

Penerapan Data Mining K-Means dalam Analisis Kematian COVID-19 di Kalimantan Timur

Marlina Yunus*¹, Masna Wati²

^{1,2}Program Studi Informatika, Universitas Mulawarman, Samarinda, Kalimantan Timur
e-mail: *¹merlynalyn10@gmail.com, masnawati@fkti.unmul.ac.id

Abstrak

Pandemi COVID-19 telah memberikan dampak yang signifikan terhadap kesehatan masyarakat, khususnya di Provinsi Kalimantan Timur. Penelitian ini bertujuan untuk menganalisis tren kematian akibat COVID-19 pada periode Maret 2020 hingga Februari 2021 menggunakan metode data mining dengan algoritma K-Means. Data yang digunakan dalam penelitian ini diperoleh dari dataset resmi dalam format CSV yang memuat jumlah pasien meninggal setiap bulannya. Proses analisis diawali dengan praproses data untuk memastikan kualitas dan konsistensi data. Selanjutnya, pada algoritma K-Means diterapkan untuk mengelompokkan data berdasarkan kemiripan tren kematian. Berdasarkan hasil analisis menggunakan metode Elbow, jumlah kluster optimal yang diperoleh adalah tiga kluster. Kluster pertama (Kluster 0) menunjukkan adanya peningkatan jumlah kematian secara moderat pada beberapa bulan. Kluster kedua (Kluster 1) menggambarkan tren kematian yang cenderung stabil dengan sedikit peningkatan. Kluster ketiga (Kluster 2) menunjukkan puncak kematian yang signifikan dengan tren peningkatan tajam, khususnya pada bulan-bulan tertentu selama pandemi. Hasil penelitian ini diharapkan dapat memberikan gambaran yang lebih jelas mengenai tren kematian akibat COVID-19 di Provinsi Kalimantan Timur, sehingga dapat menjadi bahan pertimbangan bagi pemerintah daerah dan tenaga kesehatan dalam merumuskan kebijakan penanggulangan pandemi. Selain itu, analisis tren ini juga dapat dijadikan acuan dalam merancang strategi mitigasi pada masa pandemi atau menghadapi potensi wabah penyakit serupa di masa mendatang.

Kata Kunci: COVID-19, Data Mining, K-Means, Tren Kematian, Kalimantan Timur

1. PENDAHULUAN

Pandemi COVID-19 yang bermula di Wuhan pada akhir 2019 menyebar dengan cepat ke seluruh dunia, termasuk Indonesia yang mencatat kasus pertama pada Maret 2020. Dampaknya meluas ke berbagai sektor, seperti kesehatan, ekonomi, dan sosial. Kalimantan Timur menjadi salah satu daerah terdampak dengan lonjakan kasus dan kematian yang signifikan, terutama pada akhir 2020 hingga awal 2021.

Data dari Satu Data Kalimantan Timur menunjukkan peningkatan jumlah kematian sejak awal pandemi. Kematian akibat COVID-19 menjadi indikator penting dalam menilai keparahan pandemi serta efektivitas penanganannya. Penelitian ini bertujuan untuk menganalisis tren kematian pasien COVID-19 di Kalimantan Timur pada periode Maret 2020 hingga Februari 2021. Data yang digunakan berasal dari sumber resmi pemerintah daerah, dan dianalisis menggunakan metode *data mining* dengan algoritma K-Means Clustering untuk mengelompokkan pola kematian. Hasil penelitian ini diharapkan dapat mendukung pengambilan kebijakan berbasis data yang lebih efektif. Meskipun telah banyak studi terkait COVID-19, penelitian spesifik mengenai tren kematian di Kalimantan Timur masih tergolong terbatas.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik *data mining* untuk menganalisis tren kematian pasien akibat COVID-19 di Provinsi Kalimantan Timur. Pemilihan metode ini didasarkan pada kemampuan *data mining* dalam mengidentifikasi pola tersembunyi dari data historis dalam jumlah besar secara sistematis dan efisien. Metode yang digunakan adalah algoritma **K-Means Clustering**, yang bertujuan untuk mengelompokkan wilayah berdasarkan karakteristik dan kecenderungan data kematian selama periode Maret 2020 hingga Februari 2021. Tahapan dalam penelitian ini dijabarkan sebagai berikut:

2.1. Jenis Penelitian

Penelitian ini termasuk dalam kategori penelitian kuantitatif dengan pendekatan *data mining*. Tujuan utama dari penelitian ini adalah untuk mengidentifikasi pola kematian akibat COVID-19 melalui pengelompokan data menggunakan algoritma K-Means Clustering. Pendekatan ini dipilih karena mampu mengungkap struktur tersembunyi dalam kumpulan data besar dan tidak terstruktur. Menurut Han, Kamber, dan Pei (2011), *data mining* adalah proses menemukan pola yang menarik dari data dalam jumlah besar.

