

PERFORMA ALGORITMA K-MEANS DALAM PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI PAPUA BERDASARKAN FAKTOR PENYEBAB STUNTING

NICEA ROONA PARANOAN^{1*}, CAECILIA BINTANG GIRIK ALLO¹, BOBI FRANS KUDDI¹, ARYANTO¹,
WINDA ADE FITRIYA B², DAN RITA RAYA¹

¹Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Cenderawasih, Kota Jayapura, Papua

²Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Cenderawasih, Kota Jayapura, Papua

*Penulis Korespondensi: nicearoonal2@gmail.com

ABSTRAK

Stunting merupakan permasalahan gizi kronis akibat kekurangan nutrisi jangka panjang, terutama selama 1.000 hari pertama kehidupan anak. Berdasarkan Survei Status Gizi Indonesia (SSGI) tahun 2022, prevalensi stunting di Provinsi Papua mencapai 34,6%, menunjukkan tingkat yang masih tinggi sehingga diperlukan strategi efektif untuk penanganannya. Salah satu pendekatan yang dapat dilakukan adalah mengidentifikasi karakteristik masing-masing Kabupaten/Kota melalui analisis kluster. Penelitian ini bertujuan mengelompokkan Kabupaten/Kota di Papua berdasarkan faktor-faktor yang memengaruhi kejadian stunting. Data yang digunakan berasal dari SSGI 2022 yang dilaksanakan oleh Badan Kebijakan Pembangunan Kesehatan (BKPK) Kementerian Kesehatan RI, mencakup kondisi sebelum pemekaran wilayah. Kebaruan penelitian ini terletak pada penerapan pendekatan berbasis data (*data driven*) dengan algoritma *machine learning* K-Means untuk pemetaan spasial determinan stunting secara regional, menggantikan pendekatan deskriptif dan statistik konvensional. Pendekatan ini menghasilkan gambaran empiris mengenai kesamaan karakteristik antarwilayah dan mendukung penyusunan kebijakan intervensi gizi yang berbasis bukti (*evidence based policy*). Hasil analisis menunjukkan algoritma K-Means menghasilkan dua kelompok optimal dengan nilai *Silhouette Coefficient Index* (SC) sebesar 0,34. Kelompok pertama terdiri dari 21 Kabupaten/Kota yang dipengaruhi oleh faktor bayi usia 0–5 bulan yang tidak mendapatkan ASI eksklusif dan balita dengan berat lahir kurang dari 2.500 gram. Kelompok kedua mencakup 3 Kabupaten/Kota yang lebih dipengaruhi oleh faktor tidak mendapatkan inisiasi menyusui dini, anak tidak menerima vitamin A, kurangnya asuransi kesehatan, ibu hamil dengan konsumsi tablet zat besi rendah, serta akses air minum yang tidak layak.

Kata Kunci: Algoritma K-Means, Analisis kluster, Papua, Stunting

ABSTRACT

Stunting is a chronic nutritional problem caused by prolonged nutrient deficiency, particularly during the first 1,000 days of a child's life. According to the 2022 Indonesia Nutritional Status Survey (SSGI), the prevalence of stunting among children under five in Papua Province reached 34.6%,

2020 Mathematics Subject Classification: 62H30, 62P10, 62R07

Diterima: 21-07-2025, direvisi: 29-10-2025, dimuat: 26-09-2026

indicating a persistently high rate that requires strategic intervention. One potential approach is to identify the characteristics of each district/city through a clustering analysis. This study aims to classify districts/cities in Papua based on the factors influencing stunting incidence. The data used were obtained from the 2022 SSGI conducted by the Health Development Policy Agency of the Indonesian Ministry of Health, covering the province prior to administrative division. The novelty of this study lies in the application of a data-driven machine learning approach using the K-Means algorithm for spatial mapping of stunting determinants at the regional level, as opposed to conventional descriptive or statistical analyses. This approach provides an empirical overview of interregional similarities and serves as a scientific basis for designing more targeted and evidence-based nutritional intervention policies. The results show that the K-Means algorithm formed two optimal clusters with a Silhouette Coefficient Index (SC) value of 0.34. The first cluster consists of 21 districts/cities primarily influenced by factors such as infants aged 0–5 months not receiving exclusive breastfeeding and children born with a weight below 2,500 grams. The second cluster includes 3 districts/cities influenced by factors such as delayed early breastfeeding initiation, lack of vitamin A intake, absence of health insurance, insufficient iron supplementation during pregnancy, and inadequate access to safe drinking water.

