrasi Indonesia tahun 2016 dengan membandingkan 5 metode kluster hierarki yaitu *average linkage*, *centroid*, *complete linkage*, *single linkage*, dan *ward*. Hasil penelitian tersebut adalah dari kelima metode yang digunakan, metode kluster *centroid* menunjukkan hasil terbaik. Hidayat & Jajuli (2023) melakukan penelitian *clustering* daerah rawan *stunting* di Jawa Barat menggunakan metode non hierarki (algoritma *K-Means*). Penelitian ini bertujuan untuk mengetahui hasil pengelompokan Kabupaten/Kota di Provinsi Papua berdasarkan faktor penyebab *stunting* menggunakan metode *Centroid* dan metode *K-Means* serta melakukan perbandingan antara metode *Centroid* dengan *K-Means* untuk menentukan metode terbaik dalam melakukan pengelompokan Kabupaten/Kota di Provinsi Papua berdasarkan faktor penyebab *stunting*.

Uji Multikolinearitas

Multikolinearitas merupakan sebuah kondisi dimana antara satu pasang variabel independen atau lebih memiliki korelasi yang kuat (Best & Wolf dalam Kusumawardhani, dkk, 2021). Dalam analisis kluster, adanya korelasi antar variabel bisa menyulitkan dalam menentukan pengaruh dari setiap masing-masing variabel. Oleh karena itu, sebelum melakukan pengelompokan penting untuk melakukan uji multikolinearitas. Salah satu cara yang dapat digunakan untuk mendeteksi adanya multikolinearitas adalah dengan menggunakan nilai *Variance Inflation Factor* (VIF). Menurut Best & Wolf dalam Kusumawardhani, dkk (2021), nilai VIF digunakan untuk mengukur sejauh mana variabel independen ke-*j* bergantung pada kumpulan variabel independen yang lain. Persamaan VIF dapat dilihat pada persamaan berikut:

$$VIF_j = \frac{1}{1-R_j^2} \quad \dots(1)$$

dimana,

R_j^2 : Koefisien determinasi variabel ke-*j*

m : Banyak variabel yang diamati

j : 1, 2, 3, ..., *m*

Koefisien determinasi dapat diperoleh dengan persamaan sebagai berikut:

$$R_j^2 = \frac{b_1 \sum x_1 x_j + b_2 \sum x_2 x_j + \dots + b_m \sum x_m x_j}{\sum x_j^2} \quad \dots(2)$$

dengan

$$\sum x_j^2 = \sum X_j^2 - \frac{(\sum X_j)^2}{n}$$

$$\sum x_m x_j = \sum X_m X_j - \frac{(\sum X_m)(\sum X_j)}{n}$$

dimana,

$\sum X_m$: Jumlah data dari X pada variabel ke- m

$\sum X_j$: Jumlah data dari X pada variabel ke- j

$\sum X_j^2$: Jumlah data dari X^2 pada variabel ke- j

$\sum X_m X_j$: Jumlah data dari $X_m X_j$

n : Banyak data

b_m : Koefisien regresi masing-masing variabel

Multikolinearitas juga dapat diidentifikasi dengan menghitung nilai *tolerance* dengan menggunakan persamaan berikut:

$$Tolerance = \frac{1}{VIF} \quad \dots(3)$$

Suatu model dikatakan memiliki multikolinearitas jika nilai $VIF > 10$ atau $Tolerance \leq 0,1$.

Analisis Klaster

Analisis klaster merupakan metode statistik yang mengelompokkan sampel berdasarkan kesamaan karakteristik. Tujuan dari analisis klaster adalah untuk membentuk kelompok yang berbeda dengan tingkat kesamaan tinggi di antara anggota kelompok dan tingkat perbedaan tinggi di antara kelompok yang berbeda. Pengelompokan ini didasarkan pada variabel-variabel tertentu yang ingin diteliti.

