

## KLASTERISASI PENYEBARAN VIRUS CORONA (COVID-19) DI INDONESIA MENGGUNAKAN METODE K-MEANS

Teguh Ammar Taqiyyuddin\*

Program Studi Statistika, Universitas Padjadjaran

\*email: teguh18001@mail.unpad.ac.id

Diterima: 8 September 2021, disetujui: 22 September 2021, dipublikasi: 30 Desember 2021

---

**Abstract:** *Coronavirus Disease 2019 (COVID-19) or commonly called the Corona virus is a new virus that was first confirmed to appear on December 31, 2019. In Indonesia, all of its 34 provinces have confirmed positive cases of the virus. In this study, clustering of the spread of COVID-19 was carried out as an input for the government for a better, focused handling of COVID-19 for all provinces in Indonesia. Clustering is carried out based on the parameters of total positive cases, total cases of death, Recovery Index, Case Fatality Ratio, tracking and isolation ratio, total cases of people under surveillance completed, cumulative total cases of patients under surveillance, population density, and poverty line. In this study, K-means method was used to cluster data. Based on the Elbow test, the recommended number of clusters is three, which can be translated as high, medium, and low risk clusters. The results of this study are expected to be a comparison to updated data, as well as consideration for the government's better performance in handling COVID-19 more effectively and efficiently.*

**Keywords:** *COVID-19, clustering, K-means*

**Abstrak:** *Coronavirus Disease 2019 (COVID-19) atau biasa disebut virus Corona merupakan virus baru yang pertama kali dikonfirmasi muncul pada tanggal 31 Desember 2019. Sampai saat ini terdapat kasus positif (terkonfirmasi) COVID-19 di seluruh provinsi di Indonesia. Pada penelitian ini dilakukan klusterisasi persebaran COVID-19 yang dimaksudkan sebagai masukan untuk pemerintah agar dapat melakukan penanganan COVID-19 yang lebih terarah untuk semua provinsi di Indonesia. Klusterisasi dilakukan berdasarkan parameter total kasus positif, total kasus meninggal, Recovery Index, Case Fatality Ratio, rasio lacak dan isolasi, total kasus orang dalam pengawasan selesai, kumulatif total kasus pasien dalam pengawasan, kepadatan penduduk, dan garis kemiskinan. Pada penelitian ini digunakan metode K-means dalam pengklasteran data. Berdasarkan pengujian, jumlah kluster yang direkomendasikan adalah tiga kluster yaitu resiko tinggi, sedang, dan rendah. Diharapkan hasil penelitian ini dapat menjadi pembandingan data terkini serta menjadi pertimbangan untuk meningkatkan kinerja pemerintah dalam penanganan COVID-19 yang lebih efektif dan efisien.*

**Kata Kunci:** *COVID-19, klusterisasi, K-means*

## PENDAHULUAN

Pada tanggal 31 Desember 2019, World Health Organization (WHO) mendapatkan informasi mengenai kasus pneumonia yang terjadi di kota Wuhan, Provinsi Hubei, Cina [1]. Tanggal 7 Januari 2020, otoritas Cina mengonfirmasi telah mengidentifikasi virus baru di mana dilaporkan lebih dari 2.000 kasus infeksi virus tersebut terjadi di Cina, termasuk di luar Provinsi Hubei. Pada 11 Februari 2020, WHO menamai virus Corona jenis baru tersebut sebagai Coronavirus Disease 2019 (COVID-19) [2].

Di Indonesia, Presiden Joko Widodo untuk pertama kalinya mengumumkan kasus positif virus COVID-19 pada tanggal 2 Maret 2020 [3]. Sejak saat itu penambahan jumlah penderita semakin meningkat. Banyak hal telah dilakukan pemerintah dalam menekan angka penyebaran virus COVID-19, tetapi jumlah kasus positif semakin bertambah. Angka kematian karena virus COVID-19 pun semakin bertambah, walau tidak dapat dipungkiri jika angka sembuh virus COVID-19 lebih tinggi. Terdapat sejumlah gejala dan istilah terkait penderita virus COVID-19, di antaranya terinfeksi dengan gejala-gejala ringan sampai berat; terinfeksi tanpa adanya gejala, yang biasa disebut orang tanpa gejala (OTG); berada dalam pengawasan dengan satu gejala (ODP); pasien yang berada dalam pengawasan dengan beberapa gejala (PDP); pasien meninggal; pasien yang dinyatakan sembuh; serta beberapa istilah lainnya. Setiap provinsi di Indonesia memiliki jumlah penambahan kasus yang berbeda-beda.

