

Pengelompokkan Provinsi di Indonesia Berdasarkan Indikator Kesehatan Tahun 2023 Menggunakan Analisis *Cluster*

Nadia Alma Rahmawati ¹⁾, Amanda Shofiyah Salsabila ²⁾, Sri Pingit Wulandari ³⁾

Institut Teknologi Sepuluh Nopember

SUBMISSION TRACK

Submitted : 13 November 2024
Accepted : 18 November 2024
Published : 19 November 2024

KEYWORDS

Cluster Analysis, Health Indicators, Hierarchical, Non-Hierarchical

Analisis *Cluster*, Indikator Kesehatan, Hierarki, Non-Hierarki

CORRESPONDENCE

E-mail: its.yamandaaa@gmail.com

A B S T R A C T

Health is a fundamental human right and a key indicator of well-being that must be accessible to all citizens, in line with Indonesia's national vision of achieving a high standard of health. This study aims to classify provinces in Indonesia based on health indicators in 2023 using cluster analysis to understand the similarities and differences among provinces. The methodology includes both hierarchical and non-hierarchical cluster analysis. Hierarchical methods employ a step-by-step approach to forming clusters, with results visualized in a dendrogram to determine the optimal number of clusters. Non-hierarchical methods, such as K-Means, predefine the number of clusters and iteratively partition the data until an optimal solution is achieved. The indicators used in this study include the percentage of children under five receiving complete immunizations, married couples of childbearing age, outpatient health complaints, self-medication practices, individuals reporting health complaints, and unmet service needs. The findings reveal that the data meet the assumptions for cluster analysis. Using the complete linkage method, five optimal clusters were identified in both hierarchical and non-hierarchical analyses. This indicates that Indonesia's provinces can be grouped into five clusters based on the 2023 health indicators. The percentage of married couples of childbearing age showed the widest distribution, while unmet service needs had the lowest. This study is expected to assist the government in designing more efficient and equitable health programs.

Kesehatan merupakan hak asasi manusia yang menjadi salah satu indikator kesejahteraan yang harus dirasakan oleh seluruh masyarakat sesuai dengan cita-cita bangsa Indonesia demi mewujudkan derajat kesehatan yang tinggi. Oleh karena itu, penelitian bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indikator kesehatan 2023 dengan analisis *cluster*, untuk memahami kesamaan dan perbedaan antar provinsi. Metode yang digunakan mencakup analisis *cluster* hierarki dan non-hierarki. Metode hierarki menggunakan pendekatan bertahap dalam membentuk cluster, dengan menampilkan hasil dalam dendrogram untuk menentukan jumlah *cluster*. Metode non-hierarki, seperti *K-Means*, menetapkan jumlah *cluster* di awal untuk membagi data secara iteratif hingga mencapai hasil optimal. Indikator yang digunakan meliputi persentase balita yang mendapat imunisasi lengkap, pasangan usia subur, keluhan kesehatan rawat jalan, pengobatan mandiri,

peduduk yang mempunyai keluhan kesehatan, dan *unmet need* pelayanan. Hasil penelitian menunjukkan bahwa data memenuhi asumsi uji cluster. Metode *complete linkage* dengan 5 *cluster* optimal ditemukan pada analisis hierarki dan non-hierarki, menunjukkan provinsi di Indonesia dapat dikelompokkan menjadi 5 berdasarkan indikator kesehatan 2023. Persentase pasangan usia subur memiliki penyebaran tertinggi, sementara *unmet need* pelayanan terendah. Penelitian ini diharapkan dapat membantu pemerintah dalam menyusun program kesehatan yang lebih efisien dan merata.

2024 All right reserved

This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license

Pendahuluan

Status kesehatan masyarakat Indonesia memiliki peran penting untuk meningkatkan kualitas Sumber Daya Manusia (SDM). Kualitas SDM dapat menjadi tolak ukur kemajuan suatu negara. Akan tetapi, Indonesia seringkali menghadapi tantangan dan hambatan akan kesadaran aspek kesehatan (Purwaningsih, 2023). Tantangan dan hambatan Indonesia untuk mewujudkan visi Indonesia Sehat melibatkan aspek yang kompleks, meliputi akses layanan kesehatan yang tidak merata, kualitas tenaga kesehatan, anggaran kesehatan, gaya hidup dan keserasan masyarakat, serta regulasi dan kebijakan kesehatan pemerintah. Oleh karena itu, penelitian ini dilakukan untuk mengelompokkan seluruh provinsi di Indonesia berdasarkan indikator kesehatan tahun 2023. Pengelompokan dilakukan dengan menggunakan metode analisis *cluster*, karena metode ini dapat untuk mengelompokkan objek berdasarkan karakteristik yang sama antar anggota dalam satu *cluster*.