2.2. Sumber Data

Data yang digunakan dalam penelitian ini berasal dari situs resmi Satu Data Kalimantan Timur, serta file data dalam format CSV, yang memuat informasi jumlah kematian pasien COVID-19 yang tercatat dari Maret 2020 hingga Februari 2021. Sumber data ini dipilih karena memiliki informasi yang akurat, relevan, dan diperbarui secara berkala oleh instansi resmi pemerintah daerah.

2.3. Teknik Pengumpulan dan Pengolahan Data

Data yang dikumpulkan merupakan jumlah kasus kematian akibat COVID-19 per hari, dari Maret 2020 hingga Februari 2021. Data tersebut disusun dalam tabel dengan bulan sebagai kolom dan tanggal sebagai baris, di mana setiap sel menunjukkan jumlah kumulatif kematian pada hari tertentu. Data diperoleh dari portal resmi pemerintah, kemudian dilakukan proses pembersihan dan pengolahan data. Langkah-langkah teknis yang dilakukan meliputi:

2.3.1 Pengumpulan dan Pembersihan Data:

Langkah awal dalam pengolahan data adalah memastikan integritas dan konsistensi data. Hal ini dilakukan melalui proses pembersihan data (*data cleaning*), yakni dengan menghapus data duplikat, memeriksa entri yang tidak lengkap, serta menyeragamkan format data. Duplikasi data dihapus untuk mencegah distorsi pada hasil analisis statistik, sementara entri yang tidak lengkap ditangani dengan metode khusus dalam tahap pra-pemrosesan. Proses ini penting karena data yang tidak bersih atau tidak lengkap dapat menghasilkan interpretasi yang menyesatkan (Han et al., 2011).

2.3.2 Pra-pemrosesan Data:

Setelah proses pembersihan awal, tahap berikutnya adalah pra-pemrosesan data, yang mencakup beberapa teknik sebagai berikut:

- **Normalisasi Data**

Normalisasi dilakukan untuk menyelaraskan skala antar fitur, khususnya pada variabel jumlah kematian harian akibat COVID-19. Tujuan dari normalisasi adalah untuk mencegah dominasi fitur tertentu yang memiliki rentang nilai lebih besar dalam proses klusterisasi atau pemodelan. Proses ini mengikuti praktik standar sebagaimana dijelaskan oleh Han, Kamber, dan Pei (2011), yang menekankan pentingnya penyamaan skala agar hasil analisis lebih representatif.

- **Penanganan Nilai Kosong (*Missing Values*)**

Dalam dataset yang dikumpulkan, terdapat sejumlah nilai yang hilang (*NaN*), terutama pada akhir bulan di beberapa periode. Untuk mengatasi hal ini, digunakan pendekatan imputasi dengan mengisi nilai kosong menggunakan rata-rata kolom terkait. Pendekatan ini dipilih karena dinilai efektif menjaga karakteristik distribusi data tanpa menambahkan bias yang signifikan, sebagaimana disarankan oleh Tan, Steinbach, dan Kumar (2019).

2.3.3 *Eksplorasi Data Awal:*

Sebelum dilakukan analisis lanjutan, data yang telah dibersihkan dan diproses dieksplorasi secara visual untuk memperoleh gambaran awal mengenai tren dan pola distribusi data dengan menggunakan grafik deret waktu (*time series*) dan histogram guna mempermudah identifikasi tren, anomali dan distribusi nilai kematian harian.

2.4. *Metode Analisis*

Metode utama yang digunakan adalah algoritma K-Means Clustering. Proses implementasinya terdiri atas beberapa tahapan berikut:

2.4.1 *Penentuan Jumlah Kluster Optimal*

Penentuan jumlah kluster dilakukan dengan menggunakan metode *Elbow*, yang menghitung nilai SSE (*Sum of Squared Errors*) untuk berbagai nilai k . Dimana titik “*elbow*” digunakan sebagai indikator jumlah kluster terbaik (Jain, 2010).

