

ANALISIS KLASTER KASUS AKTIF COVID-19 MENURUT PROVINSI DI INDONESIA BERDASARKAN DATA DERET WAKTU

Raditya Novidianto¹, Andrea Tri Rian Dani²

^{1,2} Pascasarjana Departemen Statistika, Fakultas Sains dan Analitika Data, Institut Teknologi Sepuluh Nopember
e-mail: ¹radit@bps.go.id

Abstrak

Corona Virus Disease 2019 (COVID-19) merupakan masalah yang sangat serius yang dihadapi oleh negara-negara di dunia, lebih dari 240 negara terjangkit virus ini. Pada 11 Maret 2020 WHO mengumumkan COVID-19 sebagai pandemi. Saat ini, penyebaran wabah COVID-19 terus bergerak signifikan, khususnya di Indonesia. Sejak mulai diumumkan pada awal Maret lalu hingga menjelang pertengahan Juli, jumlah kasus positif COVID-19 sudah mencapai 80.094 pasien yang dinyatakan positif, terkonfirmasi 3.797 kasus kematian dan 39.050 pasien yang dinyatakan sembuh. Salah satu kesulitan yang dialami pemerintah dalam penanganan COVID-19 yaitu tingkat kedaruratan dan kebijakan yang diterapkan oleh pemerintah daerah. Setiap daerah memiliki karakteristik yang berbeda-beda sehingga diperlukan pengetahuan mengenai kesamaan karakteristik daerah dalam penanganannya berdasarkan kasus COVID-19 yang berkelanjutan setiap harinya. Oleh karena itu, tujuan dari penelitian ini adalah menganalisis secara deskriptif mengenai kasus aktif COVID-19 berdasarkan data deret waktu dari setiap Provinsi di Indonesia. Selanjutnya melakukan proses pengelompokan menggunakan data kasus aktif COVID-19 di Indonesia. Proses pengelompokan menggunakan metode *agglomerative hierarchical clustering*, yaitu algoritma *single*, *complete* dan *average linkage*. Pengukuran kemiripan menggunakan *Euclidean Distance* dan *Dynamic Time Warping (DTW)*. Berdasarkan hasil analisis, dengan menggunakan ukuran kebaikan yaitu koefisien korelasi *cophenetic* menunjukkan bahwa pengukuran kemiripan yang terbaik dari ketiga algoritma yang digunakan adalah *Euclidean Distance*. Dendogram yang didapat dari hasil pengelompokan menunjukkan bahwa dengan ketiga algoritma yang digunakan menghasilkan anggota pengelompokan yang sama. Pentingnya informasi tentang hasil pengelompokan ini dapat membantu pemerintah pusat dan daerah untuk membuat strategi pencegahan penyebaran rantai virus COVID-19.

Kata kunci: *Agglomerative Hierarchical Clustering*, COVID-19, Koefisien Korelasi *Cophenetic*

Abstract

Corona Virus Disease 2019 (COVID-19) is a very serious problem faced by countries in the world, more than 240 countries have contracted this virus. On March 11, 2020 WHO declared COVID-19 a pandemic. Currently, the spread of the COVID-19 outbreak continues to move significantly, especially in Indonesia. Since it was announced in early March until mid-July, the number of positive cases of COVID-19 has reached 80.094 patients who have tested positive, 3.797 confirmed cases of death and 39.050 patients who have recovered. One of the difficulties experienced by the government in handling COVID-19 is the level of emergency and policies implemented by local governments. Each region has different characteristics so that knowledge is needed about the similarity of regional characteristics in handling it based on cases of COVID-19 that are continuous every day. Therefore, the aim of this study is to analyze descriptively the active cases of COVID-19 based on time series data from each province in Indonesia. Then carry out the grouping process using data on active COVID-19 cases in Indonesia. The grouping process uses the *agglomerative hierarchical clustering* method, namely the *single*, *complete* and *average linkage* algorithm. Similarity measurement uses *Euclidean Distance* and *Dynamic Time Warping (DTW)*. Based on the results of the analysis, using a measure of goodness, namely the *cophenetic correlation coefficient*, shows that the best similarity measurement of the three algorithms used is *Euclidean Distance*. The dendogram obtained from the grouping results shows that the three algorithms used produce the same grouping members. The importance of information about the results of this grouping can help central and local governments to create strategies to prevent the spread of the COVID-19 virus chain.

Keywords: *Agglomerative Hierarchical Clustering*, COVID-19, *Cophenetic Correlation Coefficient*

PENDAHULUAN

Pada Desember 2019, kasus *Pneumonia* misterius pertama kali dilaporkan di Wuhan, Provinsi Hubei, China. Saat itu sumber penularan kasus masih belum diketahui dengan pasti, akan tetapi munculnya masalah pertama dihubungkan dengan hewan kelelawar yang menular ke hewan lain sebelum menjangkiti manusia. Mulai akhir bulan Desember 2019 hingga awal Januari 2020 kasus ini terus meningkat. Tidak sampai satu bulan, penyakit ini telah menyebar di berbagai belahan dunia. WHO (*World Health Organization*) mengumumkan nama baru virus ini pada 11 Februari 2020 yaitu *Corona Virus Disease 2019* (COVID-19).