Analisis *cluster* atau analisis kelompok adalah teknik analisis yang bertujuan untuk mengelompokkan individu atau objek ke dalam beberapa kelompok yang memiliki sifat berbeda antar kelompok, sehingga individu atau objek yang terletak di dalam satu kelompok akan mempunyai sifat relatif homogen (Rompas, Paedong, & Salaki, 2023). Analisis cluster memiliki tujuan untuk mengelompokkan objek berdasarkan karakteristik yang sama antar anggota dalam satu *cluster* serta memiliki perbedaan yang tinggi antar *cluster* satu dengan yang lain. Analisis *cluster* memiliki dua metode utama, yaitu metode hierarki dengan penentuan banyak kelompok yang terbentuk dilihat berdasarkan hasil *dendogram* dan metode non-hierarki dengan penentuan banyak kelompok ditentukan di awal oleh peneliti (Apriliana & Widodo, 2023). Metode non-hierarki yang paling sering digunakan adalah metode *K-Means Cluster*, dimana metode tersebut dirancang untuk mengelompokkan data berdasarkan jarak dengan membagi data ke beberapa *cluster* (Puspita, 2021). Dalam penelitian ini analisis *cluster* berperan penting untuk mengelompokkan tiap provinsi di Indonesia berdasarkan kesamaan indikator kesehatan tahun 2023. Indikator kesehatan yang digunakan adalah persentase balita yang pernah mendapat imunisasi dasar lengkap, persentase pasangan usia subur, persentase penduduk yang mempunyai keluhan kesehatan dan pernah rawat jalan dalam sebulan terakhir, persentase penduduk yang mengobati sendiri selama sebulan, persentase penduduk yang mempunyai keluhan kesehatan, dan persentase *unmet need* pelayanan.

Hasil analisis *cluster* pada penelitian ini diharapkan dapat memberikan wawasan yang mendalam terkait indikator kesehatan di berbagai wilayah sehingga dapat menjadi dasar bagi pengambilan keputusan yang lebih tepat dalam merancang kebijakan pembangunan di tingkat provinsi. Penelitian ini dilakukan dengan tujuan untuk

membantu pemerintah dalam merancang sasaran program yang akan dibuat. Manfaat yang diperoleh masyarakat dan pembaca adalah sebagai informasi dan wawasan tentang pengelompokan dan pemetaan wilayah provinsi di Indonesia berdasarkan indikator kesehatan masyarakat serta dapat berpartisipasi membantu pemerintah untuk meningkatkan kesehatan masyarakat.

Metode Penelitian

Penelitian ini menggunakan pendekatan penelitian kuantitatif yang akan dikelompokkan menggunakan metode *hierarki* dan *non-hierarki* data sekunder yang diperoleh dari Badan Pusat Statistik. Data yang digunakan mencakup Persentase Balita yang Pernah Mendapat Imunisasi Dasar Lengkap, Persentase Pasangan Usia Subur, Persentase Penduduk dengan Keluhan Kesehatan dan Pernah Rawat Jalan dalam Sebulan Terakhir, Persentase Penduduk yang Mengobati Sendiri Selama Sebulan, Persentase Penduduk dengan Keluhan Kesehatan, serta Persentase *Unmet Need* Pelayanan Kesehatan. Proses yang dilakukan untuk analisis dalam penelitian ini adalah:

1. Mengumpulkan data terkait indikator kesehatan melalui *website* Badan Pusat Statistik.
2. Melakukan analisis karakteristik data pada setiap variabel.
3. Melakukan pengujian asumsi berdistribusi normal multivariat dan dependensi.
4. Melakukan analisis *cluster* hierarki menggunakan metode *complete linkage*.
 - a. Menentukan jumlah *cluster* optimum berdasarkan nilai *Pseudo-F* dan *icdrate*.
 - b. Menghitung jarak euclidean yang divisualisasikan melalui dendrogram.
 - c. Melakukan pengelompokan pada hasil *cluster* optimum yang terbentuk.
 - d. Melakukan analisis karakteristik pada masing-masing *cluster* yang terbentuk.
 - e. Melakukan pemetaan pada hasil *cluster* optimum yang terbentuk.
5. Melakukan analisis *cluster* nonhierarki menggunakan metode K-Means.
 - a. Menentukan jumlah cluster optimum berdasarkan nilai *Pseudo-F* dan *icdrate*.
 - b. Melakukan analisis perbedaan variabel pada *cluster* yang terbentuk menggunakan ANOVA.
 - c. Melakukan pengelompokan pada hasil *cluster* optimum yang terbentuk.
 - d. Melakukan analisis karakteristik pada masing-masing *cluster* yang terbentuk.
 - e. Melakukan pemetaan pada hasil *cluster* optimum yang terbentuk.
6. Interpretasi hasil analisis.
7. Menarik kesimpulan dan saran.

HASIL DAN PEMBAHASAN

1. Analisis Statistik Deskriptif

Karakteristik dari variabel indikator kesehatan di Indonesia menurut provinsi pada tahun 2023 dianalisis menggunakan statistika deskriptif mengenai mean, standar deviasi, nilai maksimum, nilai minimum, median, dan *range* dari masing-masing variabel diuraikan pada Tabel 1.