Untuk menilai tingkat kemiripan antara dua objek, diperlukan definisi tentang ukuran kemiripan tersebut. Tujuannya adalah untuk membuat matriks *proximity* yang bersifat persegi dan simetri dengan jumlah objek yang sama pada baris dan kolom. Matriks tersebut mencerminkan tingkat kemiripan atau perbedaan antara objek-objek tersebut. Untuk mengukur kemiripan antar data digunakan pengukuran jarak (*distance*). Semakin besar jarak antara dua buah data maka semakin besar perbedaan antara dua data. Terdapat berbagai jenis pengukuran jarak, tetapi yang paling sering digunakan pada pengelompokan adalah jarak *Euclidean* dengan rumus berikut:

$$d_{(i,k)} = \sqrt{\sum_{j=1}^m (x_{ij} - x_{kj})^2} \quad \dots(4)$$

dimana,

$d_{(i,k)}$: Jarak antara data ke- i dengan data ke- k

m : Banyak variabel yang diamati

x_{ij} : Data ke- i pada variabel ke- j

x_{kj} : Data ke- k pada variabel ke- j

Secara umum metode analisis klaster terbagi menjadi dua metode yaitu metode hierarki dan metode non-hierarki.

Metode Hierarki

Metode hierarki digunakan ketika jumlah kelompok yang akan dibentuk belum diketahui. Pada metode hierarki *agglomerative*, proses dimulai dengan setiap data berada dalam klaster terpisah sebanyak n klaster. Kemudian, kelompok-kelompok tersebut digabungkan dengan menggabungkan dua klaster yang paling dekat satu sama lain, selanjutnya mengevaluasi kembali kedekatan antara $n - 1$ kelompok yang baru terbentuk. Proses penggabungan ini berlanjut dengan menggabungkan dua klaster terdekat dan terus berlangsung hingga terbentuk satu kelompok tunggal yang berisi seluruh data (Nugroho, 2008). Pada penelitian ini, metode hierarki yang digunakan adalah metode *centroid*. Pada metode ini, jarak antar dua kelompok adalah jarak di antara dua *centroid* kelompok-kelompok tersebut. *Centroid* adalah vektor rata-rata jarak yang ada pada sebuah kelompok, yang didapat dengan melakukan rata-rata pada semua anggota suatu kelompok tertentu. Dengan metode ini, setiap terjadi pengelompokan baru, segera terjadi perhitungan ulang *centroid*, sampai terbentuk kelompok yang tetap. Jika \underline{x}_A dan \underline{x}_B adalah vektor rata-rata (*centroid*)

kelompok A dan B , maka jarak antar dua kelompok didefinisikan sebagai $D(A, B) = d(\underline{x}_A, \underline{x}_B)$. *Centroid* kelompok baru yang terbentuk didapat dengan rumus :

$$\underline{x}_{AB} = \frac{n_A \underline{x}_A + n_B \underline{x}_B}{n_A + n_B} \quad \dots(5)$$

dimana,

\underline{x}_{AB} : *Centroid* kelompok yang baru terbentuk

n_A : Banyaknya anggota kelompok A

n_B : Banyaknya anggota kelompok B

Untuk menentukan metode yang terbaik pada metode hierarki, Pengujian korelasi *Cophenetic* dilakukan untuk mengevaluasi pohon kluster. Korelasi *Cophenetic* untuk pohon kluster didefinisikan sebagai koefisien korelasi linier antara jarak kofenetik yang diperoleh dari pohon dengan jarak awal (ketidakmiripan) yang digunakan untuk membentuk pohon tersebut. Ini mengukur sejauh mana pohon tersebut merepresentasikan perbedaan antara pengamatan. Koefisien *Cophenetic* adalah koefisien korelasi antara elemen-elemen asli dari matriks ketidakmiripan (*dissimilarity distance*) dan elemen-elemen yang dihasilkan oleh dendrogram (matriks kofenetik berdasarkan ukuran jarak dan metode keterhubungan yang digunakan). Proses ini dimulai dengan menghitung matriks jarak asli (jarak *Euclidean*) untuk membentuk matriks jarak antar data, kemudian menerapkan metode *single linkage*, *complete linkage*, *average linkage*, metode *ward*, dan metode *centroid*. Setelah itu, buat dendrogram untuk setiap metode tersebut dan hitung matriks jarak *cophenetic* dari masing-masing dendrogram. Kemudian, hitung korelasi *cophenetic* antara matriks jarak asli dan matriks jarak *cophenetic* untuk setiap metode menggunakan persamaan berikut:

$$r_{coph} = \frac{n \sum_{i < k}^n d_{ik} d_{c_{ik}} - (\sum_{i < k}^n d_{ik})(\sum_{i < k}^n d_{c_{ik}})}{\sqrt{[n \sum_{i < k}^n x_{ik}^2 - (\sum_{i < k}^n x_{ik})^2][n \sum_{i < k}^n x_{c_{ik}}^2 - (\sum_{i < k}^n x_{c_{ik}})^2]}} \quad \dots(6)$$

dimana,

r_{coph} : Koefisien korelasi *Cophenetic*

d_{ik} : Jarak asli (jarak *Euclidean*) data ke- i dengan data ke- k

$d_{c_{ik}}$: Jarak *Cophenetic* data ke- i dengan data ke- k

n : Banyak data

Nilai korelasi *Cophenetic* antara -1 sampai 1, semakin mendekati 1 berarti solusi yang dihasilkan semakin baik (Jain & Dubes, 1988). Metode dengan nilai korelasi mendekati 1 yang akan dipilih sebagai metode terbaik dari metode-metode yang terdapat pada kluster hierarki.

Metode Non Hierarki

Metode non-hierarki diawali dengan menentukan jumlah kelompok yang diinginkan dan *centroid* di tiap kluster. Metode ini dikenal sebagai pengklasteran *K-Means*. Pengklasteran *K-Means* diawali dengan menetapkan jumlah kelompok yang diinginkan dan posisi *centroid* di setiap kluster. Dalam beberapa perangkat lunak, *centroid* awal dapat menggunakan k pengamatan pertama, namun ada juga yang menetapkan *centroid* secara acak. Setelah itu, jarak antara setiap objek dengan setiap *centroid* dihitung. Objek kemudian ditempatkan ke dalam kelompok yang sesuai berdasarkan jarak terdekat dengan *centroid* kluster yang cocok. Selanjutnya, *centroid* dihitung ulang untuk setiap kluster. Proses ini diulang hingga tidak ada lagi perpindahan objek antar kelompok (Nugroho, 2008).

Pengelompokan menggunakan *K-Means* bertujuan untuk menggambarkan algoritma yang digunakan dalam menempatkan suatu objek ke dalam kelompok tertentu berdasarkan rata-rata terdekat. Metode *K-Means* kerap digunakan karena sederhana dan mudah diimplementasikan. Berikut adalah langkah-langkahnya:

1. Tentukan nilai k sebagai jumlah kelompok yang ingin dibentuk
2. Pilih secara acak k *centroid* (titik pusat) awal
3. Hitung jarak setiap data ke masing-masing titik pusat menggunakan jarak *Euclidean*
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan pusatnya
5. Tentukan posisi titik pusat baru dengan menghitung rata-rata data dalam setiap kelompok.

$$C_{lj} = \frac{x_{1j} + x_{2j} + \dots + x_{n_{lj}}}{n_l} \quad \dots(7)$$

dimana,

C_{lj} : Nilai titik pusat baru pada data ke- l dan variabel ke- j

n_l : Banyaknya data pada kelompok ke- l

6. Hitung kembali jarak setiap data ke titik pusat yang baru terbentuk, lakukan seterusnya sampai tidak ada lagi perpindahan objek antar kelompok. Jika tidak terjadi perpindahan objek antar kelompok, maka iterasi dihentikan.

