

Analisa Kasus Covid-19 Menggunakan Metode Agglomerative Hierarchical Clustering pada Tingkat Kematian

Nurlaila*¹, Esa Fauzi²

^{1,2} Teknik Informatika, Fakultas Teknik, Universitas Widyatama, Jl. Cikutra, No. 204A, Sukapada, Cibeunying Kidul, Bandung, 40125, Indonesia

E-mail: nurlaila@widyatama.ac.id¹, esa.fauzi@widyatama.ac.id²

Abstract — Coronavirus disease was first discovered in the city of Wuhan, China in December 2019. Due to the impact of the new coronavirus infection, many infected patients have died including in the United States. Data obtained from the official web of Data Centers for Disease Control and Prevention shows a high mortality rate due to Covid-19. This study aims to analyze the Covid-19 mortality rate by region using the Agglomerative Hierarchical Clustering (AHC) method and find the optimal cluster validity using the Silhouette Index (SI) method. Clustering the Covid-19 death rate using the AHC method is needed to understand the pattern of death rates due to Covid-19 and assist in making policies for pandemic prevention and handling. This research resulted in the optimal number of clusters at n clusters = 2 with cluster-1 high mortality rate of 12,307 object and cluster-2 low mortality rate of 13,498 object. The results of this study can thoroughly analyze Covid-19 death data such as revealing some important findings and input that can be proposed to improve the quality of response to future pandemics.

Key word — Agglomerative Hierarchical Clustering, Covid-19, Silhouette Index.

Abstrak — Covid-19 pertama kali ditemukan di kota Wuhan, Tiongkok pada bulan Desember 2019. Akibat dampak infeksi virus corona baru, banyak pasien yang terinfeksi meninggal dunia termasuk di Amerika Serikat. Data yang diperoleh dari web resmi Data Centers for Disease Control and Prevention menunjukkan tingkat kematian yang tinggi akibat Covid-19. Penelitian ini bertujuan untuk menganalisis tingkat kematian Covid-19 berdasarkan wilayah dengan metode Agglomerative Hierarchical Clustering (AHC) dan mencari validitas klaster optimal menggunakan metode Silhouette Index (SI). Klasterisasi tingkat kematian Covid-19 menggunakan metode Agglomerative Hierarchical Clustering diperlukan untuk memahami pola tingkat kematian akibat Covid-19 dan membantu dalam pengambilan kebijakan pencegahan dan penanganan pandemi. Penelitian ini menghasilkan jumlah klaster optimal pada n klaster = 2 dengan klaster-1 tingkat kematian tinggi sebanyak 12.307 objek dan klaster-2 tingkat kematian rendah sebanyak 13.498 objek. Hasil penelitian ini dapat menganalisis secara menyeluruh terhadap data kematian Covid-19 ini mengungkapkan beberapa temuan penting dan masukan yang dapat diajukan untuk meningkatkan kualitas respon terhadap pandemi di masa depan.

Kata kunci— Covid-19, Klasterisasi Agglomerative Hierarchical, Silhouette Index.

I. PENDAHULUAN

Covid-19 pertama kali ditemukan di kota Wuhan, Tiongkok pada bulan Desember 2019 [1]. Covid-19 merupakan keluarga besar virus yang menyebar antara hewan dan manusia. Virus corona pada manusia memiliki beberapa jenis yang dapat menginfeksi saluran pernapasan dan menimbulkan penyakit mulai dari flu biasa sampai yang lebih serius seperti MERS dan SARS [2][3]. Virus ini membawa dampak kesehatan yang parah bagi banyak orang, dan menyebabkan kematian bagi banyak pasien yang terjangkit. Penetapan status pandemi pada penyebaran virus Covid-19 secara otomatis memberikan peringatan kepada semua negara yang terdampak untuk berperan aktif dalam menangani pandemi ini. Berbagai instruksi dan pedoman telah dikeluarkan oleh pemerintah sebagai langkah pencegahan Covid-19 [4].