Tabel 1. Karakteristik

Variabel	Mean	Standar Deviasi	Min	Max	Median	Range
X ₁	57,70	10,30	24,55	59,40	74,25	49,70
X ₂	51,61	10,86	21,20	52,34	67,30	46,10
X ₃	32,98	7,29	19,28	31,43	53,08	33,80
X ₄	78,80	5,73	62,98	79,55	89,08	26,10
X ₅	35,30	6,45	17,67	36,70	44,95	27,28

X_6	5,43	1,457	2,660	5,34	9,040	6,38
-------	------	-------	-------	------	-------	------

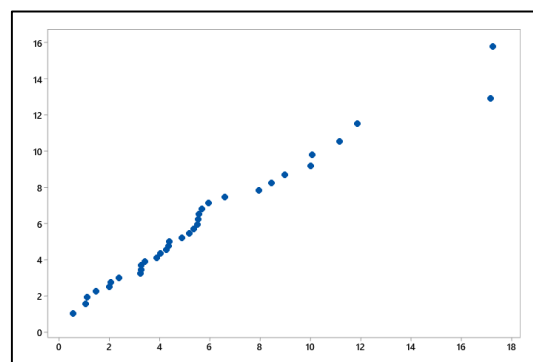
Tabel 1 menunjukkan bahwa nilai *mean* dari variabel-variabel bervariasi, di mana variabel X_1 memiliki rata-rata tertinggi yaitu sebesar 57,70, sedangkan X_6 memiliki rata-rata terendah sebesar 5,43. Standar deviasi, yang mengukur seberapa besar penyebaran data di sekitar rata-rata, menunjukkan bahwa variabel X_2 memiliki penyebaran tertinggi sebesar 10,86, dimana variasi data lebih besar dibandingkan variabel lain, sementara X_6 memiliki standar deviasi terendah sebesar 1,457. Pada segi rentang (*range*), variabel X_1 memiliki rentang paling besar yaitu 49,70, menunjukkan penyebaran data yang luas, sedangkan X_6 memiliki rentang terkecil yaitu 6,38, yang menandakan penyebaran data yang sempit. Sedangkan nilai minimum dan maksimum pada X_1 memiliki nilai minimum 24,55 dan maksimum 59,40, serta X_6 memiliki minimum 2,660 dan maksimum 5,34.

2. Uji Asumsi

Hasil pengujian asumsi yang harus dipenuhi ketika menganalisis *cluster* yaitu uji korelasi independensi dan berdistribusi normal multivariat yang akan dijelaskan sebagai berikut.

a. Distribusi Normal Multivariat

Analisis multivariat dapat digunakan apabila data memenuhi asumsi distribusi normal multivariat. Distribusi normal multivariat merupakan suatu bentuk distribusi di mana masing – masing variabel dari data memenuhi sifat normalitas (Musa, Lusiana, Mahmudi, Buwono, & Arsad, 2022). Distribusi normal multivariat adalah perluasan distribusi normal univariat dengan mean dan matriks kovarians tertentu. Penelitian ini menggunakan Q-Q plot dan uji t-proporsi untuk pemeriksaan. *Q-Q plot (quantile-quantile plot)* adalah bentuk scatter plot dengan skor quantile pada sumbu horisontal dan skor normal yang diharapkan pada sumbu vertikal. Suatu data menyebar mengikuti pola distribusi normal bila bentuk pencaran data berada sekitar garis lurus melintang. *Q-Q plot* berguna juga untuk mendeteksi adanya data *outlier* secara visual (Yamin, 2021). Pemeriksaan distribusi normal multivariat pada indikator kesehatan akan diperiksa dengan menggunakan *q-q plot* ditunjukkan pada Gambar 1.



Gambar 1. *Q-Q Plot* Uji Normalitas Multivariat

Gambar 1 menunjukkan bahwa data indikator memenuhi distribusi normal multivariat jika diperiksa dengan menggunakan *q-q plot* karena, plot-plot pada grafik *q-q plot* menyebar menyerupai garis linear. Untuk memastikan kembali hasil pemeriksaan *q-q plot*, dilakukan pengujian distribusi normal multivariat.

Hipotesis:

H_0 : Data indikator kesehatan berdistribusi normal multivariat

H_1 : Data indikator kesehatan tidak berdistribusi normal multivariat

Dengan taraf signifikan sebesar α yang bernilai 0,05 maka didapatkan daerah penolakan, tolak H_0 apabila berada di luar rentang $45\% \leq t_{\text{proporsi}} \leq 55\%$. Hasil uji distribusi normal multivariat dapat ditunjukkan pada Tabel 2.

Tabel 2. Uji Distribusi Normal Multivariat

t_{proporsi}
0,529

Tabel 2 menunjukkan bahwa nilai t_{proporsi} yang didapatkan dari hasil uji distribusi normal multivariat yaitu sebesar 0,529 yang berada di dalam rentang $45\% \leq t_{\text{proporsi}} \leq 55\%$, sehingga diputuskan gagal tolak H_0 yang berarti data indikator kesehatan berdistribusi normal multivariat.

b. Uji Independensi

Analisis multivariat mengasumsikan adanya kebebasan antar variabel respon. Oleh karena itu, sebelum dilakukan analisis multivariat terlebih dahulu dilakukan pengujian korelasi antar variabel respon Y_1, Y_2, \dots, Y_p untuk melihat apakah variabel respon saling berkorelasi satu sama lain. Pengujian ini memiliki nama yaitu uji *Bartlett Sphericity* (Marli, Kesuma, Rusdiana, Rahayu, & Fradinata, 2018). Hasil uji independensi menggunakan uji *Barlett* pada data indikator kesehatan adalah sebagai berikut.

Hipotesis:

$H_0 : \rho = \mathbf{I}$ (Matriks korelasi identik dengan matriks identitas atau antarvariabel independen)

$H_1 : \rho \neq \mathbf{I}$ (Matriks korelasi tidak identik dengan matriks identitas atau antar variabel dependen)

Taraf signifikan yang digunakan yaitu α sebesar 0,05 dan daerah penolakan, H_0 ditolak apabila $\chi^2_{\text{hitung}} > \chi^2_{(0,05,45)}$ atau $P_{\text{value}} < 0,05$. Hasil uji *Barlett* dapat ditunjukkan pada Tabel 3.