Penelitian sebelumnya yang membahas mengenai COVID-19 telah dilakukan oleh [4] yang melakukan klusterisasi menggunakan metode *K-Means* untuk data yang bersumber dari website Kaggle. Selain itu, [5] melakukan mapping tingkat kerawanan COVID-19 menggunakan analisis kluster hirarki. Namun, dalam praktiknya akan diaplikasikan ke dalam data nyata untuk setiap provinsi di Indonesia dengan menggunakan metode *K-means*.

Penelitian ini bertujuan untuk memberikan masukan untuk pemerintah agar dapat melaksanakan penanganan COVID-19 yang lebih terarah untuk semua provinsi di Indonesia. Oleh karena itu, pemerintah perlu memahami karakteristik setiap provinsi dan penanganan yang sesuai untuk provinsi tersebut, dan perlu dilakukan pengelompokan karakteristik tiap provinsi agar penanganan COVID-19 dapat dilakukan dengan tepat. Sehingga dengan hasil pengelompokan tersebut dapat menjadi pertimbangan kinerja pemerintah dalam penanganan COVID-19 yang lebih efektif dan efisien.

Berdasarkan tujuan di atas, hal tersebut dapat dicapai dengan menggunakan analisis kluster. Analisis kluster merupakan suatu metode statistika yang digunakan untuk

mengelompokkan sekumpulan objek ke dalam kelompok-kelompok berdasarkan karakteristik yang dimilikinya [6]. Proses pengklasteran dilakukan menggunakan metode *K-means* dengan jumlah klaster yang ditentukan oleh grafik Elbow [7]. Jumlah klaster dipilih berdasarkan penurunan inersia yang tidak lagi signifikan [8].

## METODOLOGI PENELITIAN

### Data

Data yang diolah merupakan data persebaran virus COVID-19 berdasarkan provinsi di Indonesia yang diperoleh dari situs resmi Badan Pusat Statistik dan kawal COVID-19 (<https://kawalCOVID19.id/>) yang diakses pada tanggal 17 Oktober 2020 pukul 16:00 WIB. Peubah yang digunakan dalam penelitian ini dipaparkan pada Tabel 1.

Tabel 1. Peubah Penelitian

Peubah	Keterangan	Jangkauan
X1	Kumulatif total terinfeksi COVID-19 (jiwa)	499 - 93.356
X2	Kumulatif total sembuh dari COVID-19 (jiwa)	368 - 77.969
X3	Total kasus positif COVID-19 aktif pada tanggal 17 Oktober 2020 (jiwa)	0 - 974
X4	Total kasus sembuh dari COVID-19 pada tanggal 17 Oktober 2020 (jiwa)	0 - 1.106
X5	Kumulatif total meninggal akibat terinfeksi COVID-19 (jiwa)	5 - 3.529
X6	Total kasus meninggal akibat COVID-19 pada tanggal 17 Oktober 2020 (jiwa)	0 - 21
X7	Recovery Index (persen)	0,38 - 0,94
X8	Case Fatality Ratio (persen)	0,01 - 0,07
X9	Rasio lacak & isolasi (persen)	0 - 13,13
X10	Kumulatif total kasus OTG (jiwa)	6 - 385.352
X11	Total kasus OTG selesai (jiwa)	6 - 284.304
X12	Kumulatif total kasus ODP (jiwa)	646 - 96.844
X13	Total kasus ODP dalam proses (jiwa)	13 - 4.519
X14	Total kasus ODP selesai (jiwa)	735 - 92.325
X15	Kumulatif total kasus PDP (jiwa)	152 - 3.781
X16	Kepadatan Penduduk (jiwa/km <sup>2</sup> )	9 - 15.900
X17	Garis Kemiskinan (Rupiah/kapita/bulan)	33.9743 - 73.2570

## Analisis Klaster

Analisis klaster merupakan suatu metode statistika yang digunakan untuk mengelompokkan sekumpulan objek ke dalam sejumlah kelompok berdasarkan karakteristik yang dimilikinya [9]. Objek diklasifikasikan ke dalam satu atau lebih klaster sehingga objek-objek yang berada di dalam klaster mempunyai kemiripan atau kesamaan karakter [10]. Solusi klaster secara keseluruhan bergantung pada peubah-peubah yang digunakan sebagai dasar untuk menilai kesamaan. Klaster sendiri didefinisikan sebagai sejumlah objek yang mirip yang dikelompokkan secara bersama.

