

Analisis Kluster untuk Data Kategorik Menggunakan Metode *K-Modes* (Studi Kasus: Data Pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung)

Shabrina Novaindah Dwiyamti¹, Khoirin Nisa^{1*}, Agus Sutrisno¹, dan Netti Herawati¹

¹Jurusan Matematika, Fakultas MIPA, Universitas Lampung
Jl. Prof. Soemantri Brojonegoro No. 1 Bandar Lampung

*Email korespondensi: khoirin.nisa@fmipa.unila.ac.id

Abstrak

Analisis kluster merupakan salah satu analisis multivariat metode interdependensi dikarenakan tidak ada perbedaan antara variabel bebas dan variabel tak bebas. Analisis kluster terdiri dari metode hierarki dan non hierarki. *K-Means* merupakan salah satu metode analisis kluster non hierarki. Namun, metode *K-Means* terbatas pada data numerik. Sehingga dibutuhkan metode untuk mengolah data kategorik. Salah satu metode non hierarki untuk data kategorik yang sering digunakan adalah *K-Modes*. Penelitian ini bertujuan untuk menerapkan analisis kluster *K-Modes* pada data pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung yang berjumlah 560 data pasien dengan variabel jenis kelamin, usia, cara masuk, dan kondisi saat keluar dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung. Dengan menggunakan Davies-Bouldin Index (DBI) dan metode *Silhouette*, diperoleh hasil nilai *k* optimal untuk analisis kluster *K-Modes* adalah sebanyak 8 kluster, yaitu terdiri dari 145 pasien kluster 1, 227 pasien kluster 2, 16 pasien kluster 3, 30 pasien kluster 4, 30 pasien kluster 5, 74 pasien kluster 6, 4 pasien kluster 7, dan 34 pasien kluster 8. Karena anggota kluster 1 dan 2 terbanyak jika dibandingkan dengan kluster lainnya, maka diperlukan penanganan yang lebih optimal untuk kluster 1 dan 2.

Kata kunci : Analisis Kluster, *K-Modes*, COVID-19, DBI, Metode *Silhouette*

Abstract

Cluster analysis is a multivariate technique with interdependence because there is no difference between independent variable and dependent variable. Cluster analysis consists of hierarchical and non hierarchical methods. *K-Means* is a non hierarchical cluster analysis method. However, the *K-Means* method is limited to numerical data. So we need a method to process the categorical data. One of the non hierarchical methods for categorical data often used is *K-Modes*. This study aims to apply *K-Modes* cluster analysis to COVID-19 patients data at RSUD Dr. H. Abdul Moeloek Lampung Province, totaling 560 patient data with variables of gender, age, admission, and their condition when leaving RSUD Dr. H. Abdul Moeloek Lampung Province. By using Davies-Bouldin Index (DBI) and *Silhouette* method, the optimal *k* value for the *K-Modes* cluster analysis is 8 clusters, consisting of 145 patients in the first cluster, 227 patients in the second cluster, 16 patients in the third cluster, 30 patients in the fourth cluster, 30 patients in the fifth cluster, 74 patients in the sixth cluster, 4 patients in the seventh cluster, dan 34 patients in the eighth cluster. Because there are more members of clusters 1 and 2 when compared to other clusters, more optimal handling is needed for clusters 1 and 2.

Keywords: Cluster Analysis, *K-Modes*, COVID-19, DBI, *Silhouette* Method.

1. Pendahuluan

Statistika merupakan ilmu pengetahuan yang berkaitan dengan data dan fakta yang tepat. Statistika meliputi teknik pengumpulan data, pengolahan, penganalisisan, penarikan kesimpulan, serta pembuatan keputusan yang tepat berdasarkan fakta yang ada [1]. Dalam mengelompokkan objek-objek berdasarkan kemiripan variabel-variabel yang diamati dapat pula digunakan analisis statistika. Salah satu analisis untuk menyelesaikan masalah tersebut, yaitu analisis kluster. Analisis kluster adalah salah satu analisis multivariat yang dapat digunakan untuk meringkas dan mempermudah interpretasi data [2]. Analisis kluster bertujuan untuk mengelompokkan objek-objek ke dalam kelompok-kelompok yang relatif homogen [3]. Metode mengelompokkan kumpulan objek ke dalam satu atau lebih kelompok sehingga data yang tergabung pada suatu kelompok memiliki tingkat kesamaan disebut analisis kluster [4].