Keywords: *K-Means Algorithm, Cluster Analysis, Papua, Stunting*

1. Pendahuluan

Masalah stunting pada balita merupakan isu gizi global yang tetap menjadi perhatian utama kesehatan masyarakat karena konsekuensi jangka panjangnya terhadap perkembangan fisik, kognitif, dan produktivitas masa dewasa [1], [2]. Estimasi internasional menunjukkan bahwa pada tahun 2020 sekitar 149,2 juta anak di bawah usia lima tahun mengalami stunting, angka yang menggambarkan beban malnutrisi kronis yang masih tinggi di banyak wilayah dunia. Tren regional menunjukkan konsentrasi kasus stunting tinggi di Asia dan Afrika, sehingga intervensi yang berskala dan kontekstual sangat diperlukan [1], [3], [4], [5]. Indonesia termasuk negara yang menghadapi tantangan stunting besar; data nasional tahun-tahun terakhir (SSGI 2021–2022) menunjukkan penurunan prevalensi stunting nasional namun tetap berada di atas ambang batas yang direkomendasikan WHO (20%), dan pemerintah menetapkan target penurunan lebih agresif dalam kebijakan nasional. Survei Status Gizi Indonesia (SSGI) 2022 melaporkan hasil lintas-nasional yang menjadi basis perencanaan intervensi gizi [6], [7]. Walaupun tren penurunan tercatat pada tingkat nasional, distribusi spasial stunting tidak merata, pada beberapa provinsi dan kabupaten/kota mencatat prevalensi jauh di atas rata-rata nasional sehingga mencerminkan ketimpangan akses layanan kesehatan, gizi, dan kondisi sosial ekonomi [6], [8].

Provinsi Papua menonjol sebagai salah satu wilayah dengan prevalensi stunting tinggi. Data SSGI dan studi lokal menemukan prevalensi stunting di Papua yang berada di atas rata-rata nasional dan menunjukkan variasi antar kabupaten/kota di dalam provinsi tersebut [6], [9], [10]. Analisis situasi wilayah menegaskan bahwa faktor struktural seperti keterbatasan akses layanan kesehatan dasar, ketidakcukupan sanitasi dan air minum layak, serta kerentanan ekonomi rumah tangga menjadi kontributor utama tingginya angka stunting di beberapa wilayah Papua [10], [11], [12]. Bukti empiris menunjukkan bahwa determinan stunting bersifat multi faktorial. Faktor yang paling konsisten berkaitan dengan stunting meliputi: status gizi ibu dan berat lahir rendah (*low birth weight*), tingkat pendidikan ibu, praktik pemberian makan (termasuk ASI eksklusif dan pemberian makanan pendamping), riwayat infeksi (misal diare dan ISPA), cakupan imunisasi, akses sanitasi/air bersih, serta kondisi sosioekonomi dan wilayah tempat tinggal (urban/rural, kawasan timur/ barat) [13]. Penelitian lintas daerah di Indonesia pada data SSGI/DHS dan studi populasi lokal juga menegaskan peran determinan tersebut dan menyoroti variasi antar provinsi serta antar kecamatan [14], [15], [16], [17].

Perbedaan karakteristik antar wilayah (provinsi, kabupaten/kota, kecamatan) menuntut pengelompokan wilayah berdasarkan determinan stunting untuk merancang intervensi yang terfokus dan efisien. Analisis kluster (*cluster analysis*) memungkinkan identifikasi kelompok wilayah yang memiliki profil risiko serupa sehingga intervensi dapat diprioritaskan sesuai kebutuhan konteks lokal [18], [19]. Dalam ranah ini, metode non-hierarki seperti K-Means banyak digunakan karena skalabilitasnya dan kemampuannya mengelompokkan *dataset* multivariat besar menjadi beberapa kluster [20], [21], [22], [23], [24], [25]. Aplikasi *K-Means* pada data kesehatan masyarakat dan masalah gizi di Indonesia telah banyak dilaporkan berdasarkan studi-studi regional dengan memanfaatkan *K-Means* untuk memetakan daerah rawan stunting, mengelompokkan kasus berdasarkan variabel ekonomi-sosial, serta menghasilkan peta kerawanan yang membantu penentuan prioritas intervensi. Evaluasi kualitas kluster sering menggunakan metrik seperti *Silhouette score*, Davies–Bouldin index, dan Calinski-Harabasz index untuk menentukan jumlah kluster optimal dan stabilitas hasil [21], [22], [23]. Kombinasi pendekatan data-driven (*machine learning*) dan analisis spasial semakin menjadi standar untuk menyusun kebijakan berbasis bukti (*evidence-based policy*) dalam penanganan stunting [23], [24], [25].