Tabel 3. Uji *Bartlett*

χ^2_{hitung}	$\chi^2_{(0,05,45)}$	P_{value}
68,508	24,996	0,001

Tabel 3 menunjukkan bahwa karakteristik indikator kesehatan memiliki χ^2_{hitung} sebesar 68,508 yang lebih besar dari nilai $\chi^2_{(0,05,15)}$ sebesar 24,996 dan diperkuat dengan P_{value} sebesar 0,001 yang lebih kecil dari α sebesar 0,05, sehingga dapat diputuskan tolak H_0 yang artinya matriks korelasi tidak identik dengan matriks identitas atau antar variabel dependen sehingga data indikator kesehatan menurut provinsi tahun 2023 dapat digunakan untuk analisis *cluster*.

3. Analisis *Cluster* Hierarki

Analisis *cluster* dengan metode hierarki digunakan sebagai pengelompokan objek secara hierarki berdasarkan kemiripan sifatnya yang belum diketahui jumlah *cluster* yang terbentuk. Metode hierarki digunakan pada data yang memiliki struktur hierarkis atau

terdiri dari beberapa sub-kelompok yang saling berkaitan dan umumnya jumlah sampel yang digunakan relatif kecil. Hasil pengelompokan dengan metode hierarki ditampilkan adalah sebuah diagram pohon yang disebut dengan *dendogram* (Apriliana & Widodo, 2023).

a. Penentuan Jumlah *Cluster* Optimum

Cluster optimum adalah jumlah klaster yang paling sesuai untuk mengelompokkan objek, di mana klaster yang terbentuk memiliki tingkat homogenitas yang tinggi di dalam setiap klaster dan tingkat heterogenitas yang tinggi antar klaster (Solikin, Kusri, & Wibowo, 2021). *Cluster* optimum akan ditentukan dengan mempertimbangkan nilai *Pseudo-F* tertinggi dan nilai *Icdrate* terendah pada setiap ukuran *cluster*. *Pseudo-F* statistic adalah ukuran untuk menentukan jumlah klaster optimal dalam analisis klaster dengan menghitung rasio variasi antar dan dalam klaster. Nilai *Pseudo-F* yang tinggi menunjukkan klaster yang terpisah dengan baik dan seragam dalam masing-masing klaster (Satyahadewi, Sinaga, & Perdana, 2023). Sedangkan *ICDrate* (internal *cluster* dispersion) menggambarkan tingkat dispersi atau perbedaan dalam klaster. Nilai *ICDrate* yang semakin kecil menunjukkan bahwa klaster tersebut semakin baik sebab antara anggota dalam satu klaster memiliki perbedaan yang rendah atau memiliki variasi yang kecil (Satyahadewi, Sinaga, & Perdana, 2023). Dalam penelitian ini akan menggunakan metode *Complete Linkage*. *Complete Linkage* adalah cara pengelompokan objek yang memiliki jarak paling jauh atau kesamaan yang sedikit (Apriliana & Widodo, 2023). Berdasarkan analisis tersebut, jumlah *cluster* optimal untuk data indikator kesehatan di Indonesia menurut provinsi pada tahun 2023 dapat diuraikan pada Tabel 4.

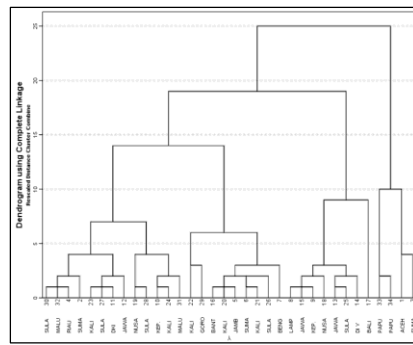
Tabel 4. Penentuan Jumlah *Cluster* Optimum

Jumlah <i>Cluster</i>	<i>Pseudo-F</i>	<i>Icdrate</i>
2	0,75	0,23
3	8,48	0,65
4	7,3	0,57
5	8,69	0,45

Tabel 4 menunjukkan bahwa jumlah *cluster* optimum untuk pengelompokan provinsi di Indonesia berdasarkan indikator kesehatan tahun 2023 yaitu 5 *cluster* yang ditunjukkan oleh nilai *Pseudo-F* terbesar sebesar 8,69 dan nilai *Icdrate* sebesar 0,45.

b. Dendogram

Dendrogram adalah diagram berbentuk pohon yang digunakan dalam analisis klaster untuk menggambarkan hubungan hierarkis antar objek atau data. Dalam *dendrogram*, objek yang lebih mirip atau memiliki kesamaan akan digabungkan dalam satu klaster lebih awal, sementara objek yang lebih berbeda akan digabungkan belakangan. *Dendrogram* membantu visualisasi seberapa dekat atau jauh hubungan antar data, dan sering digunakan dalam metode klasterisasi hierarki (Apriliana & Widodo, 2023). Berdasarkan hasil penentuan *cluster* optimum, data indikator kesehatan di Indonesia menurut provinsi pada tahun 2023 menggunakan metode *Complete Linkage* yaitu sebanyak 5 *cluster*. Anggota pada setiap *cluster* yang terbentuk dapat ditentukan dengan menggunakan nilai jarak *euclidean* yang dapat divisualisasikan sebagai berikut.