Analisis kluster adalah teknik analisis yang berfungsi untuk mengelompokkan objek-objek berdasarkan ukuran kemiripan ataupun ketidakmiripan. Semakin mirip dua objek, maka akan semakin tinggi kemungkinan berada dalam suatu kluster. Demikian juga, semakin berbeda karakteristik suatu objek, maka kemungkinan untuk berada dalam suatu kluster semakin kecil [5]. Analisis kluster terdiri dari metode hierarki dan non hierarki. Metode hierarki digunakan jika belum ada informasi jumlah kluster yang akan dipilih. Sedangkan, metode non hierarki jumlah kluster sudah ditentukan sebelumnya [6].

K-Means adalah salah satu metode analisis kluster non hierarki. Pengklasteran *K-Means* sangat sesuai jika digunakan pada data dengan ukuran yang besar dikarenakan mempunyai perhitungan yang cepat [7]. Namun, metode *K-Means* terbatas pada data numerik. Sehingga dibutuhkan metode untuk mengolah data kategorik. Salah satu metode non hierarki untuk data kategorik yang sering digunakan adalah *K-Modes*.

Metode *K-Modes* adalah metode yang paling dikenal dalam pengelompokan kumpulan data kategorik berjumlah besar ke kluster tertentu yang diwakili oleh modus atau nilai yang paling sering muncul [8]. Metode *K-Modes* mula-mula diperkenalkan oleh Huang pada tahun 1997. Metode *K-Modes* adalah modifikasi dari metode *K-Means* [9]. Metode *K-Modes* termasuk ke dalam metode partisi yang membagi objek ke-*k* kelompok yang berbeda sehingga objek yang memiliki karakteristik yang sama masuk ke dalam satu kluster yang sama dan objek yang memiliki karakteristik berbeda masuk ke dalam kluster yang lain.

Metode *K-Modes* dapat menghasilkan kluster melalui proses yang lebih detail, waktu yang dibutuhkan untuk komputasi pembentukan kluster lebih cepat, dan sangat baik digunakan dalam klusterisasi pada data berdimensi banyak [10]. Berdasarkan penelitian terdahulu dari Suciati dkk., 2021 mengkaji bahwa jumlah kluster optimal untuk pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung menggunakan metode FCM adalah 2. Sehingga, belum ada penelitian yang mengkaji mengenai analisis kluster data pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung menggunakan data kategorik.

Berdasarkan uraian di atas, penulis tertarik untuk melakukan penelitian mengenai analisis kluster menggunakan metode *K-Modes* pada data pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung, penulis menggunakan Davies-Bouldin *Index* (DBI) dan metode *Silhouette* untuk mengevaluasi jumlah kluster optimal. Tahapan yang dapat dilakukan dalam menjalankan analisis kluster metode *K-Modes* adalah sebagai berikut [11]:

1. Menentukan banyaknya kluster (*k*) yang akan dibentuk dari *r* data. Jumlah data harus lebih banyak dibandingkan dengan jumlah kluster ($r > k$).
2. Menentukan *k* modus awal dari data yang akan digunakan untuk pusat kluster (*centroid*). Banyaknya *centroid* yang dipilih harus sesuai dengan banyaknya kluster yang akan dibentuk. Contohnya, apabila ingin membentuk dua kluster, maka banyaknya *centroid* adalah dua.
3. Mencari jarak setiap objek terhadap setiap *centroid*.

Jarak dicari dengan memakai ukuran ketidakmiripan sederhana (*simple matching dissimilarity*). Asumsikan *B1* dan *B2* merupakan dua data yang mempunyai *n* variabel kategorik. Ukuran ketidakmiripan antara *B1* dan *B2* dapat didefinisikan menjadi total ketidakcocokan dari variabel-variabel kategorik yang sesuai dari kedua data. Semakin mirip kedua data, ditandai dengan nilai ketidakcocokan yang semakin kecil. Persamaan ukuran ketidakmiripan sederhana dirumuskan sebagai berikut:

$$d(V_1, V_2) = \sum_{j=1}^n \delta(x_{1j}, x_{2j}) \quad (1)$$

dengan:

$$\delta(x_{1j}, x_{2j}) = \begin{cases} 0, & x_{1j} = x_{2j} \\ 1, & x_{1j} \neq x_{2j} \end{cases}$$

