

# PENGELOMPOKKAN PROVINSI DI INDONESIA BERDASARKAN JUMLAH KASUS COVID-19 DAN FASILITAS KESEHATAN

Meinisa Fadillah Rahmi<sup>1</sup>, Paulus Satria Prasetyo<sup>2</sup>, Ratih Nurhabibah<sup>3</sup>,  
Rizky Perdana<sup>4</sup>, <sup>5</sup>Wa Ode Zuhayeni Madjida<sup>5</sup>

<sup>1,2,3,4</sup>Politeknik Statistika STIS, <sup>5</sup>Badan Pusat Statistik  
e-mail: <sup>1</sup>211709827@stis.ac.id

## Abstrak

Tingkat penyebaran COVID-19 cukup cepat, hingga 14 November 2020 tercatat jumlah kasus terkonfirmasi positif di Indonesia mencapai 463.007 jiwa. Ketersediaan fasilitas kesehatan masing-masing provinsi menentukan kesiapan daerah dalam penanganan COVID-19 sehingga penting untuk menganalisis keadaan dan distribusi provinsi-provinsi terkait kesiapannya tersebut. Penelitian ini melakukan clustering menggunakan algoritma K-Means dan K-Means with Outlier Detection untuk mengelompokkan 34 provinsi di Indonesia berdasarkan jumlah kasus COVID-19 dan data fasilitas kesehatan, lalu menentukan metode terbaiknya, serta mengidentifikasi karakteristik masing-masing kelompok berdasarkan metode terbaik. Penelitian menghasilkan tiga cluster. Cluster 1 merupakan kelompok provinsi dengan jumlah kasus COVID-19 tinggi dan fasilitas kesehatan kurang memadai, cluster 2 memiliki jumlah kasus COVID-19 tinggi dan fasilitas kesehatan memadai, sedangkan cluster 3 memiliki jumlah kasus COVID-19 rendah dan fasilitas kesehatan menengah.

**Kata kunci:** COVID-19, *Clustering*, *K-Means*, *K-Means with Outlier Detection*, Fasilitas Kesehatan

## Abstract

*The rate spread of COVID-19 is quite fast, until November 14, 2020, the number of cases in Indonesia has reached 463,007 people. The availability of health facilities for each province determines regional readiness in handling COVID-19, so it is very important to analyze the condition and distribution of provinces related to their readiness. This study conducted clustering using K-Means and K-Means with Outlier Detection algorithm to classify 34 provinces in Indonesia based on the number of COVID-19 cases and health facilities, then decide the best model, and identify the characteristics of each group based on the best model. The study resulted in three clusters. Cluster 1 is a group of provinces with a high number of COVID-19 cases and inadequate health facilities, cluster 2 has a high number of COVID-19 cases and adequate health facilities, while cluster 3 has a low number of COVID-19 cases and intermediate health facilities.*

**Keywords:** COVID-19, *Clustering*, *K-Means*, *K-Means with Outlier Detection*, *Health Facilities*

## PENDAHULUAN

*Coronavirus Disease-2019* (COVID-19) pertama kali ditemukan pada Desember 2019 di kota Wuhan, Cina yang kemudian menyebar ke seluruh dunia sehingga ditetapkan sebagai pandemi global oleh *World Health Organization*. Penyebaran awal di Indonesia terjadi pada awal Maret 2020, dengan kasus pertama di Kota Depok, Jawa Barat. Virus corona menyebar melalui berbagai macam cara, diantaranya melalui interaksi antara penderita dengan orang lain, benda-benda yang disentuh oleh penderita, bahkan dapat menyebar melalui udara. Tercatat hingga 14 November 2020 jumlah kasus terkonfirmasi positif di Indonesia mencapai 463.007 jiwa dengan total jumlah pasien sembuh sebanyak 388.094 jiwa, sedangkan jumlah kematian sebanyak 15.148 jiwa.

Menurut Muhyiddin (2020), berbagai negara telah melakukan kebijakan *lockdown* untuk membatasi penyebaran virus. Akan tetapi, mengubah kebiasaan masyarakat bukan persoalan mudah, banyak negara mengalami kendala dalam penerapan kebijakan tersebut. Sehingga negara-negara tersebut memodifikasi kebijakan *lockdown* dengan menerapkannya secara penuh, sebagian, ataupun dalam skala lokal dan seminimal mungkin. Indonesia memodifikasi kebijakan *lockdown* menjadi Pembatasan Sosial Berskala Besar (PSBB) yang diberlakukan pada setiap provinsi atau kabupaten/kota berdasarkan tingkat keparahan yang dinilai oleh Kementerian Kesehatan. Penyebaran yang begitu cepat dan jumlah kasus terkonfirmasi yang terus mengalami kenaikan membuat pemerintah menerapkan berbagai kebijakan, mulai dari *physical distancing*, penggunaan masker, *work from home*, hingga pembelajaran jarak jauh guna memutus penyebaran mata rantai virus tersebut.