COVID-19 ini dapat ditularkan dari manusia ke manusia dan saat ini telah menyebar secara luas di China dan lebih dari 240 Negara telah terjangkit jenis virus ini. Pada 11 Maret 2020 WHO mengeluarkan pemberitahuan bahwa COVID-19 ditetapkan sebagai pandemi. Hal tersebut karena COVID-19 memenuhi tiga kriteria yaitu jenis virus baru, dapat menginfeksi banyak orang dengan mudah, serta bisa menginfeksi antar manusia secara luas. Hingga tanggal 15 Juli 2020, terdapat 13.285.640 orang positif COVID-19 dan terkonfirmasi 578.110 kasus kematian (<https://covid19.go.id/>). Saat ini, penyebaran wabah COVID-19 terus menerus bergerak signifikan di Indonesia. Sejak mulai diumumkan pada awal Maret lalu hingga menjelang pertengahan Juli, jumlah kasus positif COVID-19 sudah mencapai 80.094 pasien yang dinyatakan positif, terkonfirmasi 3.797 kasus kematian dan 39.050 pasien yang dinyatakan sembuh (<https://covid19.go.id/>). Pengendalian COVID-19 ini sangat tergantung dari kewaspadaan masyarakat, kegesitan dan kesediaan sarana dan prasarana kesehatan dari manajemen pemerintah pusat maupun daerah dalam menjaga kesehatan di masyarakat serta tenaga medis dalam merawat pasien penderita COVID-19.

Analisis kluster adalah sekumpulan metode yang digunakan dengan tujuan mengelompokkan objek ke dalam sebuah

kluster berdasarkan kemiripan karakteristik yang ditemukan pada data, sehingga objek yang memiliki karakteristik yang hampir sama akan bergabung dalam satu kelompok. Analisis kluster dapat diterapkan pada data deret waktu, di mana terdapat prosedur dan algoritma pengelompokan yang sedikit berbeda apabila dibandingkan dengan pengelompokan menggunakan data *cross-section* (Dani, dkk., 2019).

Data deret waktu adalah sekumpulan observasi yang terjadi, dimana dikumpulkan berdasarkan indeks waktu secara berurutan dengan *space time* yang konstan (Aswi & Sukarna, 2006). Kluster persebaran virus COVID-19 dapat dilakukan dengan melihat dinamika data harian kasus aktif yang disajikan dalam situs resmi (<https://sinta.ristekbrin.go.id/covid/datasets>). Data yang tersedia merupakan data deret waktu yang mencatat kasus aktif tiap harinya untuk ke 34 Provinsi di Indonesia. Juru bicara tim penanganan COVID-19 dari pemerintah menyampaikan pencatatan data menjadi hal mendasar dalam pengelolaan pandemi COVID-19. Data yang dipublikasikan merupakan data kasus COVID-19, yang telah terkonfirmasi dan berasal dari hasil identifikasi serta pemeriksaan di laboratorium.

Berdasarkan latar belakang di atas, peneliti tertarik untuk membahas mengenai analisis kluster pada proses pengelompokan Provinsi di Indonesia berdasarkan kasus aktif COVID-19 menggunakan data deret waktu. Pada penelitian ini, digunakan 2 jarak pengukuran kemiripan yaitu *Euclidean Distance* dan *Dynamic Time Warping* (DTW). Metode pengelompokan yang diaplikasikan adalah pengelompokan hierarki menggunakan algoritma *single, complete* dan *average linkage*. Validitas pengukuran kemiripan yang digunakan adalah koefisien korelasi *cophenetic*. Tujuan dari penelitian ini adalah memperoleh hasil pengelompokan yang optimal pada proses pengelompokan Provinsi di Indonesia berdasarkan kasus aktif COVID-19. Pentingnya informasi tentang hasil kluster ini dapat membantu

pemerintah pusat dan daerah untuk membuat strategi pencegahan penyebaran rantai virus COVID-19.

METODE

Analisis Klaster

Supranto (2010) menjelaskan analisis klaster adalah pengelompokan data ke dalam sejumlah klaster berdasarkan kemiripan karakteristik dari data pada kelompok yang tersedia. Data atau objek yang masuk dalam batas kemiripan dengan kelompoknya akan bergabung menjadi satu kelompok, dan kemudian akan terpisah dalam kelompok yang berbeda jika keluar dari batas kemiripan dengan kelompok tersebut. Santosa (2007) menjelaskan bahwa klaster yang baik adalah klaster yang mempunyai homogenitas (kesamaan) yang tinggi antar anggota dalam satu klaster dan heterogenitas (keberagaman) yang tinggi antar klaster yang satu dengan klaster yang lainnya.

