

Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia

Rimelda Adha¹, Nana Nurhaliza², Ummi Soleha³, Mustakim⁴

^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau

Jl. HR. Soebrantas No. 155 Simpang Baru, Panam, Pekanbaru, 28293

Email: ¹11850324476@students.uin-suska.ac.id, ²11850324804@students.uin-suska.ac.id,

³11850324486@students.uin-suska.ac.id, ⁴mustakim@uin-suska.ac.id

ABSTRAK

Pada tanggal 16 Maret 2020 telah dikonfirmasi terdapat lebih dari 180.000 kasus Covid-19 di seluruh dunia, dengan lebih dari 7.000 kematian. Berbagai upaya dilakukan oleh negara-negara yang terdampak oleh Covid-19 untuk mengatasi penyebaran virus ini. Terdapat negara yang memiliki tingkat kesembuhan dan pemulihan yang tinggi, sebaliknya juga terdapat negara yang kesulitan dalam penanganan pandemi Covid-19 ini. Penelitian dilakukan untuk mengelompokkan negara-negara yang memiliki pola kasus Covid-19 di dunia. Kedepannya, hasil pengelompokan dapat dijadikan acuan dan pola gambaran negara yang memiliki tingkat pemulihan rendah dapat mengamati proses pemulihan negara yang memiliki tingkat pemulihan tinggi yang berada dalam kelompoknya. Untuk melakukan klusterisasi pada penelitian ini menggunakan algoritma DBSCAN dan K-Means. Setelah melalui beberapa percobaan diperoleh hasil bahwa K-Means lebih unggul daripada DBSCAN dalam mengelompokkan kasus Covid-19. Algoritma K-Means memiliki nilai SI terbaik sebesar 0,6902 yang terletak pada percobaan dengan nilai $k = 8$.

Kata Kunci: Covid-19, DBSCAN, K-Means, *Silhouette Index*

ABSTRACT

As of March 16, 2020 it has been confirmed that there are more than 180,000 cases of Covid-19 worldwide, with more than 7,000 deaths. Countries affected by Covid-19 have made various efforts to overcome the spread of this virus. There are countries that have high recovery and recovery rates, on the other hand, there are also countries that have difficulty handling the Covid-19 pandemic. The study was conducted to group countries that have a pattern of Covid-19 cases in the world. In the future, the results of the grouping can be used as a reference and a pattern for describing countries with a low recovery rate that can observe the recovery process for countries with a high recovery rate that are in the group. To perform clustering in this study using the DBSCAN algorithm and K-Means. After going through several experiments, it was found that K-Means was superior to DBSCAN in classifying Covid-19 cases. The K-Means algorithm has the best SI value of 0.6902 which is located in the experiment with a value of $k = 8$.

Keywords: Covid-19, DBSCAN, K-Means, *Silhouette Index*

Pendahuluan

Corona Virus Disease 2019 (Covid-19) pertama kali teridentifikasi di Wuhan, Ibukota Provinsi Hubei, China akhir Desember 2019. Virus ini tidak hanya menyerang China saja akan tetapi juga merambah ke berbagai penjuru dunia, baik negara berkembang ataupun negara maju terkena penyebaran Covid-19. Beberapa negara maju yang terkena dampak Covid-19 saat ini kewalahan dalam penanggulangannya [1]. Covid-19 disebabkan oleh sindrom pernapasan akut dari coronavirus 2 (SARS-CoV-2) yang menyebar dengan cepat secara global sehingga pada 16 Maret 2020 terdapat lebih dari 180.000 kasus Covid-19 yang dikonfirmasi di seluruh dunia, dengan lebih dari 7.000 kematian [2]. Komisi Kesehatan Nasional China melaporkan bahwa rute utama dari penularan virus corona yaitu tetesan cairan dari saluran pernapasan serta kontak langsung dari orang ke orang lainnya [3].

