

**PENERAPAN ANALISIS KLUSTER HIERARKI MENGGUNAKAN
METODE AVERAGE, SINGLE, DAN COMPLETE LINKAGE
PADA DATA PASIEN COVID-19 DI INDONESIA
Studi Kasus : Data IHSG Tahun 2016 – 2021**

Susiana Thaib¹, Rokhana Dwi Bekt^{2*}
Jurusan Statistika, Institut Sains & Teknologi AKPRIND Yogyakarta
Email : rokhana@akprind.ac.id
*corresponding author

ABSTRAK

Corona virus disease 2019 atau covid-19 merupakan virus yang menyebabkan penyakit infeksi saluran pernapasan. Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia menggunakan metode average, single, dan complete linkage pada kasus covid-19 di Indonesia. Analisis Cluster merupakan analisis pengelompokan objek-objek menjadi kelompok-kelompok yang lebih kecil berdasarkan kesamaan karakteristik diantara objek-objek. Analisis ini dilakukan menggunakan software R dan diperoleh hasil analisis 3 kelompok yang terbentuk pada masing-masing metode analisis cluster yakni average linkage, single dan complete linkage. Dari ketiga metode tersebut, ternyata memiliki kesamaan setiap anggota pada setiap kelompok. Kelompok pertama beranggotakan provinsi Jakarta, Jawa Barat, Kalimantan Timur, Yogyakarta, Banten, Riau, Bali, Sulawesi Selatan, Sumatera Utara, Sumatera Barat, Kalimantan Selatan, Nusa Tenggara Timur, Kep. Riau Kep. Bangka Belitung, Sulawesi Tengah, Kalimantan Tengah, Kalimantan Barat, Kalimantan Utara, Sulawesi Utara, Papua, Jambi, Nusa Tenggara Barat, Papua Barat, Bengkulu, Sulawesi Tenggara, Maluku, Sulawesi Barat, Maluku Utara, dan Gorontalo. Kelompok kedua beranggotakan provinsi Jawa Tengah, Sumatera Selatan, dan Aceh. Kelompok ketiga beranggotakan provinsi Jawa Timur dan Lampung. Nilai koefisien korelasi copenetic diperoleh sebesar 0,8869 untuk metode average linkage, 0,8819 untuk metode complete linkage dan single linkage sebesar 0,8651. Sehingga dipilih metode Average karena nilainya lebih tinggi dari kedua metode yang lain. Nilai index Dunn pada penelitian ini lebih baik dibuat 4 kluster karena nilainya lebih tinggi sebesar 0.6107.

Kata Kunci: Analisis Cluster, Covid-19, metode Average, Single dan Complete Linkage.

ABSTRACT

Corona virus disease 2019 or covid-19 is a virus that causes respiratory tract infections. This study aims to group the provinces in Indonesia using the average, single, and complete linkage method for Covid-19 cases in Indonesia. Cluster analysis is an analysis of grouping objects into smaller groups based on the similarity of characteristics between objects. This analysis was carried out using R software and obtained the results of the analysis of 3 groups formed in each cluster analysis method, namely average linkage, single and complete linkage. Of the three methods, it turns out that each member in each group has in common. The first group consisted of the provinces of Jakarta, West Java, East Kalimantan, Yogyakarta, Banten, Riau, Bali, South Sulawesi, North Sumatra, West Sumatra, South Kalimantan, East Nusa Tenggara, Kep. Riau Kep. Bangka Belitung, Central Sulawesi, Central Kalimantan, West Kalimantan, North Kalimantan, North Sulawesi, Papua, Jambi, West Nusa Tenggara, West Papua, Bengkulu, Southeast Sulawesi, Maluku, West Sulawesi, North Maluku and Gorontalo. The second group consists of the provinces of Central Java, South Sumatra, and Aceh. The third group consists of the provinces of East Java and Lampung. The value of the copenetic correlation coefficient is 0.8869 for the average linkage method, 0.8819 for the complete linkage method and 0.8651 for the single linkage method. So the Average method was chosen because its value was higher than the other two methods. The Dunn index value in this study is better made 4 clusters because the value is higher at 0.6107.