Uji Validasi

Uji validasi digunakan untuk menganalisa hasil yang diperoleh dari algoritma pengelompokan. Pengujian ini bertujuan untuk memperoleh jumlah kluster terbaik yang dapat menjelaskan keseluruhan struktur data. Jumlah kluster terbaik dapat memberikan hasil pengelompokan yang optimal sehingga diperoleh keputusan yang tepat (Susilowati, dkk, 2020). Pada penelitian ini akan digunakan uji *Silhouette Coefficient Index*. *Silhouette Coefficient Index* (SC) memiliki keunggulan karena hanya bergantung pada partisi sebenarnya dari objek, bukan pada algoritma pengelompokan yang digunakan untuk memperolehnya. SC dapat digunakan untuk meningkatkan hasil analisis pengelompokan atau untuk membandingkan hasil dari berbagai algoritma pengelompokan yang diterapkan pada data yang sama (Rousseeuw, 1987). Untuk menghitung jarak antara data digunakan persamaan jarak *euclidean*. Nilai *Silhouette* berada pada rentang -1 hingga 1. Semakin mendekati 1 nilai SC maka semakin baik kualitas kluster tersebut (Kaufman & Rousseeuw, 1990). Langkah-langkah menghitung nilai SC yaitu:

1. Menghitung jarak rata-rata data pada kelompok yang sama ($a(i)$). Misalkan data ke- i merupakan anggota pada kelompok ke- l maka $a(i)$ didefinisikan sebagai jarak rata-rata data ke- i dengan semua data yang berada di kelompok ke- l .

$$a(i) = \frac{1}{n_l - 1} \sum_{k_l=1}^{n_l} d(i, k_l) \quad \dots(8)$$

dimana,

$a(i)$: Jarak rata-rata data pada kelompok yang sama

n_l : Banyak data pada kelompok ke- l

$d(i, k_l)$: Jarak antara data ke- i dengan data ke- k pada kelompok l

k_l : 1,2,3,..., n_l

2. Menghitung jarak rata-rata terkecil pada kelompok yang berbeda ($b(i)$). Misalkan data ke- i merupakan anggota pada kelompok ke- l maka $b(i)$ didefinisikan sebagai jarak rata-rata terkecil antara data ke- i dengan semua data pada kelompok yang berbeda (l').

$$b(i) = \min \frac{1}{n_{l'}} \sum_{k_{l'}=1}^{n_{l'}} d(i, k_{l'}) \quad \dots(9)$$

dimana,

$b(i)$: Jarak rata-rata terkecil pada kelompok yang berbeda

$n_{l'}$: Banyak data pada kelompok yang berbeda (l')

$d(i, k_{l'})$: Jarak antara data ke- i dengan data ke- k pada kelompok l'

$k_{l'}$: 1,2,3,..., $n_{l'}$

3. Menghitung nilai *Silhouette* di tiap Kabupaten/Kota ($s(i)$)

$$s(i) = \frac{b(i) - a(i)}{(a(i), b(i))} \quad \dots(10)$$

4. Menghitung nilai SC

$$SC = \frac{1}{n} \sum_{i=1}^n s(i) \quad \dots(11)$$

dimana,

n : Banyak sampel data

2. METODE PENELITIAN

Sumber Data

Data dalam penelitian ini menggunakan data sekunder. Data yang digunakan yaitu mengenai faktor penyebab *stunting* di Provinsi Papua sebelum pemekaran. Sumber data diperoleh dari hasil Survei Status Gizi Indonesia (SSGI) pada tahun 2022 oleh Badan Kebijakan Pembangunan Kesehatan (BKPK) Kementerian Kesehatan RI.

Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah faktor-faktor penyebab terjadinya *stunting* diantaranya Pemberian ASI Tidak Eksklusif (X_1), Berat Badan Lahir Rendah (X_2), Tidak Inisiasi Menyusu Dini (X_3), Tidak Mendapatkan Vitamin A (X_4), Tidak Memiliki Kepemilikan Jaminan Kesehatan (X_5), Ibu Hamil Minum < 90 Tablet Tambah Darah (X_6), Sumber Air Minum Tidak Layak (X_7)

Metode analisis data

Analisis data menggunakan metode hierarki dan non hierarki. Tahap-tahap menganalisis data adalah sebagai berikut:

1. Menginput data
2. Melakukan uji multikolinearitas pada variabel-variabel
3. Melakukan analisis kluster menggunakan metode *Centroid* dan *K-Means*
4. Melakukan uji indeks koefisien *Silhouette* pada kedua metode
5. Interpretasi hasil pengelompokan
6. Menarik kesimpulan dan saran.