2.4.2 *Implementasi Algoritma K-Means*

Algoritma K-Means diimplementasikan menggunakan pustaka Scikit-Learn (Pedregosa et al., 2011). Proses dimulai dengan menentukan *centroid* awal secara acak dan melakukan iterasi hingga konvergen tercapai.

2.5. *Tahapan Klusterisasi*

2.5.1 *Persiapan Data*

Data yang digunakan adalah jumlah kasus kematian harian akibat COVID-19 dari Maret 2020 hingga Februari 2021. Data disusun dalam bentuk tabel dengan bulan sebagai kolom dan tanggal sebagai baris. Setiap sel menunjukkan jumlah kumulatif kematian pada hari tertentu. Data diambil dari portal resmi pemerintah dan dibersihkan serta diproses.

2.5.2 *Penentuan Jumlah Kluster (K)*

Penentuan nilai K optimal dilakukan menggunakan metode *Elbow* dan *Silhouette Score*. *Elbow Method* digunakan untuk mengidentifikasi titik optimum jumlah kluster berdasarkan penurunan nilai *inertia* terhadap penambahan jumlah kluster, di mana titik belok (*elbow*) pada grafik menunjukkan nilai K yang paling optimal. Selanjutnya, *Silhouette Score* mengevaluasi konsistensi antar kluster dengan mengukur seberapa mirip suatu data dengan kluster tempatnya berada dibandingkan dengan kluster lain. Nilai *Silhouette* yang mendekati satu menunjukkan hasil klusterisasi yang baik.

2.5.3 *Inisialisasi Centroid*

Tahapan ini melibatkan pemilihan secara acak sejumlah titik data sebanyak K sebagai pusat kluster awal atau *initial centroid*. Pemilihan *centroid* awal sangat memengaruhi hasil akhir klusterisasi karena K-Means sensitive terhadap posisi awal.

2.5.4 *Pengelompokan Data ke dalam Kluster*

Setelah centroid awal ditentukan, setiap titik data akan dihitung jaraknya terhadap seluruh centroid menggunakan metrik jarak *Euclidean*. Kemudian, tiap titik data diklasifikasikan ke dalam kluster yang memiliki jarak paling dekat terhadap centroid-nya. Proses ini memastikan bahwa objek-objek yang memiliki kemiripan karakteristik dikelompokkan bersama.

2.5.5 Pembaruan Centroid

Centroid diperbarui dengan menghitung nilai rata-rata semua data dalam satu kluster. Proses ini berulang secara iteratif hingga *centroid* tidak berubah signifikan atau jumlah iterasi maksimum tercapai. Proses ini mengikuti prinsip *Expectation-Maximization*, di mana pada setiap iterasi dilakukan pembaruan ekspektasi dan maksimisasi posisi *centroid*.

2.5.6 Evaluasi Hasil Klusterisasi

Evaluasi dilakukan menggunakan *Scatter Plot* atau *3D Plot* untuk meninjau sebaran data antar kluster. Selanjutnya, *Silhouette Coefficient* untuk mengukur tingkat kepaduan (*cohesion*) dan pemisahan (*separation*) antar kluster. Kluster yang baik memiliki nilai *Silhouette* yang tinggi dan menunjukkan pemisahan yang jelas antar kelompok data.

2.5.7 Implementasi Hasil Klusterisasi

Hasil klusterisasi diinterpretasikan untuk mengidentifikasi karakteristik masing-masing kluster. Hal ini bertujuan untuk memberikan informasi kepada pemangku kebijakan dalam merumuskan strategi penanggulangan yang lebih tepat sasaran berdasarkan tingkat risiko temporal yang teridentifikasi.

2.6 Visualisasi dan Interpretasi

Visualisasi merupakan tahap penting dalam proses analisis data, khususnya pada metode klusterisasi, karena dapat memberikan gambaran yang lebih intuitif terhadap pola dan struktur data yang telah dikelompokkan. Dalam penelitian ini, hasil dari klusterisasi divisualisasikan dengan beberapa jenis grafik, yaitu:

- **Scatter Plot**

Menampilkan sebaran data antar kluster berdasarkan dua fitur utama, memudahkan identifikasi pemisahan visual antar kelompok serta mendeteksi *outlier* atau data yang tidak sesuai dengan pola kluster yang terbentuk.