Meski begitu, masih ada kebutuhan penelitian yang menerapkan teknik kluster berbasis data SSGI yang komprehensif pada konteks Provinsi Papua (sebelum pemekaran) dengan mempertimbangkan heterogenitas geografis dan social kultural; membandingkan kinerja algoritma *K-Means* dengan parameter inisialisasi dan k-pemilihan yang berbeda serta menilai validitas kluster menggunakan beberapa metrik evaluasi; dan menghubungkan hasil kluster dengan rekomendasi intervensi yang spesifik wilayah. Oleh karena itu, penelitian ini bertujuan menganalisis performa algoritma *K-Means* dalam pengelompokan Kabupaten/Kota di Provinsi Papua berdasarkan faktor penyebab stunting, serta menginterpretasikan implikasi kebijakan dari hasil pengelompokan tersebut. Hasil penelitian diharapkan memberikan kontribusi empiris bagi pengembangan model analisis spasial stunting dan membantu pembuat kebijakan dalam merancang intervensi yang lebih tepat sasaran di wilayah dengan keragaman sosial-ekonomi dan geografi yang tinggi seperti Papua.

2. Tinjauan Pustaka

2.1. Uji Multikolinearitas

Multikolinearitas merupakan sebuah kondisi dimana antara satu pasang variabel independen atau lebih memiliki korelasi yang kuat [26]. Dalam analisis pengelompokan, 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). Nilai VIF digunakan untuk mengukur sejauh mana variabel independen ke- j bergantung pada kumpulan variabel independen yang lain. Apabila nilai VIF kurang dari 10 maka tidak terjadi kasus multikolinearitas. Persamaan VIF dapat dilihat pada persamaan berikut:

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

dimana,

R_j^2 : Koefisien determinasi variabel ke- j

m : Banyak variabel yang diamati

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

Koefisien determinasi dapat diperoleh melalui persamaan berikut:

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

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

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

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

2.2. Analisis Pengelompokan

Analisis pengelompokan adalah suatu metode statistik yang mengelompokkan sampel berdasarkan kesamaan karakteristik. Tujuan dari analisis pengelompokan 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.

Analisis pengelompokan harus memenuhi asumsi berikut:

1. Sampel yang diambil harus benar-benar bisa mewakili populasi.
2. Multikolinieritas yaitu korelasi antar objek. Sebaiknya tidak terdapat korelasi antar objek, bila ada maka besar multikolinieritas tidaklah tinggi yaitu ($< 0,5$).

Sebelum melakukan analisis pengelompokan, perlu memperhatikan data yang akan digunakan. Jika terdapat perbedaan nilai yang besar antara variabel, misalnya ada yang dalam satuan juta dan ada yang satuan puluhan atau bahkan lebih kecil, perbedaan tersebut akan menyebabkan perhitungan jarak menjadi tidak valid. Oleh karena itu, data perlu ditransformasi menggunakan uji *z-score*. Uji ini dilakukan untuk semua variabel penelitian yang memiliki perbedaan nilai yang besar sebelum dilakukan analisis pengelompokan. Tujuan dari analisis pengelompokan adalah untuk mengklasifikasikan objek ke dalam kelompok yang relatif homogen, objek-objek di dalam satu kelompok memiliki kesamaan yang lebih tinggi dibandingkan dengan objek-objek pada kelompok lain.

