

Implementasi Algoritma *K-Means Clustering* dengan Jarak *Euclidean* dalam Mengelompokkan Daerah Penyebaran *COVID-19* di Kabupaten Bogor

Muhammad Iqbal¹, Syaripuddin^{2,*}, Moh. Nurul Huda³

^{1,2,3} *Laboratorium Matematika Komputasi Jurusan Matematika FMIPA Universitas Mulawarman*

Dikirim: November 2022; Diterima: Maret 2023; Dipublikasi: Maret 2023

Alamat Email Korespondensi: syarifrahman2014@gmail.com

Abstrak

K-Means clustering merupakan algoritma pengelompokan iteratif yang melakukan partisi himpunan data ke dalam sejumlah K kluster yang sudah ditentukan di awal, sedemikian sehingga setiap data masuk dalam pusat kluster (*centroid*) terdekat. Pada penelitian dilakukan pengelompokan data menggunakan metode *K-Means clustering* dengan jarak *Euclidean* pada data daerah penyebaran *COVID-19* di Kabupaten Bogor. Penentuan jumlah kluster optimal yang ditentukan sejak awal menggunakan metode *Elbow* dan *Silhouette Index*. Penelitian bertujuan untuk mengetahui hasil dari penerapan algoritma *K-means clustering* untuk mengelompokkan daerah penyebaran *COVID-19* di Kabupaten Bogor. Adapun variabel yang digunakan yaitu kepadatan penduduk (X_1), jumlah kasus terinfeksi *COVID-19* (X_2) dan jumlah kasus meninggal *COVID-19* (X_3). Berdasarkan hasil penelitian diperoleh dua kluster tingkat risiko penyebaran *COVID-19* yaitu kluster C1 dan C2. Pada kluster C1 memiliki tingkat risiko penyebaran sedang terdiri atas 36 Kecamatan, sedangkan pada kluster C2 dengan tingkat risiko penyebaran tinggi terdiri atas 4 Kecamatan. Nilai validitas yang didapat sebesar 0.71676 yang berarti struktur dari hasil kluster tersebut kuat.

Kata Kunci:

COVID-19, Jarak Euclidean, K-Means Clustering, Metode Elbow, Silhouette Index

PENDAHULUAN

Pada awal tahun 2020 dunia dilanda pandemi virus Corona atau *COVID-19*. *COVID-19* (*Coronavirus Disease 2019*) adalah penyakit menular yang disebabkan oleh virus *SARS-CoV-2* (*Severe Acute Respiratory Syndrome Coronavirus-2*). Orang yang terinfeksi virus ini akan mengalami penyakit pernapasan ringan hingga sedang dan sembuh tanpa memerlukan perawatan khusus. Namun, beberapa kasus akan mengalami sakit yang berat dan memerlukan perhatian medis [1].

Kasus virus *pneumonia* yang tidak diketahui, sekarang dikenal *COVID-19* pertama kali dilaporkan di Kota Wuhan, Provinsi Hubei, China pada Desember 2019 [2]. Sementara di Indonesia kasus *COVID-19* pertama kali dilaporkan pada 2 Maret 2020, ketika dua orang terkonfirmasi tertular dari seorang warga negara Jepang. Setelah terkonfirmasi adanya kasus terinfeksi *COVID-19* atau virus Corona, pemerintah Indonesia melakukan berbagai cara atau upaya untuk menghindari terjadinya penularan *COVID-19* yang berkepanjangan di Indonesia. Pada tanggal 9 April 2020, pandemi sudah menyebar ke 34

provinsi, salah satunya provinsi Jawa Barat, virus ini telah menyebar ke 27 Kabupaten/Kota di Jawa Barat salah satunya adalah Kabupaten Bogor [3].

Penyebaran kasus *COVID-19* yang cukup cepat dan merata di seluruh provinsi di Indonesia berdampak negatif pada seluruh bidang. Luasnya wilayah Indonesia memungkinkan diperlukannya *clustering* (pengelompokan) berdasarkan wilayah di Indonesia khususnya Kabupaten Bogor, sebab Kabupaten Bogor merupakan Kabupaten/Kota dengan populasi penduduk terbesar di Indonesia [4]. *Clustering* ini akan menghasilkan titik-titik pusat penyebaran dan risiko penyebaran kasus *COVID-19* di Kabupaten Bogor.