3. HASIL DAN PEMBAHASAN

Uji Multikolinearitas

Berikut hasil perhitungan *R-square* (R^2_j), *Tolerance* dan VIF untuk setiap variabel penelitian dan disajikan pada Tabel 1.

Tabel 1. Nilai VIF dan Tolerance di Setiap Variabel

Variabel	R^2_j	VIF	<i>Tolerance</i>
X_1	0,40	1,67	0,60
X_2	0,42	1,73	0,58
X_3	0,46	1,86	0,54
X_4	0,75	3,94	0,25
X_5	0,46	1,84	0,54
X_6	0,60	2,50	0,40
X_7	0,34	1,52	0,66

Berdasarkan Tabel 1, diketahui bahwa nilai VIF untuk semua variabel kurang dari 10 dan nilai *tolerance* untuk semua variabel lebih dari 0,1, sehingga dapat disimpulkan bahwa tidak ada multikolinearitas di antara variabel-variabel faktor penyebab terjadinya *stunting* yang digunakan.

Penentuan Jarak Kedekatan *Euclidean*

Ukuran jarak mengindikasikan kesamaan atau ketidaksamaan antara data. Semakin besar jaraknya, semakin tidak sama datanya. Sebaliknya, semakin kecil jaraknya, semakin mirip datanya. Berikut ini adalah hasil perhitungan jarak antara setiap data.

Tabel 2. Jarak *Euclidean*

No	Kab/Kota	Merauke	Jayawijaya	Jayapura	Nabire	...	Kota Jayapura
1	Merauke	0	60,78	34,85	52,61	...	59,11
2	Jayawijaya	60,78	0,00	44,34	61,87	...	46,50
3	Jayapura	34,85	44,34	0,00	36,75	...	51,71
4	Nabire	52,61	61,87	36,75	0,00	...	45,62
5	Kepulauan Yapen	36,82	35,51	17,49	33,63	...	41,77
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
24	Kota Jayapura	59,11	46,50	51,71	45,62	...	0,00

Uji Korelasi *Cophenetic*

Uji korelasi *cophenetic* dilakukan untuk menentukan metode terbaik pada jenis metode hierarki dengan cara dilihat dari nilai koefisien korelasi *cophenetic* (r_{coph}). Hasil perhitungan koefisien korelasi *cophenetic* disajikan pada tabel berikut:

Tabel 3. Nilai Koefisien Korelasi Cophenetic

Metode	Koefisien Korelasi <i>Cophenetic</i>
Pautan Tunggal (<i>Single linkage</i>)	0,64
Pautan Lengkap (<i>Complete linkage</i>)	0,72
Pautan Rata-rata (<i>Average linkage</i>)	0,70
Metode Ward (<i>Ward's method</i>)	0,70
Metode Centroid (<i>Centroid Method</i>)	0,76

Jika dilihat pada Tabel 3, nilai koefisien korelasi *cophenetic* menggunakan metode pautan tunggal sebesar 0,64, pautan lengkap sebesar 0,72, pautan rata-rata sebesar 0,70, metode ward sebesar 0,70 dan nilai koefisien korelasi *cophenetic* menggunakan metode *centroid* sebesar 0,76. Hal ini menunjukkan bahwa penggunaan metode *centroid* memberikan solusi lebih baik dibandingkan dengan metode lainnya, karena koefisien korelasi *cophenetic* menggunakan metode *centroid* lebih besar atau lebih mendekati 1 dibandingkan menggunakan metode lainnya. Dengan demikian metode *centroid* dipilih untuk tahap analisis kluster berikutnya.

Pengelompokan Menggunakan Metode *Centroid*

Pengelompokan menggunakan metode *centroid* dimulai dengan menghitung jarak antara data menggunakan persamaan (4). Hasil perhitungan jarak antara data disajikan pada Tabel 2. Pasangan data dengan jarak terdekat akan digabungkan menjadi satu kelompok, kemudian menghitung ulang *centroid* antar kelompok dengan data lainnya. Lakukan seterusnya, sampai semua data bergabung menjadi satu kelompok.