Berbagai langkah diambil oleh pemerintahan di berbagai Negara, langkah *social distancing* dilakukan untuk mencegah penularan dari Covid-19. Negara Korea Selatan dan China mendirikan fasilitas medis yang diperlukan, rumah sakit khusus dipilih Korea Selatan dalam menangani Covid-19, untuk menghindari penularan di Rumah Sakit mereka membangun 120 ruang isolasi dan menambah jumlah peralatan medis serta ambulans khusus [4]. Oleh karena itu, untuk mengelompokkan kasus Covid-19 pada negara-negara di dunia yang dapat dijadikan rekomendasi sebagai acuan penanganan pada suatu negara dengan mengamati negara lainnya yang berada pada satu kelompok diperlukan klusterisasi. Teknik *clustering* dalam data mining dapat dilakukan untuk klusterisasi kasus Covid-19.

Clustering adalah suatu teknik dalam data mining yang bertujuan untuk mengelompokkan data-data (objek) ke dalam beberapa *cluster* atau kelompok sehingga objek yang serupa disatukan

kedalam *cluster* yang sama, dan yang berbeda harus menjadi bagian dari *cluster* yang berbeda. Berdasarkan perspektif optimasi, tujuan utama dari clustering untuk memaksimalkan homogenitas (internal) dalam sebuah cluster dan heterogenitas (eksternal) di antara *cluster* yang berbeda [5].

Pada penelitian Kamila, Khairunnisa, dan Mustakim (2019) [6] menerapkan metode *clustering* mengenai perbandingan algoritma K-Means dan K-Medoids untuk mengelompokkan data transaksi bongkar muat di Provinsi Riau yang pada kesimpulannya tidak menunjukkan adanya perbedaan yang signifikan terkait pengelompokan data. Kemudian penelitian yang dilakukan oleh Yuwono dkk. (2009) menggunakan metode *Density Based Spatial Clustering of Applications with Noise* untuk mencari arah penyebaran wabah demam berdarah. Pada penelitian ini algoritma DBSCAN mampu menjawab kebutuhan informasi tentang daerah yang memiliki tingkat kepadatan penderita penyakit demam berdarah yang tinggi ataupun rendah yang ditunjukkan dengan adanya *noise*. Penelitian yang dilakukan oleh Mustakim dkk. (2019) [7] mengenai penerapan algoritma DBSCAN dalam mengklusterisasi kutipan teks twitter Pilkada Kota Pekanbaru, kombinasi $Eps = 0.1$ dan $MinPts = 10$ dengan nilai *Silhouette Index* terbaik yaitu 0.413 menghasilkan 31 *cluster*. Frekuensi kata yang sering muncul yaitu “kpu”, diikuti “firdaus”, “kota”, “pasang”, dan “ayat”. Penelitian ini menunjukkan Firdaus-Ayat sering muncul pada kluster dan juga sesuai dengan hasil Pilkada bahwa pasangan Firdaus-Ayat terpilih menjadi Walikota dan Wakil Walikota Kota Pekanbaru dengan perolehan 33.07% suara.