Algoritma *K-Means clustering* merupakan salah satu metode *clustering* non-hierarki yang berusaha mengelompokkan data ke dalam satu kelompok yang sama [5]. Ukuran kesamaan yang digunakan adalah ukuran jarak antar objek. Kedua objek yang memiliki jarak paling dekat akan bergabung menjadi satu kelompok. Kedekatan jarak yang dimiliki menunjukkan bahwa kedua objek tersebut memiliki tingkat kesamaan karakteristik. Pada penelitian ini untuk menentukan jumlah kluster yang optimal dalam mengelompokkan daerah penyebaran *COVID-19* menggunakan metode *Elbow* dan *Silhouette Index*.

Penelitian terkait penerapan algoritma *K-Means clustering* telah dilakukan oleh beberapa peneliti sebelumnya, antara lain pada penelitian Sari dan Sukestiyarno [6]. Penelitian tersebut bertujuan untuk mengetahui tingkat persebaran kasus *COVID-19* kategori tinggi, sedang dan rendah pada masing-masing provinsi di Indonesia dengan menggunakan metode *K-Means clustering*. Berdasarkan pada beberapa penelitian sebelumnya untuk mengetahui tingkat penyebaran kasus *COVID-19* dengan tingkat risiko tinggi, sedang dan rendah di Kabupaten Bogor, ada beberapa aspek yang dapat diukur seperti kepadatan penduduk, jumlah kasus terinfeksi *COVID-19* dan jumlah kasus kematian *COVID-19*.

Berdasarkan uraian di atas, penelitian ini mengkaji tentang implementasi algoritma *K-Means clustering* dengan jarak *Euclidean* dalam mengelompokkan daerah penyebaran *COVID-19* di Kabupaten Bogor. Tujuan penelitian ini adalah mengetahui cara menentukan jumlah kluster yang paling optimal dengan menggunakan metode *Elbow* dan *Silhouette Index* dan mengetahui karakteristik kluster yang terbentuk berdasarkan tingkat risiko penyebaran *COVID-19* di Kabupaten Bogor.

METODE PENELITIAN

Penelitian ini termasuk ke dalam penelitian kuantitatif dengan populasi yang digunakan adalah seluruh data kepadatan penduduk, data jumlah kasus terinfeksi *COVID-19* dan data jumlah kasus kematian *COVID-19* Kabupaten Bogor. Adapun sampel pada penelitian ini adalah data kepadatan penduduk (km²), data jumlah kasus terinfeksi *COVID-19* (jiwa) dan data jumlah kasus kematian *COVID-19* (jiwa) pada rentang 1 sampai 31 Juli 2021. Adapun tahapan analisis data yang dilakukan pada penelitian ini adalah sebagai berikut:

1. Pengumpulan Data
2. Normalisasi Data

Menormalisasikan data dapat mempercepat tahap pembelajaran dan dapat mengatasi masalah pada perbedaan rentang nilai yang jauh pada setiap variabel [7]. Normalisasi data pada penelitian ini menggunakan metode *Min-max normalization*. Adapun rumus persamaannya sebagai berikut:

$$x' = \left(\frac{x - X_{min}}{X_{max} - X_{min}} \right), \quad (1)$$

keterangan:

- x = nilai data x ,
- x' = nilai baru data x ,
- X_{min} = nilai minimum pada variabel X ,
- X_{max} = nilai maksimum pada variabel X .

3. Menentukan Nilai K

Yaitu menentukan jumlah kluster terbaik atau paling optimal dalam mengelompokkan data menggunakan metode *Elbow* dan *Silhouette Index*. Adapapun rumus persamaan metode *Elbow* sebagai berikut:

$$SSE = \sum_{k=1}^K \sum_{x_i} |x_i - c_k|^2, \quad (2)$$

keterangan:

- K = banyak kluster,
- x_i = nilai variabel ke- i dari data x ,
- c_k = *centroid* kluster ke- k .