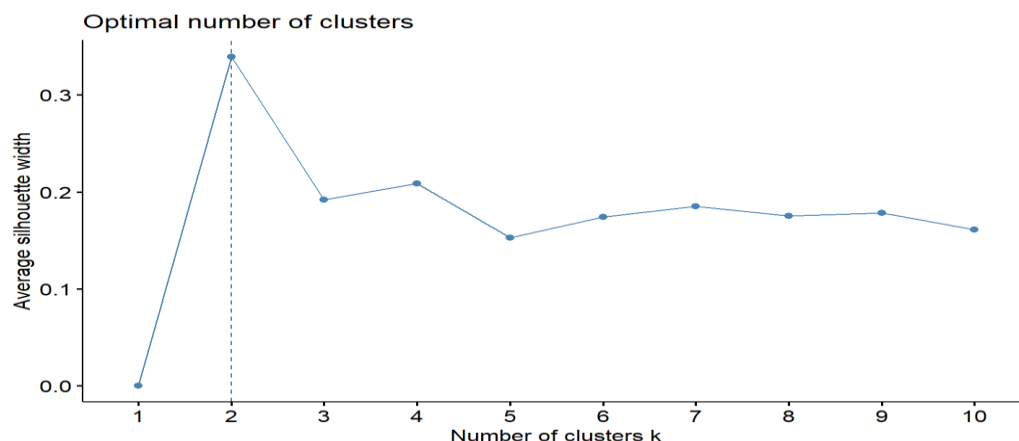
Dengan *Silhouette Coefficient Index* (SC) dapat ditentukan jumlah kelompok yang optimal pada metode *Centroid* yang terbentuk. Hasil SC diperoleh sebanyak 2 kelompok yang terbentuk dan hasil pengelompokan menggunakan metode *Centroid* disajikan pada tabel berikut:

Tabel 4. Hasil Pengelompokan Menggunakan Metode Centroid

Kelompok	Banyak Anggota	Kabupaten/Kota
1	21	Merauke, Jayawijaya, Jayapura, Nabire, Kepulauan Yapen, Biak Numfor, Paniai, Mimika, Boven Digoel, Mappi, Asmat, Pegunungan Bintang, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya, Lanny Jaya, Mamberamo Tengah, Deiyai dan Kota Jayapura.
2	3	Puncak Jaya, Yahukimo dan Tolikara.

Pengelompokan Menggunakan Metode *K-Means*

Pengelompokan *K-Means* diawali dengan menetapkan jumlah kelompok yang diinginkan serta menentukan posisi titik pusatnya. Pada penelitian ini dilakukan pengelompokan sebanyak 2 kelompok. Banyaknya kelompok dalam penelitian ini ditentukan berdasarkan grafik *Silhouette Coefficient Index* (SC) sebagai berikut:



Gambar 1. Optimal Number of Clusters Metode K-Means

Dengan menggunakan metode *K-Means*, diperoleh hasil pengelompokan yang disajikan pada tabel berikut:
Tabel 5. Hasil Pengelompokan Menggunakan Metode K-Means

Kelompok	Banyak Anggota	Kabupaten/Kota
1	21	Merauke, Jayawijaya, Jayapura, Nabire, Kepulauan Yapen, Biak Numfor, Paniai, Mimika, Boven Digoel, Mappi, Asmat, Pegunungan Bintang, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya, Lanny Jaya, Mamberamo Tengah, Deiyai dan Kota Jayapura.
2	3	Puncak Jaya, Yahukimo dan Tolikara.

Uji Validasi

Untuk menentukan jumlah kelompok optimal dan menentukan metode terbaik antara metode *Centroid* dan *K-Means* digunakan uji validasi *Silhouette Coefficient Index* (SC). Jika nilai SC semakin mendekati 1 maka hasil pengelompokan semakin baik. Berdasarkan hasil perhitungan nilai SC metode *Centroid* dan *K-Means*, maka diperoleh nilai SC dengan menggunakan metode *Centroid* sebesar 0,42 dan pada metode *K-Means* sebesar 0,34. Sehingga, dapat dinyatakan bahwa menurut uji validasi *Silhouette Coefficient Index* metode yang lebih baik yaitu metode *Centroid* karena memiliki nilai SC yang lebih tinggi dari metode *K-Means*.