x_{1j} = nilai dari variabel ke-*j* pada objek V_1

x_{2j} = nilai dari variabel ke-*j* pada objek V_2

4. Menempatkan objek berdasarkan jarak terdekat ke *centroid* menggunakan rumus jarak terdekat. Kluster mana yang paling dekat dengan objek dapat diketahui dengan nilai ukuran ketidakmiripan sederhana.
5. Apabila seluruh data sudah ditempatkan ke dalam kluster terdekat, kemudian memperbarui *centroid* setiap kluster berdasarkan modus dari setiap variabel anggota kluster yang terbentuk.
6. Mencari kembali jarak setiap data terhadap se *centroid* baru menggunakan ukuran ketidakmiripan sederhana seperti tahapan (3). Apabila pada penghitungan ulang jarak, data lebih mendekati kluster lain dibandingkan kluster saat ini, maka data ditempatkan kembali ke kluster lain yang lebih dekat tersebut.
7. Mengulangi tahapan (5) dan (6) sampai tidak ada data yang berpindah kluster.

Metode Davies-Bouldin *Index* (DBI) mula-mula diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Salah satu metode yang berfungsi untuk mengukur validitas kluster pada suatu metode pengklasteran adalah DBI. Pengukuran dengan DBI ini memaksimalkan jarak inter-kluster antara kluster C_k dan C_l dan meminimalkan jarak antartitik dalam sebuah kluster. Apabila jarak inter-kluster maksimal, artinya kesamaan karakteristik antar setiap kluster sedikit sehingga perbedaan antar-kluster terlihat lebih rinci. Apabila

jarak intra-klaster minimal artinya setiap objek dalam klaster tersebut mempunyai tingkat kesamaan karakteristik yang tinggi [12].

Metode *Silhouette* adalah metode yang berfungsi untuk mengetahui kualitas dan kekuatan dari klaster. Metode *Silhouette* adalah campuran dari dua metode yaitu metode *cohesion* yang dapat digunakan dalam menghitung seberapa dekat relasi antara objek dalam sebuah klaster, dan metode *separation* yang dapat digunakan dalam mengukur seberapa jauh sebuah klaster terpisah dengan klaster lain [13].

2. Metodologi Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder mengenai pasien COVID-19. Data tersebut diperoleh dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung. Variabel yang digunakan pada penelitian ini, yaitu jenis kelamin, usia, cara masuk, dan kondisi saat keluar dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung. Pada penelitian ini digunakan 560 data pasien, yaitu 461 data pasien pada bulan Agustus 2021, 69 data pasien pada bulan September 2021, dan 30 data pasien pada bulan Oktober 2021. Data yang digunakan adalah data kategorik. Data pada penelitian ini harus diubah ke dalam bentuk skala pengukuran nominal terlebih dahulu. Kemudian dibentuk klaster sebanyak 2, 3, 4, 5, 6, 7, dan 8. Selanjutnya dihitung nilai validitas dari hasil pengklasteran untuk melihat keefektifan dari metode *K-Modes*.

Tabel 1. Variabel yang Digunakan pada Data Penelitian

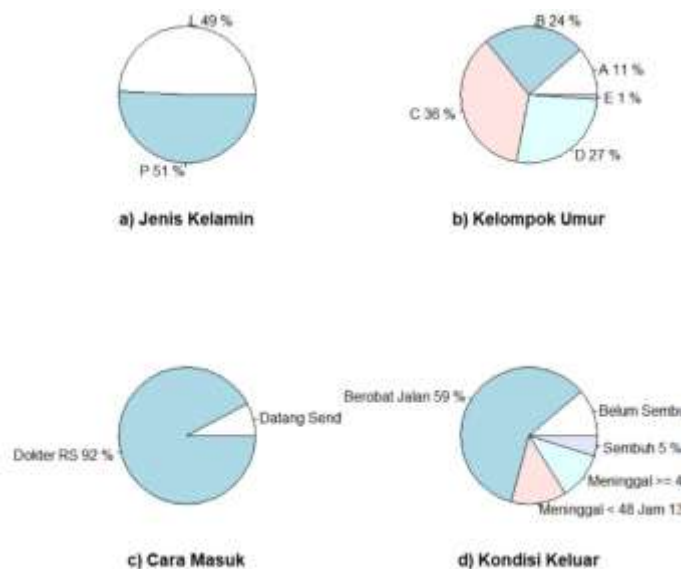
Variabel		Kategorisasi
X_1	Jenis Kelamin	0; Laki-Laki 1; Perempuan
X_2	Usia	1; 0-19 tahun 2; 20-39 tahun 3; 40-59 tahun 4; 60-79 tahun 5; 80-99 tahun
X_3	Cara Masuk	0; Rujukan dokter RS 1; Datang sendiri 1; Sembuh 2; Belum sembuh
X_4	Kondisi Saat Keluar dari RS	3; Berobat jalan 4; Meninggal < 48 jam 5; Meninggal \geq 48 jam