Penyakit virus *corona* baru telah tersebar ke berbagai negara dan wilayah pada tahun 2020. Virus ini dengan cepat menjadi ancaman global bagi seluruh penduduk dunia [5]. Termasuk Amerika Serikat menjadi salah satu wilayah utama dengan jumlah kasus pandemi yang tinggi. Pada Januari 2020 kasus pertama terkonfirmasi di AS dan pada Februari 2020 kematian pertama di AS akibat positif mengidap virus tersebut di wilayah California. Dua bulan setelahnya ribuan orang meninggal dalam satu hari. AS melaporkan lebih dari 2.000 kematian dalam satu hari untuk pertama kalinya. Pada Mei 2020 lebih dari 10 juta tes dilakukan di Amerika Serikat. Amerika Serikat telah melakukan lebih dari 10,2 juta tes virus *corona*, dengan sekitar 15% orang dinyatakan positif. Sejumlah negara bagian mengalami lonjakan kasus secara tiba-tiba, termasuk: Arizona, Alaska, Florida, Kentucky, Mississippi, Oregon, New Mexico, North Carolina, South Carolina, Arkansas, California, Tennessee, Texas dan Utah, serta Puerto Rico [6].

Penelitian ini menggunakan metode data mining, yaitu teknik pengumpulan dan pengolahan data untuk menemukan pola dan informasi tersembunyi dari kumpulan data yang besar [7]. Data diperoleh dari *web resmi Data Centers for Disease Control and Prevention* menunjukkan tingkat kematian berdasarkan wilayah tempat tinggal di Negara Amerika Serikat dengan jumlah sebanyak 25.806 data kumulatif sejak 1 Januari 2020. Jumlah dan tingkat kematian di Amerika Serikat mencakup 50 negara bagian, ditambah *District of Columbia* dan *New York City* [6].

Pada penelitian sebelumnya yang berjudul “*Covid-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm*” menjelaskan klasifikasi kasus dan kematian akibat Covid-19 di Asia Tenggara. Teknik yang digunakan adalah *K-Means clustering data mining*. Menggunakan metode ini data dibagi menjadi tiga *cluster* yaitu, tinggi (C1), sedang (C2), dan rendah (C3) [2]. Metode K-Means membutuhkan penentuan jumlah *cluster* (K) sebelumnya. Pemilihan nilai K yang salah bisa menghasilkan *cluster* yang tidak representatif atau bermakna.

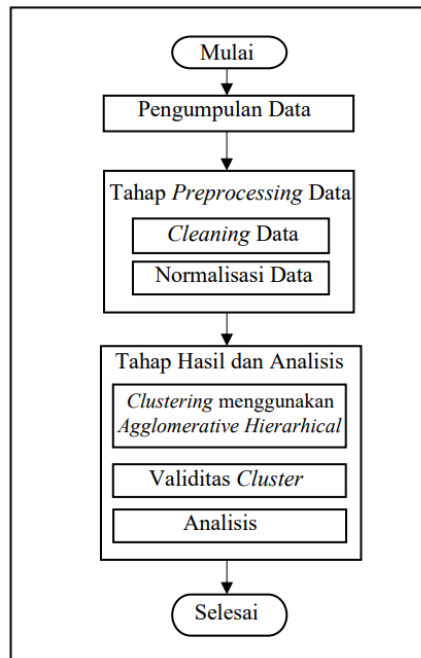
Selanjutnya penelitian yang berjudul “*Pengelompokan Data Kasus Covid-19 di Dunia Menggunakan Algoritma DBSCAN*” juga mencoba mencari solusi dengan mengelompokkan data tentang kasus Covid-19 di seluruh dunia berdasarkan wilayah. Melalui serangkaian percobaan, penelitian ini mengidentifikasi konfigurasi *cluster* yang paling optimal. Analisis tersebut mengungkapkan tiga *cluster* berbeda: *Cluster 0* terdiri dari data *outlier*, *cluster 1* mencakup negara-negara yang pertama kali mencatat peningkatan kasus Covid-19, dan *cluster 2* mencakup negara-negara yang mencatat peningkatan jumlah kasus secara signifikan. [8]. Namun penggunaan Teknik DBSCAN mungkin tidak bekerja dengan baik pada dataset yang sangat besar atau data dengan dimensi tinggi karena kompleksitas waktu yang tinggi.