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 ini 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. Pengukuran jarak ada bermacam-macam namun yang paling sering digunakan pada pengelompokan adalah jarak *Euclidean* dengan rumus berikut:

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

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

2.3. Metode Non Hierarki

Metode non-hierarki dimulai dengan menentukan terlebih dahulu jumlah kelompok yang diinginkan dan *centroid*. Metode ini biasa disebut pengelompokan *K-Means*. Pengelompokan *K-Means* dimulai dengan menetapkan terlebih dahulu jumlah kelompok yang diinginkan dan posisi titik pusat. Pada beberapa perangkat lunak, titik pusat awal dapat menggunakan k pengamatan pertama, tetapi ada juga yang menentukan titik pusat secara acak. Setelah itu, jarak antara setiap objek dengan setiap titik pusat dihitung. Objek kemudian ditempatkan ke dalam kelompok yang sesuai berdasarkan jarak terdekat dengan titik pusat. Selanjutnya hitung ulang titik pusat, proses ini diulang hingga tidak ada lagi pemindahan objek antar kelompok [27].

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-langkah algoritma *K-Means*:

1. Diberikan nilai k sebagai jumlah kelompok yang ingin dibentuk.
2. Tentukan k *centroid* (titik pusat) awal secara *random*.
3. Hitung jarak setiap data ke masing-masing titik pusat yaitu menggunakan jarak *Euclidean*.
4. Kelompokkan setiap data berdasarkan jarak terdekat antara data dengan pusatnya.
5. Tentukan posisi titik pusat baru dengan cara menghitung nilai rata-rata dari data-data yang ada pada kelompok yang sama.

$$C_{lj} = \frac{x_{1j} + x_{2j} + \dots + x_{n_{lj}}}{n_l} \quad (3)$$

dimana,

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

n_l : Banyak data pada kelompok ke- l

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

2.4. Langkah-Langkah Analisis Data

Berikut ini adalah langkah-langkah analisis data menggunakan algoritma *K-Means*:

1. Menginput data
2. Melakukan uji multikolinearitas pada seluruh variabel penelitian
3. Melakukan analisis klaster / pengelompokan menggunakan algoritma *K-Means*
4. Melakukan uji indeks koefisien *Silhouette*
5. Interpretasi hasil klaster / pengelompokan
6. Menarik kesimpulan dan saran.

3. Hasil dan Pembahasan

3.1. Sumber Data

Data dalam penelitian ini menggunakan data sekunder. Data sekunder adalah data yang diperoleh dari sumber kedua dan biasanya sudah siap dipakai sebagai keperluan untuk meneliti. Data yang digunakan yaitu mengenai faktor penyebab *stunting* di Provinsi Papua sebelum pemekaran terdiri dari 24 Kabupaten/Kota. Sumber data diperoleh dari hasil Survei Status Gizi Indonesia (SSGI) pada tahun 2022 oleh Badan Kebijakan Pembangunan Kesehatan (BKPK) Kementerian Kesehatan RI. Variabel penelitian yang digunakan dalam penelitian ini yaitu persentase bayi usia 0-5 bulan tidak mendapatkan ASI eksklusif (X_1), persentase balita dengan berat badan lahir < 2500 gram (X_2), persentase tidak mendapatkan Inisiasi Menyusui Dini (IMD) pada bayi baru lahir (X_3), persentase anak tidak mendapatkan vitamin A (X_4), persentase anggota rumah tangga yang tidak memiliki jaminan pelayanan kesehatan (X_5), persentase ibu hamil minum < 90 tablet tambah darah (X_6), persentase rumah tangga yang memiliki sarana

sumber air minum tidak layak (X_7). Data yang digunakan dalam penelitian ini adalah faktor-faktor penyebab terjadinya *stunting*, dengan variabel-variabel sebagai berikut:

Tabel 1. Variabel Penelitian

Faktor	Lambang	Keterangan	Satuan
Pemberian ASI Tidak Eksklusif	X_1	Persentase bayi usia 0-5 bulan tidak mendapatkan ASI eksklusif	Persen
Berat Badan Lahir Rendah (BBLR)	X_2	Persentase balita dengan berat badan lahir < 2500 gram	Persen
IMD	X_3	Persentase tidak mendapatkan IMD pada bayi baru lahir	Persen
Vitamin A	X_4	Persentase anak tidak mendapatkan vitamin A	Persen
Keluarga yang tidak memiliki Asuransi Kesehatan	X_5	Persentase anggota rumah tangga yang tidak memiliki jaminan pelayanan kesehatan	Persen
Minum TTD < 90 tablet	X_6	Persentase bumil minum < 90 tablet tambah darah	Persen
Sumber Air Minum Tidak layak (SAM)	X_7	Persentase rumah tangga yang memiliki sarana sumber air minum tidak layak	Persen