Pengukuran Kemiripan

Pada dasarnya proses pembentukan klaster yaitu mencari dan mengelompokkan obyek-obyek berdasarkan kemiripan dan kedekatan antar obyek yang satu dengan obyek lainnya. Kemiripan antar objek diukur dengan menggunakan ukuran jarak (*distance*). Pada penelitian ini, pengukuran kemiripan yang digunakan dalam proses pengelompokan adalah:

a. Euclidean Distance

Menurut Johnson & Wichern (2002), *Euclidean Distance* merupakan tipe pengukuran jarak dalam analisis klaster yang paling umum digunakan. *Euclidean Distance* merupakan jarak geometris antar dua objek data. *Euclidean Distance* dapat diperoleh dengan menggunakan Persamaan (1) sebagai berikut:

$$d_{EUCL} = \sqrt{\sum_{t=1}^n (x_t - y_t)^2} \quad (1)$$

dengan:

d_{EUCL} : *Euclidean Distance*

x_t : data x pada waktu ke- t

y_t : data y pada waktu ke- t

2. Dynamic Time Wrapping (DTW)

Dynamic Time Warping (DTW) adalah algoritma untuk menghitung *warping path* yang optimal antara dua data deret waktu sehingga nantinya *output* nya adalah nilai-nilai *warping path* dan jarak di antara kedua data deret waktu tersebut. Algoritma DTW dapat digunakan untuk mengukur kesamaan dua data deret waktu dengan jumlah data yang berbeda. Algoritma ini diperkenalkan pertama kali oleh Sakoe dan Chiba pada Tahun 1978.

Bayangkan apabila terdapat dua data deret waktu dengan panjang yang berbeda yaitu $x_t = x_1, x_2, \dots, x_i, \dots, x_n$ dengan

$y_t = y_1, y_2, \dots, y_j, \dots, y_m$. Langkah pertama adalah membuat matriks \mathbf{D} yang merupakan matriks jarak berukuran $n \times m$. Elemen ke-(i, j) dalam matriks \mathbf{D} dijelaskan sebagai selisih antara x_i dengan y_j yang selanjutnya ditambah dengan nilai terkecil dari tiga elemen yang berdekatan $\{d_{(i-1)(j-1)}, d_{(i-1),j}, d_{i(j-1)}\}$, di mana

$0 < i \leq m$ dan $0 < j \leq n$. Elemen ke-(i, j) dalam matriks \mathbf{D} dapat ditulis menjadi $d_{ij} = w_{ij} + \min\{d_{(i-1)(j-1)}, d_{(i-1),j}, d_{i(j-1)}\}$ (2)

Dalam hal ini nilai w_{ij} merupakan selisih antara x_i terhadap y_j dengan perhitungan dapat dituliskan pada Persamaan (3).

$$w_{ij} = |x_i - y_j| \quad (3)$$

Berdasarkan Persamaan (2) dan (3), maka jarak DTW antara data deret waktu x_i terhadap y_j dapat didefinisikan

$$d_{DTW} = \min_{\forall w \in P} \left\{ \sqrt{\sum_{i,j=1}^K d_{ij}} \right\} \quad (4)$$

di mana P merupakan sekumpulan dari keseluruhan *warping path* yang mungkin, d_{ij} adalah elemen (i, j) pada *warping path* dan K adalah panjang dari *warping path* (Montero & Vilar, 2014).

Prosedur Analisis Klaster

Dalam proses pengelompokannya, metode *Agglomerative Hierarchical Clustering* menggunakan beberapa fungsi

penghubung atau algoritma pengelompokan. Dani, dkk. (2019) dalam penelitiannya menjelaskan sebagai berikut:

Single linkages dimulai dengan memilih jarak terkecil (minimum) dalam matriks jarak, kemudian menggabungkan objek-objek yang bersesuaian misalnya A dan B untuk mendapatkan kluster (AB). Langkah selanjutnya adalah mencari jarak antara (AB) dengan kluster lainnya, misalnya C sehingga dapat dituliskan sebagai berikut:

$$d_{(AB)C} = \min(d_{AC}, d_{BC}) \quad (5)$$

Complete linkages diawali dengan memilih jarak terbesar (maksimum) dalam matriks jarak, kemudian menggabungkan objek-objek yang bersesuaian misalnya A dan B untuk mendapatkan kluster (AB). Langkah selanjutnya adalah mencari jarak antara (AB) dengan kluster lainnya, misalnya C sehingga dapat dituliskan sebagai berikut:

$$d_{(AB)C} = \max(d_{AC}, d_{BC}) \quad (6)$$

Average linkages adalah metode pengelompokan yang didasarkan pada jarak rata-rata antar objek. Algoritma *average linkages* dimulai dengan mendefinisikan matriks jarak untuk memperoleh objek-objek paling dekat, sebagai contoh A dan B, kemudian objek ini digabung ke dalam bentuk kluster (AB) dan selanjutnya jarak antara (AB) dengan kluster lainnya C, sehingga dapat dituliskan sebagai berikut:

$$d_{(AB)C} = \frac{d_{AC} + d_{BC}}{n_{AB}n_C} \quad (7)$$

Validitas Pengukuran Kemiripan

Uji validitas jarak dilakukan untuk melihat ketepatan (*goodness*) dan kualitas (*quality*) dari hasil analisis kluster. Penelitian ini akan mengaplikasikan koefisien korelasi *cophenetic* sebagai ukuran yang diterapkan untuk menguji validitas pengukuran kemiripan. Koefisien korelasi *cophenetic* adalah suatu koefisien korelasi yang diukur dari elemen-elemen asli matriks ketidakmiripan (*dissimilarity distance*) dengan elemen-elemen yang dihasilkan oleh dendrogram. Formulasi yang

digunakan untuk menghitung koefisien korelasi *cophenetic* sebagai berikut:

$$r_{coph} = \frac{\sum_{i < j}^n (d_{ij} - \bar{d})(d_{coph-ij} - \bar{d}_{coph})}{\sqrt{\left[\sum_{i < j}^n (d_{ij} - \bar{d})^2 \right] \left[\sum_{i < j}^n (d_{coph-ij} - \bar{d}_{coph})^2 \right]}} \quad (8)$$

dengan:

r_{coph} : koefisien korelasi *cophenetic*

d_{ij} : jarak asli antara objek ke- i dan ke- j

\bar{d} : rata-rata d_{ij}

$d_{coph-ij}$: jarak *cophenetic* objek ke- i dan ke- j

\bar{d}_{coph} : rata-rata $d_{coph-ij}$.

