

ANALISIS TINGKAT KEPARAHAN COVID-19 DI SUATU NEGARA MENGUNAKAN METODE K-MEDOID CLUSTERING

Syahputri Amelia¹, Eldo Yusri², Bambang Irwansyah³, Ayu Andira⁴,
Muhammad Sukri⁵, Muhammad Fauzan Yafi⁶, Muhammad Affandi⁷

Universitas Asahan, Asahan

e-mail : ¹Syahputriamalia75@gmail.com, ²Eldoyusri9@gmail.com,

³bambangirwansyah53@gmail.com, ⁴Ayuandira0034@gmail.com,

⁵Sukri7787@gmail.com, ⁶Fauzanyafi761@gmail.com, ⁷Affandisaragih01@gmail.com

Abstract: *This study aims to classify the severity of COVID-19 cases based on patient and region data using the K-Medoid Clustering method. COVID-19 has varying degrees of symptom severity, requiring cluster analysis to identify severity patterns to support decision-making in healthcare resource allocation and policy formulation. The data used included the number of positive cases, recovered cases, deaths, the average age of patients, and comorbidity levels. The results showed that the K-Medoid method was able to effectively cluster the data. In the raw dataset, the percentage of patients not infected with COVID-19 was 62.62%, while the percentage of infected patients was 37.38%. Based on sample characteristics, non-obese patients accounted for 74.54%, obese patients 25.46%, and patients with a combination of obesity and cardiovascular disease 0.57%.*

Keywords: *Covid-19, Severity, K-Medoid Clustering, Data Mining*

Abstrak: Penelitian ini bertujuan untuk mengelompokkan tingkat keparahan kasus COVID-19 berdasarkan data pasien dan wilayah menggunakan metode K-Medoid Clustering. COVID-19 memiliki variasi tingkat keparahan gejala, sehingga diperlukan analisis kluster untuk mengidentifikasi pola keparahan yang mendukung pengambilan keputusan dalam alokasi sumber daya kesehatan dan perumusan kebijakan. Data yang digunakan meliputi jumlah kasus positif, kasus sembuh, kasus meninggal, usia rata-rata pasien, serta tingkat komorbiditas. Hasil penelitian menunjukkan bahwa metode K-Medoid mampu melakukan pengelompokan data secara efektif. Pada dataset mentah, persentase pasien tidak terjangkit COVID-19 sebesar 62,62%, sedangkan pasien terjangkit sebesar 37,38%. Berdasarkan karakteristik sampel, pasien non-obesitas memiliki persentase 74,54%, pasien obesitas 25,46%, dan pasien dengan kombinasi obesitas serta penyakit kardiovaskular sebesar 0,57%.

Kata Kunci : Covid-19, Tingkat Keparahannya, K-Medoid Clustering, Data Mining

PENDAHULUAN

Pandemi COVID-19 telah menyebar ke seluruh dunia dan memberikan dampak besar terhadap sistem kesehatan global. Setiap negara mencatat jumlah kasus, kematian, dan kesembuhan yang berbeda-beda, tergantung pada kebijakan, fasilitas medis, serta kesiapan menghadapi wabah. Di tengah kompleksitas data yang besar, sangat penting untuk mengelompokkan

wilayah-wilayah berdasarkan kesamaan karakteristik COVID-19. Dengan pengelompokan ini, kita dapat memahami daerah mana yang termasuk zona krisis, mana yang masih waspada, dan mana yang relatif aman. Pendekatan ini tidak hanya berguna untuk pemetaan risiko, tetapi juga sebagai dasar pengambilan keputusan strategis seperti distribusi vaksin, tenaga medis, dan peralatan kesehatan.