Data pada penelitian ini diolah menggunakan aplikasi *R-Studio* dengan tahapan sebagai berikut:

- Menentukan nilai k , pada penelitian ini digunakan $k = 2, 3, 4, 5, 6, 7, 8$
- Menentukan *centroid* (pusat klaster) dengan cara acak dari setiap klaster.
- Mencari jarak tiap objek terhadap *centroid* menggunakan ukuran ketidakmiripan sederhana.
- Menempatkan objek berdasarkan jarak terdekat ke *centroid*.
- Mencari modus dari setiap variabel sebagai *centroid* klaster.
- Menentukan ulang jarak tiap objek terhadap *centroid* baru menggunakan ukuran ketidakmiripan sederhana.
- Menempatkan objek berdasarkan jarak terdekat ke *centroid*.
- Jika terdapat objek yang berpindah klaster, maka ulangi tahapan ke lima sampai tujuh hingga tidak ada objek yang berpindah klaster.
- Menentukan k terbaik menggunakan DBI dan metode *Silhouette*.
- Menentukan profiling dari setiap klaster.

3. Hasil dan Pembahasan

Statistika deskriptif berfungsi untuk mendeskripsikan data ke dalam bentuk yang lebih mudah dipahami. Dalam penelitian ini, statistika deskriptif disajikan dalam bentuk *pie chart*. Berikut ini merupakan *pie chart* yang menunjukkan persentase dari variabel jenis kelamin, usia, cara masuk, dan kondisi saat keluar dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung.



Gambar 1. Pie Chart Data Penelitian.

Metode *K-Modes* digunakan untuk mengelompokkan sekumpulan n objek ke dalam k kluster. Metode pengklusteran ini menggunakan *modes* atau modus sebagai pusat klasternya. Pada penelitian ini dilakukan analisis kluster dengan metode *K-Modes* dengan jumlah kluster yang akan dibentuk, yaitu 2, 3, 4, 5, 6, 7, dan 8 kluster.

Langkah-langkah yang dilakukan dalam analisis kluster metode *K-Modes* dengan menggunakan data penelitian untuk jumlah kluster yang akan dibentuk sebanyak 2 kluster adalah sebagai berikut.

1. Menentukan banyaknya kluster yang akan dibentuk, yaitu $k = 2$.
2. Memilih modus awal. Modus awal dipilih secara acak (*random*). Diasumsikan V_1 yaitu data ke-11 dan V_2 yaitu data ke-12 yang terpilih sebagai modus awal (*centroid*).

Tabel 2. Objek yang Dipilih sebagai Modes Awal

<i>Centroid</i> (V_c)	Data ke-	X_1	X_2	X_3	X_4
V_1	11	1	3	0	3
V_2	12	0	3	0	3

3. Menghitung jarak objek dengan setiap *centroid* yang telah dipilih menggunakan ukuran ketidakmiripan sederhana. Perhitungan jarak setiap objek dengan *centroid* adalah sebagai berikut:

- a. Jarak objek ke-1 dan *centroid* kluster 1

$$d(T_1, V_1) = \delta(x_{11}, v_{11}) + \delta(x_{12}, v_{12}) + \delta(x_{13}, v_{13}) + \delta(x_{14}, v_{14})$$

$$d(T_1, V_1) = \delta(0, 1) + \delta(4, 3) + \delta(0, 0) + \delta(2, 3) = 1 + 1 + 0 + 1 = 3$$

Sehingga, didapatkan nilai jarak objek ke-1 dan *centroid* kluster 1 sebesar 3.