Teknik Pengolahan data pada penelitian ini menerapkan teknik *clustering* untuk mengelompokkan tingkat kematian akibat Covid-19 berdasarkan wilayah. Metode *Agglomerative Hierarchical Clustering* diterapkan pada penelitian ini. *Agglomerative Hierarchical Clustering* adalah teknik *clustering* yang membentuk hierarki sehingga membentuk struktur pohon dan pengelompokan hierarki dengan pendekatan *bottom up* [9],[10]. Pengelompokan hierarki aglomeratif biasanya divisualisasikan menggunakan dendrogram [11]. Visualisasi dendrogram merupakan representasi visual berupa sebuah gambar atau diagram untuk menampilkan suatu informasi dari langkah-langkah analisis *cluster*, bagaimana *cluster* terbentuk, dan koefisien jarak pada setiap langkah yang menunjukkan nilainya [12][13].

Pada penelitian ini bertujuan untuk menganalisa *clustering* dan menemukan jumlah kluster kematian akibat Covid-19 yang optimal di Amerika Serikat yang mencakup 50 negara bagian dan dapat memberikan masukan kepada pihak terkait untuk meningkatkan kualitas respon pandemic di masa depan.

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metodologi yang terbagi menjadi beberapa langkah. Gambar 1 menunjukkan metodologi penelitian yang dilakukan:



Gambar 1: Flowchart Metode Penelitian

Pengumpulan data dilakukan di situs resmi *Data Centers for Disease Control and Prevention* (CDC). CDC adalah lembaga kesehatan publik nasional yang terkenal di Amerika Serikat. Data yang dikeluarkan umumnya diperoleh melalui proses pengumpulan dan validasi yang ketat, memastikan akurasi dan reliabilitas yang tinggi. Dataset yang dianalisa dalam penelitian ini adalah data kasus kematian akibat Covid-19 di Amerika Serikat mencakup 50 negara bagian, ditambah *District of Columbia* dan *New York City* [6].

Selanjutnya melakukan *preprocessing* pada yaitu *cleaning* data dan normalisasi data. Tujuan dari *cleaning* data adalah untuk membersihkan nama atribut dari elemen yang tidak relevan. Tahap selanjutnya melibatkan penggantian nilai yang hilang [14]. Normalisasi yaitu penskalaan nilai atribut data untuk menempatkannya pada rentang atau skala tertentu [11]. Penelitian ini menggunakan metode normalisasi *Min-Max scaler* untuk menganalisis tingkat kematian Covid-19. Min-max scaling adalah metode normalisasi data yang dilakukan agar data berada pada rentang yang lebih kecil seperti $[-1.1]$ atau $[0.1]$ sehingga tidak ada data yang terlalu besar atau terlalu kecil untuk memudahkan. Metode ini mengubah nilai data menjadi rentang yang telah ditentukan dengan nilai minimum dan maksimum [15].

Agglomerative Hierarchical Clustering (bottom-up) mengelompokkan setiap data sebagai *cluster* tunggal di awal dan kemudian secara berturut-turut menggabungkan (atau mengaglomerasi) pasangan *cluster* hingga semua *cluster* digabungkan menjadi satu *cluster* yang berisi semua data. Pengelompokan *top-down* memerlukan metode untuk memisahkan sebuah *cluster*. Ini dilanjutkan dengan membagi *cluster* secara rekursif hingga dokumen individual tercapai [11]. Ada beberapa metode untuk mengukur jarak antar objek, di antaranya penghitungan jarak *Euclidean* adalah pendekatan yang banyak digunakan. Jarak *Euclidean* adalah metode pemetaan objek yang mirip dengan analisis *cluster*, mengukur jarak antara dua titik. Secara matematis, jarak *Euclidean* dapat dinyatakan dengan menggunakan persamaan [16]:

$$D(i, j) = \sqrt{\sum_{k=1}^n (X_{ik} - Y_{jk})^2} \quad (1)$$

Keterangan :

$D(i, j)$ = jarak antara objek i dan j

n = jumlah variabel yang diamati

X_{ik} = nilai objek i pada variabel ke- k
 Y_{jk} = nilai objek j pada variabel ke- k