Berikut ini adalah dasar penggunaan variabel-variabel dalam penelitian ini:

1. Pemberian ASI Eksklusif Kurang Dari 6 Bulan Pertama
ASI mengandung zat gizi yang diperlukan oleh bayi untuk pertumbuhan dan perkembangan bayi (Hayati & Goejantoro, 2020). Selain itu, ASI dapat meningkatkan daya tahan tubuh, perlindungan penyakit infeksi dan perlindungan alergi sehingga apabila pemberian ASI Eksklusif kurang dari 6 bulan pertama maka akan meningkatkan risiko kejadian *stunting* (Sumarni, dkk, 2020).
2. Berat Badan Lahir Rendah (BBLR)
Bayi dengan riwayat berat badan bayi saat lahir kurang dari 2500 gram atau Berat Badan Lahir Rendah (BBLR) akan memiliki resiko lebih tinggi mengalami keterlambatan perkembangan dan pertumbuhan karena bayi BBLR telah mengalami *Intrauterine Growth Restriction* yang menyebabkan pertumbuhan dan perkembangan lebih lambat (Kamilia, 2019).
3. Inisiasi Menyusui Dini (IMD)
Penyebab masalah *stunting* salah satunya akibat dari penundaan Inisiasi Menyusui Dini (IMD). Anak yang tidak mendapatkan IMD memiliki kemungkinan 2,63 kali lebih tinggi mengalami kejadian *stunting* (Permadi, dkk, 2016).
4. Vitamin A
Kekurangan vitamin A dapat menyebabkan terganggunya fungsi pertumbuhan sehingga tinggi balita lebih rendah dari tinggi anak seusianya. Kurangnya asupan vitamin A pada balita dapat menyebabkan resiko 17,5 kali terkena *stunting* dibandingkan balita yang mendapatkan asupan vitamin A yang cukup (Sari, dkk, 2022).

5. Keluarga Yang Tidak Memiliki Jaminan Pelayanan Kesehatan

Kepemilikan jaminan pelayanan kesehatan merupakan salah satu faktor tidak langsung penyebab *stunting*. Balita yang memiliki jaminan pelayanan kesehatan memiliki kemudahan dalam mengakses pelayanan kesehatan sehingga dapat mengurangi resiko terjadinya *stunting* pada balita (Pertiwi, dkk, 2021).

6. Ibu Hamil Mengonsumsi Tablet Tambah Darah < 90 tablet

Studi yang menggunakan data dari hasil Riset Kesehatan Dasar (Riskesdas) tahun 2013, pemberian tablet tambah darah pada ibu hamil merupakan salah satu upaya percepatan pencegahan *stunting* di Indonesia. Studi menemukan bahwa ada hubungan konsumsi TTD pada ibu hamil ≥ 90 tablet dengan *stunting*. Ibu yang mengonsumsi TTD <90 tablet berpeluang 1,05 kali memiliki anak *stunting* dibanding ibu yang mengonsumsi TTD ≥ 90 tablet (Fentiana & Ginting, 2022).

7. Sumber Air Minum Tidak Layak

Sumber air bersih yang tidak memadai merupakan salah satu faktor sanitasi lingkungan yang buruk. Kondisi sanitasi yang buruk dapat meningkatkan resiko terjadinya penyakit infeksi yang dapat mengganggu proses penyerapan nutrisi sehingga dapat menyebabkan *stunting* pada balita (Angraini, dkk, 2021).

3.2. Uji Multikolinearitas

Tahapan sebelum melakukan pengelompokkan Kabupaten/Kota di Provinsi Papua adalah dilakukan pengujian multikolinearitas. Mendeteksi multikolinearitas dengan menggunakan VIF untuk masing-masing variabel penelitian disajikan pada Tabel 1.