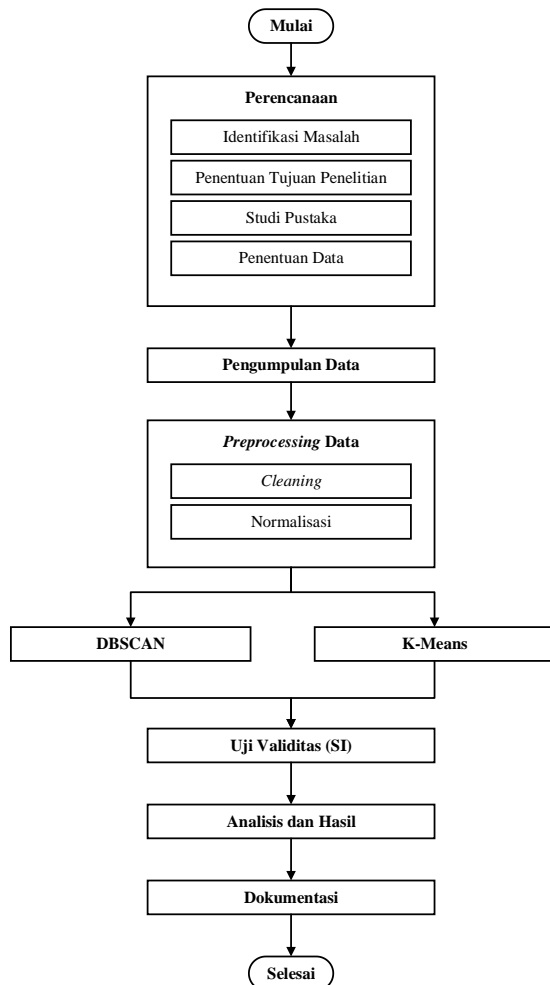
Kemudian penelitian oleh Mustakim dkk. (2021) [8] kombinasi algoritma DBSCAN dan K-Medoids untuk mengklusterisasi opini publik mengenai bencana alam di Indonesia. Pada penelitian ini algoritma DBSCAN memiliki validitas cluster dengan nilai *Silhouette Index* (SI) tertinggi yaitu 0,9140 dalam waktu eksekusi rata-rata 83,40 detik. Sementara algoritma K-Medoids memiliki nilai SI hanya 0,2258 dengan waktu eksekusi rata-rata 849,93 detik. Hasil dari penelitian ini frekuensi gempa mendominasi kategori positif, kata bencana mendominasi negatif, dan kata banjir mendominasi kategori netral. Sedangkan pada penelitian Bena Siti Ashari, Steven Christ Otniel dan Rianto [9] menerapkan metode *clustering* mengenai perbandingan kinerja K-Means dengan DBSCAN untuk metode *clustering* data penjualan online retail. Hasilnya menunjukkan penerapan K-Means menghasilkan 3 *cluster* dengan masing-masing berjumlah 103 anggota, 261 anggota dan 134 anggota. Sedangkan menggunakan DBSCAN dengan epsilon 1,005 dan *minimum points* 11 menghasilkan 3 cluster dengan masing-masing berjumlah 30 anggota, 47 anggota dan 347 anggota, serta 74 data *noise*.

Penelitian ini mengimplementasikan teknik clustering menggunakan DBSCAN dan K-Means untuk klusterisasi data Covid-19 di Dunia. Tujuan dari *clustering* adalah mengorganisasikan data sedemikian rupa kedalam kelas sehingga terdapat kesamaan intra kelas yang tinggi dan rendah [10]. Menurut Suyanto tahun 2020, DBSCAN mampu menciptakan beberapa *cluster* yang berbentuk bebas serta acak (tidak bulat) dan dapat menciptakan *cluster* dengan lebih mudah jika terdapat *noise* terhadap beberapa *cluster* tersebut. Algoritma DBSCAN dapat menemukan setiap *cluster* dengan bentuk apapun serta efektif mengidentifikasi titik-titik *noise* yang ada [11]. Sedangkan K-Means merupakan salah satu algoritma yang mudah digunakan dalam untuk melakukan klusterisasi data dalam jumlah besar dan data *outlier* dengan sangat cepat berdasarkan tingkat kemiripan dan ketidakmiripan objek [12].

Berdasarkan uraian di atas dan di dukung beberapa penelitian sebelumnya, maka pada penelitian ini akan dilakukan perbandingan algoritma DBSCAN dan K-Means untuk klusterisasi kasus Covid-19 di Dunia dengan mempertimbangkan nilai validitas cluster terbaik sebagai penentu keakuratan algoritma yang digunakan.

Metode Penelitian

Adapun metode riset dimulai dengan tahap perencanaan yang mana pada penelitian ini diperoleh topik perbandingan algoritma DBSCAN dan K-Means untuk klusterisasi kasus Covid-19 di Dunia. Tahap pengumpulan data dilakukan secara online melalui halaman website resmi *World Health Organization* (WHO) pada alamat <https://covid19.who.int>. Data yang telah diperoleh selanjutnya melalui tahap *cleaning* dan normalisasi data agar data tersebut siap untuk diolah. Selanjutnya dilakukan penerapan algoritma DBSCAN dan K-Means pada data yang sudah siap untuk diolah melalui beberapa percobaan. Pada percobaan-percobaan yang dilakukan menggunakan nilai parameter Eps dan $Minpts$ yang berbeda pada DBSCAN dan juga nilai parameter K yang berbeda pada K-Means. Pada penelitian ini digunakan *Silhouette Index* (SI) guna mengukur nilai validitas cluster. Gambaran dari metode riset dapat dilihat pada Gambar 1 berikut:



Gambar 1. Metodologi Penelitian

Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Algoritma DBSCAN dapat menemukan sampel inti dengan kepadatan tinggi dan memperluas cluster dari sampel tersebut. Terdapat dua parameter utama dari algoritma yang menentukan *cluster*: jumlah sampel minimal dan ϵ . Parameter pertama menentukan jumlah titik minimal yang dapat diklasifikasikan bersama sebagai sampel inti. Parameter ini mendefinisikan tingkat toleransi *noise* dari algoritma [13].

- 1) Tentukan nilai parameter MinPts dan Eps.
- 2) Tentukan secara acak nilai p atau titik awal.
- 3) Hitung Eps atau semua jarak titik yang *density reachable* terhadap p menggunakan rumus jarak *euclidean* berikut.

$$d_{ij} = \sqrt{\sum_a^p (x_{ia} - x_{ja})^2} \quad (1)$$

Dimana x_{ia} merupakan variabel ke-a dari obyek i ($i=1, \dots, n$; $a=1, \dots, p$) dan d_{ij} adalah nilai *euclidean distance*.

- 4) Terbentuk sebuah *cluster* ketika titik yang memenuhi Eps lebih dari MinPts dan titik p sebagai *core point*.
- 5) Lakukan pengulangan langkah 3 – 4 hingga dilakukan proses pada semua titik. Jika p merupakan titik border dan tidak ada titik yang *density reachable* terhadap p, maka proses dilanjutkan ke titik yang lain.

K-Means

K-Means merupakan salah satu metode pengelompokan data yang dapat melakukan pengelompokan data yang ada dalam bentuk dua atau lebih kelompok. Metode ini melakukan pengelompokan data sehingga data yang memiliki karakteristik serupa akan dimasukkan dalam kelompok yang sama, dan data dengan karakteristik berbeda akan dikelompokkan dalam kelompok yang berbeda pula [14]. K-means merupakan algoritma yang paling sederhana dan algoritma clustering yang paling umum digunakan [15]. Dalam melakukan perhitungan jarak ke-I (x_i) pada pusat cluster ke-k (c_k), yang diberi nama (d_{ik}), dapat digunakan rumus Euclidean dengan persamaan (2):

$$d_{ik} = \sqrt{\sum_{j=1}^m (c_{ij} - x_{ik})^2} \quad (2)$$

Silhouette Index (SI)

Menurut Aini dkk, validasi menggunakan metode Silhouette Index merupakan salah satu ukuran dari validasi yang berbasis kriteria internal [16]. Metode validaitas cluster ini berbasis kriteria internal, dimana evaluasi penempatan objek didalam setiap cluster dilakukan dengan membandingkan jarak rata-rata objek baik dalam cluster yang sama maupun cluster berbeda [17]. Untuk menghitung nilai Silhoutte Index dapat menggunakan persamaan berikut:

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

Dimana $a(i)$ merupakan jarak rata-rata sampel i ke sampel lain dalam cluster, dan $b(i)$ mewakili jarak minimum sampel dari sampel i ke *cluster* lain.

Hasil dan Pembahasan

Pembahasan dari penelitian ini meliputi tahap pengumpulan dan preprocessing data, penerapan algoritma DBSCAN dan K-Means, pengujian dan perbandingan nilai avaliditas cluster. Adapun pembahasannya sebagai berikut sebagai berikut :

Pengumpulan dan Preprocessing Data

Sumber data dalam penelitian ini diperoleh dari situs World Health Organization (WHO) yaitu <https://covid19.who.int>. Data yang diambil berupa

data kasus Covid-19 pada 197 Negara di dunia. Atribut yang terdapat pada data antaralain new case (NC), new date (ND), total case (TC), dan total death (TD). Pada tahap preprocessing data, dilakukan cleaning terhadap atribut NC dan ND karena memiliki rentang yang terlalu jauh dan memiliki banyak noise, sehingga atribut yang digunakan adalah TC dan TD. Data hasil cleaning yang digunakan pada penelitian ini seperti pada Tabel 1.