Adapun algoritma pada pengelompokan hierarki yang diterapkan yaitu pengelompokan *Complete Linkage*. *Complete Linkage* adalah prosedur aglomerasi berdasarkan jarak terjauh antar objek. Algoritma pencocokan penuh dimulai dengan mengidentifikasi jarak terbesar dari jarak *Euclidean* yang dihitung, sehingga menghasilkan matriks $D = \{D_{ij}\}$, Kemudian menggabungkan objek serupa seperti U dan V untuk membentuk sebuah *cluster* (UV). Pada langkah selanjutnya, jarak antara (UV) dan *cluster* lain seperti W ditentukan, yang dapat dinyatakan [10]:

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \quad (2)$$

Keterangan :

d_{UW} = jarak objek terjauh dari *cluster* U dan W
 d_{VW} = jarak objek terjauh dari *cluster* V dan W

Setelah semua hasil matriks jarak antar objek dan antar *cluster* didapatkan. Selanjutnya dilakukan evaluasi *cluster* optimal dengan teknik *Silhouette Index*. Evaluasi *cluster* digunakan untuk mengetahui keakuratan suatu kelompok data. Untuk menghitung *Silhouette Index* data ke- i menggunakan persamaan [8]:

$$SI = \frac{1}{n} \sum_{i=1}^n \left(\frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \right) \quad (3)$$

Keterangan :

$a(i)$ = jarak rata-rata sampel i ke sampel lain dalam *cluster*
 $b(i)$ = jarak minimum sampel dari sampel i ke *cluster* lain

Pada tahap terakhir melakukan analisis yang mencakup deskripsi yang mendalam tentang klaster yang dihasilkan dan implikasi yang mungkin.

III. HASIL DAN PEMBAHASAN

Data pada penelitian ini merupakan kumpulan informasi mengenai tingkat kematian yang disebabkan oleh virus Covid-19 di berbagai daerah di Amerika Serikat yang menjadi dasar utama dalam penelitian ini. Data dikumpulkan melalui situs *web* resmi *Data Centers for Disease Control and Prevention* menunjukkan tingkat kematian, dan persen dari total kematian berdasarkan wilayah tempat tinggal dengan jumlah sebanyak 25.806 data dan 14 atribut. Tabel 1 merupakan sample dataset mentah tingkat kematian yang disebabkan oleh virus Covid-19 di berbagai daerah di Amerika Serikat.

Tabel 1: Dataset tingkat kematian akibat Covid-19

No	Data_as_of	Jurisdiction_residence	Group	Data_period_start	Data_period_end	COVID_deaths	...
1	10/19/2023	Region 1	Total	1/1/2020	3/21/2020	22	...
2	10/19/2023	Region 1	Total	1/1/2020	3/28/2020	141	...
3	10/19/2023	Region 1	Total	1/1/2020	4/4/2020	630	...
4	10/19/2023	Region 1	Total	1/1/2020	4/11/2020	1873	...
5	10/19/2023	Region 1	Total	1/1/2020	4/18/2020	3782	...
...
25.806	10/19/2023	Region 8	3 Month period	7/23/2023	10/14/2023	26	...

Data dikelompokkan berdasarkan periode waktu yang berbeda termasuk periode 3 bulan, mingguan, dan total (kumulatif sejak 1 Januari 2020). Jumlah dan tingkat kematian di Amerika Serikat mencakup 50 negara bagian, ditambah *District of Columbia* dan *New York City* [7]. Tabel 2 merupakan negara-negara bagian berdasarkan wilayah di Amerika Serikat.

Tabel 2: Negara Bagian Berdasarkan Wilayah

Wilayah	Negara Bagian
Wilayah 1	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
Wilayah 2	New Jersey, New York, Kota New York, Puerto Rico
Wilayah 3	Delaware, Distrik Columbia, Maryland, Pennsylvania, Virginia, Virginia Barat
Wilayah 4	Alabama, Florida, Georgia, Kentucky, Mississippi, Carolina Utara, Carolina Selatan, Tennessee
Wilayah 5	Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin
Wilayah 6	Arkansas, Louisiana, New Mexico, Oklahoma, Texas
Wilayah 7	Iowa, Kansas, Missouri, Nebraska
Wilayah 8	Colorado, Montana, Dakota Utara, Dakota Selatan, Utah, Wyoming
Wilayah 9	Arizona, California, Hawaii, Nevada
Wilayah 10	Alaska, Idaho, Oregon, Washington