4. Metode *K-Means Clustering*

Mengelompokkan set data dengan menggunakan algoritma *K-Means Clustering*. Adapapun algoritma *K-Means clustering* adalah sebagai berikut:

- i. Inisialisasi: menentukan nilai K sebagai jumlah kluster yang diinginkan dan metrik ketidakmiripan (jarak). Pada penelitian ini digunakan metrik ketidakmiripan jarak *Euclidean*. Adapapun rumus persamaannya sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (3)$$

keterangan:

- $d(x, y)$ = jarak data x ke data y ,
- x_i = nilai variabel ke- i dari data x ,
- y_i = nilai variabel ke- i dari data y .

- ii. Memilih K data dari set data X sebagai *centroid* atau pusat kluster.
- iii. Mengalokasikan semua data ke *centroid* terdekat dengan metrik jarak yang sudah ditetapkan (memperbarui kluster pada setiap data).
- iv. Menghitung kembali *centroid* C berdasarkan data yang mengikuti kluster masing-masing.
- v. Mengulangi langkah ke iii dan iv hingga kondisi konvergen tercapai, yaitu
 - a. Perubahan fungsi objektif sudah di bawah ambang batas yang diinginkan; atau
 - b. Tidak ada data yang berpindah kluster; atau
 - c. Perubahan posisi *centroid* sudah di bawah ambang batas yang ditetapkan.

5. Menghitung Nilai Validitas

Melakukan perhitungan hasil *clustering* dengan *Silhouette Index* untuk mengukur validitas hasil kluster. Adapapun untuk menghitung nilai *SI* dari sebuah data ke- i , ada dua komponen yaitu a_i dan b_i . a_i adalah rata-rata jarak data ke- i terhadap semua data lainnya dalam satu kluster, sedangkan b_i adalah rata-rata jarak data ke- i terhadap semua data dari kluster lain yang tidak dalam satu kluster dengan data ke- i , kemudian diambil yang terkecil. Adapapun rumus persamaannya sebagai berikut:

$$a_i^j = \frac{1}{m_j - 1} \sum_{\substack{r=1 \\ r \neq i}}^{m_j} d(x_i^j, x_r^j), \quad i = 1, 2, \dots, m_j. \quad (4)$$

$d(x_i^j, x_r^j)$ adalah jarak data ke- i dengan data ke- r dalam satu kluster j , sedangkan m_j adalah jumlah data dalam kluster ke- j . Sementara rumus b_i^j :

$$b_i^j = \min_{\substack{n=1, \dots, k \\ n \neq j}} \left\{ \frac{1}{m_n} \sum_{\substack{r=1 \\ r \neq i}}^{m_n} d(x_i^j, x_r^n) \right\}, \quad i = 1, 2, \dots, m_n, \quad (5)$$

untuk mendapatkan *Silhouette Index* (SI) data ke- i menggunakan persamaan berikut:

$$SI_i^j = \frac{b_i^j - a_i^j}{\max\{a_i^j, b_i^j\}}. \quad (6)$$

Sementara nilai SI global didapatkan dengan menghitung rata-rata SI dari semua kluster seperti pada persamaan berikut:

$$SI = \frac{1}{k} \sum_{j=1}^k SI_j. \quad (7)$$

Interpretasi nilai dari *Silhouette Index* menurut Kauffman dan Rousseeuw [8] ditunjukkan pada Tabel 1.

Tabel 1. Interpretasi Nilai *Silhouette Index*

<i>Silhouette Index</i>	Interpretasi
0.71 – 1.00	Struktur yang dihasilkan kuat
0.51 – 0.70	Struktur yang dihasilkan baik
0.26 – 0.50	Struktur yang dihasilkan lemah
≤ 0.25	Tidak terstruktur

HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini berjumlah 40 data, masing-masing data merepresentasikan Kecamatan di Kabupaten Bogor. Adapun Variabel X_1 adalah data kepadatan penduduk yang diambil dari data Badan Pusat Statistik Kabupaten Bogor [9]. Sementara X_2 adalah data jumlah kasus terinfeksi *COVID-19* dan X_3 adalah jumlah kasus kematian *COVID-19* yang diambil dari laman Geoportal Kabupaten Bogor [10]. Data tersebut dapat dilihat pada Tabel 2.