- **Bar Chart**

Menyajikan jumlah data dalam setiap kluster untuk menunjukkan distribusi dan dominasi kluster tertentu. Visualisasi ini memberikan informasi kuantitatif mengenai proporsi tiap kluster terhadap keseluruhan data.

2.7 Teori K-Means

K-Means adalah algoritma pembelajaran tanpa pengawasan (*unsupervised learning*) untuk mengelompokkan data. Proses ini bertujuan membuat setiap kluster lebih homogen dan berbeda dengan yang lain. Algoritma ini terdiri atas dua langkah, yaitu *Assignment Step*: menempatkan titik data ke dalam kluster dengan *centroid* terdekat, dan *Update Step*: memperbaharui posisi *centroid* berdasarkan rata-rata data dalam setiap kluster. Iterasi dilakukan hingga tidak ada perubahan signifikan. Keberhasilan algoritma bergantung pada pemilihan nilai *K* dan posisi awal *centroid*. Meski memiliki keterbatasan, K-Means tetap populer karena kesederhanaan dan efisiensinya.

2.8 SSE (Sum of Square Error)

SSE atau *Within-Cluster Sum of Squares* (WCSS) merupakan metrik untuk menilai kualitas klusterisasi. SSE dihitung dengan menjumlahkan kuadrat jarak *Euclidean* antara tiap

titik data dengan *centroid* kluster-nya. Matematis, ini diwakili sebagai total kuadrat dari norma Euclidean untuk semua titik di setiap kluster, yang kemudian dijumlahkan untuk seluruh kluster. Nilai SSE yang rendah menunjukkan bahwa titik-titik data dalam kluster cenderung berdekatan dengan *centroid*-nya, menandakan kluster yang padat. Sebaliknya, nilai SSE yang tinggi berarti ada lebih banyak variasi di dalam kluster, yang bisa berarti klasterisasi tidak optimal. Ditunjukkan pada persamaan (1)

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

Semakin kecil nilai SSE, maka semakin baik klasterisasi karena menunjukkan bahwa titik-titik data semakin dekat dengan pusat kluster masing-masing. SSE juga digunakan dalam *Elbow Method* untuk menentukan jumlah kluster optimal, di mana nilai k dipilih pada titik di mana penurunan nilai SSE mulai melambat.

2.9 Silhouette Score

Silhouette Score merupakan sebuah metrik komprehensif yang digunakan untuk mengevaluasi validitas dan kualitas hasil klasterisasi dalam algoritma pembelajaran tanpa pengawasan. Metrik ini mengkuantifikasi seberapa baik setiap objek data cocok dengan klusternya sendiri dibandingkan dengan kluster tetangga. Nilai *Silhouette Score* berkisar antara -1 hingga $+1$, di mana nilai yang mendekati $+1$ mengindikasikan performa klasterisasi yang superior. Secara spesifik, perhitungan *Silhouette Score* untuk setiap objek melibatkan perbandingan antara jarak rata-rata objek tersebut ke objek lain dalam klusternya sendiri (kohesi, a) dengan jarak rata-rata objek tersebut ke objek-objek dalam kluster tetangga terdekat (separasi, b). *Silhouette Coefficient* (s) untuk objek individual kemudian dihitung menggunakan persamaan (2).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2)$$

Nilai s yang mendekati 1 menunjukkan objek lebih mirip dengan objek dalam klusternya sendiri dibandingkan kluster lain. Nilai mendekati -1 menunjukkan kemungkinan salah klasifikasi ke kluster tetangga. *Silhouette Score* rata-rata menunjukkan validitas hasil klasterisasi. Semakin tinggi nilainya, semakin baik kluster terpisah. Nilai dekat 0 berarti ada tumpang tindih antar kluster dan klasifikasi yang kurang jelas.

3 HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil analisis data dengan metode klasterisasi K-Means untuk mengelompokkan objek berdasarkan karakteristik tertentu tanpa label. Proses analisis mencakup pra-pemrosesan data, penentuan jumlah kluster optimal melalui metode *Elbow*, serta evaluasi performa model menggunakan metrik SSE dan *Silhouette Score*. Nilai SSE yang rendah menunjukkan kluster yang padat dan homogen, sedangkan *Silhouette Score* mengukur kohesi dan separasi antar kluster dalam rentang -1 hingga 1 . Hasil evaluasi menunjukkan bahwa model berhasil membentuk kluster secara efektif, dan temuan ini akan dibahas lebih lanjut melalui visualisasi serta interpretasi hasil sesuai konteks penelitian.