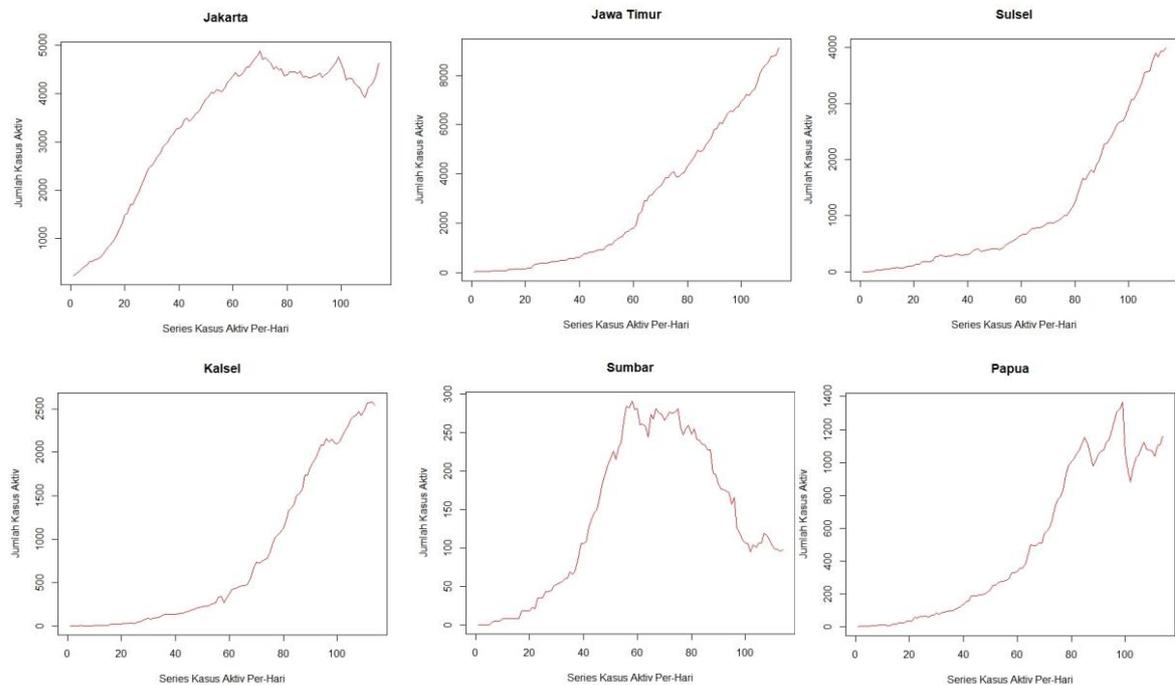
Nilai r_{coph} berkisar antara -1 dan 1, nilai r_{coph} mendekati 1 berarti jarak yang digunakan dalam proses *clustering* cukup baik (Saracli, dkk., 2013).

HASIL DAN PEMBAHASAN

Statistika Deskriptif

Data kasus aktif COVID-19 yang digerombolkan terdiri dari 3.876 Kasus berdasarkan 34 Provinsi di Indonesia, dari tanggal 21 Maret 2020 hingga tanggal 12 Juli 2020. Selanjutnya diperlukan gambaran pergerakan data kasus aktif COVID-19 yang dihasilkan secara berkelanjutan untuk mengetahui karakteristik data kasus aktif COVID-19 pada suatu Provinsi. Kepala daerah memiliki hak dalam mengambil dan memutuskan kebijakan yang diterapkan di daerahnya masing-masing dengan persetujuan serta tetap memperhatikan kebijakan pemerintah pusat sedemikian sehingga tetap mengedepankan sinergitas dalam penanganan COVID-19.

Adapun pemerintah kemudian menerbitkan suatu kebijakan yaitu PSBB guna penyelesaian rantai penyebaran COVID-19 yang kemudian tertuang dalam Peraturan Pemerintah (PP) Nomor 21 Tahun 2020 dan Keputusan Presiden Nomor 11 Tahun 2020. Saat ini, beberapa daerah telah menerapkan PSBB di wilayahnya dengan membatasi aktivitas masyarakat, mulai dari sekolah, bekerja dan beribadah yang dimana disarankan



Gambar 1. Grafik Deret Waktu Kasus Aktif COVID-19

dilakukan di rumah masing-masing serta melarang masyarakat berkerumun.