Ketersediaan infrastruktur dan layanan kesehatan yang terdapat pada masing-masing provinsi menentukan kesiapan daerah dalam penanganan pasien terkonfirmasi dan penyebaran berkelanjutan COVID-19. Berdasarkan data

Ikatan Dokter Indonesia (IDI), jumlah dokter di Indonesia pada tahun 2020 sebanyak 183.605 dokter. Secara umum, jumlah tersebut mencukupi untuk seluruh masyarakat. Namun, permasalahannya terjadi pada persebaran yang tidak merata. Persebaran dokter bertumpuk di kota-kota besar dan Pulau Jawa, sehingga kebutuhan dokter di berbagai wilayah Indonesia masih belum terpenuhi. Begitupun dengan infrastruktur kesehatan dan tenaga medis lainnya. Ditambah lagi, hingga penghujung tahun 2020, belum terdapat kepastian mengenai vaksin untuk mencegah penyebaran COVID-19.

Noviyanto (2020) telah melakukan pengelompokan jumlah kematian penderita COVID-19 di Benua Asia dengan menerapkan teknik *data mining*. Hasil dari penelitian ini menunjukkan jumlah kematian penderita COVID-19 dibagi kedalam 3 *cluster*. Terdapat 4 negara yang masuk kedalam *cluster* dengan tingkat kematian tinggi, yaitu Turki, Iran, India dan China. Negara yang masuk kedalam *cluster* dengan tingkat kematian sedang sebanyak 4 negara, yaitu Pakistan, Indonesia, Jepang, dan Piliphina. Terakhir, negara yang masuk kedalam *cluster* dengan tingkat kematian rendah adalah 41 negara lainnya.

Penelitian Rembulan dkk. (2020) mengenai kebijakan Pemerintah terkait COVID-19 di setiap Provinsi dengan menggunakan analisis *cluster* menghasilkan 3 *cluster* optimal. *Cluster* 1 adalah *cluster* dengan risiko tinggi karena memiliki variabel jumlah kasus aktif dan jumlah kasus kematian per satu juta penduduk yang tertinggi. *Cluster* 2 adalah *cluster* dengan risiko rendah karena memiliki variabel dengan jumlah kasus kesembuhan tertinggi dan jumlah kasus aktif terendah. *Cluster* 3 adalah *cluster* dengan risiko sedang karena memiliki variabel jumlah kesembuhan terendah dan jumlah kasus aktif sedang. Kebijakan pemerintah pada *cluster* 1 hendaknya memprioritaskan variabel jumlah kasus aktif dan jumlah kasus kematian per satu juta penduduk. Untuk *cluster* 2 harus memprioritaskan variabel jumlah kasus kematian, sedangkan untuk *cluster* 3 harus

memprioritaskan variabel jumlah kasus aktif.

Berdasarkan pemaparan di atas, sangat diperlukan penelitian mengenai pengelompokan provinsi di Indonesia untuk mengidentifikasi tingkat penyebaran dan kapasitas penanganannya di masing-masing daerah. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan 34 provinsi berdasarkan jumlah kasus COVID-19 dan fasilitas kesehatan, menentukan metode yang terbaik, dan mengidentifikasi karakteristik masing-masing kelompok berdasarkan metode terbaik. Diharapkan dengan dibentuknya kelompok-kelompok tersebut dapat membantu pemerintah dalam menentukan kebijakan dan langkah yang harus diambil kedepannya guna menangani penyebaran COVID-19 di Indonesia.

## METODOLOGI

### 1. Tinjauan Referensi

#### *Data Mining*

*Data Mining* adalah sebuah proses pencarian informasi secara otomatis dalam tempat penyimpanan data yang berukuran besar untuk menemukan pola tertentu. Teknik yang digunakan pada *data mining* tersebut dapat dikelompokkan ke dalam dua kategori, yaitu *descriptive mining* dan *predictive mining*. *Descriptive mining* adalah proses menemukan karakteristik tertentu dari basis data, meliputi *clustering*, *asosiation*, dan *sequential mining*. *Predictive mining* adalah proses menemukan pola dari data menggunakan variabel-variabel lain di masa depan, meliputi *classification*.