Pandemi *Coronavirus Disease*

2019 (COVID-19) yang disebabkan oleh virus SARS-CoV-2 telah memberikan tekanan luar biasa pada sistem kesehatan global. Karakteristik klinis virus ini sangat variatif, mulai dari pasien asimtomatik (tanpa gejala), gejala ringan menyerupai flu, hingga kondisi kritis yang menyebabkan *Acute Respiratory Distress Syndrome* (ARDS) dan kegagalan multiorgan. Heterogenitas gejala ini memunculkan tantangan besar bagi fasilitas kesehatan, terutama ketika terjadi lonjakan kasus yang tidak sebanding dengan ketersediaan tenaga medis, ruang isolasi, dan peralatan vital seperti ventilator. Dalam situasi krisis sumber daya ini, proses triase atau pengelompokan pasien berdasarkan tingkat kegawatdaruratan menjadi kunci utama untuk menekan angka mortalitas Arora, P., et al. (2022).

Metode K-Medoid Clustering menjadi pilihan tepat dalam proses ini karena menggunakan data nyata (medoid) sebagai pusat kluster. Dibanding metode lain seperti K-Means, K-Medoid lebih tahan terhadap nilai ekstrim (outlier), yang sangat umum pada data pandemi global. Dengan pendekatan ini, kita dapat mengidentifikasi pola-pola yang tersembunyi dari kumpulan data global COVID-19, dan mengelompokkan negara-negara atau wilayah ke dalam kluster berdasarkan kesamaan jumlah kasus, kematian, dan tingkat kesembuhan.

Penerapan K-Medoids agar dapat diketahui pola pemilihan penentuan pengelompokan penyebaran covid-19 di berbagai wilayah di Indonesia. K-Medoids merupakan metode partisional clustering dimana bertujuan untuk menemukan satu set k-cluster di antara data yang paling mencirikan objek dalam kumpulan suatu data, Kaur, P., & Sharma, M. (2023).

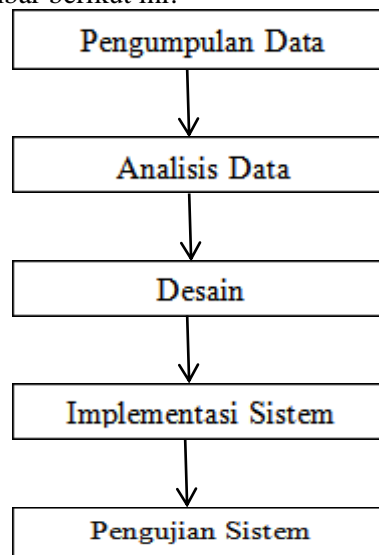
Secara konvensional, penentuan tingkat keparahan pasien dilakukan melalui diagnosis klinis manual oleh dokter berdasarkan pemeriksaan fisik dan tanda-tanda vital. Namun, metode manual ini memiliki keterbatasan, terutama risiko *human error* akibat kelelahan tenaga medis dan subjektivitas dalam

menginterpretasikan kombinasi parameter vital yang kompleks. Oleh karena itu, pendekatan berbasis teknologi informasi, khususnya *Data Mining* dan *Machine Learning*, menjadi solusi strategis untuk membantu tenaga medis dalam mengambil keputusan yang cepat, objektif, dan berbasis data (*data-driven decision making*), Hassan, M. R., et al. (2022).

Melalui proses ini, hasil pengelompokan memberikan gambaran visual dan terstruktur mengenai tingkat keparahan COVID-19 di berbagai wilayah, yang akan sangat membantu pemerintah, organisasi kesehatan, dan masyarakat dalam memahami dan merespons pandemi secara lebih tepat.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik *data mining*. Kami menerapkan algoritma K-Medoid Clustering untuk mengekstraksi pola tersembunyi dari data COVID-19, yang bertujuan untuk memetakan Struktur kerja yang dimaksud dapat dilihat pada gambar berikut ini:



Gambar 1 Kerangka Kerja

Tahapan Analisis

Proses penelitian ini diawali dengan mengumpulkan data rekam medis pasien COVID-19 yang meliputi atribut

vital seperti usia, saturasi oksigen (SpO₂), laju pernapasan, dan skor komorbiditas. Data tersebut kemudian dibersihkan dari nilai yang tidak lengkap (*missing values*) dan dinormalisasi agar perbedaan skala antar atribut tidak mengganggu akurasi perhitungan jarak. Setelah data siap, kami menerapkan algoritma K-Medoids menggunakan pengukuran jarak *Manhattan Distance* melalui beberapa iterasi hingga ditemukan titik pusat (medoid) yang tidak lagi berubah, menandakan pengelompokan tingkat keparahan pasien telah stabil.