3.1. Deskripsi Dataset dan Eksplorasi Awal

Dataset yang digunakan mencakup jumlah kematian akibat COVID-19 di Provinsi Kalimantan Timur selama periode Maret 2020 hingga Februari 2021. Data bersifat *time-series*

28	1	1.0	3	7	28	155	312	471	587	731	976	1168.35
29	1	1.0	3	4.34	30	160	332	472	592	738	987	1168.35
30	1	1.0	3	4.34	31	161	249.46	482	533.93	743	996	1168.35

Setelah data dibersihkan, langkah selanjutnya adalah menentukan jumlah kluster (K) untuk algoritma K-Means. Nilai K membagi data menjadi dua kelompok yang menunjukkan pola distribusi kematian yang berbeda. Klusterisasi dilakukan dengan K-Means, dan evaluasi kluster diukur dengan SSE dan *Silhouette Score*.

3.2. Tabel Kluster

Berdasarkan hasil *preprocessing* data yang telah dilakukan, langkah selanjutnya yang krusial dalam analisis *time series clustering* adalah penentuan jumlah kluster (k) yang optimal untuk segmentasi pola temporal. Sehingga ditampilkan hasil klustering pada Gambar 1.

```

Hasil klusterisasi K-Means berdasarkan tanggal:
Tanggal 1 masuk ke kluster 2
Tanggal 2 masuk ke kluster 2
Tanggal 3 masuk ke kluster 8
Tanggal 4 masuk ke kluster 8
Tanggal 5 masuk ke kluster 8
Tanggal 6 masuk ke kluster 4
Tanggal 7 masuk ke kluster 4
Tanggal 8 masuk ke kluster 4
Tanggal 9 masuk ke kluster 4

```

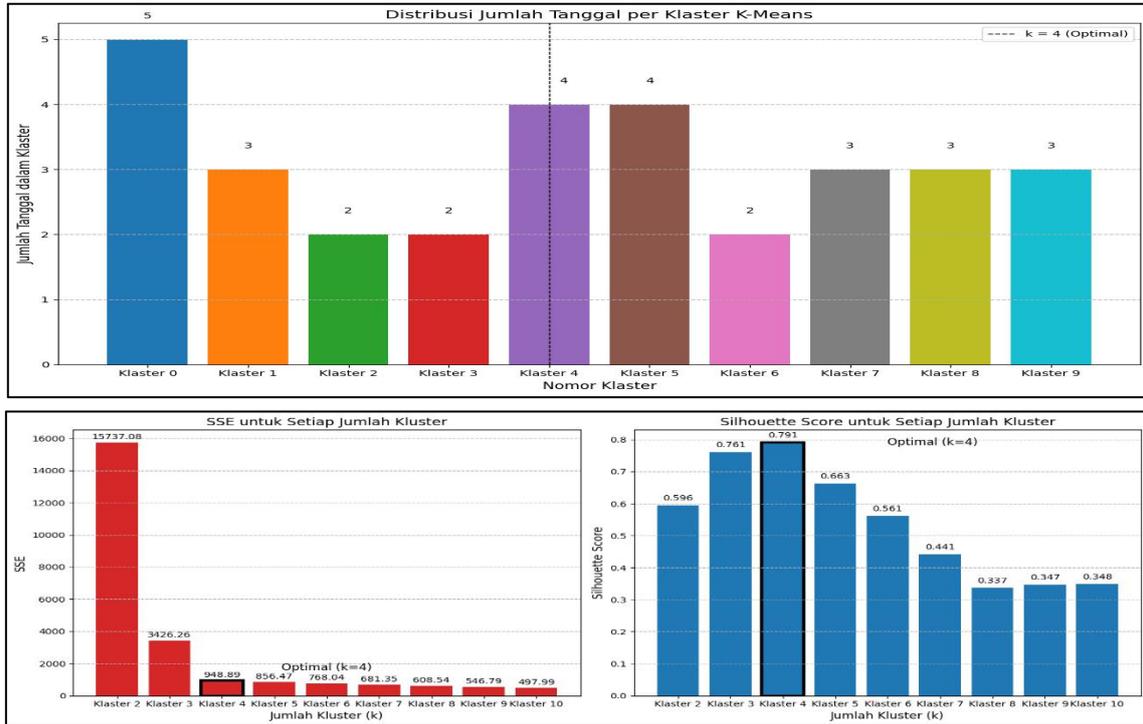
Gambar 1. Hasil Clustering Data Kematian COVID-19