Gambar 2. Dendrogram Complete Linkage

Gambar 2 menunjukkan bahwa dendrogram dari analisis *cluster* menggunakan *Complete Linkage* menunjukkan terdapat 5 *cluster* yang terbentuk. *Cluster* pertama terbentuk 2 provinsi, *cluster* kedua terbentuk 13 provinsi, *cluster* ketiga terbentuk 9 provinsi, *cluster* keempat terbentuk 8 provinsi, dan *cluster* kelima terbentuk 2 provinsi.

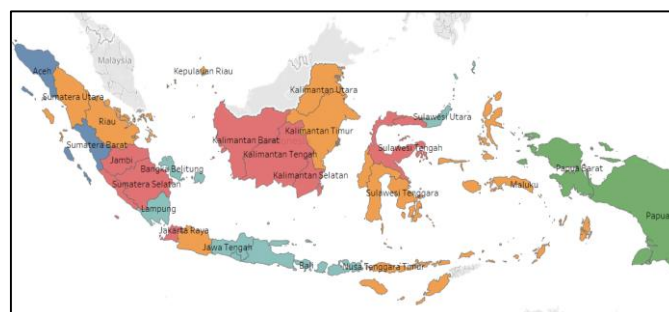
c. Hasil Cluster Hierarki

Hasil pengelompokan provinsi di Indonesia tahun 2023 berdasarkan indikator-indikator kesehatan menggunakan metode *cluster hierarki* yang akan memberikan gambaran tentang pembagian provinsi-provinsi menurut kesamaan indikator kesehatan diuraikan pada Tabel 5.

Tabel 5. Hasil Cluster Hierarki

Cluster	Jumlah	Provinsi
1	2	Aceh, Sumatera Barat
2	13	Sumatera Utara, Riau, Kep. Riau, DKI Jakarta, Jawa Barat, Nusa Tenggara Timur, Kalimantan Timur, Kalimantan Utara, Sulawesi Selatan, Sulawesi Tenggara, Sulawesi Barat, Maluku, Maluku Utara
3	9	Jambi, Sumatera Selatan, Bengkulu, Banten, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Tengah, Gorontalo
4	8	Lampung, Kep. Bangka Belitung, Jawa Tengah, DI Yogyakarta, Jawa Timur, Bali, Sulawesi Utara, Nusa Tenggara Barat
5	2	Papua, Papua Barat

Tabel 5 menunjukkan hasil pengelompokan provinsi di Indonesia pada tahun 2023 berdasarkan indikator kesehatan menggunakan metode *complete linkage*. *Cluster* dengan anggota terbanyak adalah *cluster* 2, yang terdiri dari 12 provinsi. Pengelompokan ini memberikan gambaran pola distribusi indikator kesehatan di berbagai wilayah Indonesia. Untuk memvisualisasikan hasil pengelompokan ini, pemetaan geografis akan digunakan sebagai berikut. Pemetaan ini diharapkan dapat memberikan perspektif yang lebih jelas mengenai distribusi *cluster* berdasarkan lokasi provinsi ditunjukkan pada Gambar 3.



Gambar 3. Cluster Complete Linkage

d. Karakteristik *Cluster* Terbaik

Karakteristik data dari setiap *cluster* yang terbentuk pada pengelompokan provinsi di Indonesia tahun 2023 berdasarkan indikator kesehatan menggunakan metode *Complete Linkage* ditunjukkan pada Tabel 6.

Tabel 6. Hasil *Cluster* Hierarki

Variable	Cluster	Mean	Minimum	Maksimum
X ₁	1	33,74	24,55	42,94
	2	58,06	45,15	67,48
	3	54,95	44,8	61,73
	4	68,43	62,67	74,25
	5	48,78	45,8	51,75
X ₂	1	45,195	44,3	46,09
	2	45,78	34,44	58,16
	3	60,98	53,72	67,3
	4	58,69	51,33	66,1
	5	25,42	21,2	29,63
X ₃	1	45,49	45,15	45,83
	2	31,08	24,53	41,49
	3	26,51	19,28	31,26
	4	39,23	33,94	53,08
	5	36,97	34,74	39,2
X ₄	1	71,06	66,21	75,91
	2	79,124	76,95	81
	3	84,31	78,92	89,08
	4	76,26	62,98	82,61
	5	69,72	66,16	73,29
X ₅	1	37,23	33,11	41,34
	2	37,38	24,47	44,95
	3	33,8	17,67	42,14
	4	32,03	23,51	39,65
	5	39,77	38,52	41,01
X ₆	1	5,23	4,67	5,79
	2	5,458	3,29	7,5
	3	6,158	4,53	8,72
	4	5,279	2,75	9,04
	5	2,815	2,66	2,97