Tabel 2. Nilai VIF untuk Setiap Variabel Penelitian

Variabel	VIF
X ₁	1,67
X ₂	1,73
X ₃	1,86
X ₄	3,94
X ₅	1,84
X ₆	2,50
X ₇	1,52

Berdasarkan Tabel 2, diketahui bahwa nilai VIF yang diperoleh pada semua variabel <10 sehingga dapat dikatakan bahwa tidak terdapat multikolinearitas pada semua variabel pada variabel faktor penyebab terjadinya *stunting* yang digunakan.

3.3. Penentuan Jarak Kedekatan *Euclidean*

Ukuran jarak menunjukkan kesamaan atau ketidaksamaan antara data. Semakin besar jarak menunjukkan ketidaksamaan antar data, sebaliknya semakin kecil jarak menunjukkan kesamaan antar data. Pengukuran jarak pada penelitian ini menggunakan jarak *euclidean* pada persamaan (2).

Diberikan contoh perhitungan manual jarak antara Kabupaten Merauke (data ke-1) dan Kabupaten Jayawijaya (data ke-2).

$$d_{1,2} = \sqrt{(x_{1,1} - x_{1,2})^2 + \dots + (x_{7,1} - x_{7,2})^2}$$

$$= \sqrt{(16,00 - 30,00)^2 + \dots + (15,80 - 65,80)^2} = 60,78$$

Sehingga kedekatan antara data ke-1 dan ke-2 adalah 60,78. Selanjutnya menghitung kedekatan Kabupaten Merauke (data ke-1) dan Kabupaten Jayapura (data ke-3):

$$d_{1,3} = \sqrt{(x_{1,1} - x_{1,3})^2 + \dots + (x_{7,1} - x_{7,3})^2}$$

$$= \sqrt{(16,00 - 26,00)^2 + \dots + (15,80 - 27,70)^2} = 34,85$$

Sehingga kedekatan antara data ke-1 dan ke-3 adalah 34,85. Selanjutnya menghitung kedekatan Kabupaten Jayawijaya (data ke-2) dan Kabupaten Jayapura (data ke-3):

$$d_{2,3} = \sqrt{(x_{1,2} - x_{1,3})^2 + \dots + (x_{7,2} - x_{7,3})^2}$$

$$= \sqrt{(30,00 - 26,00)^2 + \dots + (65,80 - 27,70)^2} = 44,34$$

Sehingga untuk kedekatan antara data ke-2 dan ke-3 adalah 44,34.

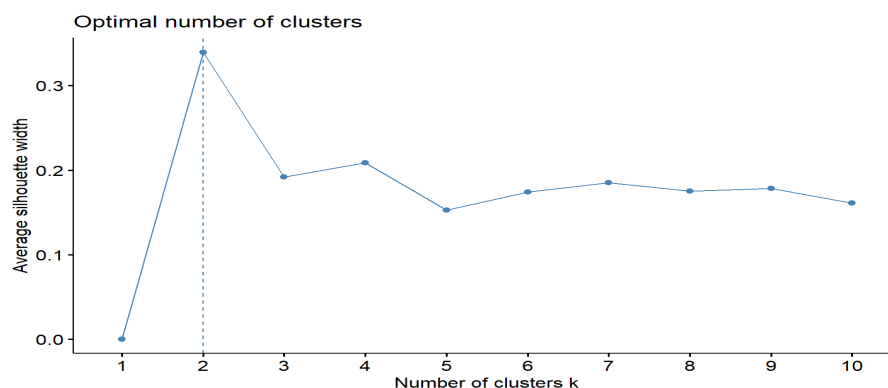
Berdasarkan tiga contoh yang diberikan terlihat bahwa jarak *euclidean* Kabupaten Merauke dan Kabupaten Jayawijaya adalah 60,78. Sedangkan jarak *euclidean* Kabupaten Merauke dan Kabupaten Jayapura adalah 34,85 dan untuk jarak Kabupaten Jayawijaya dan Kabupaten Jayapura adalah 44,34. Hal ini menunjukkan bahwa Kabupaten Merauke memiliki kemiripan karakteristik yang lebih mirip dengan Kabupaten Jayapura dari pada Kabupaten Merauke dengan Kabupaten Jayawijaya. Dengan cara yang sama dan dengan menggunakan *software* SPSS diperoleh hasil perhitungan jarak antara masing-masing data. Berikut ditampilkan jarak *Euclidean* beberapa Kabupaten/Kota dalam Tabel 3.