Keywords: Cluster Analysis, Covid-19, Average Method, Single and Complete Linkage

1. PENDAHULUAN

Corona virus disease 2019 atau covid-19 merupakan virus yang menyebabkan penyakit infeksi saluran pernapasan. Penyakit ini menyebar diantara orang-orang melalui tetesan pernapasan dari batuk dan bersin. Virus corona tetap dapat bertahan hingga tiga hari dengan plastik dan stainless steel atau dalam aerosol selama tiga jam (Kemendagri, 2020:3). Oleh karena itu, corona virus hanya bisa berpindah melalui perantara dengan media tangan, baju atau yang terkena tetesan batuk dan bersin. Indonesia menjadi salah satu negara positif virus corona atau covid-19 pertama kali dialami oleh dua warga Depok, Jawa Barat. Hal tersebut diumumkan langsung oleh Presiden Joko Widodo di istana kepresidenan, Jakarta 2 Maret 2020. Namun, banyak juga masyarakat yang sudah sembuh dari virus covid-19. Keadaan covid-19 di Indonesia saat ini masih terus bertambah walaupun sudah dilakukan vaksinasi terhadap masyarakat Indonesia. Kita perlu melakukan pengelompokan terhadap kasus covid-19 agar dapat melakukan strategi pencegahan penyebaran rantai virus covid-19 dan kebijakan terhadap setiap provinsi yang memiliki kemiripan kasus covid-19 dirawat, sembuh dan meninggal yang ada di Indonesia.

Analisis cluster merupakan analisis pengelompokan objek atau kasus menjadi kelompok-kelompok yang lebih kecil. Setiap kelompok berisi objek yang mirip dengan objek lainnya (Supranto, 2004). Analisis cluster dibagi menjadi dua metode yaitu metode hierarki dan metode non-hierarki. Metode hierarki dibagi lagi menjadi dua yaitu metode aglomeratif dan metode divisif. Pada metode aglomeratif, masing-masing objek pengamatan dijadikan sebagai kelompok yang memiliki satu anggota setiap kelompok. Sedangkan pada metode divisif, semua objek pengamatan dianggap sebagai satu kelompok kemudian dipisah sampai terbentuk kelompok-kelompok dengan anggota satu. Penggunaan cluster hierarki untuk data yang kecil sedangkan cluster non-hierarki digunakan pada data yang lebih besar. Pada penelitian ini, akan digunakan penelitian analisis cluster Hierarki dengan metode aglomeratif yaitu metode Average, Single dan Complete linkage karena pada penelitian ini hanya menggunakan data yang kecil sehingga metode analisis cluster hierarki adalah metode yang cocok dalam penelitian ini. Penggunaan metode average, Single dan complete linkage akan membantu untuk mengetahui pengelompokan provinsi-provinsi di Indonesia yang memiliki kemiripan berdasarkan jumlah pasien kasus covid-19 yang dirawat, sembuh dan meninggal di Indonesia melalui dendrogram yang dihasilkan dari proses analisis menggunakan software R Studio. Adapun variabel-variabel yang akan dianalisis yaitu proporsi pasien covid-19 yang dirawat X_1 , proporsi pasien covid-19 yang sembuh X_2 , proporsi pasien covid-19 yang meninggal X_3 .

2. METODE PENELITIAN

2.1 Metode dan Objek Penelitian

Metode pada penelitian ini menggunakan metode Average, Single dan Complete linkage. Objek penelitian ini adalah kasus covid-19 di Indonesia.

2.2 Jenis dan Sumber Data

Data yang diperoleh dari skripsi ini merupakan data sekunder. Data yang digunakan adalah data proporsi pasien yang dirawat, proporsi pasien sembuh dan proporsi pasien meninggal akibat corona virus disease 2019 atau covid-19 di Indonesia yang diambil dari website resmi <https://www.antaraneews.com/covid-19>.