Tabel 2. Data *COVID-19* Kabupaten Bogor

No.	Kecamatan	X_1	X_2	X_3
1.	Babakan Madang	2,619.84	312	3
2.	Bojonggede	10,145.32	1,671	52
3.	Caringin	2,777.83	236	0
⋮	⋮	⋮	⋮	⋮
38.	Tanjungsari	667.54	109	1
39.	Tenjo	880.51	114	1
40.	Tenjolaya	1,539.12	51	3

Normalisasi Data

Selanjutnya dilakukan normalisasi data pada Tabel 2 dengan rentang 0 sampai 1 menggunakan Persamaan (1)

$$\begin{aligned}
 x'_{1,1} &= \left(\frac{x_{1,1} - X_{min}}{X_{max} - X_{min}} \right) \\
 &= \left(\frac{2,619.84 - 303.33}{10,145.32 - 303.33} \right) = 0.23537 \\
 &\vdots \\
 x'_{40,1} &= \left(\frac{x_{40,1} - X_{min}}{X_{max} - X_{min}} \right) \\
 &= \left(\frac{1,539.12 - 303.33}{10,145.32 - 303.33} \right) = 0.12556
 \end{aligned}$$

Perhitungan yang sama dilakukan seterusnya pada setiap data sampai dengan data ke-40 dan dilanjutkan untuk semua variabel. Berikut adalah hasil normalisasi data yang diperoleh, dapat dilihat pada Tabel 3.

Tabel 3. Data Normalisasi

No.	X_1	X_2	X_3
1.	0.23537	0.09859	0.03947
2.	1.00000	0.57711	0.68421
3.	0.25142	0.07183	0.00000
⋮	⋮	⋮	⋮
38.	0.03701	0.02711	0.01316
39.	0.05864	0.02887	0.01316
40.	0.12556	0.00669	0.03947

Menentukan Nilai K yang Optimal

Pada penelitian ini untuk menentukan nilai K atau jumlah kluster yang optimal menggunakan dua metode yakni Metode *Elbow* dan *Silhouette Index*, kedua metode ini digunakan untuk memperkuat dan memastikan jumlah kluster yang paling optimal pada data di penelitian.

1. Metode *Elbow*

Grafik hasil perhitungan nilai *SSE* (*Sum of Squared Error*) menggunakan *syntax* Python dapat dilihat pada Gambar 1.

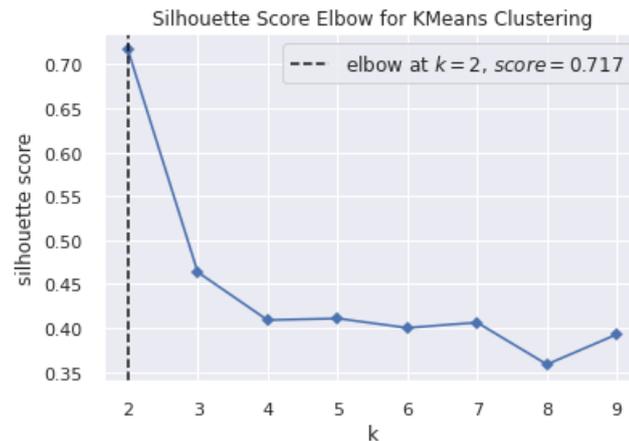


Gambar 1. Grafik Hasil Perhitungan *SSE* Metode *Elbow*

Berdasarkan metode *Elbow* didapatkan jumlah kluster terbaik yang dilakukan pada penelitian ini adalah sebanyak dua kluster. pada saat $K = 2$ nilai *SSE* sebesar 2.02272, setelah melewati $K = 2$ nilai *SSE* cenderung stabil, serta terdapat sedikit sudut siku pada saat $K = 2$. Maka pada penelitian digunakan $K = 2$ atau membagi data ke dalam dua kluster.

2. *Silhouette Index*

Sementara itu, hasil perhitungan nilai *Silhouette Index* dengan *syntax* pemrograman Python dapat dilihat pada Gambar 2.



Gambar 2. Grafik Rata-rata Nilai *Silhouette Index*

Berdasarkan Gambar 2, maka didapatkan jumlah kluster paling optimal menggunakan *Silhouette Index* sebanyak dua kluster. Karena nilai rata-rata *Silhouette Index* tertinggi adalah ketika $k = 2$.