Tabel 3. 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

3.4. Pengelompokan Menggunakan Metode *K-Means*

Pengelompokan *K-Means* dimulai dengan menetapkan terlebih dahulu jumlah kelompok yang diinginkan dan posisi titik pusat. 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* Algoritma *K-Means*

Berdasarkan Gambar 1, terlihat bahwa dengan membentuk 2 kelompok paling optimal. Hal ini dilandasi oleh nilai SC yang didapatkan yaitu 0,34 tertinggi dibandingkan banyak kelompok

lainnya. Hasil pengelompokan dengan menggunakan algoritma *K-Means* disajikan pada tabel berikut:

Tabel 4. Hasil Pengelompokan Menggunakan Algoritma *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.

Hasil pengelompokan menggunakan algoritma K-Means menghasilkan dua klaster utama, masing-masing mencerminkan perbedaan karakteristik sosial, ekonomi, dan kesehatan masyarakat antar Kabupaten/Kota di Provinsi Papua. Klaster pertama, yang terdiri dari 21 Kabupaten/Kota seperti Merauke, Nabire, Biak Numfor, Mimika, dan Jayapura, umumnya memiliki ketersediaan infrastruktur kesehatan yang relatif lebih baik, akses terhadap pendidikan dan layanan dasar yang lebih tinggi, serta tingkat urbanisasi yang lebih cepat. Daerah-daerah ini cenderung memiliki prevalensi stunting yang lebih rendah karena faktor penunjang seperti peningkatan cakupan imunisasi, sanitasi layak, serta program gizi masyarakat yang lebih aktif.

Sebaliknya, klaster kedua yang mencakup Kabupaten Puncak Jaya, Yahukimo, dan Tolikara menggambarkan wilayah dengan tantangan geografis dan sosial-ekonomi yang lebih kompleks. Ciri khas daerah ini adalah akses transportasi yang terbatas, ketergantungan terhadap pangan lokal non-fortifikasi, serta minimnya intervensi kesehatan dan gizi yang terdistribusi secara merata. Kondisi tersebut berpotensi memperburuk tingkat stunting di wilayah tersebut, terutama pada masyarakat pedalaman yang sulit dijangkau program pemerintah.

3.5. Interpretasi Hasil Pengelompokan

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

Tabel 5. Rata-rata Variabel pada Setiap Kelompok Menggunakan Algoritma K-Means

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

Berdasarkan Tabel 5, terlihat bahwa hasil pengelompokan dengan algoritma *K-Means* memiliki karakteristik yang berbeda pada masing-masing Kelompok. Berikut adalah penjelasan untuk setiap kelompok.

1. Kelompok 1

Kelompok 1 mempunyai nilai rata-rata yang tinggi untuk variabel bayi usia 0-5 bulan tidak mendapatkan ASI eksklusif (X₁) dan variabel balita dengan berat badan lahir < 2500 gram (X₂).

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 mempunyai nilai rata-rata yang tinggi untuk variabel 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 < 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.

Setelah memperoleh hasil pengelompokan masing-masing klaster yang terbentuk, maka dilakukan proses perhitungan SC pada algoritma *K-Means*. Berikut ini merupakan hasil perhitungan nilai SC dengan menggunakan algoritma *K-Means*:

Tabel 6. Nilai SC pada Algoritma *K-Means*

Jumlah Kelompok	Nilai SC
2	0,34

Berdasarkan Tabel 6, dengan menggunakan algoritma *K-Means* diperoleh nilai SC sebesar 0,34.