Setelah melalui proses pembersihan data yang mencakup penghapusan nilai yang tidak valid dan pengisian nilai yang hilang, langkah selanjutnya adalah menentukan jumlah kluster (k) yang optimal untuk analisis klusterisasi. Penentuan nilai k ini dilakukan dengan menggunakan berbagai metode evaluasi, seperti metode *Elbow*, SSE dan *Silhouette Score* yang bertujuan untuk mendapatkan pembagian data yang paling representatif dan bermakna. Adapun hasil dari penentuan jumlah kluster (k) yang digunakan dalam analisis pada Tabel 4.

Tabel 4. Data Clustering

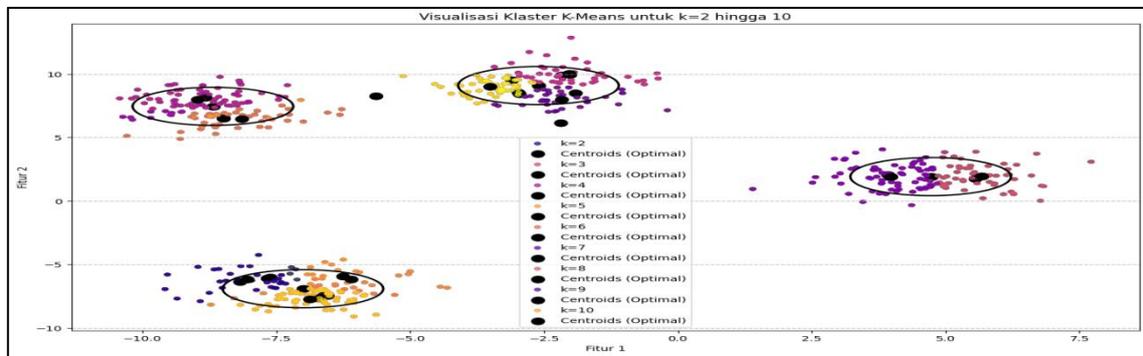
in	2020										2021		Cluster
	Mar	Apr	Mei	Jun	Jul	Agu	Sep	Okt	Nov	Des	Jan	Feb	
0	0	1.0	1	3	7	32	164	352	487	593	755	1010	2
1	0	1.0	1	3	7	36	165	362	489	594	759	1018	2
2	0	1.0	1	3	8	40	192	364	493	597	763	1031	8
3	0	1.0	2	3	8	41	199	367	494	602	769	1045	8
4	0	1.0	2	3	8	42	207	376	495	602	777	1056	8
5	0	1.0	2	3	9	42	214	379	495	604	782	1079	4
...
28	1	1.0	3	7	28	155	312	471	587	731	976	1168.35	7
29	1	1.0	3	4.34	30	160	332	472	592	738	987	1168.35	7
30	1	1.0	3	4.34	31	161	249.46	482	533.93	743	996	1168.35	7