Tabel 6 menunjukkan bahwa pengelompokan provinsi di Indonesia tahun 2023 berdasarkan indikator kesehatan. Pada *cluster* 4 menunjukkan persentase tertinggi balita yang mendapat imunisasi dasar lengkap sebesar 68,43% dengan nilai minimum 62,67% dan maksimum 74,25%, sedangkan *cluster* 1 memiliki persentase terendah 33,74% dengan nilai minimum 24,55% dan maksimum 42,94%. Persentase pasangan usia subur tertinggi terdapat di *cluster* 3 sebesar 60,98% dengan nilai minimum 53,72% dan maksimum 67,3%, sedangkan terendah di *cluster* 5 sebesar 25,42% dengan nilai minimum 21,2% dan maksimum 29,63%. Persentase penduduk yang memiliki keluhan kesehatan tertinggi ada di *cluster* 1 sebesar 45,49% dengan minimum 45,15% dan maksimum 45,83%, sementara *cluster* 3 memiliki persentase terendah sebesar 26,51% dengan minimum 19,28% dan maksimum 31,26%. Persentase penduduk yang mengobati sendiri tertinggi terdapat di *cluster* 3 84,31% dengan minimum 78,92% dan maksimum 89,08%, sedangkan *cluster* 5 memiliki persentase terendah 69,72% dengan minimum 66,16% dan maksimum 73,29%. Untuk persentase unmet need pelayanan, *cluster* 3

memiliki persentase tertinggi sebesar 6,16% dengan minimum 4,53% dan maksimum 8,72%, sementara *cluster* 5 memiliki persentase terendah sebesar 2,82%) dengan minimum 2,66% dan maksimum 2,97%.

4. Analisis Cluster Non-Hierarki

Metode non-hierarki digunakan untuk pengelompokan objek dimana banyaknya *cluster* yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur penggerombolan. Kemudian, metode ini dapat diterapkan pada data yang lebih besar dibandingkan metode hirarki (Rompas, Paedong, & Salaki, 2023).

a. Penentuan Jumlah Cluster Optimum

Analisis *cluster* non hierarki dilakukan untuk menentukan jumlah *cluster* yang optimal dengan mempertimbangkan nilai *Pseudo-F* tertinggi dan nilai *Icdrate* terendah pada setiap ukuran *cluster*. Metode *K-Means* digunakan dalam proses ini. *K-means* merupakan salah satu metode pengelompokan data non-hierarki yang mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang di set dalam suatu kelompok dan memaksimalkan variasi antar kelompok (Asroni, Fitri, & Prasetyo, 2018). Berdasarkan analisis tersebut, jumlah *cluster* optimal untuk data indikator kesehatan di Indonesia menurut provinsi pada tahun 2023 dapat diuraikan pada Tabel 7.

Tabel 71. Jumlah Cluster Optimum

Jumlah Cluster	<i>Pseudo-F</i>	<i>Icdrate</i>
2	10,80	0,75
3	9,81	0,61
4	7,94	0,56
5	11,08	0,40

Tabel 7 menunjukkan bahwa jumlah *cluster* optimum untuk pengelompokan provinsi di Indonesia berdasarkan indikator kesehatan tahun 2023 yaitu 5 *cluster* karena memiliki nilai *Pseudo-F* terbesar sebesar 11,08 dan nilai *Icdrate* terkecil sebesar 0,40.

b. ANOVA

Analisis varians (analysis of variance, ANOVA) adalah suatu metode untuk menguraikan keragaman total data menjadi komponen-komponen yang mengukur berbagai sumber keragaman. suatu metode analisis statistika yang termasuk ke dalam cabang statistika inferensi (Setiawan, 2019). Uji signifikansi dilakukan untuk mengetahui perbedaan yang ditimbulkan oleh indikator pada masing-masing *cluster* yang terbentuk. Pengujian signifikansi menggunakan uji ANOVA adalah sebagai berikut.

Hipotesis

$H_0 : \beta_1 = \beta_2 = \dots = \beta_6 = 0$ (Variabel prediktor ke-*j* tidak berpengaruh signifikan terhadap pengelompokan provinsi di Indonesia tahun 2023 berdasarkan indikator kesehatan)

$H_1 : \beta_1 = \beta_2 = \dots = \beta_6 \neq 0$ (Variabel prediktor ke-*j* berpengaruh signifikan terhadap pengelompokan provinsi di Indonesia tahun 2023 berdasarkan indikator kesehatan)

Ditetapkan taraf signifikan 5%, maka diperoleh daerah penolakan, yaitu tolak H_0 jika $F_{hitung} > F_{(0,05;4;29)}$ dan $P\text{-Value} < 0,05$. Hasil perhitungan uji ANOVA dapat diuraikan pada Tabel 8.

Tabel 82. ANOVA

Variabel	F_{hitung}	$F_{(0,05;4;29)}$	P-Value	Keputusan
β_1	27,44		0,000	Tolak H_0
β_2	7,67		0,000	Tolak H_0
β_3	9,44	2,70	0,000	Tolak H_0
β_4	8,75		0,000	Tolak H_0
β_5	4,04		0,010	Tolak H_0
β_6	3,66		0,016	Tolak H_0

Tabel 8 menunjukkan bahwa semua variabel prediktor memiliki nilai F_{hitung} yang lebih besar dari $F_{(0,05;4;29)}$ sebesar 2,70 dan diperkuat dengan nilai $P\text{-Value}$ lebih kecil dari 0,05 sehingga dapat diputuskan tolak H_0 untuk ke semua variabel, yang artinya semua variabel prediktor berpengaruh signifikan terhadap pengelompokan provinsi di Indonesia tahun 2023 berdasarkan indikator kesehatan.

