



Hierarchic Cluster Analysis of the Distribution of Covid-19 Cases by Province in Indonesia

Euis Sartika, Sri Murniati and Agus Binarto

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

October 31, 2022

ANALISIS CLUSTER HIERARKI PADA PERSEBARAN KASUS COVID-19 BERDASARKAN PROVINSI DI INDONESIA

Euis Sartika¹, Sri Murniati², Agus Binarto³

^{1,2,3}Politeknik Negeri Bandung

sartikaeuis1965@gmail.com; sri.murniati@polban.ac.id, agus.binarto@polban.ac.id

ABSTRACT

This study aims to apply Hierarchical cluster analysis to the distribution of Covid-19 cases by province in Indonesia. The data used is secondary data for the 2021 period. The variables used are the number of confirmed patients, the number of recovered patients, the number of patients who died, the population, population density, the number of elderly people, and health facilities. The method in this research is hierarchical cluster analysis with agglomeration process, Between groups linkage, with the concept of square Euclidean distance. The research step begins with data exploration, standardization, multicollinearity assumption test and Barlett KMO test, hierarchical cluster analysis, and interpretation. The results showed that the clustering process in Hierarchical analysis can be determined based on the desired number of clusters and the results are strengthened by the Dendogram representation. The distribution of Covid-19 cases in 34 provinces is divided into 2,3, and 4 clusters based on the variables used. Prov. DKI Jakarta has different characteristics from the prov. West Java, Central Java, East Java. Likewise with the prov. Banten sd. Prov. Gorontalo has similar characteristics in the case of the spread of Covid-19.

Keywords: covid-19, Cluster Analysis, Hierarchy cluster,

INTISARI

Penelitian ini bertujuan menerapkan analisis cluster Hierarki pada persebaran kasus Covid-19 berdasarkan provinsi di Indonesia. Data yang digunakan adalah data sekunder periode tahun 2021. Variabel yang digunakan adalah jumlah pasien terkonfirmasi, Jumlah pasien sembuh, jumlah pasien meninggal, jumlah penduduk, kepadatan penduduk, jumlah penduduk lansia, dan sarana kesehatan. Metode pada penelitian ini adalah Analisis cluster Hierarki dengan proses Agglomeration, Between groups Linkage, dengan konsep jarak Square Euclidean distance. Langkah penelitian dimulai dengan eksplorasi data, standarisasi, uji asumsi multikolinearitas dan uji KMO Barlett, analisis cluster Hierarki, dan interpretasi. Hasil penelitian menunjukkan bahwa proses pengklasteran pada analisis Hierarki dapat ditentukan hasilnya berdasarkan jumlah cluster yang diinginkan dan hasilnya diperkuat dengan representasi Dendogram. Pada persebaran kasus Covid-19 di 34 provinsi dibedakan kedalam 2,3, dan 4 cluster berdasarkan variable-variabel yang digunakan. Prov. DKI Jakarta mempunyai karakteristik yang berbeda dengan prov. Jawa Barat, Jawa Tengah, Jawa Timur. Begitu juga dengan prov. Banten sd. Prov. Gorontalo mempunyai karakteristik yang mirip dalam kasus persebaran Covid-19.