2.3 Variabel Penelitian

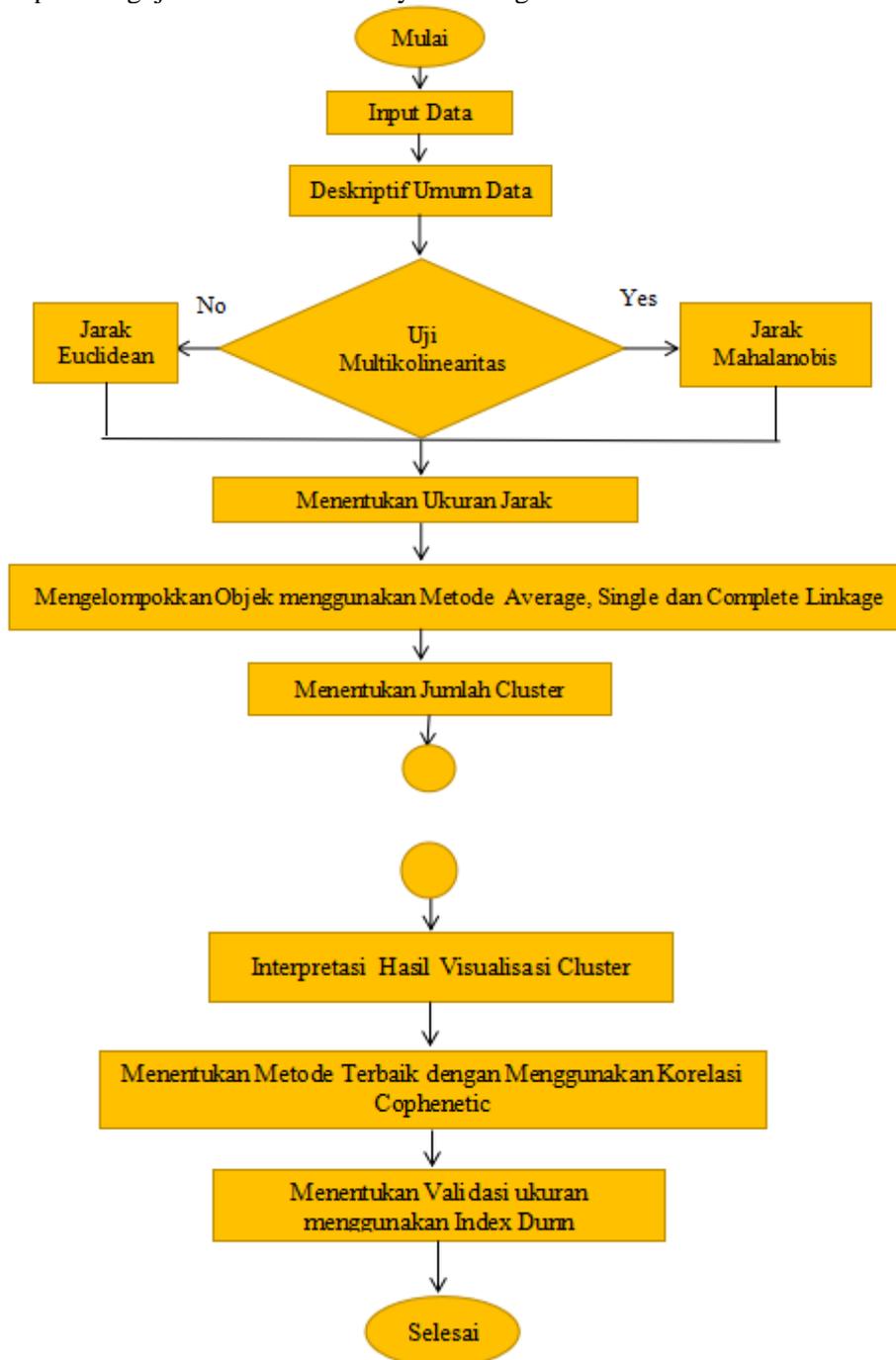
Variabel yang digunakan dalam penelitian ini adalah data proporsi pasien yang dirawat X_1 , proporsi pasien sembuh X_2 , dan proporsi pasien meninggal X_3 akibat corona virus disease 2019 atau covid-19 di Indonesia. Sumber data dari <https://www.antaraneews.com/covid-19>, dengan satuan masing-masing variabel adalah jiwa.

2.4 Metode Analisis Data

Metode analisis data yang digunakan adalah metode Average, Single, dan Complete linkage untuk mengelompokkan kasus covid-19 berdasarkan provinsi di Indonesia yang memiliki kemiripan.

2.5 Cluster Hierarki

Tahapan Pengujian Cluster Hirarki yaitu sebagai berikut:



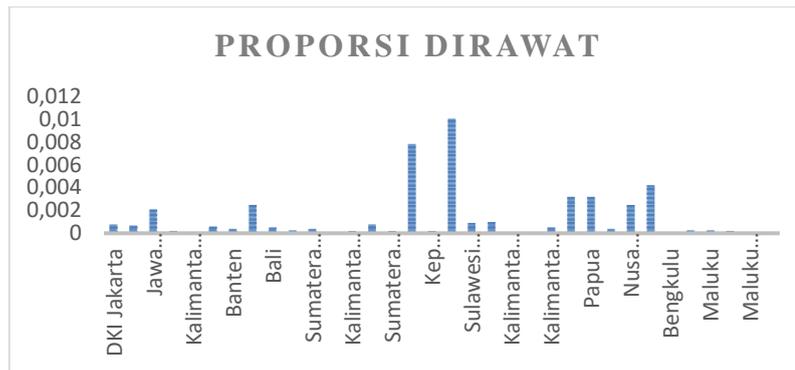
Gambar 2.1 Flowchart Cluster Hierarki

3. HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif

3.1.1 Proporsi Dirawat

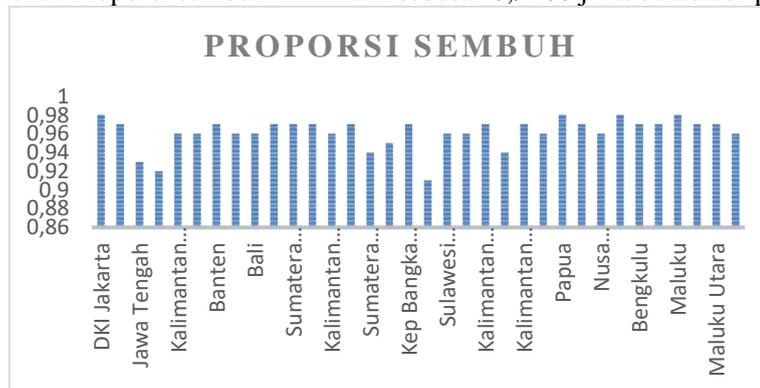
Proporsi dirawat untuk pasien covid-19 di Indonesia sejak 2 Maret 2020 hingga 6 Januari 2022 dapat dilihat pada diagram dibawah ini. Nilai rata-rata 0,0013 jiwa, proporsi dirawat maksimum sebesar 0,0100 jiwa berada di provinsi Lampung, proporsi dirawat minimum sebesar 0,0000 jiwa berada di provinsi Bengkulu, Maluku Utara dan Gorontalo.



Gambar 3.1 Grafik Proporsi Dirawat

3.1.2 Proporsi Sembuh

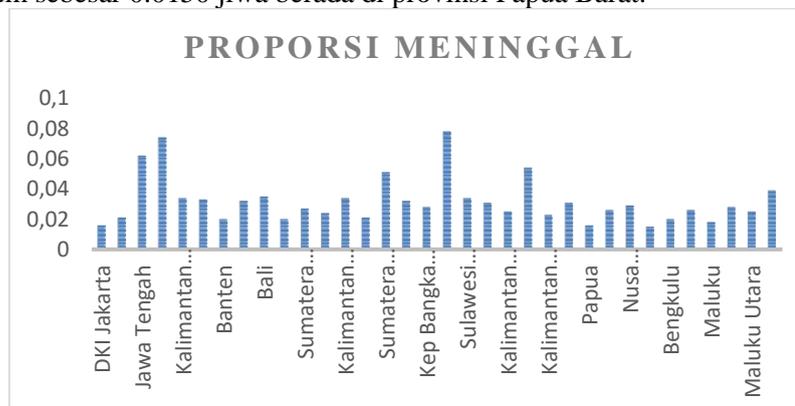
Proporsi sembuh untuk pasien covid-19 di Indonesia sejak 2 Maret 2020 hingga 6 Januari 2022 dapat dilihat pada diagram dibawah ini. Nilai rata-rata 0,9615 jiwa, proporsi sembuh maksimum sebesar 0,9800 jiwa berada di provinsi DKI Jakarta, Papua, Papua Barat dan Maluku. Proporsi sembuh minimum sebesar 0,9100 jiwa berada di provinsi Lampung.



Gambar 3.2 Grafik Proporsi Sembuh

3.1.3 Proporsi Meninggal

Proporsi meninggal untuk pasien covid-19 di Indonesia sejak 2 Maret 2020 hingga 6 Januari 2022 dapat dilihat pada diagram dibawah ini. Nilai rata-rata 0,0318 jiwa, proporsi meninggal maksimum sebesar 0,0780 jiwa berada di provinsi Lampung, proporsi meninggal minimum sebesar 0,0150 jiwa berada di provinsi Papua Barat.