Tabel 1. Data kasus Covid-19 hasil *cleaning*

Lokasi	Tanggal					
	01/05/2020		...	31/08/2020		
	TC	TD	...	TC	TD	
Afghanistan	2171	64	...	38162	1402	
Australia	6762	92	...	25670	611	
Bulgaria	1506	66	...	16190	613	
Burundi	15	1	...	445	1	
Cambodia	122	0	...	274	0	
Cameroon	1832	61	...	19142	411	
Finland	4995	211	...	8077	335	
Guyana	82	9	...	1234	37	
Haiti	81	8	...	8209	201	
Jordan	453	8	...	1966	15	
Kazakhstan	3504	25	...	130673	1781	
Kenya	396	17	...	34057	574	
...	
Zimbabwe	34	4	...	6412	196	

Normalisasi Data

Normalisasi yang diterapkan pada data kasus Covid-19 pada penelitian ini adalah min-max normalization. Hasil normalisasi data dapat dilihat pada tabel 2.

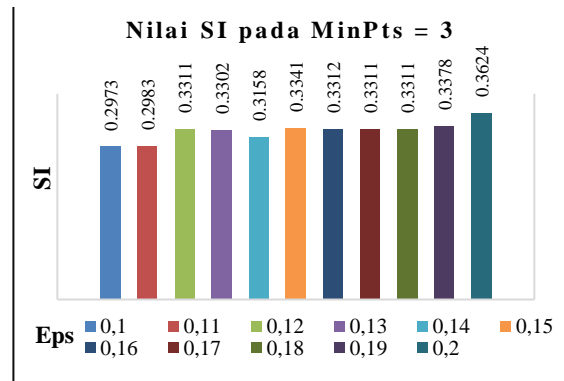
Tabel 2. Data Normalisasi

Lokasi	Tanggal					
	01/05/2020		...	31/08/2020		
	TC	TD	...	TC	TD	
Afghanistan	0,2141	0,0808	...	0,0610	0,0724	
Australia	0,6681	0,1162	...	0,0410	0,0316	
Bulgaria	0,1483	0,0833	...	0,0259	0,0317	
Burundi	0,0009	0,0013	...	0,0007	0,0001	
Cambodia	0,0115	0,0000	...	0,0004	0,0000	
Cameroon	0,1806	0,0770	...	0,0306	0,0212	
Finland	0,4934	0,2664	...	0,0129	0,0173	
Guyana	0,0075	0,0114	...	0,0020	0,0019	
Haiti	0,0074	0,0101	...	0,0131	0,0104	
Jordan	0,0442	0,0101	...	0,0031	0,0008	
Kazakhstan	0,3459	0,0316	...	0,2090	0,0920	
Kenya	0,0386	0,0215	...	0,0545	0,0296	
...	
Zimbabwe	0,0028	0,0051	...	0,0102	0,0101	

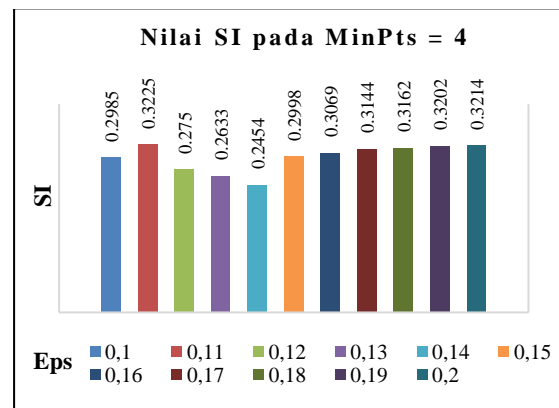
Klasterisasi dengan Algoritma DBSCAN

Proses klasterisasi menggunakan algoritma DBSCAN dengan percobaan cluster sebanyak 22 kali dimana setiap percobaan menggunakan nilai

epsilon (Eps) dan MinPoints (MinPts) yang berbeda. Nilai Eps yang digunakan pada penelitian ini yaitu pada rentang 0,1 hingga 0,2, sementara nilai MinPts 3 dan 4, selanjutnya dilakukan uji validitas cluster menggunakan SI. Nilai SI berdasarkan hasil klasterisasi menggunakan DBSCAN dapat dilihat pada Gambar 2 dan Gambar 3.