Kata kunci : covid-19, Analisis Cluster, cluster Hierarki,

1. PENDAHULUAN

Di Indonesia, kasus terkonfirmasi positif hingga 8 Februari 2021 adalah 1.157.837 orang dengan jumlah kematian 31.556 orang, dengan persentase tertinggi adalah kelompok usia 46-59 tahun. Daerah dengan tingkat kepadatan tinggi menduduki tempat tertinggi kasus terkonfirmasi, provinsi tersebut antara lain : Jawa barat, Jawa tengah, Jawa Timur, dan DKI Jakarta. Kenyataannya, provinsi-provinsi tersebut mempunyai jumlah penduduk yang tinggi dengan tingkat kepadatan yang tinggi juga. Hal inilah yang menjadi alasan jumlah penduduk dan tingkat kepadatan penduduk menjadi variabel yang digunakan dalam penelitian ini (Ganasegeran *et al.*, 2021). Digunakannya factor demografi sebagai salah satu variabel yang memengaruhi persebaran jumlah kasus Covid-19 didasarkan pada hasil penelitian di beberapa negara, selain factor lingkungan (Aabed and Lashin, 2021). Secara geografis, daratan Indonesia yang mencapai luas 1,91 juta km² dengan 17.504 buah pulau, ditambah jumlah penduduk yang relative besar, menjadi tugas pemerintah untuk lebih serius lagi dalam mencari upaya memutus mata rantai penyebaran virus. Upaya tersebut adalah mendapatkan informasi mengenai karakteristik data pasien Covid-19 di Indonesia berdasarkan provinsi- provinsi yang dikelompokkan menurut pasien yang terjangkit Covid- 19 ke dalam karakteristik yang sama, sehingga dapat diketahui ciri khas dari masing-masing kelompok (Sari and Sukestiyarno, 2021). Analisis Cluster merupakan teknik multivariate yang mempunyai tujuan mengelompokkan objek- objek berdasarkan kesamaan karakteristik. Analisis Cluster mengelompokkan objek-objek yang paling dekat kesamaannya dengan objek lain yang berada dalam cluster yang sama. Pengelompokkan yang baik mempunyai karakteristik homogenitas yang tinggi antar anggota dalam satu cluster (*within cluster*) dan heterogenitas yang tinggi antar cluster (*between cluster*) (A.GoreJr, 2000). Keterbaruan penelitian ini adalah digunakannya data terbaru yakni data kasus Covid- 19 pada tahun 2021 dan pada penelitian sebelumnya telah

digunakan variable demografi, selain faktor demografi juga faktor usia lansia (Dowd *et al.*, 2020) dan sarana kesehatan. Analisis cluster merupakan salah satu analisis multivariat yang bertujuan mengelompokkan objek-objek atau variable-variabel berdasarkan karakteristik yang dimiliki. Tujuan Analisis cluster adalah mengelompokkan objek ke dalam suatu cluster yang saling bebas sehingga objek-objek didalam satu cluster memiliki kesamaan karakteristik, dan objek-objek yang berada dalam cluster yang berlainan tidak mempunyaikesamaan. Ukuran yang digunakan pada analisis cluster adalah ukuran jarak . Ukuran jarak menghasilkan kesamaan atau kedekatan antar objek, jarak yang digunakan ialah jarak euclidian (Rachmatin, 2014). Tujuan dari clustering adalah untuk menggolongkan obyek yang memiliki kemiripan pada klaster yang sama. Sehingga beberapa ukuran yang diperlukan untuk mengetahui seberapa mirip atau berbeda obyek-obyek tersebut. Pendekatan yang paling sering digunakan adalah mengukur kemiripan yang dinyatakan dalam jarak (distance) antar obyek. Semakin besar nilai ukuran kemiripan atau jarak antar obyek, maka semakin semakin besar perbedaan antar obyek tersebut. Sehingga semakin cenderung untuk tidak menganggapnya ke dalam kelompok yang sama (Yulianto & Hidayatullah, 2014). Adapun cluster yang baik mempunyai ciri-ciri sebagai berikut (Santoso, Menguasai Statistika Multivariat Konsep Dasar dan Aplikasi SPSS, 2015) : a) memiliki kesamaan (homogenitas) yang tinggi antar objek dalam satu cluster (within cluster), b) memiliki perbedaan (heterogenitas) yang tinggi antar cluster yang satu dengan cluster yang lainnya (between cluster). Proses clustering pada dasarnya adalah mencari dan mengelompokkan data yang mirip. Proses pengolahan data sehingga sekumpulan data mentah dapat dikelompokkan menjadi satu atau beberapa cluster adalah sebagai berikut (Santoso, Statistika Multivariat dengan SPSS, 2018) :

- i) Menentukan ukuran jarak antar data, jarak dapat diukur dengan : korelasi antara pasangan objek dari berbagai variable, jarak (distance) antara dua objek, dan hubungan antar objek.
- ii) Melakukan proses standarisasi data.
- iii) Melakukan proses clustering.
- iv) Melakukan identifikasi cluster-cluster yang terbentuk
- v) Melakukan validasi dan profiling cluster.