Gambar 3.3 Grafik Proporsi Meninggal

3.2 Uji Multikolinearitas

Adapun hasil uji multikolinearitas berdasarkan hasil analisis pada lampiran 3 sebagai berikut:

$$VIF = \frac{1}{1-R_j^2}$$

Tabel 3.1 Hasil Uji Multikolinearitas

| | Dirawat | Sembuh | Meninggal |
|-----------|---------|---------|-----------|
| Dirawat | 1,0000 | -0,4235 | 0,3208 |
| Sembuh | -0,4235 | 1,0000 | -0,9783 |
| Meninggal | 0,3208 | -0,9783 | 1,0000 |

Dari hasil uji multikolinearitas diatas, diketahui bahwa variabel-variabel bebas mengalami multikolinearitas karena ada nilai koefisien korelasi lebih dari 0.8, sehingga akan dilakukan penentuan jarak Mahalanobis.

3.3 Menentukan Ukuran Jarak Mahalanobis

Berdasarkan hasil analisis, didapatkan jarak Mahalanobis yang terdapat pada lampiran. Penentuan jarak pada analisis ini dengan menggunakan jarak Mahalanobis seperti dibawah ini:

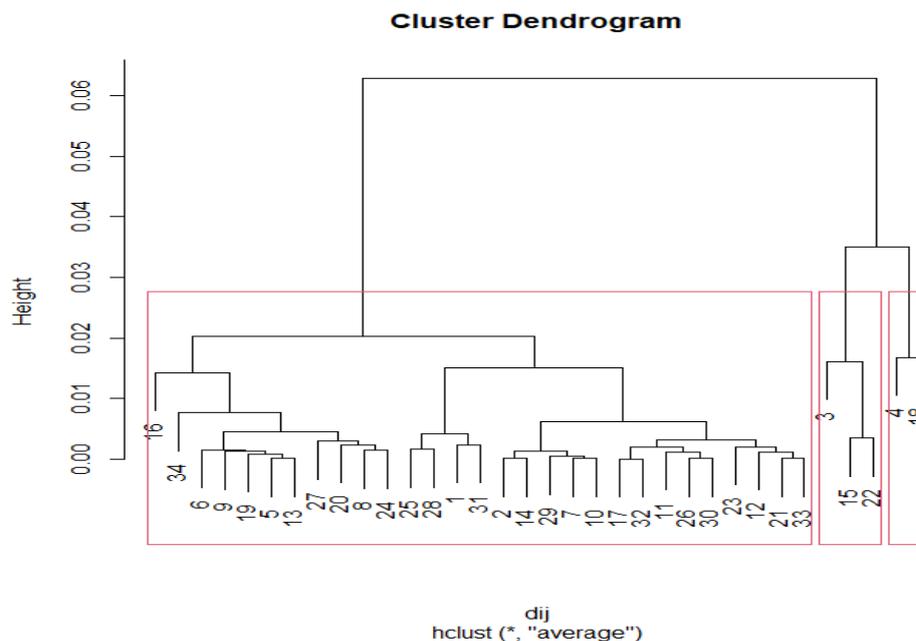
$$d_{ijm} = \sqrt{(x_i - x_j)^T S^{-1} (x_i - x_j)}$$

Sehingga diperoleh matriks sebagai berikut:

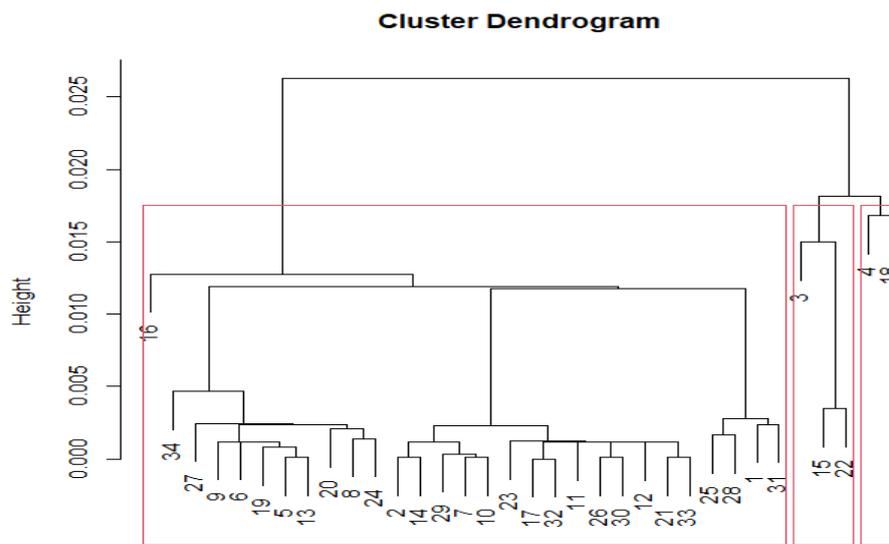
$$d_{ij} = \begin{matrix} & \dots & 16 & 17 & \dots & 31 & 32 & \dots & 34 \\ \begin{matrix} \vdots \\ 16 \\ 17 \\ \vdots \\ 31 \\ 32 \\ \vdots \\ 34 \end{matrix} & \begin{pmatrix} \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ \vdots & 0 & 0.0251 & \vdots & 0.0392 & 0.0251 & \dots & 0.0679 \\ \vdots & 0.0251 & 0 & \vdots & 0.0163 & 0.0000 & \dots & 0.0049 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ \vdots & 0.0392 & 0.0163 & \vdots & 0 & 0.0163 & \dots & 0.0335 \\ \vdots & 0.0251 & 0.0000 & \vdots & 0.0163 & 0 & \dots & 0.0172 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & 0.0679 & 0.0049 & \dots & 0.0335 & 0.0172 & \dots & 0 \end{pmatrix} \end{matrix}$$