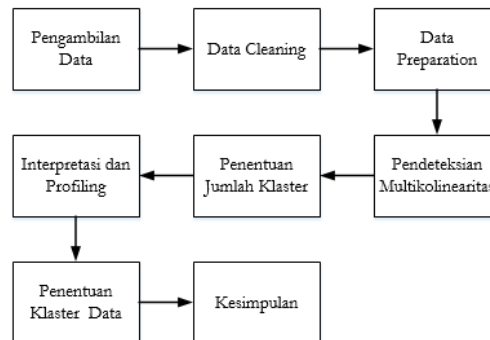
## Metode *K-Means*

Metode *K-means* digunakan sebagai alternatif metode klaster untuk data dengan ukuran yang lebih besar [11]. Hal ini dikarenakan metode ini memiliki kecepatan yang lebih tinggi jika dibandingkan dengan metode hirarki [12]. Metode *K-means* dapat digunakan untuk menjelaskan algoritma dalam penentuan suatu objek ke dalam klaster tertentu berdasarkan rataan terdekat [13]. Dalam prosedur pembentukan klaster *K-means*, langkah-langkah yang dilakukan adalah sebagai berikut:

1. Penentuan banyaknya klaster ( $k$ ) yang akan dibentuk.
2. Pembangkitan  $k$  *centroid* awal (rata-rata setiap klaster).
3. Perhitungan jarak antara setiap objek dengan setiap *centroid*; objek tersebut dimasukkan ke dalam klaster yang sesuai berdasarkan jarak terdekat.
4. Penentuan *centroid* dari klaster yang baru.
5. Langkah 3 dan 4 diulangi sampai tidak ada lagi pemindahan objek antar klaster.

## Tahapan Analisis

Penerapan metode *K-Means* pada klasterisasi persebaran COVID-19 di Indonesia dilakukan dalam beberapa tahap, seperti disajikan pada Gambar 1.



Gambar 1. Langkah-langkah Analisis

### 1) *Pengambilan Data*

Detail mengenai data yang diolah beserta peubahnya telah dipaparkan pada Subbab Data.

### 2) *Data Cleaning*

Proses ini penting dilakukan dalam meningkatkan kualitas data yang juga berpengaruh terhadap analisis secara keseluruhan. Data yang tidak akurat dapat berpengaruh buruk terhadap akurasi dan performa model [14]. Sebagian data yang tidak akurat dan tidak dibutuhkan dibuang sehingga akan tersedia data yang berkualitas [15]. Pada proses ini dilakukan penanganan data hilang (*missing value*) yaitu menghilangkan data hilang tersebut atau mengganti nilai yang hilang dengan nilai lain, seperti rata-rata dari peubah tersebut (mean), nilai yang paling sering muncul (modus), atau nilai tengah (median) [16].

### 3) *Data Preparation*

Setelah dataset dibersihkan, masih ada beberapa tahap yang perlu dilakukan agar dataset benar-benar siap untuk digunakan. Pada proses ini dilakukan pendeteksian pencilan (*outlier*) dengan melakukan visualisasi serta pembakuan/standarisasi yang berfungsi menyamakan skala nilai dari tiap atribut pada data. Pembakuan terhadap data perlu dilakukan jika terdapat perbedaan satuan yang menyebabkan data antar peubah terlihat kontras [17]. Proses pembakuan (*standardization*) dapat dilakukan dengan menggunakan

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$$

dengan :