Jarak objek ke-2 dan *centroid* kluster 1

$$d(T_2, V_1) = \delta(x_{21}, v_{21}) + \delta(x_{22}, v_{22}) + \delta(x_{23}, v_{23}) + \delta(x_{24}, v_{24})$$

$$d(T_2, V_1) = \delta(0, 1) + \delta(4, 3) + \delta(1, 0) + \delta(5, 3) = 1 + 1 + 1 + 1 = 4$$

Sehingga, didapatkan nilai jarak objek ke-2 dan *centroid* kluster 1 sebesar 4.

⋮

dan selanjutnya sampai objek ke-560 ($d(T_{560}, V_1)$).

- b. Jarak objek ke-1 dan *centroid* kluster 2

$$d(T_1, V_2) = \delta(x_{11}, v_{21}) + \delta(x_{12}, v_{22}) + \delta(x_{13}, v_{23}) + \delta(x_{14}, v_{24})$$

$$d(T_1, V_2) = \delta(0, 0) + \delta(4, 3) + \delta(0, 0) + \delta(2, 3) = 0 + 1 + 0 + 1 = 2$$

Sehingga, didapatkan nilai jarak objek ke-1 dan *centroid* kluster 2 sebesar 2.

Jarak objek ke-2 dengan *centroid* kluster 2

$$d(T_2, V_2) = \delta(x_{21}, v_{21}) + \delta(x_{22}, v_{22}) + \delta(x_{23}, v_{23}) + \delta(x_{24}, v_{24})$$

$$d(T_2, V_2) = \delta(0, 0) + \delta(4, 3) + \delta(1, 0) + \delta(5, 3) = 0 + 1 + 1 + 1 = 3$$

Sehingga, didapatkan nilai jarak objek ke-2 dengan *centroid* kluster 2 sebesar 3.

⋮

dan seterusnya sampai objek ke-560 ($d(T_{560}, V_2)$).

4. Setelah diperoleh nilai jarak dari setiap objek dengan *centroid*, tahap selanjutnya adalah mengelompokkan objek berdasarkan jarak terdekat ke *centroid*.

Apabila proses pengklasteran mencapai konvergen, maka didapatkan hasil pengklasteran untuk setiap k adalah sebagai berikut:

Tabel 3. Hasil Pengklasteran $k = 2, 3, 4, 5, 6, 7, 8$

Iterasi	k	Klaster	Jumlah Anggota (n_a)	<i>Centroid</i> (V_a)
2	2	1	285	1, 3, 0, 3
		2	275	0, 3, 0, 3
2	3	1	283	1, 3, 0, 3
		2	261	0, 3, 0, 3
		3	16	0, 4, 1, 3
2	4	1	251	1, 3, 0, 3
		2	261	0, 3, 0, 3
		3	16	0, 4, 1, 3
		4	32	1, 3, 0, 2
		1	172	1, 3, 0, 3
2	5	2	261	0, 3, 0, 3
		3	16	0, 4, 1, 3
		4	82	1, 2, 0, 3
		5	28	1, 3, 0, 5
		1	147	1, 3, 0, 3
2	6	2	261	0, 3, 0, 3
		3	16	0, 4, 1, 3
		4	32	1, 3, 0, 2
		5	30	1, 3, 0, 5
		6	74	1, 2, 0, 3
		1	145	1, 3, 0, 3
		2	261	0, 3, 0, 3
		3	16	0, 4, 1, 3
2	7	4	30	1, 3, 0, 2
		5	30	1, 3, 0, 5
		6	74	1, 2, 0, 3
		7	4	1, 1, 1, 2
		1	145	1, 3, 0, 4
		2	227	0, 3, 0, 3
		3	16	0, 4, 1, 3
2	8	4	30	1, 3, 0, 2
		5	30	1, 3, 0, 5
		6	74	1, 2, 0, 3
		7	4	1, 1, 1, 2
		8	34	1, 4, 0, 3

Saat menentukan jumlah klaster terbaik digunakan metode DBI. Apabila semakin kecil nilai DBI, maka akan semakin baik pula klaster yang dihasilkan. Berikut ini disajikan contoh perhitungan nilai DBI untuk $k = 2$.

Berdasarkan Tabel 1, akan dihitung jarak antara *centroid* klaster 1 dan 2

$$d(V_1, V_2) = \delta(1, 0) + \delta(3, 3) + \delta(0, 0) + \delta(3, 3) = 1 + 0 + 0 + 0 = 1$$

Sehingga, didapatkan nilai jarak antara *centroid* klaster 1 dan 2 sebesar 1.