3.3. Visualisasi Kluster



Gambar 2. Visualisasi Kluster

3.4. Visualisasi Scatter Plot



Gambar 3. Visualisasi Scatter Plot

3.5. Sum of Square Error (SSE) & Silhouette Score

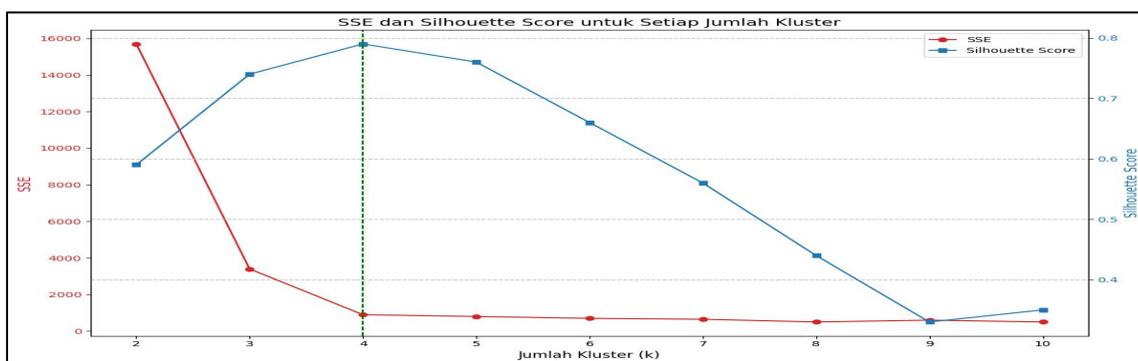
Berdasarkan hasil evaluasi menggunakan SSE dan *Silhouette Score*, hasil pengelompokkan data dengan nilai $k = 2$, sehingga $k = 10$, bisa disimpulkan di mana evaluasi dilakukan terhadap pemilihan jumlah kluster (k) menggunakan dua metrik utama, yaitu *Sum of Squared Errors* (SSE) dan *Silhouette Score*, dengan variasi nilai k dari 2 hingga 10. Hasil evaluasi ditampilkan pada Tabel 5.

Tabel 5. Jumlah Kluster, SSE dan Silhouette Score

index	Jumlah Kluster (k)	SSE	Silhouette Score
0	2	15737.083578	0.595521
1	3	3486.257238	0.761348
2	4	948.890343	0.791104

index	Jumlah Kluster (k)	SSE	Silhouette Score
3	5	856.471889	0.663349
4	6	768.038998	0.561280
5	7	681.351538	0.441483
6	8	608.538032	0.337241
7	9	546.790597	0.346955
8	10	497.991486	0.348140

Analisis metrik SSE dan *Silhouette Score* untuk jumlah kluster 2 hingga 10 menunjukkan bahwa enam kluster adalah yang paling optimal. SSE menurun seiring bertambahnya jumlah kluster, dengan penurunan terbesar dari $k = 2$ dan $k = 3$ menandakan pemisahan data yang signifikan. *Silhouette Score* mencapai maksimum pada $k = 6$, menunjukkan keseimbangan terbaik antara kedekatan sampel dalam kluster dan jarak antar kluster. Sehingga menghasilkan data seperti pada Gambar 4.



Gambar 4. Hasil Evaluasi SSE dan Silhouette Score

Dalam kode tersebut, SSE mengukur total kuadrat jarak antara setiap titik data dan *centroid* klusternya. Semakin rendah SSE, semakin baik titik-titik data tersebut mengelompok dalam klusternya. Idealnya, kita mencari *elbow* pada plot SSE, di mana SSE menurun secara signifikan sebelum titik tersebut, dan penurunannya menjadi lebih lambat setelahnya. Titik *elbow* ini menunjukkan jumlah kluster yang memberikan pengelompokan terbaik dengan kompleksitas yang tidak terlalu tinggi.

4 KESIMPULAN

Berdasarkan analisis nilai *Sum of Squared Errors* (SSE) dan *Silhouette Score*, dapat disimpulkan bahwa jumlah kluster optimal untuk data tersebut adalah empat kluster ($k=4$), yang ditunjukkan oleh penurunan signifikan pada nilai SSE setelah $k=4$ dan nilai *Silhouette Score* tertinggi pada $k = 4$ (0.761). Evaluasi jumlah kluster menggunakan metode *Elbow* pada nilai *Sum of Squared Errors* (SSE) menunjukkan penurunan yang tajam hingga mencapai empat kluster, setelah itu penurunan nilai SSE menjadi tidak signifikan, mengindikasikan bahwa penambahan jumlah kluster lebih lanjut tidak memberikan perbaikan yang substansial dalam mengurangi inersia kluster. Lebih lanjut, analisis nilai *Silhouette Score* untuk setiap jumlah kluster mengidentifikasi bahwa jumlah kluster optimal adalah empat ($k=4$) dengan nilai *Silhouette Score* tertinggi sebesar 0.761, yang mengindikasikan bahwa data terbagi menjadi kluster-kluster yang padat dan terpisah dengan baik. Oleh karena itu, berdasarkan kombinasi visualisasi penurunan SSE dan nilai *Silhouette Score* yang optimal, dapat disimpulkan bahwa empat kluster merupakan representasi struktur data yang paling sesuai dan bermakna.