#### *Clustering*

*Clustering* merupakan pengelompokan data untuk membentuk kelas baru dengan mengidentifikasi kelompok-kelompok yang memiliki kesamaan karakteristik tertentu. Objek-objek pada kelompok (*cluster*) yang terbentuk akan memiliki kemiripan satu sama lain dan berbeda dengan objek pada kelompok lain. Terdapat beberapa algoritma pada *clustering* diantaranya *K-Means*, *Fuzzy C-Means*, *DBSCAN*, dan *K-*

*Medoids*. Setiap algoritma memiliki fungsi yang berbeda dan kelebihan serta kekurangannya masing-masing.

#### *K-Means*

*K-Means Clustering* merupakan salah satu algoritma dalam *clustering* yang paling sederhana. *K-means clustering* melakukan proses pemodelan menggunakan *unsupervised learning* dan mengelompokkan data menggunakan sistem partisi. Secara umum, terdapat dua jenis *clustering* yang digunakan dalam proses pengelompokan data, yaitu *Hierarchical* dan *Non-Hierarchical*. *K-Means* merupakan salah satu metode *clustering Non-Hierarchical* atau *Partitional Clustering* yang secara umum dilakukan dengan algoritma sebagai berikut:

1. Menentukan jumlah *cluster* ( $k$ ).
2. Menentukan titik pusat (*centroid*) *cluster* awal secara *random*.
3. Menghitung jarak *centroid* dengan setiap data. Ukuran jarak yang digunakan adalah *Euclidean distance* dengan rumus berikut.

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

Dimana  $D(i, j)$  = Jarak data ke- $i$  ke pusat *cluster*  $j$ ,  $X_{ki}$  = Data ke- $i$  pada atribut data ke- $k$ ,  $X_{kj}$  = *Centroid* ke- $j$  pada atribut ke- $k$

4. Data akan dimasukkan ke dalam *cluster* berdasarkan jarak *centroid* terdekat. Lalu hitung kembali *centroid cluster* yang baru. *Centroid cluster* merupakan rata-rata semua data dalam sebuah *cluster*, dapat dihitung menggunakan rumus berikut.

$$V(i, j) = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \quad (2)$$

dimana:  $V(i, j)$  = *Centroid* rata-rata pada *cluster*  $i$  untuk atribut data ke- $j$

$N_i$  = Jumlah anggota *cluster*  $i$

5. Ulangi penghitungan jarak *centroid* dengan setiap data (kembali pada langkah ke-3) jika masih terdapat data

yang berpindah *cluster* atau apabila perubahan nilai centroid berada di atas nilai *threshhold* yang dibentuk atau jika perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan.

Kembali ke langkah 3, jika masih terdapat data yang berpindah *cluster* atau apabila perubahan nilai centroid berada di atas nilai *threshold* yang dibentuk atau jika perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan.

### ***K-Means with Outlier Detection***

Menurut Chawla dan Gionis (2013), algoritma *k-means* sangat sensitif terhadap adanya *outlier* dalam data dikarenakan menggunakan rata-rata sebagai pusat *cluster*-nya sehingga seharusnya selain melakukan *clustering*, harus dilakukan deteksi

*outlier* juga dalam prosesnya. Hal ini dapat menghasilkan pengelompokan yang lebih tepat. Metode ini menggunakan metode *minimum covariance determinant* (MCD) untuk menentukan *outlier*.

## **2. Metode Analisis**

Penelitian ini bersifat kuantitatif yang mencakup 34 Provinsi di Indonesia dengan menggunakan data sekunder yang terdiri dari data kasus COVID-19 dan data fasilitas kesehatan. Data kasus COVID-19 diperoleh dari *website* resmi COVID-19 oleh Gugus Tugas dengan *update* kasus pada tanggal 14 November 2020. Sedangkan data infrastruktur kesehatan diperoleh dari publikasi Rekapitulasi SDM Kesehatan yang Didayagunakan di Fasyankes pada Tahun 2019 oleh Kementerian Kesehatan.

Proses pengolahan data dilakukan dengan bantuan *software Microsoft Excel 2019* dan *RStudio*. Sebelumnya, data yang didapat dilakukan tahapan *preprocessing* untuk mempersiapkan variabel-variabel yang akan digunakan. Tahapan *preprocessing* yang dilakukan adalah sebagai berikut.

1. Melakukan *data cleaning* meliputi pengecekan *missing data*, *outlier*, dan inkonsistensi. Data yang digunakan

bebas dari *missing data* dan inkonsistensi, namun banyak terdapat *outlier*. Dalam kasus ini, penulis mempertahankan *outlier* dikarenakan masih ingin mempertahankan nilai yang sebenarnya pada tiap-tiap variabel dan diakomodasi dengan metode *k-means with outlier detection*.

2. Melakukan *data reduction*, yaitu mereduksi variabel yang tidak perlu dan melakukan agregat pada variabel-variabel yang menunjukkan karakteristik yang sama.