3.4 Analisis Cluster Hierarki dengan Metode Average Linkage

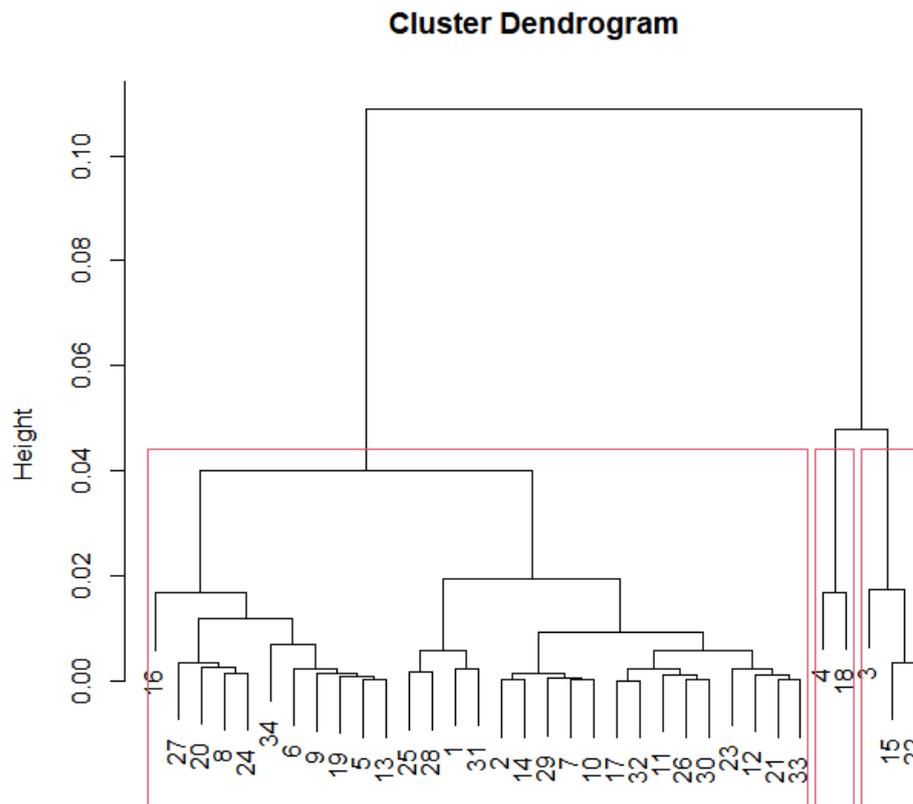
Dengan bantuan software R, didapatkan hasil 3 kelompok provinsi di Indonesia yang memiliki kemiripan menggunakan metode average, single dan complete linkage pada data covid-19 di Indonesia sebagai berikut:



Gambar 3.4 Dendrogram dari metode Average linkage



Gambar 3.5 Dendrogram dari metode Single linkage



Gambar 3.6 Dendrogram dari metode Complete linkage

Dari ketiga dendrogram diatas, diketahui bahwa anggota setiap kelompok pada setiap metode adalah sama.