Selanjutnya akan dihitung jarak dari setiap anggota klaster terhadap *centroid* setiap klaster

a. Jarak antara setiap anggota klaster 1 terhadap *centroid* klaster 1

$$d(T_1, V_1) = \delta(1, 1) + \delta(3, 3) + \delta(0, 0) + \delta(2, 3) = 0 + 0 + 0 + 1 = 1$$

Sehingga, didapatkan nilai jarak antara data ke-1 (anggota klaster 1) terhadap *centroid* klaster 1 sebesar 1.

⋮

dan seterusnya sampai objek ke-285

b. Jarak antara setiap anggota klaster 2 terhadap *centroid* klaster 2

$$d(T_3, V_2) = \delta(0, 0) + \delta(4, 3) + \delta(0, 0) + \delta(2, 3) = 0 + 1 + 0 + 1 = 2$$

Sehingga, didapatkan nilai jarak antara data ke-3 (anggota klaster 2) terhadap *centroid* klaster 2 sebesar 2.

⋮

dan seterusnya sampai objek ke-275

Menghitung ukuran dispersi setiap kluster

a. Ukuran dispersi pada kluster 1

$$s_1 = \left[\frac{1}{285} \times (1^2 + 1^2 + 1^2 + \dots + 1^2) \right]^{\frac{1}{2}} = 1,354$$

Sehingga, didapatkan nilai dispersi kluster 1 sebesar 1,354.

b. Ukuran dispersi pada kluster 2

$$s_2 = \left[\frac{1}{275} \times (2^2 + 3 + 1^2 + \dots + 1^2) \right]^{\frac{1}{2}} = 1,340$$

Sehingga, didapatkan nilai dispersi kluster 2 sebesar 1,340.

Dikarenakan dalam penentuan jumlah kluster optimal membutuhkan nilai DBI terkecil, maka ukuran dispersi yang akan dicari pun harus yang terkecil. Ukuran kemiripan antara kluster 1 dan 2 diberikan oleh

$$R_{1,2} = \frac{s_1 + s_2}{d(V_1, V_2)} = \frac{1,354 + 1,340}{1} = 2,694$$

Sehingga, didapatkan nilai kemiripan antara kluster 1 dan 2 sebesar 2,694.

Jika dibuat dalam bentuk matriks, maka

$$R = \begin{bmatrix} - & 2,694 \\ 2,694 & - \end{bmatrix}$$

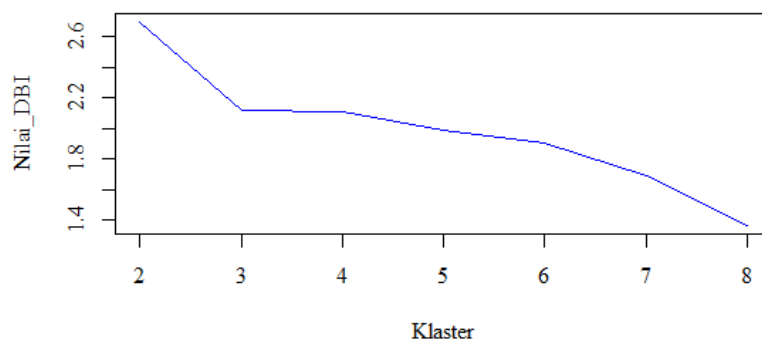
Sehingga nilai DBI untuk $k = 2$ adalah:

$$DBI = \frac{1}{2} \sum_{a=1}^2 R_a = \frac{1}{2} (2,694 + 2,694) = 2,694$$

Jadi, didapatkan nilai DBI untuk $k = 2$ adalah 2,694.