3. Melakukan *data transformation*, yaitu mengubah nilai variabel yang berupa jumlah infrastruktur menjadi rasio infrastruktur terhadap 1.000 penduduk yang ada di provinsi yang bersangkutan agar lebih menggambarkan kondisi yang sebenarnya dan lebih mudah dibandingkan, lalu melakukan normalisasi data menggunakan *z-score* dikarenakan variabel-variabel yang digunakan memiliki satuan yang berbeda-beda.

Berdasarkan hasil *preprocessing* yang telah dilakukan, didapatkan sebanyak 11 variabel yang siap dilakukan analisis, yaitu rasio jumlah puskesmas terhadap 1.000 penduduk (V1), rasio jumlah rumah sakit terhadap 1.000 penduduk (V2), rasio jumlah klinik terhadap 1.000 penduduk (V3), rasio jumlah posyandu terhadap 1.000 penduduk (V4), rasio jumlah dokter terhadap 1.000 penduduk (V5), rasio jumlah perawat terhadap 1.000 penduduk (V6), rasio jumlah bidan terhadap 1.000 penduduk (V7), jumlah kasus terkonfirmasi positif COVID-19 (V8), *recovery rate* (V9), *fatality rate* (V10), dan jumlah kasus aktif atau belum sembuh yang masih dalam perawatan atau isolasi mandiri (V11) menurut provinsi.

Tahapan analisis yang dilakukan selanjutnya adalah mengecek asumsi nonmultikolinearitas pada variabel-variabel yang digunakan dengan menghitung matriks korelasi antarvariabel. Kemudian, menentukan jumlah optimal *cluster* dengan menggunakan *by majority rule*, yaitu menggunakan *package NbClust* dimana terdapat 30 indeks yang melakukan *voting*

terhadap jumlah *cluster* yang optimal. Lalu, melakukan pengelompokkan objek dengan metode *k-means* dan *k-means with outlier detection* dengan ukuran ketidaksamaan berupa jarak Euclidean. Kemudian, menentukan metode terbaik dengan membandingkan nilai *R-squared* yang dihitung dari pembagian nilai varians *between* dengan varians total hasil *clustering*. Menurut Jhonson (2007), semakin besar nilai *R-squared*, maka semakin baik pengelompokkan yang dilakukan. Terakhir, melakukan identifikasi karakteristik masing-masing *cluster* dengan menghitung *centroid* berdasarkan metode terbaik.

## HASIL DAN PEMBAHASAN

Setelah dilakukan tahapan *preprocessing data* dan didapatkan variabel-variabel yang siap digunakan untuk analisis, dilakukan tahap analisis *cluster*.

### Pengujian Asumsi Nonmultikolinearitas

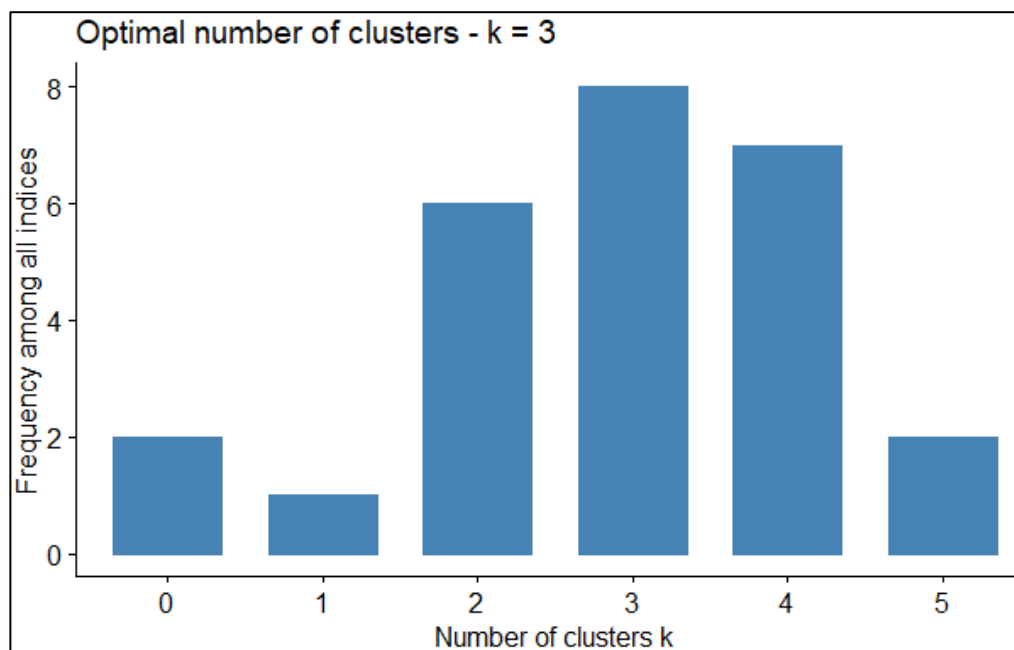
Analisis *cluster* harus memenuhi asumsi nonmultikolinearitas agar tidak ada variabel yang bersifat duplikasi atau merepresentasikan hal yang sama. Oleh karena itu, dilakukan pengujian dengan cara menghitung matriks korelasi antarvariabel.