1.1 Analisis Cluster

Analisis Cluster dibagi menjadi dua, yakni : metode hirarki dan metode nonhirarki. Pembentukan cluster hirarki lebih bersifat ilmiah, sehingga banyak digunakan di berbagai bidang penelitian. Metode cluster hirarki dibedakan menjadi dua analisis, yaitu metode agglomerative dan metode divisive. Metode agglomerative dibagi lagi menjadi single linkage, complete linkage, average linkage, dan metode Ward. Sedangkan metode divisive adalah kebalikan dari metode agglomerative (Gudono, 2011). Metode hierarki adalah metode yang dilakukan secara bertahap atau berjenjang. Langkah metode ini akan membentuk tahapan berjenjang seperti struktur pohon serta dapat dihasilkan bentuk dendogram. Dendogram merupakan gambaran visual dari langkah proses analisis cluster yang terbentuk dan nilai koefisien jarak pada setiap tahap. Nilai atau Angka yang berada disebelah kanan dendogram merupakan objek penelitian, digambarkan garis yang menghubungkan objek-objek penelitian tersebut sehingga membentuk satu cluster (Simamora, 2005). Algoritma analisis cluster metode hierarki adalah sebagai berikut :

- a. Menentukan jumlah cluster yang ingin dibentuk, yakni k.
- b. Setiap data objek dianggap sebagai cluster. Kalau $n =$ jumlah data dan $c =$ jumlah cluster, berarti $n = c$.
- c. Menghitung jarak antara cluster.
- d. Menggabungkan dua cluster yang mempunyai jarak antar cluster paling kecil (berarti $n = c - 1$).
- e. Jika $n > k$, Kembali ke Langkah c.

Metode Hierarki dibedakan dalam dua macam, yaitu : metode penggabungan (agglomerative) dan metode pemisah (divisive) (Handoyo, 2014).

- A) Konsep metode Agglomeratif, menganggap setiap objek merupakan cluster yang berlainan. Setiap objek yang memiliki jarak terdekat akan digabungkan kedalam satu cluster dan objek yang memiliki jarak terdekat ketiga digabungkan kedalam cluster pertama atau bergabung dengan objek lain yang memiliki jarak terdekat yang sama sehingga membentuk cluster baru. Tahap tersebut berlangsung sampai terbentuknya cluster-cluster yang terdiri dari keseluruhan objek. Metode Agglomeratif (agglomerative method) terdiri dari lima metode, yaitu:
- 1) Metode Pautan Tunggal (Single Linkage Method) Single linkage method (jarak terdekat) atau tautan tunggal dapat dilakukan dengan mengelompokkan data berdasarkan jarak paling dekat (nearest neighbour) (Amponsah, et al, 2013). Dalam ukuran jarak terdekat mendefinisikan jarak antara dua cluster merupakan jarak terkecil antara cluster pertama dengan cluster kedua (Odilia, 2015). Metode single linkage method dapat dihitung dengan persamaan berikut (Rencher, 2002).

$$D(A, B) = \min \{d(y_i, y_j), \quad \text{untuk } y_i \text{ dalam } A \text{ dan } y_j \text{ dalam } B\}$$

2) Metode Pautan Lengkap (Complete Linkage Method)

Proses Complete linkage method (jarak jauh) dilakukan dengan menggabungkan data berdasarkan jarak paling jauh. Konsep ukuran ini sama dengan ukuran single linkage, namun yang membedakannya pada metode ini memerlukan jarak terjauh antara cluster (Odilia, 2015). Metode complete linkage method dapat dihitung dengan persamaan berikut (Rencher, 2002) :

$$D(A, B) = \max \{d(y_i, y_j), \quad \text{untuk } y_i \text{ dalam } A \text{ dan } y_j \text{ dalam } B\}$$