Gambar 2. Nilai SI pada MinPts 3

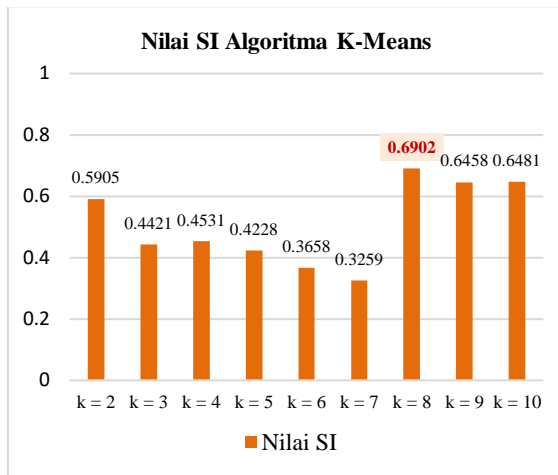


Gambar 3. Nilai SI pada MinPts 4

Berdasarkan nilai SI terbaik, *cluster* terbaik pada algoritma DBSCAN terletak pada percobaan dengan nilai Eps 0,2 dan nilai MinPts 3. Untuk nilai SI diperoleh 0,3624 disusul percobaan dengan nilai Eps 0,15 dan MinPts 3 diperoleh nilai SI 0,3341.

Klasterisasi dengan Algoritma K-Means

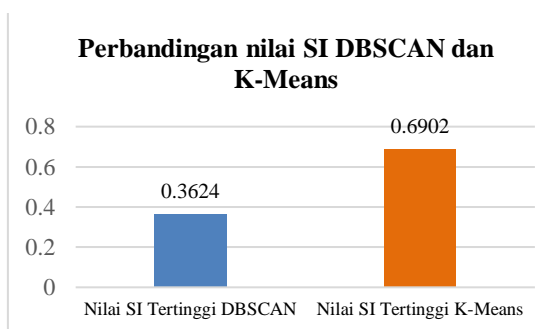
Proses klasterisasi menggunakan algoritma K-Means dengan percobaan cluster 2, 3, 4, 5, 6, 7, 8, 9, dan 10 berdasarkan penelitian Insanul Kamila, Ulya Khairunnisa, dan Mustakim (2019), Kemudian juga dilakukan validitas cluster menggunakan Silhouette Index (SI). Nilai SI berdasarkan hasil klasterisasi data menggunakan algoritma K-Means dapat dilihat Gambar 4 berikut.



Gambar 4. Grafik nilai SI hasil klusterisasi K-Means

Perbandingan Algoritma DBSCAN dan K-Means

Perbandingan algoritma DBSCAN dan K-Means dilakukan dengan menentukan jumlah cluster yang paling optimal dilihat dari nilai Silhouette Index (SI). Grafik perbandingan nilai SI tertinggi pada setiap algoritma dapat dilihat pada Gambar 5 berikut.



Gambar 5. Grafik perbandingan nilai SI DBSCAN dan K-Means

Hasil Pengujian validitas cluster terhadap algoritma DBSCAN dan K-Means menggunakan nilai Silhouette Index (SI) sudah dilakukan. Klusterisasi menggunakan DBSCAN yang memiliki nilai SI terbaik yaitu 0.3624 dengan nilai Eps 0,2 dan nilai MinPts 3. Sementara dari hasil klusterisasi menggunakan K-Means, nilai SI terbaik diperoleh percobaan k=8 dengan nilai 0,6902. Maka pada penelitian ini, algoritma K-Means memiliki nilai validitas cluster lebih baik dibandingkan algoritma DBSCAN. Dengan demikian, pada diperoleh cluster paling optimal yaitu percobaan menggunakan algoritma K-Means dengan nilai k=8.