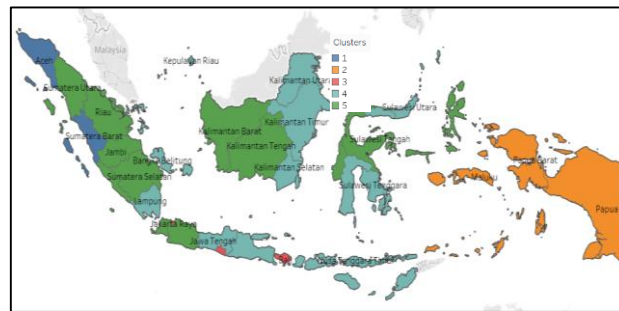
c. Hasil Cluster

Hasil pengelompokan provinsi di Indonesia tahun 2023 berdasarkan indikator-indikator kesehatan menggunakan metode *K-Means* yang akan memberikan gambaran tentang pembagian provinsi-provinsi menurut kesamaan indikator kesehatan dapat diuraikan pada Tabel 9.

Tabel 93. Hasil Cluster Hierarki

Cluster	Jumlah	Provinsi
1	2	Aceh, Sumatera Barat
2	3	Maluku, Papua Barat, Papua
3	3	DKI Jakarta, DI Yogyakarta, Bali
4	14	Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo
5	12	Sumatera Utara, Riau, Jambi, Sumatera Selatan, Bengkulu, Jawa Barat, Banten, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tengah, Sulawesi Barat, Maluku Utara

Tabel 9 menunjukkan hasil pengelompokan provinsi di Indonesia pada tahun 2023 berdasarkan indikator kesehatan menggunakan metode *K-Means*. Cluster dengan anggota terbanyak adalah cluster 4, dengan jumlah 14 provinsi. Pengelompokan ini memberikan gambaran pola distribusi indikator kesehatan di berbagai wilayah Indonesia. Untuk memvisualisasikan hasil pengelompokan ini, pemetaan geografis akan digunakan sebagai berikut. Pemetaan ini diharapkan dapat memberikan perspektif yang lebih jelas mengenai distribusi cluster berdasarkan lokasi provinsi akan ditunjukkan pada Gambar 4.



Gambar 4. Cluster K-Means

d. Karakteristik Cluster yang Terbentuk

Karakteristik data dari setiap *cluster* yang terbentuk pada pengelompokan provinsi di Indonesia tahun 2023 berdasarkan indikator kesehatan menggunakan metode *K-Means* ditunjukkan pada Tabel 10.

Tabel 10. Hasil Cluster Non-Hierarki

Variable	Cluster	Mean	Minimum	Maksimum
X ₁	1	33,74	24,55	42,94
	2	50,53	45,8	54,05
	3	70,61	63,37	74,25
	4	64,42	59,27	69,11
	5	52,41	44,8	61,73
X ₂	1	45,195	44,3	46,09
	2	28,42	21,2	34,44
	3	50,28	48,16	51,36
	4	54,26	40,73	67,3
	5	55,71	43,18	63,74
X ₃	1	45,49	45,15	45,83
	2	33,67	27,08	39,2
	3	45,1	40,72	53,08
	4	32,2	22,68	41,56
	5	28,61	19,28	38,8
X ₄	1	71,06	66,21	75,91
	2	72,21	66,16	77,17
	3	71,08	62,98	76,95
	4	80	72,06	89,08
	5	82,257	78,92	87,57
X ₅	1	37,23	33,11	41,34
	2	39,993	38,52	41,01
	3	36,6	31,29	39,65
	4	30,86	17,67	42,5
	5	38,67	33,01	44,95
X ₆	1	5,23	4,67	5,79
	2	4,05	2,66	6,51
	3	3,39	2,75	4,13
	4	6,004	4,33	9,04
	5	5,655	4,53	7,23

Tabel 10 menunjukkan bahwa *Cluster* 3 memiliki persentase tertinggi balita yang mendapat imunisasi lengkap sebesar 70,61% dengan minimum 63,37% dan maksimum 74,25%, sementara *cluster* 1 terendah sebesar 33,74% dengan minimum 24,55% dan maksimum 42,94%. Persentase pasangan usia subur tertinggi ada di *cluster* 5 sebesar 55,71% dengan minimum 43,18% dan maksimum 63,74%, dan terendah di *cluster* 2

sebesar 28,42% dengan minimum 21,2% dan maksimum 34,44%). Persentase keluhan kesehatan tertinggi ada di *cluster* 1 sebesar 45,49% dengan minimum 45,15% dan maksimum 45,83%, sedangkan terendah di *cluster* 5 sebesar 28,61% dengan minimum 19,28% dan maksimum 38,8%). Persentase pengobatan sendiri tertinggi di *cluster* 5 sebesar 82,26% dengan minimum 78,92% dan maksimum 87,57%, terendah di *cluster* 1 sebesar 71,06% dengan minimum 66,21% dan maksimum 75,91%. *Cluster* 2 mencatat persentase tertinggi penduduk dengan keluhan kesehatan sebesar 39,99% dengan minimum 38,52% dan maksimum 41,01%, terendah di *cluster* 4 sebesar 30,86% dengan minimum 17,67% dan maksimum 42,5%. Unmet need pelayanan tertinggi di *cluster* 4 sebesar 6,00% dengan minimum 4,33% dan maksimum 9,04%, terendah di *cluster* 3 sebesar 3,39% dengan minimum 2,75% dan maksimum 4,13%.