3) Metode Pautan Rata-Rata (Average Linkage Method) Average linkage method merupakan metode yang dilakukan dengan mengelompokkan data berdasarkan jarak rata-rata antar keseluruhan data. Metode average linkage method dapat dihitung dengan persamaan berikut (Rencher, 2002).

$$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} d(y_i, y_j)$$

4). Metode Ward (Ward's Method) Ward's method adalah pengclusteran dengan memaksimalkan kesamaan dalam satu cluster serta memakai perhitungan yang lengkap. Pada setiap tahap, jarak antara dua cluster yang dapat dibentuk ialah Sum of Square Error (SSE) dalam dua cluster terkecil digabungkan (Rencher, 2002). Metode ward merupakan bagian dari metode hierarki yang membentuk pengclusteran sejumlah objek n ke dalam n, n-1, n-2, sampai seterusnya hingga membentuk satu cluster. Sum of Square Error (SSE) dapat dilakukan jika setiap cluster beranggotakan lebih dari satu objek. Sum of Square Error (SSE) akan bernilai nol jika terdapat cluster yang beranggotakan satu objek saja. Metode ward dapat dihitung dengan persamaan berikut (Sarfia, 2016),

$$SSE = \sum_{i=1}^n (y_i - \bar{y}) (y_i - \bar{y})$$

dimana :

y_i = nilai objek ke-i dengan $i=1, 2, 3, \dots, n$

\bar{y} = rata-rata nilai objek dalam cluster n = banyaknya cluster

5. Metode Pusat (Centroid Method)

Metode pusat disebut juga metode titik pusat, dimana jarak antar cluster pada metode centroid merupakan jarak antar centroid. Jika terjadi pembentukan cluster baru maka akan terjadi perhitungan ulang (Rencher, 2002). Jarak antara dua kelompok didefinisikan sebagai berikut :

$$D(A, B) = d(\bar{y}_A, \bar{y}_B)$$

dimana \bar{y}_A dan \bar{y}_B adalah vector rata-rata untuk pengamatan cluster y_A dan y_B yang dirumuskan sebagai :

$$\bar{y}_A = \sum_{i=1}^{n_A} \frac{y_A}{n_A}$$

Dua cluster dengan jarak terkecil antara centroid pada setiap langkah. Setelah dua cluster A dan B bergabung, pusat dari cluster AB dapat diberikan dengan rata-rata sebagai berikut:

$$\bar{y}_{AB} = \frac{n_A \cdot \bar{y}_A + n_B \cdot \bar{y}_B}{n_A + n_B}$$

6. Metode Divisif (*Devisive Method*)

Proses ini dimulai hanya dengan satu cluster saja, namun anggota cluster tersebut mencakup semua objek penelitian. Kemudian objek memiliki perbedaan yang cukup besar akan pisahkan dan digabungkan ke dalam cluster yang sama. Proses ini berlangsung sampai terbentuknya sejumlah cluster yang diinginkan.

METODOLOGI PENELITIAN

Data dalam penelitian ini adalah data sekunder yakni, data kasus terkonfirmasi positif Covid-19 (X1), jumlah pasien dirawat (X2), jumlah pasien meninggal (X3) diperoleh dari Gugus Tugas Percepatan Covid-19, dan data jumlah penduduk, kepadatan penduduk, jumlah penduduk lansia, serta sarana Kesehatan diperoleh dari BPS

tahun 2021. Langkah-langkah dalam penelitian ini adalah sebagai berikut :