Tabel 4. Nilai DBI untuk $k = 2, 3, 4, 5, 6, 7, 8$

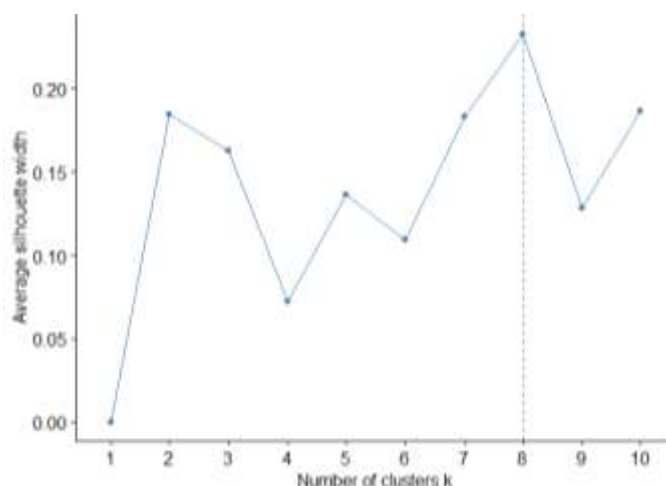
k	DBI
2	2,694
3	2,121
4	2,105
5	1,992
6	1,910
7	1,693
8	1,368



Gambar 2. Grafik Nilai DBI.

Berdasarkan Tabel 4 dan Gambar 2, didapatkan nilai DBI terkecil, yaitu 1,368 yang dimiliki oleh $k = 8$. Karena semakin kecil nilai DBI, maka akan semakin baik pula kluster yang dihasilkan.

Pada metode *Silhouette*, untuk menentukan k optimal, dilakukan dengan cara melihat nilai rata-rata *Silhouette* yang paling tinggi. Berikut ini merupakan grafik yang menunjukkan nilai rata-rata *Silhouette*.



Gambar 3. Visualisasi k Optimal Metode *Silhouette*.

Berdasarkan Gambar 3, terlihat bahwa nilai rata-rata *Silhouette* yang paling tinggi saat $k = 8$.

Dari kedua metode yang digunakan dalam menentukan banyak kluster, didapatkan hasil nilai k optimal untuk analisis kluster *K-Modes* adalah sebanyak 8 kluster.

Tabel 5. Karakteristik Setiap Kluster untuk $k = 8$

Kluster	Modus Setiap Variabel			
	X_1	X_2	X_3	X_4
1	Perempuan	40-59 tahun	Rujukan dokter RS	Meninggal < 48 jam
2	Laki-laki	40-59 tahun	Rujukan dokter RS	Berobat jalan
3	Laki-laki	60-79 tahun	Datang sendiri	Berobat jalan
4	Perempuan	40-59 tahun	Rujukan dokter RS	Belum sembuh
5	Perempuan	40-59 tahun	Rujukan dokter RS	Meninggal \geq 48 jam
6	Perempuan	20-39 tahun	Rujukan dokter RS	Berobat jalan
7	Perempuan	0-19 tahun	Datang sendiri	Belum sembuh
8	Perempuan	60-79 tahun	Rujukan dokter RS	Berobat jalan

Berdasarkan hasil pengklusteran di atas, didapatkan beberapa informasi mengenai karakteristik pasien COVID-19 yang ada di RSUD Dr. H. Abdul Moeloek Provinsi Lampung. Pada variabel usia, terdapat kesamaan *centroid* pada kluster 1, 2, 4, dan 5, yaitu pada usia 40-59 tahun. Sehingga, untuk masyarakat yang berada di usia pada interval tersebut, diharapkan untuk lebih menjaga kesehatan dan menerapkan protokol kesehatan agar terhindar dari COVID-19. Selanjutnya, pada *centroid* kluster 1, 2, 4, 5, 6, dan 8 variabel cara masuk, cara masuk dengan rujukan dari dokter RSUD Dr. H. Abdul Moeloek Provinsi Lampung mempunyai frekuensi yang lebih tinggi jika dibandingkan dengan *centroid* kluster lainnya, yaitu datang sendiri. Sehingga, apabila masyarakat telah melakukan kontak erat dengan seseorang yang menderita COVID-19, diharapkan segera melakukan *swab* di fasilitas pelayanan kesehatan agar dapat segera mendapatkan penanganan lebih lanjut.

Kebijakan yang dapat diambil oleh RSUD Dr. H. Abdul Moeloek Provinsi Lampung adalah penanganan yang lebih optimal untuk kluster 1 dan 2, hal ini dikarenakan anggota kluster 1 dan 2 mempunyai anggota kluster terbanyak jika dibandingkan kluster lainnya. Kebijakan yang diambil dapat berupa mengutamakan vaksinasi untuk kelompok usia 40-59 tahun untuk mengurangi gejala yang dialami pasien sehingga dapat mengurangi angka positif dan juga kondisi pasien saat keluar jadi lebih membaik.