Pada tahap *preprocessing cleaning data*, dilakukan penghapusan beberapa kolom yang tidak relevan dan mengisi nilai yang hilang dengan nilai rata-rata. Gambar 2 merupakan hasil proses *cleaning data*:

	Jurisdiction_Residence	COVID_deaths	COVID_pct_of_total	crude_COVID_rate	aa_COVID_rate	---
0	Region 1	0.0	0.0	0.0	0.0	---
1	Region 1	0.0	0.0	0.0	0.0	---
2	Region 1	0.0	0.0	0.0	0.0	---
3	Region 1	0.0	0.0	0.0	0.0	---
4	Region 1	0.0	0.0	0.0	0.0	---
---	---	---	---	---	---	---
25800	Region 3	295.0	1.8	3.4	3.0	---
25801	Region 10	275.0	2.1	3.6	3.2	---
25802	Region 3	106.0	2.2	5.9	4.2	---
25803	Region 5	133.0	1.2	2.3	1.8	---
25804	Region 8	26.0	2.3	4.5	4.0	---

Gambar 2: *Cleaning Data*

Data dinormalisasi menggunakan teknik normalisasi *min-max scaling* yang memetakan data ke rentang antara 0 dan 1. Gambar 3 merupakan hasil proses data yang telah dinormalisasi:

	Jurisdiction_Residence	COVID_deaths (Normalized)	COVID_pct_of_total (Normalized)	crude_COVID_rate (Normalized)	aa_COVID_rate (Normalized)	...
0	0	0.000000	0.000000	0.000000	0.000000	...
1	0	0.000000	0.000000	0.000000	0.000000	...
2	0	0.000000	0.000000	0.000000	0.000000	...
3	0	0.000000	0.000000	0.000000	0.000000	...
4	0	0.000000	0.000000	0.000000	0.000000	...
...
25800	3	0.000257	0.026866	0.006709	0.006951	...
25801	1	0.000239	0.031343	0.007103	0.007414	...
25802	3	0.000092	0.032836	0.011642	0.009731	...
25803	5	0.000116	0.017910	0.004538	0.004171	...
25804	8	0.000023	0.034328	0.008879	0.009268	...

Gambar 3. Normalisasi Data

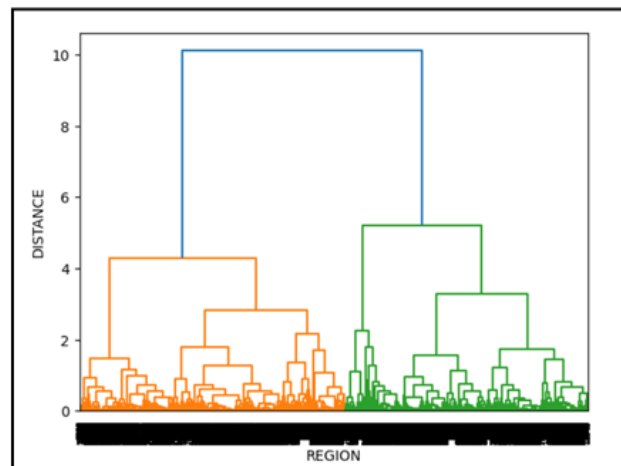
Sebelum melakukan proses *Agglomerative Hierarchical Clustering* perlu mengubah kolom objek menjadi numerik. Melakukan proses kalkulasi matriks jarak antara data menggunakan teknik *eulidean distance*. Berikut merupakan perhitungan jarak antara data ke-16 dan data ke-17 dengan teknik *eulidean distance*:

$$D(i, j) = \sqrt{\sum_{k=1}^n (X_{ik} - Y_{jk})^2}$$

$$D(16,17) = \sqrt{((1873 - 3782)^2 + (4,4 - 8)^2 + (12,4 - 25,1)^2 + (9,4 - 18,8)^2 + \dots + (0,001 - 0,002)^2)}$$