Didapatkan bahwa nilai korelasi mutlak antarvariabel yang digunakan memiliki rentang antara 0,00 hingga 0,70. Nilai korelasi terendah terdapat antara V2 dan V4. Sedangkan, nilai korelasi tertinggi terdapat antara variabel V8 dan V11. Dengan demikian, tidak terdapat nilai korelasi antarvariabel yang melebihi  $|0,8|$ . Artinya, variabel-variabel yang digunakan telah memenuhi asumsi nonmultikolinearitas dan dapat dilanjutkan ke tahap selanjutnya.

### Penentuan Jumlah Cluster Optimal

Dalam pembentukan *cluster* ini, digunakan ukuran ketidaksamaan antarobjek berupa jarak Euclidean. Selanjutnya, dilakukan pemilihan jumlah *cluster* yang optimal untuk data yang digunakan menggunakan *by majority rule* dengan *package NbClust* di *software R*. Didapatkan hasil pada Gambar 1.

Berdasarkan hasil penentuan jumlah *cluster* yang optimal seperti tergambar pada Gambar 1, diperoleh bahwa jumlah *cluster* yang paling banyak direkomendasikan adalah 3 dengan jumlah *voting* sebanyak 8. Dengan kata lain, jumlah *cluster* yang optimal dalam kasus ini adalah 3. Oleh karena itu, akan dibentuk *cluster* yang berjumlah 3 untuk mengelompokkan provinsi-provinsi di Indonesia.



Gambar 1. Penentuan Jumlah *Cluster* Optimal

## Pengelompokan Provinsi

Dalam penelitian ini, pengelompokan provinsi dilakukan dengan menggunakan dua metode, yaitu metode *k-means* dan *k-means with outlier detection* untuk dibandingkan sehingga diperoleh metode yang terbaik untuk kasus ini.

Untuk metode *k-means*, diperoleh hasil pengelompokan yang dilakukan dengan jumlah *cluster* 3, yaitu 6 provinsi di *cluster* 1, yaitu Bangka Belitung, Kepulauan Riau, Sulawesi Utara, DKI Jakarta, DI Yogyakarta, dan Bali. Sedangkan, *cluster* 2 ada 12 provinsi, yaitu Sumatera Utara, Riau, Sumatera Selatan, Lampung, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Barat, Kalimantan Barat, Kalimantan Selatan, dan Sulawesi Barat. Kemudian, 16 provinsi lainnya di *cluster* 3, yaitu Aceh, Sumatera Barat, Jambi, Bengkulu, Nusa Tenggara Timur, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua Barat, dan Papua.

Hasil pengelompokan dengan menggunakan metode *k-means with outlier detection* dengan jumlah *cluster* 3 dan *specified outlier* sebanyak 4 adalah terdapat 12 provinsi di *cluster* 1, yaitu Sumatera Utara, Sumatera Barat, Riau, Sumatera

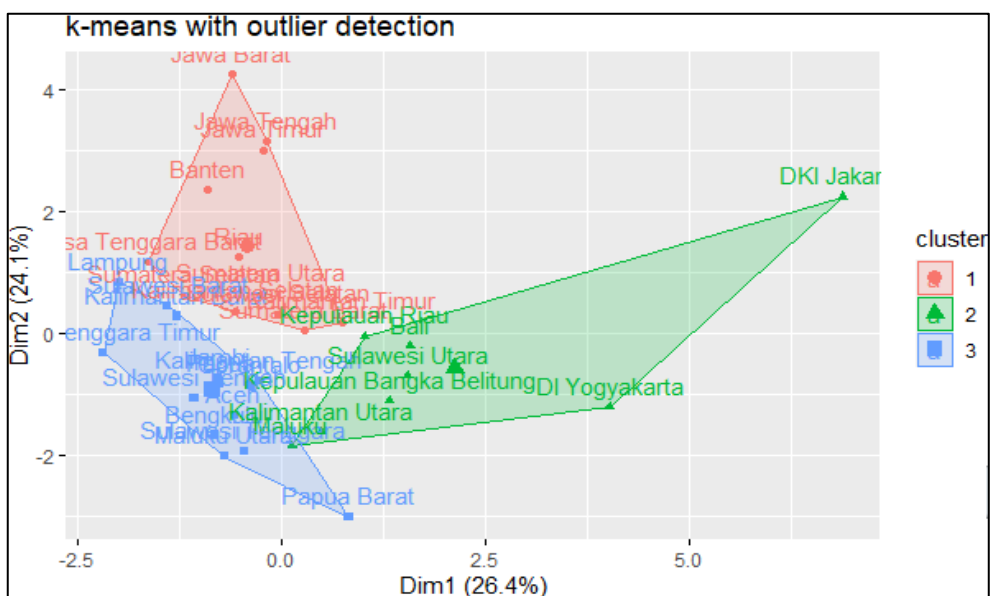
Selatan, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Barat, Kalimantan Selatan, Kalimantan Timur, dan Sulawesi Selatan. Sedangkan, 8 provinsi di *cluster* 2 adalah Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Bali, Kalimantan Utara, Sulawesi Utara, dan Maluku. Kemudian, 14 provinsi lainnya masuk ke *cluster* 3, yaitu Aceh, Jambi, Bengkulu, Lampung, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua Barat, dan Papua.