Sebelum dilakukan proses pengelompokkan, akan dilakukan analisis secara deskriptif untuk melihat gambaran dari perubahan, pola serta fluktuasi dari data. Dengan melihat dinamika data harian kasus aktif yang disajikan dalam situs resmi (<https://sinta.ristekbrin.go.id/covid/datasets>) maka ditampilkan grafik deret waktu dari kasus aktif tiap harinya beberapa Provinsi di Indonesia.

Gambar 1 merupakan grafik deret waktu dari 6 Provinsi di Indonesia yang digunakan sebagai contoh trend kasus aktif COVID-19 yang cenderung meningkat setiap harinya. Secara umum data jumlah kasus aktif COVID-19 memiliki rentang nilai yang relatif jauh jika dibandingkan antar daerah untuk setiap pulau di Indonesia. Sebagai contoh Pulau Jawa cenderung memiliki trend yang cukup tinggi dibandingkan pulau-pulau lainnya di Indonesia. Provinsi Sumatera Selatan memiliki rata-rata kasus aktif COVID-19 tertinggi di Pulau Sumatera yaitu sebesar 456,7 kasus. Provinsi Bali memiliki rata-rata kasus aktif tertinggi di Pulau Bali yaitu sebesar 222,6 kasus. Provinsi Kalimantan Selatan memiliki rata-rata kasus aktif

tertinggi di Pulau Kalimantan yaitu sebesar 790,25 kasus. Provinsi Sulawesi Selatan memiliki rata-rata kasus aktif tertinggi di Pulau Sulawesi yaitu sebesar 1.763 kasus. Provinsi Papua memiliki rata-rata kasus aktif tertinggi di Kepulauan Maluku dan Papua yaitu sebesar 494,2 kasus. Pola persebaran kasus aktif COVID-19 antar Provinsi memiliki pola yang berbeda-beda. Diduga dipengaruhi dari pola mobilitas penduduk setiap daerah antar Provinsi atau dalam daerah di Provinsi tersebut.

Proses Pengelompokkan dengan *Time Series Based Clustering*

Analisis kluster menggunakan data deret waktu dalam penelitian ini bertujuan untuk mengelompokkan Provinsi di Indonesia berdasarkan kasus aktif COVID-19. Untuk mendapatkan hasil tersebut, terlebih dahulu dilakukan pengukuran kemiripan. Ukuran jarak yang digunakan adalah ukuran *Euclidean Distance* dan *Dynamic Time Warping (DTW)*. Selanjutnya dilakukan proses pengelompokkan menggunakan beberapa fungsi penghubung atau algoritma sebagai berikut:

1. *Single Linkage*

Proses pengelompokkan pada tahapan ini menggunakan algoritma *single linkage*. Nilai koefisien korelasi *cophenetic* pada masing-masing ukuran jarak ditampilkan sebagai berikut:

Tabel 1. Koefisien korelasi *cophenetic* berdasarkan jarak pengukuran kemiripan dengan algoritma *single linkage*

No	Jarak	Korelasi <i>Cophenetic</i>
1	<i>Euclidean Distance</i>	0,9757
2	<i>Dynamic Time Warping (DTW)</i>	0,8736

Berdasarkan Tabel 1 terlihat *Euclidean Distance* menghasilkan nilai korelasi *cophenetic* terbesar yaitu 0,9757. Berdasarkan hasil pengukuran dari koefisien korelasi *cophenetic* menggunakan algoritma *single linkage*, didapat bahwa jarak pengukuran kemiripan yang optimal dalam mengelompokkan Provinsi di Indonesia berdasarkan kasus aktif COVID-19 adalah *Euclidean Distance*.

2. Complete Linkage

Proses pengelompokkan pada tahapan ini menggunakan algoritma *complete linkage*. Nilai koefisien korelasi *cophenetic* pada masing-masing ukuran jarak ditampilkan sebagai berikut:

Tabel 2. Koefisien korelasi *cophenetic* berdasarkan jarak pengukuran kemiripan dengan algoritma *complete linkage*

No	Jarak	Korelasi <i>Cophenetic</i>
1	<i>Euclidean Distance</i>	0,9816
2	<i>Dynamic Time Warping (DTW)</i>	0,9778

Berdasarkan Tabel 2 didapatkan bahwa *Euclidean Distance* menghasilkan nilai korelasi *cophenetic* terbesar yaitu 0,9816. Dari hasil pengukuran dari koefisien korelasi *cophenetic* menggunakan algoritma *complete linkage*, didapat bahwa jarak pengukuran kemiripan yang optimal dalam mengelompokkan Provinsi di Indonesia berdasarkan kasus aktif COVID-19 adalah *Euclidean Distance*.

3. Average Linkage

Proses pengelompokkan pada tahapan ini menggunakan algoritma *average linkage*. Nilai koefisien korelasi *cophenetic* pada masing-masing ukuran jarak ditampilkan sebagai berikut:

Tabel 3. Koefisien korelasi *cophenetic* berdasarkan jarak pengukuran kemiripan dengan algoritma *average linkage*

No	Jarak	Korelasi <i>Cophenetic</i>
1	<i>Euclidean Distance</i>	0,9848
2	<i>Dynamic Time Warping (DTW)</i>	0,9800

Berdasarkan tabel 3 didapatkan bahwa *Euclidean Distance* menghasilkan nilai korelasi *cophenetic* terbesar yaitu 0,9848. Dari hasil pengukuran dari koefisien korelasi *cophenetic* menggunakan algoritma *average linkage*, didapat bahwa jarak pengukuran kemiripan yang optimal dalam mengelompokkan Provinsi di Indonesia berdasarkan kasus aktif COVID-19 adalah *Euclidean Distance*.