Interpretasi Hasil Pengelompokan

Berdasarkan hasil pengelompokan menggunakan metode *Centroid* dan *K-Means* diperoleh nilai rata-rata setiap variabel pada seluruh data dan rata-rata variabel di setiap kelompok yang sama dan disajikan pada tabel berikut:

Tabel 6. Rata-rata Variabel pada Setiap Kelompok Metode Centroid dan K-Means

Hasil	Variabel	Rata-Rata Variabel	Rata-Rata Variabel di Setiap Kelompok	
			Kelompok 1	Kelompok 2
	X ₁	34,43	36,79	17,97
	X ₂	3,70	3,88	2,47
	X ₃	84,95	83,43	95,53
	X ₄	42,60	36,91	82,43
	X ₅	82,25	81,14	90,00
	X ₆	86,42	84,48	100,00
	X ₇	31,38	24,68	78,33

pengelompokan antara metode *Centroid* dan metode *K-Means* menunjukkan kesamaan disebabkan oleh

anggota kelompok 1 pada metode *Centroid* dan anggota Kelompok 1 pada metode *K-Means* sama, demikian pula dengan anggota Kelompok 2 untuk kedua metode. Berdasarkan tabel 6, dapat dilihat karakteristik di masing-masing kelompok. Berikut ini adalah penjelasan untuk setiap kelompok.

1. Kelompok 1

Kelompok 1 memiliki nilai rata-rata yang tinggi untuk variabel bayi usia 0-5 bulan tidak mendapatkan ASI eksklusif (X_1) dan variabel balita dengan berat badan lahir < 2500 gram (X_2). Hal ini berarti bahwa kedua variabel tersebut merupakan faktor utama terjadinya *stunting* pada Kelompok 1 karena rata-rata variabel pada Kelompok 1 lebih besar dari rata-rata variabel pada data keseluruhan.

2. Kelompok 2

Kelompok 2 memiliki nilai rata-rata yang tinggi untuk variabel tidak mendapatkan inisiasi menyusui dini (IMD) pada bayi baru lahir (X_3), anak tidak mendapatkan vitamin A (X_4), anggota rumah tangga yang tidak memiliki jaminan pelayanan kesehatan (X_5), ibu hamil minum < 90 tablet tambah darah (X_6) dan rumah tangga yang memiliki akses sarana air minum tidak layak (X_7). Hal ini berarti bahwa variabel-variabel tersebut merupakan faktor utama terjadinya *stunting* pada Kelompok 2 karena rata-rata variabel pada kelompok 2 lebih besar dari rata-rata variabel pada data keseluruhan.

4. SIMPULAN DAN SARAN

Hasil pengelompokan menggunakan metode *K-Means* diperoleh jumlah kelompok optimal adalah dua kelompok berdasarkan uji validasi *Silhouette Coefficient Index*. Kelompok pertama terdiri dari 21 Kabupaten/Kota yaitu Merauke, Jayawijaya, Jayapura, Nabire, Kepulauan Yapen, Biak Numfor, Paniai, Mimika, Boven Digoel, Mappi, Asmat, Pegunungan Bintang, Sarmi, Keerom, Waropen, Supiori, Mamberamo Raya, Lanny Jaya, Mamberamo Tengah, Deiyai dan Kota Jayapura dimana kelompok ini lebih dipengaruhi oleh faktor bayi usia antara 0 sampai 5 bulan yang tidak mendapatkan ASI eksklusif (X_1) dan balita dengan berat badan lahir kurang dari 2500 gram (X_2). Sedangkan kelompok kedua terdiri dari 3 Kabupaten/Kota yaitu Puncak Jaya, Yahukimo dan Tolikara, dimana kelompok ini lebih dipengaruhi oleh faktor tidak mendapatkan inisiasi menyusui dini pada bayi baru lahir (X_3), anak tidak mendapatkan vitamin A (X_4), anggota rumah tangga yang tidak memiliki jaminan pelayanan kesehatan (X_5), ibu hamil minum kurang dari 90 tablet tambah darah (X_6), dan rumah tangga yang memiliki akses sarana air minum tidak layak (X_7).