4. Kesimpulan

Berdasarkan hasil dan pembahasan yang telah diuraikan, maka dapat diambil kesimpulan bahwa analisis kluster metode *K-Modes* merupakan metode yang dapat digunakan untuk tipe data kategorik. Pengelompokkan data pasien COVID-19 di RSUD Dr. H. Abdul Moeloek Provinsi Lampung terbagi menjadi 8 kluster berdasarkan jenis kelamin, usia, cara masuk, dan kondisi saat keluar dari RSUD Dr. H. Abdul Moeloek Provinsi Lampung. Dari hasil penelitian, dominasi kluster 1 dan kluster 2 cukup signifikan karena setiap beranggotakan 145 dan 227 dari 560 pasien. Lebih jauh, penelitian dapat dikembangkan dengan meninjau perbandingan antara analisis kluster menggunakan metode *K-Modes* dan metode *ROCK (Robust Clustering Using Links)*. Hal ini dikarenakan metode *ROCK* juga dapat digunakan untuk pengklusteran menggunakan tipe data kategorik.

Daftar Pustaka:

- [1] Wanitaningsih, S. K. 2012. Program Statistik: Upaya Memotivasi Mahasiswa untuk Senang Belajar Statistika. *Media Bina Ilmiah*. 6(4): 62-67.
- [2] Santosa, S., Panjidang, W. L., & Santosa, Y. P. 2019. Klasterisasi Kecerdasan Majemuk Siswa Berbasis Jaringan Syaraf Kohonen Guna Mendukung Adaptive Elearning. *Jurnal Teknologi Informasi*. 15(2): 123-141.
- [3] Wibowo, R. A., Nisa, K., Faisol, A., & Setiawan, E. 2020. Simulasi Pemilihan Metode Analisis Cluster Hirarki Agglomerative Terbaik Antara Average Linkage dan Ward pada Data yang Mengandung Masalah Multikolinearitas. *Jurnal Siger Matematika*. 1(2): 49-55.
- [4] Rohmatullah, A., Rahmalia, D., & Pradana, M. S. 2019. Klasterisasi Data Pertanian di Kabupaten Lamongan Menggunakan Algoritma K-Means dan Fuzzy C-Means. *Jurnal Ilmiah Teknosains*. 5(2): 86-93.
- [5] Chrisinta, D., Sumertajaya, I. M., & Indahwati. 2020. Evaluasi Kinerja Metode Klaster Ensemble dan Latent Class Klastering pada Peubah Campuran. *Indonesian Journal of Statistics and Its Applications*. 4(3): 448-461.
- [6] Johnson, R.A. & Wichern, D. 2007. *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey.
- [7] Ananta, A. Y. 2017. Seleksi Notifikasi Serangan Berbasis IDS Snort Menggunakan Metode K-Means. *SMARTICS Journal*. 3(2): 31-38.
- [8] Khan, S. S. & Ahmad, A. 2013. Cluster Center Initialization Algorithm For K-Modes Clustering. *Expert Systems with Applications*. 40(18): 7444-7456.
- [9] Aranganayagi, S. & Thangavel, K. 2009. Improved K-Modes for Categorical Clustering Using Weighted Dissimilarity Measure. *International Journal of Computer and Information Engineering*. 3(3): 729-735.
- [10] Kartikasari, Y. A. P., Pranoto, Y. A., & Rudhistiar, D. 2021. Penerapan Metode K-Modes untuk Proses Penentuan Penerima Bantuan Langsung Tunai (BLT). *Jurnal Mahasiswa Teknik Informatika*. 5(1): 389-397.
- [11] Az-zahra, A. A., Marsaoly, A. F., Lestyani, I. P., Salsabila., & Madjida, W. O. Z. 2021. Penerapan Algoritma K-Modes Klastering dengan Validasi Davies Bouldin Index pada Pengelompokan Tingkat Minat Belanja Online di Provinsi Daerah Istimewa Yogyakarta. *Jurnal Matematika dan Statistika serta Aplikasinya*. 9(1): 24:36.
- [12] Wani, M. A. & Riyaz, R. 2017. A Novel Point Density Based Validity Index for Clustering Gene Expression Datasets. *Int. J. Data Mining and Bioinformatics*. 17(1): 66-84.
- [13] Pramesti, D. F., Furqon, M. T., & Dewi, C. 2017. Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*. 1(9): 723-732.