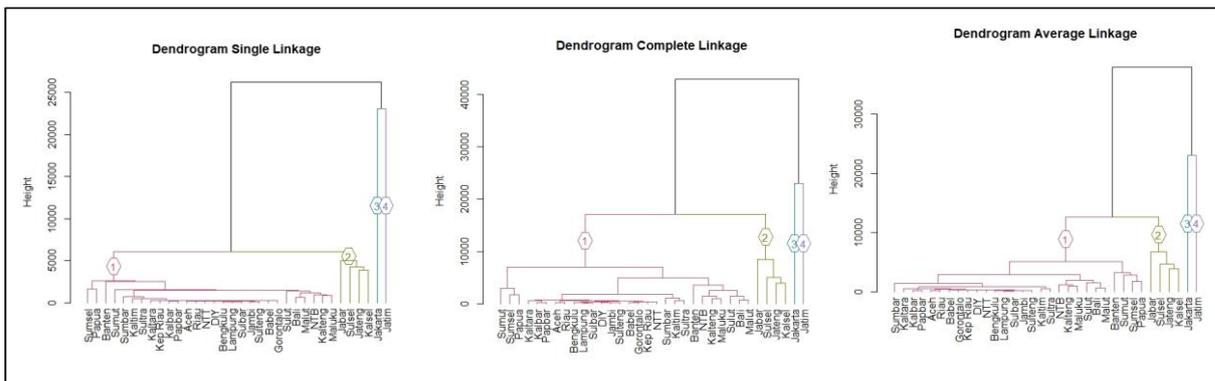
4. Penentuan Algoritma Terbaik Secara Keseluruhan

Perbandingan algoritma dalam proses *agglomerative hierarchical clustering* dan hasil pengukuran kemiripan dengan menggunakan koefisien korelasi *cophenetic* ditampilkan pada tabel berikut:

Tabel 4. Penentuan algoritma terbaik secara keseluruhan

Jarak	<i>Single Linkage</i>	<i>Complete Linkage</i>	<i>Average Linkage</i>
Jarak <i>Euclidean</i>	0,9756	0,9816	0,9848
<i>Dynamic Time Warping (DTW)</i>	0,8736	0,9778	0,9800

Pada Tabel 4, menunjukkan perbandingan koefisien korelasi *cophenetic* yang diukur berdasarkan algoritma *agglomerative hierarchical clustering*, dimana koefisien korelasi *cophenetic* digunakan untuk menghitung korelasi antara dua matriks jarak *cophenetic*. Semakin tinggi nilai koefisien korelasi *cophenetic* dari pohon kluster maka



Gambar 2. Dendrogram Hasil Pengelompokan berdasarkan Algoritma *Single*, *Complete* dan *Average Linkage*

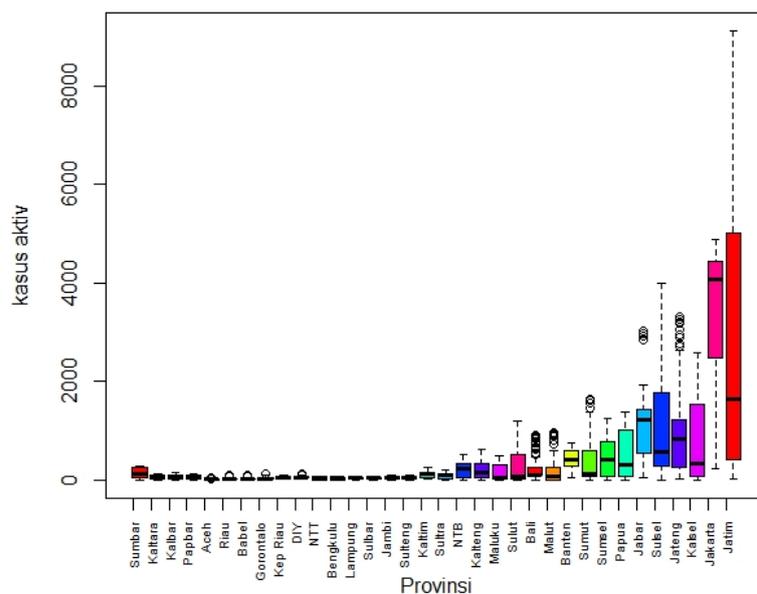
semakin baik algoritma *agglomerative hierarchical clustering* mengelompokkan objek.

Berdasarkan hasil pengukuran dari koefisien korelasi cophenetic secara keseluruhan didapat bahwa jarak pengukuran kemiripan yang optimal dalam mengelompokkan Provinsi di Indonesia berdasarkan kasus aktif COVID-19 adalah *Euclidean Distance*. Adapun hasil pengelompokan ditampilkan dalam Dendrogram seperti pada Gambar 2.