Implementasi K-Medoid

Dalam pengerjaannya, kami membagi data ke dalam 3 kluster tingkat keparahan (Berat, Sedang, dan Ringan). Secara teknis, setiap data rekam medis pasien dihitung jarak kedekatannya dengan objek pusat (*medoid*) menggunakan rumus *Manhattan Distance*:"

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

Alat dan Evaluasi

Analisis data dilakukan menggunakan bantuan bahasa pemrograman Python untuk menjamin presisi perhitungan algoritma yang kompleks. Hasil akhir dari

pengelompokan ini kemudian divalidasi untuk melihat seberapa tegas pemisahan karakteristik antara pasien kondisi Ringan, Sedang, dan Berat, sehingga manajemen rumah sakit dapat mengambil keputusan triase yang tepat sasaran.

HASIL DAN PEMBAHASAN

Analisis Data

Menentukan Data Cluster

Salah satu langkah krusial dalam pendekatan K-Medoids Clustering adalah menetapkan jumlah kluster (K) yang ideal. Kemudian, memilih objek data aktual secara acak sebagai titik pusat kluster awal (medoid) sebanyak k. Dari dataset di atas terpilih 3 Kluster awal diantaranya:

Tabel 1 Data Cluster

			281644
Kluster 1	148011	1325804	4
Kluster 2	4838	58173	37292
Kluster 3	0	86	13

Dataset

Setelah data atribut sudah di dapat langkah selanjutnya menentukan data latih. Dalam khusus ini prolehan data latih dapat dilihat pada tabel dibawah ini:

Tabel 2 Dataset

No	Negara	KEMATIAN	SEMBUH	AKTIF
1	Afghanistan	1269	25198	9796
2	Albania	144	2745	1991
3	Algeria	1163	18837	7973
4	Andorra	52	803	52
5	Angola	41	242	667
6	Antigu and Barbuda	3	65	18
7	Argentina	3059	72575	91782
8	Armenia	711	26665	10014
9	Australia	167	9311	5825
10	Austria	713	18242	1599
11	Azerbaijan	423	23242	6781
12	Bahamas	11	91	280
13	Bahrain	141	36110	3231
14	Bangladesh	2965	125683	97577

15 Barbados	7	94	9
16 Belarus	538	60492	6221
17 Belgium	9822	17452	39154
18 Belize	2	26	20
19 Benin	35	1036	699
20 Bhutan	0	86	13
.....
100 US	148011	1325804	2816444

Menentukan Jumlah Kluster

Salah satu langkah krusial dalam pendekatan K-Medoids Clustering adalah menetapkan jumlah kluster (k) yang ideal. Kuantitas kluster ini akan mempengaruhi cara data pasien dikelompokkan sesuai dengan kesamaan tingkat keparahan klinis. Untuk menentukan nilai k yang paling sesuai, penelitian ini menerapkan metode Silhouette Coefficient (atau *Elbow Method*), yang sering dipakai dalam analisis validasi kluster untuk mengukur kualitas pemisahan data. Adapun cluster yang akan dibentuk antara lain:

Tabel 3 Cluster

A	Cluster 1 (C1)	Wilayah krisis
B	Cluster 2 (C2)	Wilayah waspada
C	Cluster 3 (C3)	Wilayah Aman

Kemudian, membangkitkan nilai random untuk pusat kluster awal (centroid) sebanyak k. Dari dataset diatas terpilih 3 kluster pusat diantaranya