$Z_{ij}$  : Data hasil pembakuan observasi ke-i peubah ke-j

$x_{ij}$  : Observasi ke-i peubah ke-j

$\bar{x}_j$  : Rataan peubah ke-j

$s_j$  : Simpangbaku peubah ke-j

#### 4) Pendeteksian Multikolinearitas

Uji multikolinearitas adalah uji yang dilakukan untuk memastikan apakah di dalam sebuah model regresi ada interkorelasi atau kolinearitas antar peubah bebas [18]. Interkorelasi adalah hubungan yang linear atau hubungan yang kuat antara satu peubah bebas dengan peubah bebas lainnya di dalam sebuah model regresi. Interkorelasi dapat dilihat dengan nilai koefisien korelasi antara peubah bebas, nilai VIF dan Tolerance, nilai Eigenvalue dan Condition Index, serta nilai galatbaku (*standard error*) koefisien beta atau koefisien regresi parsial. Nilai VIF yang melebihi angka 10 merupakan indikasi adanya multikolinearitas dalam model regresi [19]. Sedangkan nilai TOL sama dengan 1 merupakan indikasi tidak adanya multikolinearitas.

$$VIF_k = (1 - R^2_k)^{-1}$$

$$TOL = \frac{1}{VIF}$$

dengan  $k = 1, 2, \dots, p - 1$

#### 5) Penentuan Jumlah Klaster

Klasterisasi dapat dilakukan setelah data dibakukan. Dalam penentuan jumlah klaster ini digunakan metode Elbow [20]. Metode elbow bertujuan untuk menentukan siku (*elbow*), yaitu jumlah K yang optimal [21]. Penentuan elbow dilakukan secara manual, yaitu dengan melihat titik dimana penurunan inersia tidak lagi signifikan.

## 6) Interpretasi dan Profiling Hasil Klaster

Nilai *centroid* klaster dapat digunakan untuk mengetahui kondisi dari 34 provinsi di Indonesia berdasarkan peubah bebas. Tanda positif ataupun negatif dikarenakan adanya pengaruh pembakuan data sebelum klasterisasi dilakukan.

## HASIL DAN PEMBAHASAN

### Pendeteksian Multikolinearitas

Hasil uji multikolinearitas dapat dilihat pada Tabel 2. Nilai VIF melebihi 10 pada 12 peubah bebas, sehingga dapat disimpulkan bahwa terdapat multikolinearitas di antara 12 dari 17 peubah bebas yang digunakan; artinya terdapat hubungan linear antara masing-masing peubah bebas. Maka perlu dilakukan suatu perlakuan (*treatment*) untuk mengatasi multikolinearitas tersebut sehingga dapat memenuhi asumsi non multikolinearitas. Perlakuan yang dapat digunakan untuk mengatasi multikolinearitas ini yaitu dengan cara mengeliminasi peubah bebas yang mempunyai nilai VIF paling besar dari pengujian sebelumnya. Setelah dilakukan iterasi pengujian multikolinearitas, didapatkan hasil pada Tabel 3.

Tabel 2. Asumsi Multikolinearitas (1)

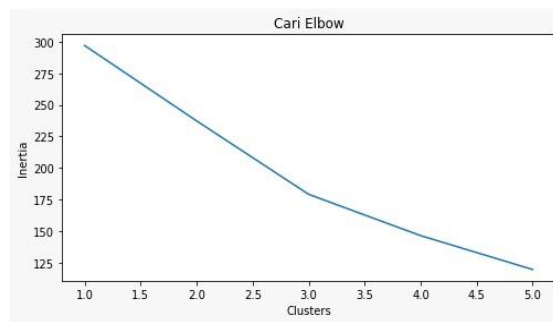
Peubah	VIF	result
X1	1135.702	<i>mulicolinearity</i>
X2	923.061	<i>mulicolinearity</i>
X3	55.821	<i>mulicolinearity</i>
X4	101.179	<i>mulicolinearity</i>
X5	185.215	<i>mulicolinearity</i>
X6	11.097	<i>mulicolinearity</i>
X7	2.218	<i>non mulicolinearity</i>
X8	3.989	<i>non mulicolinearity</i>
X9	4.727	<i>non mulicolinearity</i>
X10	512.638	<i>mulicolinearity</i>
X11	467.527	<i>mulicolinearity</i>
X12	116.030	<i>mulicolinearity</i>
X13	16.005	<i>mulicolinearity</i>
X14	94.871	<i>mulicolinearity</i>
X15	5.645	<i>non mulicolinearity</i>
X16	48.918	<i>mulicolinearity</i>
X17	4.348	<i>non mulicolinearity</i>