Kesimpulan

Berdasarkan hasil yang dijelaskan pada analisis dan pembahasan didapatkan kesimpulan dari penelitian sebagai berikut.

1. Persentase pasangan usia subur memiliki penyebaran tertinggi, sedangkan yang terendah adalah persentase unmet need pelayanan. Persentase balita yang pernah mendapat imunisasi dasar lengkap memiliki penyebaran data terluas, sedangkan yang tersempit adalah persentase unmet need pelayanan.
2. Data indikator kesehatan telah memenuhi asumsi uji *cluster*, yaitu mengikuti distribusi normal multivariat dan antar variabel dependen sehingga dapat dianalisis lebih lanjut.
3. Hasil analisis *cluster hierarki* menunjukkan bahwa metode *complete linkage* yang paling optimum dengan jumlah *cluster* sebanyak 5. Hal tersebut mengandung arti bahwa berdasarkan kedekatan indikator kesehatan 2023, provinsi di Indonesia dapat dikelompokkan menjadi 5.
4. Hasil analisis *cluster non-hierarki* menunjukkan bahwa metode *complete linkage* yang paling optimum dengan jumlah *cluster* sebanyak 5. Hal tersebut mengandung arti bahwa berdasarkan kedekatan indikator kesehatan 2023, provinsi di Indonesia dapat dikelompokkan menjadi 5.

Saran

Saran bagi pembaca dapat membaca terlebih dahulu teori mengenai analisis *cluster* agar dapat lebih cepat memahami isi laporan ini dan menggunakan informasi yang didapat dari penelitian ini untuk sesuatu yang positif. Saran bagi pemerintah dapat menggunakan hasil penelitian ini untuk melaksanakan program yang telah dibuat mengenai pemerataan kesehatan di seluruh provinsi di Indonesia yang diutamakan pada *cluster* dengan indikator kesehatan minim.

Daftar Pustaka

- Apriliansa, T., & Widodo, E. (2023). Analisis *Cluster Hierarki* untuk Pengelompokan Provinsi di Indonesia berdasarkan Jumlah Base Transceiver Station dan Kekuatan Sinyal. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*.
- Asroni, Fitri, H., & Prasetyo, E. (2018). Penerapan Metode *Clustering* dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik). *SEMESTA TEKNIKA*, 60.

- Budiwanto, S. (2017). *METODE STATISTIKA UNTUK MENGOLAH* . Malang: UNIVERSITAS NEGERI MALANG.
- Gozali, I. (2016). *Aplikasi Analisis Multivariete Dengan Program IBM SPSS 23 (Edisi 8)*. Semarang: Badan Penerbit Universitas Diponegoro.
- Hamdi, A. S., & Bahrudin, E. (2015). *Metode Penelitian Kuantitatif Aplikasi dalam Pendidikan*. Yogyakarta: Deepublish.
- Marli, Z., Kesuma, Rusdiana, S., Rahayu, L., & Fradinata, E. (2018). *PENGANTAR BIOSTATISTIKA DAN APLIKASINYA PADA STATUS KESEHATAN GIZI REMAJA*. Banda Aceh: Syiah Kuala University Press.
- Musa, M., Lusiana, E. D., Mahmudi, M., Buwono, N. R., & Arsad, S. (2022). *Analisis Multivariat Terapan untuk Penelitian Ekologi Kuantitatif*. Malang: Universitas Brawijaya Press.
- Purwaningsih, E. (2023). Kebijakan Terkait Krisis Kesehatan: Analisa Kebutuhan Tenaga . *JURNAL KEBIJAKAN KESEHATAN INDONESIA*, 66-73.
- Puspita, R. N. (2021). Analisis K-Means *Cluster* pada Kabupaten/Kota di Provinsi Banten Berdasarkan Indikator Indeks Pembangunan Manusia. *Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, 271-272.
- Rompas, G., Paedong, M., & Salaki, D. (2023). Penerapan Analisis *Cluster* Hierarki dalam Pengelompokan Kecamatan Berdasarkan Produksi Buah-Buahan di Kabupaten Minahasa Selatan Tahun 2020. *d'CartesiaN (Jurnal Matematika dan Aplikasi)*.
- Satyahadewi, N., Sinaga, S., & Perdana, H. (2023). HIERARCHICAL *CLUSTER* ANALYSIS OF DISTRICTS/CITIES IN NORTH SUMATRA PROVINCE BASED ON HUMAN DEVELOPMENT INDEX INDICATORS USING PSEUDO-F . *Barekeng*.
- Setiawan, K. (2019). Buku ajar metodologi penelitian.
- Solikin, A., Kusri, & Wibowo, F. (2021). Evaluasi *Cluster* Data Interkomparasi Anak Timbangan Dengan Algoritma Self Organizing Maps . *JURNAL SISFOTENIKA* .
- Sugiarto, & Setio, H. (2021). *STATISTIKA TERAPAN UNTUK BISNIS DAN EKONOMI*. Yogyakarta: ANDI.
- Supriadi, I., abadi, R., & Maghfiroh, R. U. (2023). *statistika Ekonomi dan Bisnis*. Yogyakarta: Jejak Pustaka.
- Yamin, S. (2021). *TUTORIAL STATISTIK SPSS, LISREL, WARPPLS, & JASP (MUDAH & APLIKATIF)*. Depok: Dewangga Energi Internasional Publishing.