5. Profilisasi dan Interpretasi Hasil Pengelompokan

Pembentukan box-plot digunakan untuk memberikan gambaran ukuran tendensi sentral dan kesimetrisan data pengamatan dalam melihat perbandingan pola persebaran data kasus aktif COVID-19 di 34 Provinsi Indonesia. Gambar 3 merupakan box-plot persebaran dari data kasus aktif COVID-19. Secara umum, persebaran data kasus aktif COVID-19 untuk setiap Provinsi memiliki perbedaan. Provinsi Jawa Timur dan Jakarta terlihat memiliki nilai yang cukup tinggi diantara provinsi-provinsi lainnya, ini mengindikasikan bahwa jumlah kasus aktif

Boxplot Kasus Aktif COVID-19 Per-Provinsi di Indonesia



Gambar 3. Boxplot Kasus Aktif COVID-19 Per-Provinsi di Indonesia

COVID-19 di Provinsi Jawa timur dan DKI Jakarta sangat tinggi.

Rata-rata kasus aktif COVID-19 yang ditemukan pada klaster pertama, antara 10,97 kasus pada Provinsi Aceh dan 494,2 kasus pada Provinsi Papua. Pada klaster kedua, rata-rata kasus aktif pada Provinsi Kalimantan Selatan yaitu sebesar 790,25 kasus dan pada Provinsi Sulawesi Selatan yaitu sebesar 1.104 kasus. Pada klaster ketiga, rata-rata kasus aktif pada Provinsi Jakarta yaitu sebesar 3.339 kasus dan pada Provinsi Jawa Timur yaitu sebesar 4.996 kasus.

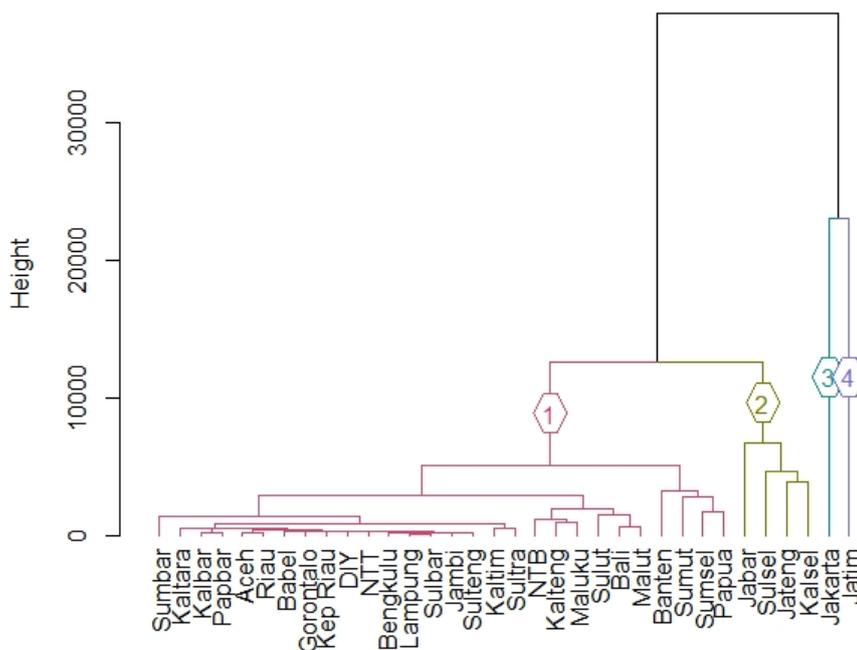
Langkah selanjutnya adalah melakukan profilisasi dan interpretasi hasil pengelompokkan. Koefisien korelasi *cophenetic* menunjukkan bahwa pengukuran kemiripan terbaik berdasarkan algoritma *single*, *complete* dan *average linkage* adalah *Euclidean Distance*, kemudian diperoleh bahwa keseluruhan hasil pengelompokkan yang paling tepat digunakan untuk membentuk melihat karakteristik kasus aktif COVID-19 yaitu algoritma *average linkage*. Dendrogram

hasil pengelompokkan ditampilkan pada Gambar 4.

Provinsi Jawa Timur dan DKI Jakarta sampai saat ini masih merupakan Provinsi dengan kasus terkonfirmasi kasus yang tertinggi, disusul dengan Kalimantan Selatan, Jawa Tengah, Sulawesi Selatan dan Jawa Barat. Pengendalian pandemi ini sangat tergantung dari kewaspadaan masyarakat dan ketersediaan sarana dan prasarana kesehatan sejak dini dari pemerintah pusat maupun daerah dalam menjaga kesehatan dilingkungannya.

Secara umum penerapan kebijakan daerah dalam rangka pencegahan penularan COVID-19 pada setiap klaster memiliki penanganan yang sama seperti pembentukan tim daerah tanggap COVID-19, memperketat akses masuk antar daerah dengan pengecekan suhu tubuh, memperketat protokol kesehatan di ruang publik, membatasi izin keramaian, meniadakan kegiatan belajar mengajar di Sekolah, peniadaan sementara kegiatan keagamaan di rumah ibadah dan kebijakan-kebijakan lainnya.