Metode *K-Means Clustering*

Pada tahap ini dilakukan proses utama yaitu penerapan algoritma *K-Means clustering*. Berdasarkan metode *Elbow* dan *Silhouette Index* jumlah kluster pada penelitian ini menggunakan dua kluster, kluster pertama direpresentasikan dengan C1 dan kluster kedua direpresentasikan dengan C2. Selanjutnya dipilih dua data sebagai *centroid* atau pusat kluster yang terdapat pada Tabel 4.

Tabel 4. Data *Centroid* Awal

C1				C2	
0.18895	0.44930	0.32895	1.00000	0.57711	0.68421

Selanjutnya adalah menghitung jarak antara data dengan kedua *centroid* menggunakan jarak *Euclidean* pada Persamaan (3).

$$\begin{aligned}
 d_{1,c1} &= \sqrt{\sum_{i=1}^3 (x_i - C_1)^2}, \\
 &= \sqrt{(0.23537 - 0.18895)^2 + (0.09859 - 0.44930)^2 + (0.03947 - 0.32895)^2}, \\
 &= \sqrt{0.20895}, = 0.45710, \\
 d_{1,c2} &= \sqrt{\sum_{i=1}^3 (x_i - C_2)^2},
 \end{aligned}$$

$$= \sqrt{(0.23537 - 0.25142)^2 + (0.09859 - 0.07183)^2 + (0.03947 - 0.00000)^2},$$

$$= \sqrt{1.22933} = 1.10875.$$

Hasil perhitungan data ke-1 didapatkan jarak data ke *centroid* terdekat adalah ke C1 yakni sebesar 0.45710, maka data pertama masuk kedalam kelompok atau kluster C1. Selanjutnya dilakukan perhitungan yang sama pada seluruh data ke setiap *centroid* C1 dan C2, hasil perhitungan iterasi ke-1 dapat dilihat pada Tabel 5.

Tabel 5. Perhitungan Iterasi ke-1

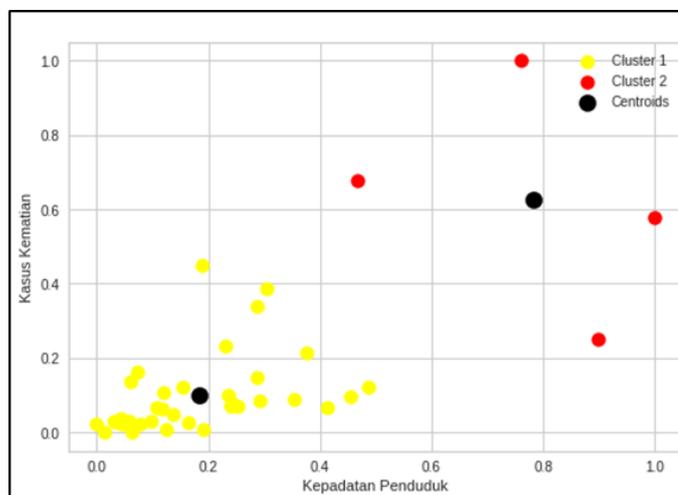
No.	Jarak Data ke <i>Centroid</i>		Cluster
	C1	C2	
1.	0.45710	1.10875	C1
2.	0.89462	0.00000	C2
3.	0.50457	1.13306	C1
⋮	⋮	⋮	⋮
38.	0.54868	1.29621	C1
39.	0.54172	1.27946	C1
40.	0.53265	1.22707	C1

Selanjutnya dihitung rata-rata data tiap kluster untuk mendapatkan nilai *centroid* baru. Lalu kembali hitung jarak tiap data dengan *centroid* baru sampai keadaan konvergen. Pada penelitian ini didapatkan hasil konvergen pada iterasi ke-3. Pada kluster C1 terdapat 36 Kecamatan sementara pada kluster C2 terdiri atas 4 Kecamatan, hasil tersebut dapat dilihat pada Tabel 6.