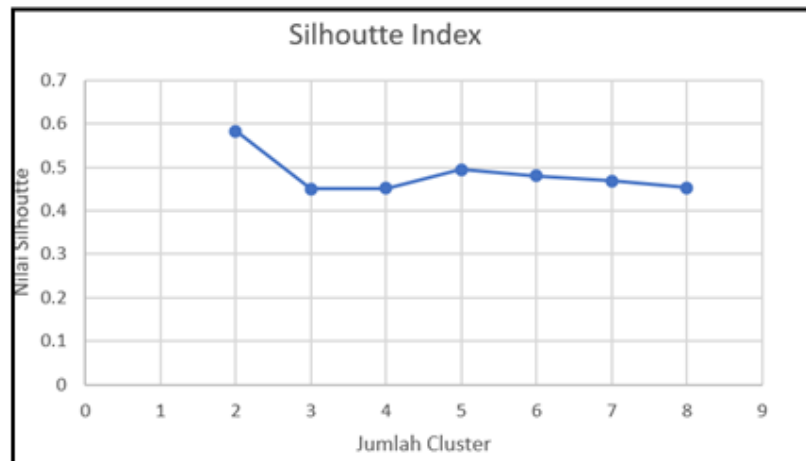
$$D(16,17) = 9,64$$

Jika jarak antar objek memiliki nilai jarak paling kecil dibandingkan jarak antar objek lainnya dalam matriks jarak *Euclidean*, maka gabungkan dua *cluster* dan pengelompokkan antar data dengan metode *complete linkage*. Pengelompokkan hierarki aglomeratif melibatkan proses bertahap penggabungan titik data dari *N cluster* awal menjadi satu *cluster*. *N* mewakili jumlah total titik data atau jumlah *cluster* awal dalam algoritma pengelompokan hierarki aglomeratif. Perbarui matriks jarak sesuai dengan teknik pengelompokan *agglomerative method* sampai hanya tersisa satu *cluster*. Gambar 4 merupakan dendrogram yang menunjukkan hasil *clustering*:



Gambar 4: Dendrogram Cluster

Berdasarkan visualisasi dendogram dari banyaknya *cluster* yang terbentuk, dendogram memotong pohon hierarki menjadi dua kelompok, yang menghasilkan dua kelompok data yang terpisah dari *cluster*-1 disimbolkan dengan warna *orange* sebanyak 12.307 data dan *cluster*-2 disimbolkan dengan warna hijau sebanyak 13.498 data. Selanjutnya evaluasi hasil *clustering* dilakukan setelah tahap sebelumnya untuk menghasilkan jumlah *cluster* yang paling cocok. *Silhouette Index* (SI) digunakan untuk menemukan nilai validitas *cluster*. Gambar 5 merupakan hasil validitas *cluster*:



Gambar 5. *Silhouette Index*

Hasil *Silhouette Index* bernilai fluktuasi. *Cluster* optimal diidentifikasi dengan nilai SI yang paling dekat dengan 1. Berdasarkan perhitungan *Silhouette Index* dengan beberapa percobaan *cluster* sebanyak 8 kali, maka didapatkan *cluster* optimal pada $n=2$ yaitu sebesar 0,58 yang diterapkan pada proses *clustering* dan analisa tingkat kematian Covid-19. Berdasarkan hasil *clustering* maka didapatkan wilayah-wilayah negara bagian pada *cluster* 1 yaitu tingkat kematian tinggi dan *cluster* 2 tingkat kematian rendah akibat Covid-19 di Negara Amerika Serikat. Tabel 3 merupakan anggota dari setiap *cluster* yang dihasilkan:

Tabel 3: Hasil Anggota *Cluster*

Cluster	Wilayah	Negara Bagian	Angka Kematian (dinormalisasi)
1	5	Illinois, Indiana, Michigan, Minnesota, Ohio, Wisconsin	0,117
	6	Arkansas, Louisiana, New Mexico, Oklahoma, Texas	0,114
	7	Iowa, Kansas, Missouri, Nebraska	0.131
	8	Colorado, Montana, North Dakota, South Dakota, Utah, Wyoming	0,110
	9	Arizona, California, Hawaii, Nevada	0,112
2	10	Alaska, Idaho, Oregon, Washington	0,114
	1	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont	0,102
	2	New Jersey, New York, New York City, Puerto Rico	0,08
	3	Delaware, District of Columbia, Maryland, Pennsylvania, Virginia, West Virginia	0,104
	4	Alabama, Florida, Georgia, Kentucky, Mississippi, North Carolina, South Carolina, Tennessee	0,110