Dendrogram Average Linkage



Gambar 4. Dendrogram Hasil Pengelompokkan berdasarkan Algoritma Average Linkage

Tabel 5. Pengelompokan Provinsi Berdasarkan Kasus Aktif COVID-19

Klaster	Provinsi
Klaster 1	Sumatera Barat, Kalimantan Utara, Kalimantan Barat, Papua Barat, Aceh, Riau, Bangka Belitung, Gorontalo, Kepulauan Riau, Yogyakarta, NTT, Bengkulu, Lampung, Sulawesi Barat, Jambi, Sulawesi Tengah, Kalimantan Timur, Sulawesi Utara, NTB, Kalimantan Tengah, Maluku, Sulawesi Utara, Bali, Maluku Utara, Banten, Sumatera Utara, Sumatera Selatan, Papua
Klaster 2	Jawa Barat, Sulawesi Selatan, Jawa Tengah, Kalimantan Selatan
Klaster 3	Jakarta dan Jawa Timur

Beberapa provinsi dengan kasus aktif COVID-19 yang tinggi menerapkan Pembatasan Sosial Berskala Besar (PSBB) diantaranya Provinsi Sumatera Barat, Riau, Banten dan Gorontalo pada klaster pertama. Sulawesi Selatan dan Jawa Barat pada klaster kedua dan DKI Jakarta serta Jawa Timur pada klaster ketiga.

Berdasarkan dendrogram, dapat kita ketahui bahwa Provinsi-Provinsi besar memiliki kasus aktif COVID-19 yang cenderung tinggi, dikarenakan memiliki mobilitas yang tinggi sehingga menyebabkan kecenderungan penularan yang semakin cepat. Hal tersebut diyakinkan, dengan adanya kecenderungan provinsi dengan sebagian besar wilayahnya memiliki wilayah perdesaan dan mobilitas rendah cenderung berada pada klaster pertama, sedangkan wilayah dengan perkotaan dan mobilitas sedang serta tinggi berada pada klaster kedua dan tiga. Selanjutnya berdasarkan dendrogram yang diperoleh dari hasil pengelompokan diharapkan dapat menjadi bahan masukan untuk menghentikan rantai penyebaran COVID-19, salah satunya menghambat laju mobilitas penduduk hingga ditemukannya obat atau vaksin yang bisa mencegah penularan COVID-19 secara efektif.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil analisis, diperoleh beberapa kesimpulan dari penelitian ini yaitu:

1. Gambaran umum pola penyebaran jumlah kasus aktif COVID-19 sepanjang 21 Maret 2020 hingga 17 Juli 2020 menunjukkan titik *epicentrum* pola

penyebaran yang masih fluktuatif di Indonesia, terdapat pola naik turun yang cenderung belum stabil dan pola dari setiap Provinsi cenderung berbeda satu sama lain. Hal inilah yang menarik, dimana setiap daerah memiliki karakteristik yang berbeda-beda dan terdapat persamaan pola sehingga dijadikan menjadi klaster.

2. Pengelompokan Provinsi di Indonesia berdasarkan kasus aktif COVID-19 menggunakan metode *agglomerative hierarchical clustering* dengan beberapa algoritma pengelompokan dan beberapa pengukuran kemiripan tanpa melihat intervensi pemerintah daerah atau pusat mengenai penanganan COVID-19 menghasilkan tiga kelompok yang memiliki kemiripan dalam pembentukan klaster.
3. Koefisien korelasi *cophenetic* menunjukkan bahwa pengukuran kemiripan yang terbaik berdasarkan ketiga algoritma pengelompokan yang digunakan adalah *Euclidean Distance*, sehingga secara keseluruhan hasil pengelompokan yang paling tepat digunakan untuk membentuk melihat karakteristik kasus aktif COVID-19 yaitu metode *average linkage* dengan pengukuran kemiripan menggunakan *Euclidean Distance*.

Saran

Saran yang di ajukan oleh penulis untuk penelitian selanjutnya mengenai proses pengelompokan dengan *Time Series Based Clustering* adalah:

1. Menggunakan dan membandingkan hasil pengukuran kemiripan lainnya

seperti Mahalanobis *Distance* atau Pearson Correlation *Distance* sehingga bisa melihat fenomena pola penyebaran jumlah kasus aktif COVID-19 setiap daerah di Indonesia.

2. Melakukan proses *updating* data yang terbaru mengenai kasus aktif COVID-19 di Indonesia untuk mengetahui pergerakan penanganan setiap harinya di Indonesia.
3. Melakukan dan membandingkan pemilihan jumlah kluster yang optimal menggunakan *Silhouette Coefficient*.

DAFTAR PUSTAKA

- Dani, A.T.R., Wahyuningsih, S. & Rizki, N. A. (2019). Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu. *Jambura Journal of Mathematics* 1(2):64-78.
- Johnson, R. A. & Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis, Fifth Edition*. New Jersey: Pearson Prentice Inc.
- Montero, P. & Villar, J. A. (2014). TSclust: An R Packages for Time Series Clustering. *Journal of Statistical Software*. 62(1):01-43.
- Santosa, E. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Saracli, S., Dogan, N. & Dogan, I. (2013). Comparison of Hierarchical Cluster Analysis Methods by Cophenetic Correlation. *Journal of Inequalities and Applications*. <https://doi.org/10.1186/1029-242X-2013-203>.
- Supranto. (2010). *Statistik: Teori dan Aplikasinya Edisi 8*. Jakarta: Erlangga.