Hasil pengelompokan menggunakan metode *Centroid* dan metode *K-Means* menunjukkan kesamaan. Berdasarkan nilai *Silhouette Coefficient Index* (SC), metode *Centroid* terbukti lebih baik daripada metode *K-Means*. Hal tersebut terlihat dari nilai SC pada metode *Centroid* yang lebih tinggi yakni sebesar 0,42, dibandingkan dengan nilai SC pada metode *K-Means* yang hanya sebesar 0,34.

Berdasarkan kesimpulan di atas, maka terdapat saran yaitu diharapkan pada penelitian selanjutnya digunakan dua metode yaitu *Centroid* dan *K-Means*. Untuk peneliti selanjutnya, penulis memberikan saran agar dapat menggunakan metode lain dalam analisis pengelompokan non hierarki seperti *Fuzzy C-Means* atau *K-Medoids*. Selain itu, peneliti selanjutnya dapat menambahkan beberapa variabel lain yang dapat mempengaruhi terjadinya *stunting* seperti anak yang tidak mendapatkan imunisasi dasar lengkap, Pemberian Makanan Tambahan (PMT) pada anak yang kurus dan PMT pada ibu hamil kekurangan energi kronik.

DAFTAR PUSTAKA

- Bappenas. 2019. *Pembangunan Gizi Indonesia*. Jakarta. Direktorat Kesehatan dan Gizi Masyarakat.
- Ghaisani, S. Y. Hikmah, N. Prasetyo, A. H. dan Widodo. E. 2019. *Analisis cluster hirarki untuk pengelompokan provinsi di Indonesia berdasarkan indikator demokrasi Indonesia tahun 2016*. Prosiding Konferensi Nasional Penelitian Matematika dan Pembelajarannya (KNPMP) IV Universitas Muhammadiyah Surakarta, 27 Maret 2019.
- Hidayat, T. dan Jajuli. M. 2023. *Clustering daerah rawan stunting di Jawa Barat menggunakan algoritma K-Means*. INFOTECH: Jurnal Informatika & Teknologi. Vol. 4. <http://jurnal.stmcileungsi.ac.id/index.php/infotech>. DOI:10.37373/infotech.v4i2.642.
- Jain, A. K. and Dubes. R. C. 1988. *Algorithms for clustering data*. America. Prentice-Hall, Inc.

- Kaufman, L. and Rousseeuw, P. J. 1990. *Finding groups in data: an introduction to cluster analysis*. America. John Wiley & Sons.
- Kementrian Kesehatan RI. 2018. *Buletin Jendela Data dan Informasi: Situasi Balita Pendek (Stunting) Di Indonesia*. Jakarta. <https://fliphtml5.com/qwklg/gvdc/basic>.
- Kementrian Kesehatan RI. 2023. *Buletin Jendela Data dan Informasi: Prevalensi Stunting di Indonesia Turun ke 21,6% dari 24,4%*. Jakarta. <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20230125/3142280/prevalensi-stunting-di-indonesia-turun-ke-216-dari-244/>.
- Kusumawardhani, R. Rizqiena, Z. D. dan Astuti, S. P. 2021. *Ekonometrika Suatu Pengantar*. Yogyakarta. Gerbang Media Aksara.
- Nugroho, S. 2008. *Statistika Multivariat Terapan*. Bengkulu. UNIB Press Bengkulu.
- Rousseeuw, P. J. 1987. *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*. Journal of computational and applied mathematics.
- Salsabila, A. Fitriyan, D. A. Rahmiati, H. Sekar, M. dkk. 2021. *Upaya Penurunan Stunting Melalui Peningkatan Pola Asuh Ibu*. *Jurnal Pengabdian Kesehatan Masyarakat (Pengmaskesmas)*. Vol.1. DOI: doi.org/10.31849/pengmaskesmas.v1i2/5739
- Susilowati, T. Sugiarto, D. dan Mardianto, I. 2020. *Uji Validasi Algoritma Self-Organizing Map (SOM) dan K-Means untuk Pengelompokan Pegawai*. *Rekayasa Sistem dan Teknologi Informasi*. Vol. 4 <http://jurnal.iaii.or.id>.