Tabel 6. Hasil *K-Means Clsuetring*

Cluster	
C1	<ul style="list-style-type: none"> Babakan Madang, Caringin, Cariu, Ciampea, Ciawi, Cibungbulang, Cigombong, Cigudeg, Cijeruk, Cileungsi, Cisarua, Ciseeng, Citeureup, Dramaga, Gunungsindur, Jasinga, Jonggol, Kemang, Klapanunggal, Leuwiliang, Leuwisadeng, Megamendung, Nanggung, Pamijahan, Parung, Parungpanjang, Rancabungur, Rumpin, Sukajaya, Sukaraja, Sukamakmur, Tajurhalang, Tamansari, Tanjungsari, Tenjo, Tenjolaya
C2	<ul style="list-style-type: none"> Bojonggede, Cibinong, Ciomas, Gunung Putri

Hasil algoritma *K-Means clustering* pada penelitian dapat dilihat dalam bentuk visualisasi grafik *scatter*, visualisasi sebaran *dataset* dapat dilihat pada Gambar 3.

Gambar 3. Grafik Plot Data Hasil *K-Means Clustering*

Data kluster C1 direpresentasikan dengan warna kuning, sementara data kluster C2 direpresentasikan dengan warna merah dan warna hitam adalah *centroid*.

Daerah pada kluster C1 memiliki rata-rata kepadatan penduduk 2,108.49 orang/km², rata-rata kasus terinfeksi *COVID-19* sebanyak 315.44 orang dan rata-rata kasus kematian 6.03 orang. Sementara daerah pada kluster C2 memiliki rata-rata kepadatan penduduk 7,994.66 orang/km², rata-rata kasus terinfeksi *COVID-19* sebanyak 1,809 orang dan rata-rata kasus kematian 48.05 orang. Berdasarkan nilai rata-rata tersebut daerah pada kluster C1 memiliki kepadatan penduduk, kasus terinfeksi *COVID-19* dan kasus kematian lebih rendah dibandingkan dengan daerah pada kluster C2. Maka dapat diartikan bahwa daerah kluster C1 termasuk dalam Kecamatan yang memiliki tingkat risiko penyebaran *COVID-19* yang sedang, sementara daerah pada kluster C2 memiliki tingkat risiko penyebaran *COVID-19* yang tinggi.

Menghitung Nilai Validitas dengan *Silhouette Index*

Pada tahap ini dilakukan validasi hasil kluster yang telah dilakukan menggunakan *Silhouette Index*. Menghitung nilai rata-rata *SI* secara manual dengan jumlah kluster $k = 2$. Pertama adalah mencari nilai a_i untuk data yang berada pada kluster C2 dengan Persamaan (4)

$$\begin{aligned} a_1^{C2} &= \frac{1}{4-1} \sum_{r=1}^4 d(x_1^1, x_r^1), \\ &= \frac{1}{4-1} (d(x_1^1, x_1^1) + d(x_1^1, x_2^1) + d(x_1^1, x_3^1) + d(x_1^1, x_4^1)), \\ &= \frac{1}{3} (0.00000 + 0.57926 + 0.57443 + 0.54389) = 0.56586. \end{aligned}$$

Hal yang sama dilakukan pada data lainnya di kluster C2. Perhitungan yang sama juga dilakukan pada data di kluster C1.

Selanjutnya yaitu menghitung nilai b_i untuk data yang berada pada kluster C2 menggunakan Persamaan (5)

$$\begin{aligned} b_1^{C2} &= \min_{n=1} \left\{ \frac{1}{36} \sum_{r=1}^{36} d(x_1^{C2}, x_r^{C1}) \right\}, \\ &= \min \left\{ \frac{1}{36} (1.10875 + 1.13306 + \dots + 1.22707) \right\}, \\ &= \min\{1.13016\} = 1.13016. \end{aligned}$$

Hal yang sama dilakukan pada data lainnya di kluster C2. Perhitungan yang sama juga dilakukan pada data di kluster C1.

Selanjutnya yaitu menghitung nilai SI_i (*Silhouette Index*) untuk data yang berada pada kluster C2 dengan menggunakan Persamaan (6)

$$\begin{aligned} SI_1^{C2} &= \frac{b_1^{C2} - a_1^{C2}}{\max\{a_1^{C2}, b_1^{C2}\}}, \\ &= \frac{1.13016 - 0.56586}{\max\{1.13016, 0.56586\}}, \\ &= \frac{1.13016 - 0.56586}{1.13016} = 0.49931. \end{aligned}$$

Hal yang sama dilakukan pada data lainnya di kluster C2. Perhitungan yang sama juga dilakukan pada data di kluster C1.