Kluster 1: Kluster pertama terdiri dari 29 provinsi yang ada di Indonesia termasuk dalam

kategori rendah. Provinsi-provinsi tersebut adalah Jakarta, Jawa Barat, Kalimantan Timur, Yogyakarta, Banten, Riau, Bali, Sulawesi Selatan, Sumatera Utara, Sumatera Barat, Kalimantan Selatan, Nusa Tenggara Timur, Kep. Riau Kep. Bangka Belitung, Sulawesi Tengah, Kalimantan Tengah, Kalimantan Barat, Kalimantan Utara, Sulawesi Utara, Papua, Jambi, Nusa Tenggara Barat, Papua Barat, Bengkulu, Sulawesi Tenggara, Maluku, Sulawesi Barat, Maluku Utara, dan Gorontalo.

Kluster 2: Kluster kedua terdiri dari 3 provinsi yang ada di Indonesia termasuk dalam kategori sedang. Provinsi-provinsi tersebut adalah Jawa Tengah, Sumatera Selatan, dan Aceh.

Kluster 3: Kluster ketiga terdiri dari 2 provinsi yang ada di Indonesia termasuk dalam kategori tinggi. Provinsi-provinsi tersebut adalah Jawa Timur, dan Lampung.

3.5 Menentukan Metode Terbaik Berdasarkan Korelasi Cophenetic

Berikut ini adalah nilai korelasi cophenetic dari ketiga metode average, single dan complete linkage:

Tabel 3.2 Tabel Hasil Korelasi Cophenetic

| Metode Cluster | Nilai Korelasi Cophenetic |
|------------------------|---------------------------|
| Average Linkage | 0,8869 |
| Single Linkage | 0,8651 |
| Complete Linkage | 0,8819 |

Berdasarkan nilai koefisien korelasi cophenetic diatas diketahui semua metode baik dalam penelitian ini karena sudah lebih dari 0.80 atau mendekati 1. Namun, akan dipilih satu yang paling tinggi diantaranya yaitu metode Average linkage dengan nilai 0,8869 lebih tinggi dari nilai koefisien korelasi cophenetic dari metode Complete linkage dan Single linkage dengan nilai 0,8819 dan 0,8651.

3.6 Index Dunn

Berikut ini adalah nilai index Dunn dari metode Average linkage:

Tabel 3.3 Hasil Index Dunn

| Jumlah Cluster Average Linkage | Index Dunn |
|--------------------------------|---------------|
| 2 Kluster | 0.5484 |
| 3 Kluster | 0.4533 |
| 4 Kluster | 0.6107 |

Berdasarkan hasil analisis dan tabel index Dunn diatas maka diketahui nilai index Dunn dari metode average linkage dengan jumlah 2 kluster sebesar 0.5484, dengan 3 kluster sebesar 0.4533 dan 4 kluster sebesar 0.6107. Sehingga disimpulkan bahwa penelitian ini lebih baik jika 4 kluster karena nilai index Dunnnya lebih besar.

3.7 Interpretasi Hasil

Dari analisis diatas, maka dapat diinterpretasikan sebagai berikut:

Tabel 4.4 Profiling Kluster

| Variabel | Kluster | | |
|--------------------|---------|--------|--------|
| | 1 | 2 | 3 |
| Proporsi Dirawat | 0.0011 | 0.0008 | 0.0051 |
| Proporsi Sembuh | 0.9672 | 0.9367 | 0.9150 |
| Proporsi Meninggal | 0.0263 | 0.0556 | 0.0760 |
| Rata-rata | 0.3315 | 0.3310 | 0.3320 |

Keterangan:

Merah : Proporsi kasus Covid-19 Tinggi

Kuning : Proporsi kasus Covid-19 Sedang

Hijau : Proporsi kasus Covid-19 Rendah

Kluster pertama: Beberapa provinsi yang terdapat pada kluster pertama dengan nilai rata-rata 0.3315 memiliki kasus proporsi dirawat cenderung sedang, dengan kasus proporsi sembuh yang tinggi dan kasus proporsi meninggal yang rendah.

Kluster kedua: Terdapat 3 provinsi yang ada pada kluster kedua yakni Jawa Tengah, Sumatera Selatan dan Aceh dengan nilai rata-rata 0.3310 memiliki kasus proporsi dirawat yang rendah, dengan kasus proporsi sembuh dan kasus proporsi meninggal yang sedang.

Kluster ketiga: Terdapat 2 provinsi yang ada pada kluster ketiga yakni, Jawa Timur dan Lampung dengan nilai rata-rata 0.3320 memiliki kasus proporsi dirawat yang tinggi, dengan kasus proporsi sembuh yang rendah dan kasus proporsi meninggal yang tinggi.