Berdasarkan hasil anggota *cluster* yang disajikan, terdapat dua klaster wilayah yang diidentifikasi berdasarkan angka kematian (yang dinormalisasi) akibat Covid-19 di berbagai negara bagian di Amerika Serikat. Berikut adalah analisis mendalam untuk kedua klaster tersebut:

Analisis Klaster 1:

- a. Angka kematian di klaster ini berkisar antara 0,110 hingga 0,131. Wilayah dengan angka kematian tertinggi adalah Wilayah 7 (0,131) yang mencakup Iowa, Kansas, Missouri, dan Nebraska. Sedangkan angka kematian terendah terdapat di Wilayah 8 (0,110).
- b. Wilayah-wilayah ini sebagian besar mencakup negara bagian dengan populasi yang tersebar dan beragam secara geografis, mulai dari negara bagian di Midwest (seperti Illinois, Indiana) hingga negara bagian di Barat (seperti California, Oregon). Faktor seperti aksesibilitas layanan kesehatan, kepadatan populasi, dan kebijakan kesehatan mungkin mempengaruhi tingkat kematian.

Analisis Klaster 2:

- a. Angka kematian di klaster ini berkisar antara 0,080 hingga 0,110. Angka kematian terendah terdapat di Wilayah 2 (0,080) yang mencakup New Jersey, New York, New York City, dan Puerto Rico. Angka kematian tertinggi ada di Wilayah 4 (0,110).
- b. Wilayah ini termasuk area dengan kepadatan penduduk yang tinggi dan sistem kesehatan yang lebih berkembang, terutama di Northeastern US (seperti New York City dan sekitarnya). Hal ini mungkin berkontribusi pada penurunan angka kematian karena akses yang lebih baik ke fasilitas kesehatan dan upaya mitigasi yang lebih efektif.

Metode *Agglomerative Hierarchical Clustering* tidak memerlukan penentuan jumlah *cluster* di awal proses sehingga dapat dilakukan evaluasi *cluster* untuk mendapatkan *cluster* yang optimal. Dibandingkan metode *K-Means* yang membutuhkan jumlah *cluster* yang harus ditentukan sebelumnya dan sulit jika jumlah klaster optimal tidak diketahui, sehingga dapat menghasilkan jumlah *cluster* kurang optimal.

IV. SIMPULAN

Berdasarkan metodologi yang dilakukan, beberapa kesimpulan utama dapat diambil; Klaster 2 memiliki angka kematian yang sedikit lebih rendah dibandingkan dengan Klaster 1. Hal ini dapat disebabkan oleh beberapa faktor seperti perbedaan dalam kebijakan kesehatan publik, infrastruktur kesehatan, serta akses dan kualitas perawatan medis. Faktor-faktor seperti tingkat urbanisasi, demografi populasi (usia, ras, dan faktor kesehatan dasar), serta respons pemerintah terhadap pandemi dapat berperan dalam perbedaan angka kematian antar klaster. Penelitian lebih lanjut diperlukan untuk mengidentifikasi faktor-faktor spesifik yang berkontribusi pada perbedaan ini dan untuk mengembangkan strategi yang dapat mengurangi tingkat kematian di masa depan.

UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih kepada semua orang yang berpartisipasi dan memberikan dukungan penuh dalam proses ini. Apresiasi tulus juga kami sampaikan kepada dosen pembimbing kami di Universitas Widyatama atas bimbingan dan arahnya yang sangat berharga selama proses penelitian ini.

DAFTAR PUSTAKA

- [1] R. Dahlia, S. Hadiani, R. Dahlia, N. Wuryani, W. Gata, and A. Selawati, "Penerapan Data Mining Terhadap Data Covid-19 Menggunakan Algoritma Klasifikasi," *Sekol. Tinggi Ilmu Komput. Nusa Mandiri Jakarta*, vol. 21, no. 1, pp. 44–52, 1045.
- [2] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D.