Langkah terakhir yaitu menghitung rata-rata nilai *SI (Silhouette Index)* dengan menggunakan Persamaan (7)

$$\begin{aligned} SI &= SI = \frac{1}{40} \sum_{j=1}^{40} SI_j, \\ &= \frac{1}{40} (0.82522 + 0.49931 + \dots + 0.83617), \\ &= \frac{1}{40} (28.67048) = 0.71676. \end{aligned}$$

Pada penelitian didapatkan hasil nilai rata-rata *SI (Silhouette Index)* sebesar 0.71676, maka berdasarkan Tabel 2.1 dapat diartikan bahwa implementasi algoritma *K-Means clustering* pada penelitian ini memiliki struktur kluster yang kuat.

PENUTUP

Berdasarkan hasil penerapan algoritma *K-Means clustering* pada penelitian ini didapatkan hasil 36 Kecamatan masuk ke dalam kluster C1 dan 4 Kecamatan masuk ke dalam kluster C2, Karakteristik kluster C1 memiliki tingkat risiko penyebaran sedang dengan rata-rata kepadatan penduduk 2,108.49 orang/km², rata-rata kasus terinfeksi *COVID-19* sebanyak 315.44 orang dan rata-rata kasus kematian 6.03 orang. Sementara daerah pada kluster C2 memiliki tingkat risiko penyebaran tinggi dengan rata-rata kepadatan penduduk 7,994.66 orang/km², rata-rata kasus terinfeksi *COVID-19* sebanyak 1,809 orang dan rata-rata kasus kematian 48.05 orang. Nilai validitas *clustering* dengan *Silhouette Index* pada penelitian ini didapatkan hasil sebesar 0.71676 berdasarkan hasil tersebut, maka struktur yang dihasilkan algoritma *K-means clustering* pada penelitian ini memiliki struktur yang kuat.

DAFTAR PUSTAKA

- [1] World Health Organization (WHO). (2020). *Coronavirus disease 2019 (COVID-19): Situation Report*.
- [2] Yang, Y., Peng F., Wang, R., Guan, K., Jiang, T., Xu, G., Sun, J. & Chang, C. (2020). The Deadly Coronaviruses: The 2003 SARS Pandemic and The 2020 Novel Coronavirus Epidemic in China. *J Autoimmun* 2020:102434.
- [3] Ramadanti, E. & Muslih, M. (2021). Analisis Persebaran Kasus COVID-19 di Jawa Barat Menggunakan Metode K-Means Clustering. *Jurnal SISMATIK Nusa Putra*, 319-326.
- [4] Kementerian Dalam Negeri RI (2022). Rilis Data Kemendagri: Jumlah Penduduk Jabar Terbanyak, Kaltara Tersedikit. Diakses pada 30 Juni 2022, dari <https://www.dukcapil.kemendagri.go.id/berita/baca/1033/rilis-data-kemendagri-jumlah-penduduk-jabar-terbanyak-kaltara-tersedikit>.
- [5] Prasetyo, Eko. (2013). *Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Penerbit ANDI.
- [6] Sari, D. N. P. & Sukestiyarno, Y. L. (2021). Analisis Cluster Dengan Metode K-Means Pada Persebaran Kasus COVID-19 Berdasarkan Provinsi di Indonesia. *PRISMA; Prosiding Seminar Nasional Matematika 2021. Vol 4, 602-610*.

- [7] Li, W. & Liu, Z. (2011). A Method of SVM with Normalization in Intrusion Detection. *Procedia Environ. Sci*, vol. 11, pp. 256–262, 2011.
- [8] Kauffman, L. & Rousseeuw, P. (1990). *Finding Groups in Data*. Kanada: John Willey & Sons, Inc.
- [9] Badan Pusat Statistik Kabupaten Bogor (BPS). (2021). *Kabupaten Bogor Dalam Angka*. Bogor: Badan Pusat Statistik Kabupaten Bogor.
- [10] Geoportal Kabupaten Bogor. (2021). Info COVID-19 Kabupaten Bogor. Diakses pada 13 Oktober 2021, dari <https://geoportal.bogorkab.go.id/covid19/>.