Tabel 3. Asumsi Multikolinearitas (2)

Peubah	VIF	result
X3	3.217	<i>non multicolinearity</i>
X6	3.936	<i>non multicolinearity</i>
X7	1.344	<i>non multicolinearity</i>
X8	1.470	<i>non multicolinearity</i>
X9	1.321	<i>non multicolinearity</i>
X14	2.023	<i>non multicolinearity</i>
X15	1.592	<i>non multicolinearity</i>
X16	2.415	<i>non multicolinearity</i>
X17	2.279	<i>non multicolinearity</i>

Nilai VIF kurang dari 10 pada sembilan peubah bebas yang ada pada Tabel 3. Dapat disimpulkan bahwa tidak terdapat multikolinearitas di antara kesembilan peubah bebas ini, artinya tidak terdapat hubungan linear antara masing-masing peubah bebas.

### Penentuan Jumlah Kluster



Gambar 2. Grafik Elbow.

Berdasarkan Gambar 2 tersebut, nilai Elbow berada pada nilai  $k = 3$ , karena penurunan inersia pada  $k$  seterusnya tidak lagi signifikan (perubahan nilainya kecil). Sehingga jumlah kluster yang optimal untuk kasus ini adalah 3.

### Interpretasi dan Profiling Hasil Kluster

Menurut [22], tanda positif menunjukkan ke arah yang kuat dan negatif ke arah yang lemah. Nilai  $z$  ( $z$ -score) pada setiap peubah di setiap kluster ditampilkan pada Tabel 4. Berdasarkan tabel ini, dapat diketahui bahwa kluster satu merupakan kluster dengan



provinsi-provinsi yang memiliki nilai yang besar pada peubah  $X_3$ ,  $X_6$ ,  $X_{15}$ , dan  $X_{17}$ . Klaster dua merupakan klaster dengan provinsi-provinsi yang memiliki nilai yang besar pada peubah  $X_6$ ,  $X_9$ ,  $X_{14}$ , dan  $X_{15}$ . Sedangkan klaster tiga merupakan klaster dengan provinsi-provinsi yang memiliki nilai yang besar pada peubah  $X_7$ ,  $X_8$ , dan  $X_{16}$ .

Tabel 4. Klaster pusat

	Klaster		
	1	2	3
Zscore( $X_3$ )	0.6474	-0.1560	-0.3419
Zscore( $X_6$ )	0.3154	0.1160	-0.2070
Zscore( $X_7$ )	-0.2106	-1.3140	0.3986
Zscore( $X_8$ )	-0.3501	-0.9314	0.3987
Zscore( $X_9$ )	-0.1516	1.5824	-0.2453
Zscore( $X_{14}$ )	-0.0522	1.3769	-0.2596
Zscore( $X_{15}$ )	0.3110	1.0227	-0.3954
Zscore( $X_{16}$ )	-0.2443	-0.1018	0.1628
Zscore( $X_{17}$ )	0.9753	-0.4138	-0.4775

Tabel 5. Analisis Ragam pada Klaster 1, 2, dan 3

	Kuadrat Tengah	Nilai-p
Zscore( $X_3$ )	3.465	0.026
Zscore( $X_6$ )	0.981	0.387
Zscore( $X_7$ )	5.210	0.003
Zscore( $X_8$ )	3.920	0.015
Zscore( $X_9$ )	5.707	0.001
Zscore( $X_{14}$ )	4.447	0.008
Zscore( $X_{15}$ )	4.109	0.012
Zscore( $X_{16}$ )	0.601	0.563
Zscore( $X_{17}$ )	7.741	0.000

Hasil analisis ragam (ANOVA) terhadap ketiga klaster ditampilkan pada Tabel 5. Jika nilai  $p$  ( $p - value$ )  $< \alpha(0.05)$  maka  $H_0$  ditolak, yang artinya terdapat perbedaan yang signifikan antara ketiga klaster. Dengan demikian, peubah yang signifikan adalah

X3, X7, X8, X9, X14, X15, dan X17. Sedangkan peubah yang tidak signifikan, atau peubah yang tidak memiliki perbedaan yang signifikan antara ketiga klaster, adalah X6 dan X16.