Tabel 4 Matriks Penjumlahan Tiap Baris

		132580		
Kluster 1	148011	4	2816444	
Kluster 2	4838	58173	37292	
Kluster 3	0	86	13	

Menghitung Setiap Jarak Dataset

Menghitung jarak setiap data pasien dalam dataset terhadap masing-masing titik pusat (*medoid*) yang telah ditentukan menggunakan rumus jarak *Manhattan (Manhattan Distance)*:

$$d(x_i, \mu_j) = \sqrt{\sum(x_i - \mu_j)^2} \quad (1)$$

sehingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Berikut adalah persamaan Euclidian Distance:

$$D_{(1,1)} = \sqrt{(1269 - 148011)^2 + (25198 - 1325804)^2 + (9796 - 2816444)^2}$$

$$= -9,712.020.271.410$$

$$D_{(1,2)} = \sqrt{(1269 - 4838)^2 + (25198 - 58173)^2 + (9796 - 37292)^2}$$

$$= -4,798.161.174$$

$$D_{(1,3)} = \sqrt{(1269 - 0)^2 + (25198 - 86)^2 + (9796 - 13)^2}$$

$$= 28,698$$

Tabel 5 Hasil Cluster Tahap 1

NO	Negara	Parameter			Distance			Kluster
		Keuntungan	Sembuh	Aktif	C1	C2	C3	
1	Afghanistan	1269	25198	9796	3.096.834.219	430.826.926	269.801.778	Kluster 3
2	Albania	144	2745	1991	3113437.884	65882.148	3317.15254	Kluster 3
3	Algeria	1163	18837	7973	3101166.947	49192.8587	20403.7783	Kluster 3
4	Andorra	82	803	52	3118020.443	68864.1337	219.940275	Kluster 3
5	Angola	41	242	667	3115203.749	68705.1861	673.597064	Kluster 3
6	Antigua	3	65	18	3116367.985	69204.5372	21.7944947	Kluster 3
7	Argentina	4059	22575	91782	4002561.743	56389.2059	116985.307	Kluster 2
8	Armenia	711	26665	10014	3096047.283	41879.2965	28407.1956	Kluster 3
9	Australia	167	9311	5825	3107183.746	58305.1059	10904.488	Kluster 3
10	Austria	713	18246	1599	3107210.485	53713.2897	18243.0645	Kluster 3
11	Azerbaijan	423	23242	6781	3100428.383	46589.5708	24128.512	Kluster 3
12	Bahamas	11	91	280	3116119.162	69041.3412	267.273268	Kluster 3
13	Bahrain	141	36110	3231	3098280.738	40853.2679	36167.7202	Kluster 3
14	Bangladesh	2965	125683	97577	2975494.322	90528.3903	159066.438	Kluster 2
15	Barbados	7	94	9	3116362.992	69184.2591	11.3578167	Kluster 3
16	Belarus	538	60492	6221	3085468.541	31452.7392	60726.5473	Kluster 2
17	Belgium	9822	17452	39154	3073145.115	41067.1053	3393.5337	Kluster 2
18	Belize	2	26	20	3116382.216	69236.2802	60.340053	Kluster 3
19	Benin	35	1036	699	3115337.39	68020.2413	1172.31438	Kluster 3
20	Bhutan	0	86	13	3116363.113	69189.8089	0	Kluster 3
100	US	148011	1325804	2816444	3059922.21	5059922.21	3059922.21	Kluster 1

Menghitung Iterasi 1

Setelah memperbaharui Nilai

centroid baru di peroleh dari rata-rata *cluster* yang bersangkutan dengan

menggunakan rumus:

$$\mu_j(t+1) = \frac{1}{N_j} \sum_{i \in S_j} x_i \quad (2)$$

$\mu_j(t+1)$: centroid baru pada iterasi ke-(t+1), N_j : banyak data pada cluster S_j .