- mengumpulkan semua artikel tentang penggunaan analisis cluster,
- mengidentifikasi bahan-bahan kajian yang belum dan telah dipublikasi, dengan memperhatikan penerapan yang paling dibutuhkan dan memungkinkan untuk dilakukan sebuah penelitian,
- mengkaji tentang penggunaan analisis *Cluster* pada persebaran kasus Covid-19 dengan menggunakan data tahun 2021 dan variable yang digunakan dalam penelitian, yang belum pernah digunakan dalam penelitian lain yang serupa,
- data penelitian adalah data sekunder, berupa : kasus terinfeksi Covid-19, pasien sembuh, dan pasien meninggal diperoleh dari Gugus Tugas Percepatan Penanganan Covid-19 untuk tahun 2021. Variable jumlah penduduk, kepadatan penduduk, penduduk usia lansia, dan sarana kesehatan diperoleh dari BPS tahun 2021. Analisis statistik dalam penelitian ini adalah analisis deskriptif berdasarkan variable-variabel yang terlibat penelitian ini dan analisis *Cluster*.
- pengumpulan dan eksplorasi data, selanjutnya dilakukan analisis deskripsi berdasarkan variable- variabel yang terlibat.
- pengujian asumsi data, yakni uji multikolinearitas. Jika data tidak multikolinearitas maka dilakukan transformasi data, penghilangan sebagian data, atau eliminasi data, sampai diperoleh kondisi multikolinearitas,
- melakukan standardisasi data,
- pengujian analisis Cluster,
- interpretasi hasil analisis data.

HASIL DAN PEMBAHASAN

Analisis Deskriptif

Tabel 1. Statistika Deskriptif Data Berdasarkan Variabel

	Minimum	Maximum	Mean	Std. Deviation
X1	13901	1250715	178085.71	290260.743
X2	13397	1234305	173376.97	284335.839
X3	291	33290	4609.24	7921.124
X4	713600	48782400	7658373.53	11346397.77
X5	9	15978	744.26	2721.058
X6	114176	11867100	1802459.91	2964648.041
X7	11	329	73.29	82.515

Tabel1 menunjukkan bahwa pasien terkonfirmasi maksimum sebanyak 1250715 orang dan minimum sebanyak 13901 orang. Sementara pasien sembuh maksimum mencapai 1234305 orang dan minimum mencapai 13397 orang . Jumlah pasien yang meninggal, minimum sebanyak 291 orang dan maksimum sebanyak 33290 orang. Jadi secara persentase, pasien terkonfirmasi dan dinyatakan sembuh cukup tinggi mendekati 99% .

- **Uji Multikolinearitas**

Uji Multikolinearitas dilakukan untuk menyelidiki apakah terdapat korelasi antar variable-variabel bebas. Jika data tidak multikolinearitas, maka dilakukan penyembuhan dengan cara transformasi, eliminasi, atau penghilangan sebagian data. Hasil pengujian diperlihatkan pada table 2.

Tabel 2. Hasil Pengujian Multikolinearitas

Variabel	Tolerance	VIF
X1	0,122	8,228
X2	0,115	6,448
X3	0,142	7,027
X4	0,271	3,692
X5	0,571	1,934
X6	0,101	9,878
X7	0,112	8,908

Tabel 2 menunjukkan bahwa semua nilai tolerance lebih besar dari 0,10 dan nilaiVIF dari semua variabel adalah lebih kecil dari 10. Hal ini menunjukkan bahwa tidak terjadi multikolinearitas diantara variable- variabel bebas.

- **Uji KMO (Kaiser-Meyer-Olkin)**

Uji KMO (Kaiser-Meyer-Olkin) berfungsi untuk menentukan kelayakan (*appropriateness*) dari suatu analisis faktor. Kisaran nilai uji KMO antara 0 sampai 1. Apabila nilai KMO kurang dari 0.5, analisis faktor tidak layak untuk dilakukan. Jika nilai KMO lebih dari 0.5, analisis faktor layak dilakukan. Hasil analisis data menunjukkan nilai KMO sebesar 0.743 dan terletak diantara 0 dan 1, maka proses analisis faktor dapat dilanjutkan. Hasil analisis menunjukkan bahwa nilai *Bartlett's test of sphericity* atau nilai signifikannya

mencapai nilai 0.000 dan kurang dari 0,05, artinya variabel-variabel yang digunakan berkorelasi dan dapatdijalankan ke proses selanjutnya.