## Perbandingan Metode *K-means* dan *K-means With Outlier Detection*

Dilakukan perbandingan nilai *R-squared* antara metode *k-means* dan *k-means with outlier detection* untuk menentukan metode yang terbaik. Berdasarkan output pada *software R*, diperoleh nilai *R-squared* masing-masing metode adalah sebagai berikut.

Tabel 1. Nilai *R-squared* Metode *K-means* dan *K-means With Outlier Detection*

Metode	Nilai R-squared
<i>k-means</i>	0,3151
<i>k-means with outlier detection</i>	0,5899



Gambar 2. Visualisasi Hasil Pengelompokan

Berdasarkan perbandingan nilai *R-squared* kedua metode pada Tabel 1, didapatkan hasil bahwa nilai *R-squared* metode *k-means with outlier detection* sebesar 58,99%, nilai ini lebih besar daripada nilai *R-squared* metode *k-means*, yaitu hanya sebesar 31,51%. Hal ini menunjukkan bahwa hasil pengelompokan metode *k-means with outlier detection* lebih baik daripada metode *k-means*. Oleh karena itu, metode *k-means with outlier detection* adalah metode yang terbaik untuk mengelompokkan provinsi-provinsi di Indonesia dalam kasus ini. Berikut adalah visualisasi hasil pengelompokan dengan *k-means with outlier detection*.

Berdasarkan visualisasi hasil pengelompokan pada Gambar 2, terlihat bahwa pengelompokan dengan metode *k-means with outlier detection* telah mengelompokkan objek berdasarkan jarak terdekatnya dan cukup jelas batas-batas antar-*cluster*-nya. Selain itu, jumlah anggota setiap *cluster* telah terdistribusi sesuai karakteristiknya masing-masing dan tidak terdapat *cluster* yang memiliki jumlah anggota yang sangat minim atau bahkan satu. Hal ini dapat mengindikasikan bahwa *cluster* telah terbentuk dengan cukup baik.

### Identifikasi Karakter Cluster

Berdasarkan hasil pengelompokan dengan metode yang terbaik, yaitu metode *k-means with outlier detection*, untuk melakukan identifikasi karakteristik *cluster*, maka harus memprofilkan hasil *cluster* melalui nilai-nilai *centroid* tiap *cluster*. Berikut hasil perhitungan *centroid* tiap variabel untuk masing-masing *cluster*.

Berdasarkan hasil perhitungan *centroid* masing-masing *cluster* pada Tabel 2, *cluster* 1 memiliki infrastuktur kesehatan dan tenaga medis yang secara umum sangat minim dan merupakan yang paling sedikit dibandingkan *cluster* lain. Hal ini

ditunjukkan dari rata-rata rasio puskesmas (V1), rasio rumah sakit (V2), rasio perawat (V6), dan rasio bidan (V7) yang dimiliki oleh anggota-anggota dalam *cluster* 1 merupakan yang terendah dibandingkan 2 *cluster* lainnya. Walaupun rasio klinik (V3) dan rasio dokternya (V5) menengah, serta rasio posyandunya (V4) tertinggi, hal ini masih menunjukkan *cluster* 1 memiliki fasilitas kesehatan yang kurang memadai. Selain itu, *cluster* 1 juga memiliki jumlah kasus terkonfirmasi positif (V8), *fatality rate* (V10), dan jumlah kasus *active* (V11) yang tertinggi dibandingkan 2 *cluster* lain, walaupun *recovery rate*-nya (V9) merupakan yang tertinggi juga. Hal ini cukup mengkhawatirkan mengingat fasilitas kesehatan yang tersedia sangat minim sehingga pasien COVID-19 tidak dapat ditangani dengan baik. Oleh karena itu, *cluster* 1 merupakan kelompok provinsi dengan jumlah kasus COVID-19 tinggi, namun fasilitas kesehatan kurang memadai.