Kluster 1 (Kematian)

$$= 1269 + 144 + 1163 + 52 + 41 + 3 + 3059 + 711 + 167 + 713 + 423 + 11 + 141 + 2965 + 7 + 538 + 98 + 22 + 2 + 35 + 0 + 39 = 21.266$$

Kluster 1 (Sembuh) =

$$= 25198 + 2745 + 18837 + 803 + 242 + 65 + 7257 + 5 + 26665 + 9311 + 18246 + 23242 + 91 + 36110 + 125683 + 94 + 60492 + 17452 + 26 + 1036 + 86 + 39 = 415.673$$

Kluster 1 (Aktif)

$$= 9796 + 1991 + 7973 + 52 + 667 + 18 + 91782 + 10014 + 5825 + 1599 + 6781 + 280 + 3231 + 975$$

$$= 77 + 9 + 6221 + 39154 + 20 + 699 + 13 / 39 = 283,689$$

Tabel 6 Hasil Iterasi 1

	KEMATIAN	SEMBUH	AKTIF
C1	148011	1325804	495499
C2	4458	77144	30983
C3	0	94	1

Menunjukkan hasil iterasi pertama dari proses klusterisasi data nilai COVID 19 berdasarkan tiga komponen penilaian, yaitu Kematian, Sembuh, dan Aktif. Setiap kluster (C1, C2, dan C3) merepresentasikan rata-rata nilai dari masing-masing kelompok Pasien yang dikelompokkan berdasarkan kemiripan data

Tabel 7 Hasil Cluster Tahap 2

No	Negara	Distance	Distance	Distance	Kluster
		C1	C2	C3	
1	Afghanistan	27064,94	27064,94	27064,94	Klaster 3
2	Albania	3394,089	3394,089	3394,089	Klaster 3
3	Algeria	20487,9	20487,9	20487,9	Klaster 3
4	Andorra	806,3603	806,3603	806,3603	Klaster 3
5	Angola	710,7278	710,7278	710,7278	Klaster 3
6	Antigua and Barbuda	67,51296	67,51296	67,51296	Klaster 3
7	Argentina	117048,8	117048,8	117048,8	Klaster 2
8	Armenia	28492,24	28492,24	28492,24	Klaster 3
9	Australia	10984,23	10984,23	10984,23	Klaster 3
10	Austria	18329,8	18329,8	18329,8	Klaster 3
11	Azerbaijan	24214,7	24214,7	24214,7	Klaster 3
12	Bahamas	294,6218	294,6218	294,6218	Klaster 3
13	Bahrain	36254,54	36254,54	36254,54	Klaster 3
14	Bangladesh	159142,3	159142,3	159142,3	Klaster 2
15	Barbados	94,68896	94,68896	94,68896	Klaster 3
16	Belarus	60813,42	60813,42	60813,42	Klaster 2
17	Belgium	43978,17	43978,17	43978,17	Klaster 2
18	Belize	32,86335	32,86335	32,86335	Klaster 3
19	Benin	1250,249	1250,249	1250,249	Klaster 3
20	Bhutan	86,97701	86,97701	86,97701	Klaster 3
...
100	US	3116411	3116411	3116411	Klaster 1

Proses pengumpulan data dilakukan sebanyak 3 tahap, dimana tahap pertama, kedua ada perubahan dan pada tahapan

ketiga data sama (tidak berubah) dengan tahap yang kedua. Hasil Clustering dari 1 dan 2 :

Cost Akhir 5516721,139
 Rumus mencari nilai akhir = cost akhir - cost awal nilai akhir 390090,7015
 Kesimpulan
 jika nilai akhir $C > 0$ maka proses literasi berhenti, jika nilai akhir $C < 0$ maka literasi dilanjutkan lagi

PEMBAHASAN

Dalam menjalankan sistem yang penulis, berikut langkah-langkah yang harus dilaksanakan :

1. Menu login merupakan menu yang akan muncul saat admin menjalankan aplikasi, selanjutnya input username dan password, berikut tampilan menu login saat menu login dipilih.