4. KESIMPULAN

Berdasarkan hasil dan pembahasan dapat disimpulkan sebagai berikut:

1. Sejak awal pandemi di Indonesia 2 Maret 2020 hingga 6 Januari 2022, proporsi pasien dirawat terbanyak atau yang tinggi terdapat di provinsi Lampung sebesar 0.0100. Proporsi pasien dirawat paling sedikit atau rendah terdapat di provinsi Bengkulu, Maluku Utara dan Gorontalo sebesar 0.0000. Proporsi pasien sembuh terbanyak atau yang tinggi terdapat pada provinsi DKI Jakarta, Papua, Papua Barat dan Maluku sebesar 0.9800. Proporsi pasien sembuh paling sedikit atau rendah terdapat di provinsi Lampung sebesar 0.9100. Proporsi pasien meninggal terbanyak atau yang tinggi terdapat pada provinsi Lampung sebesar 0.0780. Proporsi pasien meninggal paling sedikit atau rendah terdapat di provinsi Papua Barat sebesar 0.0150.
2. Hasil pengelompokan provinsi di Indonesia pada kasus covid-19 untuk setiap metode adalah sama. Kelompok pertama terdiri dari provinsi Jakarta, Jawa Barat, Kalimantan Timur, Yogyakarta, Banten, Riau, Bali, Sulawesi Selatan, Sumatera Utara, Sumatera Barat, Kalimantan Selatan, Nusa Tenggara Timur, Kep. Riau, Kep. Bangka Belitung, Sulawesi Tengah, Kalimantan Tengah, Kalimantan Barat, Kalimantan Utara, Sulawesi Utara, Papua, Jambi, Nusa Tenggara Barat, Papua Barat, Bengkulu, Sulawesi Tenggara, Maluku, Sulawesi Barat, Maluku Utara, dan Gorontalo. Kelompok kedua terdiri dari 3 provinsi yaitu provinsi Jawa Tengah, Sumatera Selatan, dan Aceh. Kelompok ketiga terdiri dari 2 provinsi yaitu provinsi Jawa Timur dan Lampung.
3. Ketiga metode termasuk metode terbaik dalam penelitian ini karena nilai yang diperoleh melebihi 0,80 atau sudah mendekati 1. Namun, akan dipilih satu metode yaitu metode average linkage karena nilai korelasi copheneticnya lebih tinggi dari kedua metode yang lain, yaitu 0,8869.

DAFTAR PUSTAKA

- Afandi, M.I. (2020). *Analisis Cluster Hierarki dengan Metode Complete Linkage pada Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indikator Kemiskinan*. (Skripsi). Universitas Islam Negeri Maulana Malik Ibrahim Malang.
- Afriansah, D. (2020). *Perbandingan Analisis Cluster Metode Single Linkage, Average Linkage dan Ward's untuk Pengelompokan Kabupaten di Jawa Tengah Berdasarkan Hasil Produksi Tanaman Pangan*. (Skripsi). Institut Sains & Teknologi AKPRIND Yogyakarta.
- Antaraneews.com. (2022). *Covid19.go.id*. Akses pada 6 Januari 2022, dari <https://www.antaraneews.com/covid-19>.
- Dewanti. (2013). *Perbandingan Metode Cluster Validity pada Jenis Data Numerik dan Kategori*. (Skripsi). Bogor : Institut Pertanian Bogor.
- Ghaisani, S.Y., dkk. (2016). *Analisis Cluster Hirarki untuk Pengelompokan Provinsi di*

Indonesia Berdasarkan Indikator Demokrasi Indonesia Tahun 2016. *Konferensi Nasional Penelitian Matematika dan Pembelajarannya (KNPMP) IV Universitas Muhammadiyah Surakarta*.

Malikhatin, H., Rusgiyono, A., & Maruddani, DAI. (2021). Penerapan K-Modes Clustering dengan Validasi Dunn Index pada Pengelompokan Karakteristik Calon TKI Menggunakan R-Gui. *Jurnal Gaussian*, 10, 359-366.

Matdoan, M.Y & Delsen, M.S.N.V. (2020). Penerapan Analisis Cluster dengan Metode Hierarki untuk Klasifikasi Kabupaten/Kota di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia. *Jurnal Statistika dan Matematika*, 2, 123-130.