4. Kesimpulan

Hasil pengelompokan menggunakan algoritma *K-Means* diperoleh jumlah kelompok optimal adalah dua kelompok dengan perolehan nilai SC sebesar 0,34. 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). Untuk peneliti selanjutnya, penulis memberikan saran agar 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

- [1] World Health Organization, *Levels and Trends in Child Malnutrition: UNICEF/WHO/World Bank Group Joint Child Malnutrition Estimates*, Geneva: WHO, 2021.
- [2] Global Nutrition Report, *2021 Global Nutrition Report — Executive Summary*, 2021.
- [3] Global Nutrition Report, *2020 Global Nutrition Report — Launch Presentation*, 2020.
- [4] UNICEF & WHO, *Joint Malnutrition Estimates: Levels and Trends in Child Malnutrition (JME)*, New York: UNICEF, 2020.
- [5] Food and Agriculture Organization (FAO), *State of Food Security and Nutrition in the World 2022 (SOFI): Global Nutrition Targets Trends*, Rome: FAO, 2022.
- [6] Kementerian Kesehatan Republik Indonesia, *Survei Status Gizi Indonesia (SSGI) 2022*

- *Data dan Buku Saku*, Jakarta: Stunting.go.id, 2023.
- [7] D. Azriani, et al., “Risk factors associated with stunting incidence in under-five children,” *PMC Journal*, 2024.
 - [8] R. D. Wulandari, et al., “A secondary data analysis of the 2022 Indonesian National Health Survey,” *PMC Journal*, 2025.
 - [9] Y. S. Arief, et al., “Social and environmental determinants of childhood stunting,” *PMC Journal*, 2025.
 - [10] H. Ashar, et al., “Factors related to stunting in children under 2 years old in Indonesia,” *PMC Journal*, 2024.
 - [11] D. Kurniawati, et al., “Geographically weighted regression model of stunting in Indonesia,” *The Journal of Maternal and Child Health*, vol. 10, no. 2, pp. 121–130, 2025.
 - [12] H. Wahyudianto and A. Prasetyo, “Addressing the issue of stunting in Papua Province,” in *Proceedings of the International Conference on Health and Social Sciences*, Atlantis Press, 2024.
 - [13] S. P. Manongga, “Determinant factors of malnutrition in Papuan children,” *SciSpace Journal*, 2023.
 - [14] A. D. Laksono, et al., “Stunting among children under two years in Indonesia: Does maternal education matter?,” *PLoS ONE*, vol. 17, no. 8, e0274361, 2022.
 - [15] “Penerapan K-Means untuk clustering data stunting pada Kecamatan Kota Waingapu,” *Jurnal Inovatif*, Universitas Kristen Wira Wacana Sumba, 2024.
 - [16] “Implementation of K-Means and Fuzzy C-Means in stunting mapping (Gunungkidul),” *Barekeng Journal of Mathematics and Applications*, Universitas Pattimura, 2024.
 - [17] “Application of K-Means for clustering stunting cases in West Sumatra,” *Semantic Scholar Database*, 2023.
 - [18] I. P. Sari, “Implementation of data classification using K-Means,” *Jurnal Teknologi Informasi UMSU*, vol. 5, no. 2, pp. 45–54, 2023.
 - [19] M. Iqbal, et al., “Determination of characteristics of stunting risk families using clustering approach,” in *ITM Web of Conferences*, 2024.
 - [20] Handayani, et al., “Performance analysis of clustering algorithms for health data classification,” *Semantic Scholar Database*, 2023.
 - [21] “Comparative analyses of clustering quality using Silhouette and Davies-Bouldin Index (DBI),” *PMC Journal*, 2023.
 - [22] “Optimization and evaluation studies of K-Means with Davies–Bouldin Index (K-optimization),” *DJournals Proceedings*, 2023.
 - [23] “Recent regional K-Means studies (Sambas, Asahan, Gunungkidul, and others): Applications for stunting data clustering,” *Jurnal Politeknik Ganesha Medan*, 2024–2025.
 - [24] “Review and synthesis of stunting determinants in Indonesia: Evidence from SSGI and DHS data,” *International Journal of Public Health (IJPH)*, 2024.
 - [25] N. Juniarti, et al., “Determinant factors influencing stunting prevention behaviors among working mothers in West Java Province,” *BMC Public Health*, vol. 25, no. 4, 2025.
 - [26] R. Kusumawardhani, Z. D. Rizqiena, and S. P. Astuti, *Ekonometrika: Suatu Pengantar*, Yogyakarta: Gerbang Media Aksara, 2021.
 - [27] S. Nugroho, *Statistika Multivariat Terapan*, Bengkulu: UNIB Press Bengkulu, 2008.