*Cluster* 2 memiliki infrastruktur kesehatan dan tenaga medis yang secara umum sangat memadai dan merupakan yang paling banyak dibandingkan *cluster* lain. Hal ini terlihat dari rasio rumah sakit (V2), rasio klinik (V3), rasio dokter (V5), dan rasio perawat (V6) yang tertinggi dibandingkan 2 *cluster* lainnya. Walaupun rasio puskesmas (V1) dan rasio bidannya (V7) menengah, serta rasio posyandu (V4) merupakan yang tertinggi, hal ini masih menunjukkan *cluster* 1 memiliki fasilitas kesehatan yang memadai. Akan tetapi, jumlah kasus terkonfirmasi positif (V8) dan jumlah kasus *active*-nya (V11) masih tergolong tinggi, walaupun tidak setinggi di *cluster* 1. Beruntungnya, hal ini dapat ditangani dengan baik mengingat fasilitas kesehatan yang tersedia sangat memadai sehingga *recovery rate* (V9) cukup tinggi dan *fatality rate* (V10) merupakan yang terendah dibandingkan 2 *cluster* lainnya.

Tabel 1. Nilai *R-squared* Metode *K-means* dan *K-means With Outlier Detection*

<i>Cluster</i>	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
1	-0,64	-0,45	-0,06	0,18	-0,29	-0,67	-0,26	0,27	0,37	0,69	0,25
2	0,06	0,88	-0,04	-0,50	0,66	0,77	-0,26	-0,38	0,33	-0,49	-0,48
3	0,49	-0,34	-0,35	0,05	-0,43	-0,00	0,63	-0,47	-0,32	-0,15	-0,50

Oleh karena itu, *cluster 2* merupakan kelompok provinsi dengan jumlah kasus COVID-19 tinggi dan fasilitas kesehatan yang cukup memadai.

*Cluster 3* memiliki infrastruktur kesehatan dan tenaga medis yang secara umum tidak terlalu tinggi dan tidak terlalu rendah dibandingkan 2 *cluster* lainnya. Hal ini ditunjukkan dengan rasio puskesmas (V1) dan rasio bidan (V7) yang merupakan tertinggi, tetapi rasio klinik (V3) dan rasio dokternya (V5) merupakan yang terendah. Kemudian, rasio rumah sakit (V2), rasio posyandu (V4), dan rasio perawat (V6) pada *cluster 3* termasuk menengah. Sedangkan jumlah kasus terkonfirmasi positif (V8), jumlah kasus *active* (V11), dan *recovery rate cluster 3* merupakan yang terendah. Untuk *fatality rate* (V10) *cluster 3* menengah dibandingkan 2 *cluster* lainnya. Oleh karena itu, *cluster 3* adalah kelompok provinsi dengan jumlah kasus COVID-19 rendah dan fasilitas kesehatan menengah.

## KESIMPULAN DAN SARAN

### Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan dapat diambil kesimpulan sebagai berikut:

1. Untuk melihat metode *clustering* mana yang lebih baik, dapat dilakukan perbandingan nilai *R-squared* antarmetode. Dalam penelitian ini, nilai *R-squared* metode *k-means* (0,3151) lebih kecil dibandingkan metode *k-means with outlier detection* (0,5899), sehingga metode *k-means with outlier detection* menghasilkan *clustering* yang lebih baik. Dengan menggunakan metode *clustering k-means with outlier detection*, diperoleh jumlah *cluster* optimum sebanyak 3 *cluster*.
2. *Cluster 1* memiliki infrastuktur kesehatan dan tenaga medis yang secara umum sangat minim dan merupakan yang paling sedikit. Oleh karena itu, *cluster 1* merupakan kelompok provinsi dengan jumlah kasus COVID-19 tinggi, namun fasilitas kesehatan kurang memadai. *Cluster* ini terdiri dari Provinsi Sumatera Utara, Sumatera Barat, Riau,

Sumatera Selatan, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Barat, Kalimantan Selatan, Kalimantan Timur, dan Sulawesi Selatan.

3. *Cluster 2* memiliki infrastruktur kesehatan dan tenaga medis yang secara umum sangat memadai dan merupakan yang paling banyak. Oleh karena itu, *cluster 2* merupakan kelompok provinsi dengan jumlah kasus COVID-19 tinggi dan fasilitas kesehatan yang cukup memadai. *Cluster* ini terdiri dari Provinsi Bangka Belitung, Kepulauan Riau, DKI Jakarta, DI Yogyakarta, Bali, Kalimantan Utara, Sulawesi Utara, dan Maluku.
4. *Cluster 3* memiliki infrastruktur kesehatan dan tenaga medis yang secara umum tidak terlalu tinggi dan tidak terlalu rendah dibandingkan 2 *cluster* lainnya. Oleh karena itu, *cluster 3* adalah kelompok provinsi dengan jumlah kasus COVID-19 rendah dan fasilitas kesehatan menengah. *Cluster* ini terdiri dari Provinsi Aceh, Jambi, Bengkulu, Lampung, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku Utara, Papua Barat, dan Papua.