Tabel 3. Uji KMO dan Barletts

Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy	0,743
Bartlett's Test of Sphericity (Sig)	0,000

Berdasarkan tabel 3, dapat ditunjukkan bahwa nilai Signifikansi (probabilitas) sebesar 0,000 kurang dari 0,05 sehingga dapat dilakukan langkah analisis selanjutnya. Setelah asumsi analisis cluster dipenuhi, selanjutnyadilakukan langkah analisis cluster dengan tahapan sebagai berikut :

- **Standardisasi Data**

Standardisasi data digunakan jika terjadi perbedaan satuan yang signifikan antar variable, jika tidak terjadi perbedaan satuan yang tidak cukup mencolok, tidak perlu dilakukan. Penghitungan proses standardisasi digunakan rumus z-score sebagai berikut :

$$Z = \frac{x_i - \bar{x}}{S_x}$$

- Hasil Analisis Cluster

Proximity Matrix memperlihatkan matrik jarak antara variabel satu dengan variabel yang lain yang ditunjukkan dengan jarak Euclidean. Semakin kecil nilai jarak, maka karakteristik kedua variable semakin mirip, selanjutnya kedua variabel tersebut akan membentuk kelompok (cluster).

Tabel 2. Hasil Clustering dengan Between Group Linkage

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		NextStage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	33	34	.004	0	0	3
2	20	21	.013	0	0	9
3	32	33	.015	0	1	10
4	12	13	.016	0	0	14
5	30	31	.017	0	0	8
6	17	18	.019	0	0	9
7	22	26	.024	0	0	11
8	29	30	.030	0	5	11
9	17	20	.033	6	2	13
10	24	32	.038	0	3	15
11	22	29	.061	7	8	13
12	14	19	.063	0	0	14
13	17	22	.076	9	11	16
14	12	14	.076	4	12	17
15	24	28	.089	10	0	16
16	17	24	.129	13	15	22
17	12	23	.140	14	0	19
18	11	15	.143	0	0	23
19	12	25	.155	17	0	22
20	6	8	.157	0	0	24
21	7	10	.189	0	0	24
22	12	17	.322	19	16	25
23	11	16	.322	18	0	26
24	6	7	.623	20	21	26
25	12	27	.694	22	0	28
26	6	11	.751	24	23	28
27	3	4	1.041	0	0	31
28	6	12	1.090	26	25	29
29	5	6	4.026	0	28	30
30	5	9	5.888	29	0	32
31	2	3	14.496	0	27	32
32	2	5	58.393	31	30	33
33	1	2	96.461	0	32	0

Tabel 2 merupakan hasil proses clustering dengan metode Between Group Linkage. Pengelompokan, dilakukan secara bertingkat, setelah jarak Euclidean ditentukan dan hasilnya sebagai berikut :

Stage 1 : Terbentuk 1 cluster yang beranggotakan prov. Maluku Utara dan prov. Gorontalo dengan jarak 0,004 (kolom Coefficients). Karena proses aglomerasi dimulai dari 2 obyek yang terdekat, maka jarak tersebut adalah yang terdekat dari sekian kombinasi jarak 34 obyek yang ada. Selanjutnya pada kolom terakhir (Next Stage), terlihat angka 3. Hal ini berarti clustering selanjutnya dilakukan dengan melihat stage 3, dengan penjelasan berikut :

Baris ke-3 (stage 3) terlihat obyek ke-32 (Prov. Sulawesi Barat) membentuk cluster dengan prov. Maluku Utara. Selanjutnya, cluster terdiri dari 3 obyek yaitu prov. Maluku Utara, Gorontalo , dan Maluku Utara. Sedangkan jarak sebesar 0,15 merupakan jarak rata-rata obyek terakhir yang bergabung dengan 2 obyek sebelumnya, seperti tampak dalam Proximity matrix dan dapat dihitung sebagai berikut :

- Jarak prov. Maluku Utara dan Gorontalo = 0,004
- Jarak prov. Maluku Utara dan Sulawesi Barat=0,015
- Jarak rata-rata = (0,004 + 0,015) / 2 = 0,0095

Stage 2 : Terjadi pembentukan cluster prov. Sulawesi Tengah dan Kalimantan Tengah, berjarak 0,013 yang kemudian berlanjut ke stage 9.

Demikian seterusnya dari stage 9 dilanjutkan ke stage 13, sampai ke stage terakhir.