Tabel 6. Banyaknya anggota pada setiap klaster

<b>Klaster</b>	<b>1</b>	11
	<b>2</b>	4
	<b>3</b>	19
<i>Valid</i>		34
<i>Missing</i>		0

Banyaknya anggota pada setiap klaster diperlihatkan pada Tabel 6. Berdasarkan tabel tersebut, diperoleh bahwa 34 provinsi yang ada di Indonesia dapat digolongkan ke dalam 3 klaster, dengan pembagian jumlah klaster satu sebanyak 11 provinsi, klaster dua sebanyak empat provinsi, dan klaster tiga sebanyak 19 provinsi. Klasterisasi per provinsi ditampilkan pada Tabel 7.

Tabel 7. Klasterisasi per provinsi

Provinsi	Klaster	Jarak ke Pusat Klaster
Aceh	1	1.559
Bali	3	1.352
Banten	1	1.184
Bangka Belitung	1	2.366
Bengkulu	3	1.045
DI Yogyakarta	3	1.314
DKI Jakarta	1	5.524
Jambi	2	3.253
Jawa Barat	1	2.786
Jawa Tengah	1	4.697
Jawa Timur	3	6.342
Kalimantan Barat	2	3.754
Kalimantan Timur	3	1.777
Kalimantan Tengah	3	0.601
Kalimantan Selatan	3	1.295
Kalimantan Utara	3	2.363
Kepulauan Riau	3	1.095

Nusa Tenggara Barat	3	1.451
Sumatera Selatan	3	1.162
Sumatera Barat	1	2.481
Sulawesi Utara	3	1.106
Sumatera Utara	3	1.287
Sulawesi Tenggara	1	2.045
Sulawesi Selatan	1	1.955
Sulawesi Tengah	3	1.561
Lampung	3	1.315
Riau	2	4.686
Maluku Utara	3	1.493
Maluku	3	2.173
Papua Barat	3	2.245
Papua	2	2.521
Sulawesi Barat	1	1.865
Nusa Tenggara Timur	1	1.954
Gorontalo	3	1.853

Nilai jarak setiap provinsi ke pusat kluster ditampilkan pula pada Tabel 7. Dapat diartikan bahwa semakin kecil nilai jaraknya maka semakin dekat dengan pusat klasternya, sedangkan semakin besar nilai jaraknya maka semakin jauh dengan pusat klasternya.

### **Pembahasan Hasil Penelitian**

Penelitian ini memiliki kontribusi praktis dalam membantu penyajian data kasus COVID-19 secara lebih baik. Hasil penelitian ini menyajikan hasil klasterisasi data kasus COVID-19 di Indonesia yang ditampilkan per provinsi. Pada penelitian ini dilakukan pengujian jumlah kluster terbaik menggunakan metode Elbow. Hasil pengujian menunjukkan bahwa jumlah kluster yang direkomendasikan adalah tiga kluster. Pengelompokan provinsi sesuai kluster yang telah dibuat dapat dilihat pada Tabel 8.

Tabel 8. Hasil klasterisasi data menurut provinsi

<b>Klaster</b>	<b>Provinsi</b>
1	Aceh, Banten, Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, Sumatera Barat, Sulawesi Tenggara, Sulawesi Selatan, Sulawesi Barat, dan Nusa Tenggara Timur
2	Jambi, Kalimantan Barat, Riau, dan Papua

3	Bali, Bengkulu, DI Yogyakarta, Jawa Timur, Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Kepulauan Riau, Nusa Tenggara Barat, Sumatera Selatan, Sulawesi Utara, Sumatera Utara, Sulawesi Tengah, Lampung, Maluku Utara, Maluku, Papua Barat, dan Gorontalo
---	---

## KESIMPULAN

Dari hasil analisis dan pembahasan yang telah diuraikan pada bab sebelumnya, diperoleh kesimpulan sebagai berikut:

1. Berdasarkan nilai Elbow diperoleh jumlah klaster untuk kasus ini adalah sebanyak 3. Hal ini diketahui karena penurunan inersia setelah  $k = 3$  tidak lagi signifikan (perubahan nilainya kecil).
2. Berdasarkan nilai Elbow tersebut, pengklasteran dibuat ke dalam 3 klaster. Hasil pengklasteran menunjukkan bahwa terdapat 11 provinsi pada klaster satu, empat provinsi pada klaster dua, dan 19 provinsi pada klaster tiga. Rincian dari ketiga klaster tersebut adalah sebagai berikut.
  - a. Klaster satu memiliki nilai yang besar pada empat peubah yaitu Total Kasus Positif, Total Kasus Meninggal, Kumulatif Total Kasus PDP, dan Garis Kemiskinan. Klaster ini dapat digolongkan sebagai klaster dengan resiko tinggi. Sebagai perbandingan, disarankan kepada pemerintah untuk memantau kesebelas provinsi ini, apakah masih sesuai dengan data terkini. Provinsi dengan resiko tinggi tergolong riskan, dan diharapkan kasus COVID-19 ini pada klaster satu ini dapat segera diminimalkan. Selain itu tingginya garis kemiskinan pada provinsi-provinsi di klaster ini perlu menjadi perhatian pemerintah agar dapat memperbaiki tatanan ekonomi pada provinsi-provinsi terkait.
  - b. Klaster dua memiliki nilai yang besar pada 4 peubah yaitu total Kasus Meninggal, Rasio Lacak dan Isolasi, Total Kasus ODP Selesai, dan Kumulatif Total Kasus PDP. Klaster dua tergolong pada angka kasus yang cukup besar dilihat dari kasus ODP dan PDP serta rasio lacak dan isolasi. Klaster ini dapat digolongkan sebagai klaster dengan resiko sedang. Terkait angka kematian yang tinggi, diharapkan pemerintah mampu mengatasi hal tersebut.
  - c. Klaster tiga memiliki nilai yang besar pada tiga peubah yaitu Recovery Index, Case Fatality Ratio, dan Kepadatan Penduduk. Klaster tiga tergolong cukup baik dilihat dari tingkat kesembuhan pasien, namun proporsi kematian akibat COVID-19 yang tergolong tinggi ini dapat disebabkan pula oleh kepadatan

penduduk yang tinggi di sejumlah provinsi yang termasuk ke dalam klaster ini. Klaster ini dapat digolongkan sebagai klaster dengan resiko rendah. Diharapkan pemerintah mengatasi kepadatan penduduk pada sejumlah provinsi yang termasuk ke dalam klaster ini.

#### DAFTAR PUSTAKA

- [1] Handayani, D., Hadi, D.R., Isbaniah, F., Burhan, E., Agustin, H. (2020). COVID-19 virus disease 2019. *Jurnal Respirologi Indonesia*. **40(2)**: 119–129.
- [2] Putri, R.N. (2020). Indonesia dalam menghadapi pandemi COVID-19. *Jurnal Ilmiah Universitas Batanghari Jambi*. **20(2)**: 705–709.
- [3] Hayati, H.N., Yoedtadi, M.G. (2020). Konstruksi berita COVID-19 di Kompas. com dan Tribunnews. com. *Koneksi*. **4(2)**: 243–250.
- [4] Indraputra, R.A., Fitriana, R. (2020). K-Means clustering data COVID-19. *Jurnal Teknik Industri*. **10(3)**: 275–282.
- [5] Rahmi, N.S. (2020). Mapping tingkat kerawanan COVID-19 dan faktor akselerator penyebaran virus antar daerah di Indonesia dengan menggunakan analisis cluster hierarki. Seminar Nasional Official Statistics; 2020 Sep 23-24; Jakarta, Indonesia [webinar]. Jakarta: Politeknik Statistika STIS. pp. 73–81.
- [6] Ramdhani, F., Hoyyi, A., Mukid, M.A. (2015). Pengelompokan provinsi di Indonesia berdasarkan karakteristik kesejahteraan rakyat menggunakan metode K-Means Cluster. *Jurnal Gaussian*. **4(4)**: 875–884.
- [7] Nainggolan, R., Lumbantoruan, G. (2018). Optimasi performa cluster K-Means menggunakan Sum of Squared Error (SSE). *METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi*. **2(2)**: 103–108.
- [8] Khakim, M.A., Rahmadhani, L., Purnomo, E.S.B., Idayani, R.W., Rakhmawati, N.A. (2020). Analisa jaringan dokumentasi dan informasi hukum kementerian tentang teknologi informasi menggunakan metode K-Means clustering. *Fountain of Informatics Journal*. **5(1)**: 27–34.
- [9] Hair, J.F., Anderson R.E., Tatham R.L., Black W.C. (1998). *Multivariate Data Analysis*. 5<sup>th</sup> edition. Upper Saddle River, New Jersey: Prentice Hall.
- [10] Yulianto, S., Hidayatullah, K.H. (2014). Analisis klaster untuk pengelompokan kabupaten/kota di provinsi Jawa Tengah berdasarkan indikator kesejahteraan rakyat. *Jurnal Statistika Universitas Muhammadiyah Semarang*. **2(1)**:56-63.