DAFTAR PUSTAKA

- [1] W. Adisasmito, "Strategi mitigasi pandemi COVID-19 di Indonesia," *J. Kesehat. Masy. Nas.*, vol. 15, no. 2, pp. 59–68, 2020.
- [2] Z. Arifin, "Analisis data kematian COVID-19 menggunakan metode statistik," *Stat. Indones.*, vol. 10, no. 1, pp. 45–53, 2021.
- [3] E. Budiarto, "Pengaruh mobilitas penduduk terhadap penyebaran COVID-19," *J. Epidemiol. Indones.*, vol. 5, no. 3, pp. 123–130, 2020.
- [4] Dinas Kesehatan Provinsi Kalimantan Timur, "Data dan Informasi COVID-19 Kaltim." [Online]. Available: <https://dinkes.kaltimprov.go.id>.
- [5] R. Fauzan, "Implementasi algoritma K-Means dalam analisis data COVID-19," *J. Tek. Inf. Komput.*, vol. 8, no. 2, pp. 98–105, 2021.
- [6] T. Hidayat, "Penerapan machine learning untuk prediksi kasus COVID-19," *J. Inform. Indones.*, vol. 9, no. 1, pp. 67–75, 2021.
- [7] H. Kusuma, "Kapasitas layanan kesehatan dan angka kematian COVID-19," *J. Adm. Kesehat. Indones.*, vol. 8, no. 4, pp. 210–218, 2020.
- [8] D. Lestari, "Peran edukasi masyarakat dalam pencegahan COVID-19," *J. Pendidik. Kesehat.*, vol. 13, no. 1, pp. 34–42, 2021.
- [9] S. Mulyadi, "Dampak kepadatan penduduk terhadap tingkat kematian COVID-19," *J. Demogr. Kesehat.*, vol. 7, no. 2, pp. 89–97, 2020.
- [10] A. Nugroho, "Analisis spasial penyebaran COVID-19 di Indonesia," *J. Geogr.*, vol. 13, no. 3, pp. 145–153, 2021.
- [11] Y. Pratama, "Evaluasi efektivitas pembatasan sosial berskala besar (PSBB)," *J. Kebijakan Publik*, vol. 6, no. 2, pp. 101–110, 2020.
- [12] A. Putri, "Keterbatasan fasilitas kesehatan dan angka kematian COVID-19," *J. Kesehat. Glob.*, vol. 9, no. 1, pp. 56–64, 2021.
- [13] F. Rahman, "Kepadatan penduduk dan mobilitas sebagai faktor risiko COVID-19," *J. Epidemiol. Kesehat. Masy.*, vol. 12, no. 2, pp. 78–86, 2021.
- [14] B. Santoso, "Perbandingan metode clustering dalam analisis data COVID-19," *J. Sist. Inform.*, vol. 14, no. 2, pp. 112–120, 2020.
- [15] Satu Data Kalimantan Timur, "Portal Data Terbuka Pemerintah Provinsi Kalimantan Timur." [Online]. Available: <https://data.kaltimprov.go.id>
- [16] I. Setiawan, "Pengaruh karakteristik geografis terhadap penyebaran COVID-19," *J. Ilmu Lingkungan*, vol. 16, no. 1, pp. 23–31, 2022.
- [17] M. Siregar, "Analisis tren kematian akibat COVID-19 di Indonesia," *J. Kesehat. Masy.*, vol. 14, no. 3, pp. 201–209, 2021.
- [18] D. Susanto, "Penerapan data mining untuk prediksi kasus COVID-19," *J. Tek. Inform.*, vol. 11, no. 2, pp. 88–96, 2020.
- [19] S. Utami, "Faktor-faktor yang mempengaruhi tingkat kematian COVID-19," *J. Kesehat. Indones.*, vol. 10, no. 1, pp. 15–23, 2021.
- [20] H. Wijaya and S. Nugroho, "Dampak pelanggaran pembatasan sosial terhadap lonjakan kasus COVID-19," *J. Kebijakan. Kesehat.*, vol. 9, no. 2, pp. 134–142, 2020.
- [21] E. Yulianti, "Analisis temporal penyebaran COVID-19 di Indonesia," *J. Stat. Komput.*, vol. 5, no. 1, pp. 49–57, 2021.
- [22] N. Zahra, "Peran protokol kesehatan dalam menurunkan angka kematian COVID-19," *J. Kesehat. Masy. Nas.*, vol. 15, no. 1, pp. 12–20, 2020.