Langkah aglomerasi ini bersifat lengkap, terutama perhitungan koefisien yang melibatkan obyek lebih banyak lagi. Langkah aglomerasi pada akhirnya akan menyatukan semua obyek menjadi satu cluster. Namun, prosesnya menghasilkan beberapa cluster dengan masing-masing anggotanya, tergantung jumlah cluster yang dibentuk. Perincian jumlah cluster dengan anggota yang terbentuk dapat dilihat pada tabel output berikut ini :

Tabel 3. Cluster Membership

Case	4 Clusters	3 Clusters	2 Clusters
1:DKI Jakarta	1	1	1
2:Jawa Barat	2	2	2
3:Jawa Tengah	3	2	2
4:Jawa Timur	3	2	2
5:Banten	4	3	2
6:Daerah Istimewa Yogyakarta	4	3	2
7:Kalimantan Timur	4	3	2
8:Bali	4	3	2
9:Sumatera Utara	4	3	2
10:Riau	4	3	2
11:Sulawesi Selatan	4	3	2
12:Sumatera Barat	4	3	2
13:Nusa Tenggara Timur	4	3	2
14:Kalimantan Selatan	4	3	2
15:Sumatera Selatan	4	3	2
16:Lampung	4	3	2
17:Kepulauan Riau	4	3	2
18:Kepulauan Bangka Belitung	4	3	2
19:Kalimantan Barat	4	3	2
20:Sulawesi Tengah	4	3	2
21:Kalimantan Tengah	4	3	2
22:Sulawesi Utara	4	3	2
23:Papua	4	3	2
24:Kalimantan Utara	4	3	2
25:Aceh	4	3	2
26:Jambi	4	3	2
27:Nusa Tenggara Barat	4	3	2
28:Papua Barat	4	3	2
29:Bengkulu	4	3	2
30:Sulawesi Tenggara	4	3	2
31:Maluku	4	3	2
32:Sulawesi Barat	4	3	2
33:Maluku Utara	4	3	2
34:Gorontalo	4	3	2

Tabel 3 memperlihatkan bahwa :

- Jika diharapkan pembentukan 4 cluster, maka :
 - Anggota cluster 1 adalah Prov. DKI Jakarta
 - Anggota cluster 2 adalah Prov. Jawa Barat

DAFTAR PUSTAKA

- A.GoreJr, P. (2000) *Handbook of Applied Multivariate Statistics and Mathematical Modeling*. ACADEMIC PRESS. doi: <https://doi.org/10.1016/B978-012691360-6/50012-4>.
- Aabed, K. and Lashin, M. M. A. (2021) 'An analytical study of the factors that influence COVID-19spread', *Saudi Journal of Biological Sciences*, pp. 1177–1195. doi: 10.1016/j.sjbs.2020.11.067.
- Dowd, J. B. *et al.* (2020) 'Demographic science aids in understanding the spread and fatality rates of COVID- 19', *Leverhulme Centre for Demographic Science, Nuffield College, University of Oxford, Oxford OX1 3UQ, United Kingdom*. doi: <https://doi.org/10.1073/pnas.2004911117>.
- Ganasegeran, K. *et al.* (2021) 'Influence of population density for covid-19 spread in malaysia: Anecological study', *International Journal of Environmental Research and Public Health*, 18(18). doi: 10.3390/ijerph18189866.
- Sari, D. N. P. and Sukestiyarno, Y. L. (2021) 'Analisis Cluster dengan Metode K-Means pada PersebaranKasus Covid-19 Berdasarkan Provinsi di Indonesia', *PRISMA, Prosiding Seminar Nasional Matematika*, 4, pp. 602–610. Available at: <https://journal.unnes.ac.id/sju/index.php/prisma/>.
- Yulianto, S. and Hidayatullah, K. H. (2016) 'Analisis Klaster Untuk Pengelompokan Kabupaten/Kota DiProvinsiJawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat', *Statistika*, 2(1), pp. 56–63. Available at: <https://jurnal.unimus.ac.id/index.php/statistik/article/view/1115>.