- [11] Johnson, R.A., Wichern, D. W. (2014). *Applied Multivariate Statistical Analysis* 6<sup>th</sup> edition. London: Pearson.
- [12] Soemartini, S., Supartini, E. (2017). Analisis K-Means cluster untuk pengelompokan kabupaten/kota di Jawa Barat berdasarkan indikator masyarakat. Prosiding Konferensi Nasional Penelitian Matematika dan Pembelajarannya (KNPMP) II 2017; 2017 Mar 18; Surakarta, Indonesia. Surakarta: Pendidikan Matematika FKIP Universitas Muhammadiyah Surakarta. pp. 144-154.
- [13] Abdillah, G., Putra, F.A., Renaldi, F. (2016). Penerapan data mining pemakaian air pelanggan untuk menentukan klasifikasi potensi pemakaian air pelanggan baru di PDAM Tirta Raharja menggunakan algoritma K-means. Seminar Nasional Teknologi Informasi dan Komunikasi 2016 (SENTIKA 2016); 2016 Mar 18-19; Yogyakarta, Indonesia. Yogyakarta: Universitas Atma Jaya Yogyakarta. pp. 498–506.
- [14] Menarianti, I. (2015). Klasifikasi data mining dalam menentukan pemberian kredit bagi nasabah koperasi. *Jurnal Ilmiah Teknosains*. **1(1)**: 36-45.
- [15] Lantang, O.A. (2010). Proses extract, transform, dan load pada data warehouse. *TEKNO*. **8(53)**: 32-36.
- [16] Khotimah, B.K. (2020). Optimasi pemodelan Hybrid Naive Bayes berbobot dengan menggunakan algoritma genetika pada data missing value [disertasi]. Surabaya: Universitas Airlangga.
- [17] Setiawan, R. (2017). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus: Politeknik Lp3i Jakarta). *Jurnal Lentera Ict*, **3(1)**, 76–92.
- [18] In, A.W.K., Asyik, N.F. (2019). Pengaruh kompetensi dan bebassi terhadap kualitas audit dengan etika auditor sebagai peubah pemoderasi. *Jurnal Ilmu Dan Riset Akuntansi (JIRA)*. **8(8)**: 1-15.
- [19] Tjeleni, I. E. (2013). Kepemilikan manajerial dan institusional pengaruhnya terhadap kebijakan hutang pada perusahaan manufaktur di Bursa Efek Indonesia. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis Dan Akuntansi*. **1(3)**: 129-139.
- [20] Hadi, A.F., Bagus C.W., D., Hasan, M. (2017). Text mining pada media sosial Twitter; Studi kasus: Masa tenang Pilkada DKI 2017 putaran 2. Seminar Nasional Matematika dan Aplikasinya; 2017 Oct 21; Surabaya, Indonesia. Surabaya: Universitas Airlangga. pp. 324-331.

- [21] Winarta, A., Kurniawan, W.J. (2021). Optimasi cluster K-means menggunakan metode Elbow pada data pengguna narkoba dengan pemrograman Python. *JTIK (Jurnal Teknik Informatika Kaputama)*. **5(1)**: 113–119.
- [22] Simamora, B. (2005). Analisis multivariat pemasaran. Gramedia Pustaka Utama.