Gambar 1 Halaman Tampilan Login

2. Menu utama berisikan menu data training, data testing, kriteria, kriteria nilai dan administrator, menu-menu tersebut berfungsi untuk melakukan pengolahan data yang ada di sistem. Berikut tampilan dari menu utama dari aplikasi yang dirancang.



Gambar 2 Halaman Tampilan Utama Admin

3. Menu Data Pasien COVID-19 Di sini, Anda dapat melihat dan mengakses informasi mengenai rekam medis vital pasien, yang terdiri dari data usia, saturasi oksigen (SpO2), laju pernapasan, dan suhu tubuh.



Gambar 3 Halaman Menu Nilai Awal

4. Menu Iterasi 1 menyajikan parameter klinis yang digunakan untuk mengevaluasi tingkat keparahan pasien COVID-19. Kriteria ini mencakup indikator vital utama, yaitu usia, saturasi oksigen (SpO2), laju pernapasan, suhu tubuh, dan komorbiditas.



Gambar 4 Menu Data Iterasi 1



Gambar 5 Menu Data Iterasi 2

Hasil Akhir Menyangkut semua data yang dikelola oleh dataset dan di akumulasikan dan mendapatkan hasil yang maksimal



Gambar 6 Menu Hasil akhir

SIMPULAN

Berdasarkan hasil analisis dan pengujian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma *K-Medoids Clustering* terbukti efektif

dalam memetakan tingkat keparahan pasien COVID-19. Melalui perhitungan jarak *Manhattan Distance*, penelitian ini berhasil mengidentifikasi tiga kelompok prioritas penanganan yang distingtif, yaitu kategori Ringan, Sedang, dan Berat, berdasarkan variabel vital utama seperti saturasi oksigen, laju pernapasan, dan komorbiditas. Keunggulan metode K-Medoids dalam menangani data pencilan (*outlier*) terbukti mampu menghasilkan struktur klaster yang lebih stabil dan representatif dibandingkan metode konvensional, sehingga meminimalisir kesalahan diagnosis pada pasien dengan gejala anomali. Selain itu, hasil validasi menggunakan *Silhouette Coefficient* menunjukkan tingkat kohesi dan separasi yang baik, menegaskan bahwa sistem yang dibangun layak diimplementasikan sebagai alat bantu pendukung keputusan yang objektif bagi tenaga medis dalam mengoptimalkan proses triase dan manajemen sumber daya rumah sakit.

DAFTAR PUSTAKA

- Qomariyah, & Siregar, M. U. (2022). Comparative Study of K-Means Clustering Algorithm and K-Medoids Clustering in Student Data Clustering. *JISKA (Jurnal Informatika Sunan Kalijaga)*, <https://doi.org/10.14421/jiska.2022.7>
- Sindi, S., Ningse, W. R. O., Sihombing, I. A., R.H.Zer, F. I., & Hartama, D. (2020). Analisis Algoritma K-Medoids Clustering Dalam Pengelompokan Penyebaran Covid-19 Di Indonesia. *Jurnal Teknologi Informasi*, 4(1), 166–173. <https://doi.org/10.36294/jurti.v4i1.1296>
- Singh, A. K., Mittal, S., Malhotra, P., & Srivastava, Y. V. (2020). Clustering Evaluation by Davies-Bouldin Index(DBI) in Cereal data using K-Means. 83 *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020, Iccmc*, 306–310.
- Marlina, D. et al. 'Implementasi Algoritma K Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak', *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 4(2), p. 64. doi: 10.24014/coreit.v4i2.4498. 2018.
- Deswiasqa, K., Darmawan, E., & Sugiyarto, S. (2022). Application of K-Means for Clustering Based on the Severity of COVID-19 in Indonesian Private Hospitals. *EKSAKTA: Journal of Sciences and Data Analysis*, 3(2), 95– 102. <https://doi.org/10.20885/eksakta.vol3.iss2.art5>