- Panjaitan, "COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, pp. 0–6, 2021, doi: 10.1088/1742-6596/1783/1/012027.
- [3] C. Guevara and M. S. Penas, "Surveillance Routing of COVID-19 Infection Spread Using an Intelligent Infectious Diseases Algorithm," *IEEE Access*, vol. 8, pp. 201925–201936, 2020, doi: 10.1109/ACCESS.2020.3036347.
- [4] R. Nurlistiani, N. Purwati, and S. Yanto, "Audit E-Learning Dengan Framework Cobit 5.0 Di Masa Pandemi Covid-19," *J. Inform.*, vol. 21, no. 1, pp. 90–103, 2021, doi: 10.30873/ji.v21i1.2873.
- [5] V. Crnogorac, M. Grbic, M. Dukanovic, and D. Matic, "Clustering of European countries and territories based on cumulative relative number of COVID 19 patients in 2020," *2021 20th Int. Symp. INFOTEH-JAHORINA, INFOTEH 2021 - Proc.*, pp. 17–19, 2021, doi: 10.1109/INFOTEH51037.2021.9400670.
- [6] NCHS and DVS, "Provisional COVID-19 death counts, rates, and percent of total deaths, by jurisdiction of residence," data.cdc.gov. Accessed: Nov. 19, 2023. [Online]. Available: https://data.cdc.gov/NCHS/Provisional-COVID-19-death-counts-rates-and-percen/mpx5-t7tu/about_data
- [7] F. Dikarya and S. Muharni, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Universitas Terbaik Di Dunia," *J. Inform.*, vol. 22, no. 2, pp. 124–131, 2022, doi: 10.30873/ji.v22i2.3324.
- [8] N. Nurhaliza and M. Mustakim, "Pengelompokan Data Kasus Covid-19 di Dunia Menggunakan Algoritma DBSCAN," *Indones. J. Inform. Res. Softw. Eng.*, vol. 1, no. 1, pp. 1–8, 2021, doi: 10.57152/ijirse.v1i1.42.
- [9] D. I. Yulianti, T. I. Hermanto, and M. Defriani, "RESOLUSI: Rekayasa Teknik Informatika dan Informasi Analisis Clustering Donor Darah dengan Metode Agglomerative Hierarchical Clustering," *Media Online*, vol. 3, no. 6, p. 441, 2023, [Online]. Available: <https://djournals.com/resolusi>
- [10] A. T. R. Dani, S. Wahyuningsih, and N. A. Rizki, "Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu," *Jambura J. Math.*, vol. 1, no. 2, pp. 64–78, 2019, doi: 10.34312/jjom.v1i2.2354.
- [11] K. Pratama Simanjuntak and U. Khaira, "Hotspot Clustering in Jambi Province Using Agglomerative Hierarchical Clustering Algorithm," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 7–16, 2021.
- [12] P. Purnomo, M. P. E. Sutadji, and M. Pd, *Analisis Data Multivariat*.
- [13] N. Purwati, R. Nurlistiani, N. Purwati, R. Nurlistiani, and O. Devinsen, "Data Mining Dengan Algoritma Neural Network Dan Visualisasi Data," *J. Inform.*, vol. 20, no. 2, pp. 156–163, 2020, [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/JurnalInformatika/article/view/2273>
- [14] N. I. Khomali, "Penerapan Data Mining Dalam Menemukan Pola Data Anggaran Pendapatan Belanja Daerah Kota Malang Metode Fp-Growth," *J. Inform.*, vol. 23, no. 2, pp. 214–228, 2023, doi: 10.30873/ji.v23i2.3925.
- [15] M. YENI, "Analisis K-means pada Pengelompokan Kabupaten-Kota Provinsi Jawa Timur Berdasarkan Kasus Kesembuhan dan Kasus Kematian Covid-19," pp. 1–88, 2022.
- [16] S. K. Dirjen, P. Riset, D. Pengembangan, R. Dikti, R. O. Pratikto, and N. Damastuti, "Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir," vol. 3, no. 1, pp. 13–20, 2018.