### Saran

1. Kebijakan pemerintah pada *cluster 1* hendaknya meningkatkan kualitas dan kuantitas infrastruktur kesehatan dalam penanganan kasus COVID-19 yang cukup tinggi agar lebih maksimal dan mengalokasikan vaksin terlebih dahulu ke *cluster 1*. Untuk *cluster 2* harus mempertahankan kualitas dan kuantitas infrastruktur kesehatan yang memadai dalam penanganan kasus COVID-19 yang cukup tinggi. Sedangkan, untuk *cluster 3*, kualitas infrastruktur kesehatan harus ditingkatkan agar penanganan kasus COVID-19 lebih maksimal, walaupun saat ini jumlah kasusnya masih tergolong rendah.
2. Dapat dilakukan pengelompokkan menggunakan metode *clustering* lain dan indikator perbandingan antarmetode



lainnya selain *R-squared* serta penambahan *series* data terbaru agar kualitas *cluster* yang dihasilkan lebih baik.

## DAFTAR PUSTAKA

- Ariawan, Pasek Agus. 2019. Optimasi Pengelompokan Data Pada Metode - *K-Means* dengan Analisis *Outlier*. *Jurnal Nasional Teknologi dan Sistem Informasi*, Vol. 5, No. 2.
- Charrad, M., dkk.. 2014. NbClust Package for Determining the Best Number of Clusters. <https://rdrr.io/cran/NbClust/man/NbClust.html#:~:text=NbClust%20package%20provides%2030%20indices,distance%20measures%2C%20and%20clustering%20methods>. (Diakses 13 Desember 2020)
- Chawla, Sanjay dan Gionis, Aristides. (2013). *k-means--: A unified approach to clustering and outlier detection*.
- Dwitri, Nayuni, dkk. 2020. Penerapan Algoritma *K-Means* dalam Menentukan Tingkat Penyebaran Pandemi *Covid-19* di Indonesia, *Jurnal Teknologi Infomasi*, Vol. 4, No. 1, hal. 128-132.
- Ikatan Dokter Indonesia (IDI). 2020. Statistik Anggota. <http://www.idionline.org/statistik/> (Diakses 12 Desember 2020).
- Johnson, Richard A. dan Wichern, Dean W. (2007). *Applied Multivariate Statistical Analysis Sixth Edition*. United States of America : Pearson Education, Inc.
- Kementerian Kesehatan. 2019. Rekapitulasi SDM Kesehatan yang didayagunakan di Fasyankes pada tahun 2019. [http://bppsdmk.kemkes.go.id/info\\_sdmk/history/](http://bppsdmk.kemkes.go.id/info_sdmk/history/) (Diakses 11 Desember 2020).
- Khomarudin, Agus Nur. 2018. Teknik Data Mining : Algoritma *K-means Clustering*. <https://ilmukomputer.org/wp-content/uploads/2018/05/agus-k-means-clustering.pdf> . (Diakses 13 Desember 2020)
- Muhyiddin. 2020. Covid-19 *New Normal* dan Perencanaan Pembangunan di Indonesia. *The Indonesian Journal of Development Planning*, Vol. 4, No. 2.
- Noviyanto. 2020. Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia, *Jurnal Informatika dan Komputer*, Vol. 22, No. 2.
- Patel, Maulik. 2016. Clustering Based Outlier Detection Technique. <https://rpubs.com/maulikpatel/228345> (Diakses 12 Desember 2020).
- Rembulan, Glisina Dwinoor. dkk. 2020. Kebijakan Pemerintah Mengenai *Coronavirus Disease* (COVID-19) di Setiap Provinsi di Indonesia Berdasarkan Analisis Klaster, *Jurnal of Industrial Engineering and Management Systems*, Vol. 13, No. 2.
- Sindi, Sukma. dkk. 2020. Analisis Algoritma K-Medoids Clustering dalam Pengelompokan Penyebaran Covid-19 di Indonesia, *Jurnal Teknologi Informasi*, Vol. 4, No. 1.
- Solichin, Achmad dan Khansa Khairunnisa. 2020. Klasterisasi Persebaran Virus Coroana (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means, *Fountain of Informatics Journal*, Vol. 5, No. 2.