Kesimpulan

Berdasarkan hasil penelitian dan pengolahan data yang telah dilakukan, pengolahan data menggunakan algoritma DBSCAN diperoleh *cluster* terbaik dengan nilai Eps 0,2 dan MinPts 3, sementara

pada algoritma K-Means diperoleh *cluster* terbaik pada k = 8. Hasil klusterisasi tersebut sudah di validasi menggunakan *Silhouette Index* (SI). Berdasarkan hasil pengujian validasi *cluster* terhadap hasil klusterisasi data kasus Covid-19 di dunia menggunakan algoritma DBSCAN dan K-Means, maka pada penelitian ini K-Means lebih unggul daripada DBSCAN dengan nilai SI terbaik yaitu 0,6902 dengan nilai k = 8. Pola dari hasil penelitian dapat dijadikan sebagai acuan dalam menggambarkan model *clustering* Covid-19 di Dunia.

Daftar Pustaka

- [1] A. Bénassy-Quéré and B. Weder di Mauro, "Europe in the time of Covid-19: A new crash test and a new opportunity," *Eur. Time Covid-19*, p. 1, 2020.
- [2] P. Dashraath *et al.*, "Coronavirus disease 2019 (COVID-19) pandemic and pregnancy," *Am. J. Obstet. Gynecol.*, vol. 222, no. 6, pp. 521–531, 2020, doi: 10.1016/j.ajog.2020.03.021.
- [3] H. Qi *et al.*, "COVID-19 transmission in Mainland China is associated with temperature and humidity: a time-series analysis," *Sci. Total Environ.*, vol. 728, p. 138778, 2020.
- [4] M. Her, "Repurposing and Reshaping of Hospitals During the COVID-19 Outbreak in South Korea," *One Heal.*, vol. 10, no. 1, pp. 1–3, 2020, doi: 10.1016/j.onehlt.2020.100137.
- [5] A. C. Benabdellah, A. Benghabrit, and I. Bouhaddou, "A survey of clustering algorithms for an industrial context," *Procedia Comput. Sci.*, vol. 148, pp. 291–302, 2019, doi: 10.1016/j.procs.2019.01.022.
- [6] I. Kamila, U. Khairunnisa, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau," *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, p. 119, 2019, doi: 10.24014/rmsi.v5i1.7381.
- [7] Mustakim *et al.*, "DBSCAN algorithm: Twitter text clustering of trend topic pilkada pekanbaru," *J. Phys. Conf. Ser.*, vol. 1363, no. 1, 2019, doi: 10.1088/1742-6596/1363/1/012001.
- [8] Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, "Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms," *J. Phys. Conf. Ser.*, vol. 1783, no. 1, p. 012016, 2021, doi: 10.1088/1742-6596/1783/1/012016.

- [9] B. S. Ashari, S. C. Otniel, and R. Rianto, "Perbandingan Kinerja K-Means dengan DBSCAN untuk Metode Clustering data Penjualan Online Retail," *J. Siliwangi Seri Sains dan Teknol.*, vol. 5, no. 2, pp. 64–67, 2019.
- [10] A. Malav, K. Kadam, and P. Kamat, "Prediction of heart disease using kmeans and artificial neural network as hybrid approach to improve accuracy," *Int. J. Eng. Technol.*, vol. 9, no. 4, pp. 3081–3085, 2017.
- [11] W. Jing, C. Zhao, and C. Jiang, "An improvement method of DBSCAN algorithm on cloud computing," *Procedia Comput. Sci.*, vol. 147, pp. 596–604, 2019, doi: 10.1016/j.procs.2019.01.208.
- [12] B. M. Metisen and H. L. Sari, "Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila," *J. Media Infotama*, vol. 11, no. 2, pp. 110-118. ISSN: 1858 – 2680, 2015.
- [13] M. Pietrzykowski, "Comparison of mini-models based on various clustering algorithms," *Procedia Comput. Sci.*, vol. 176, pp. 3563–3570, 2020, doi: 10.1016/j.procs.2020.09.030.
- [14] Rahmawati and M. Faisal, "Analisis Cluster untuk Pengelompokan Desa Berdasarkan Indikator Penyakit Diare," *Saintifik*, vol. 5, no. 1, pp. 75–80, 2019.
- [15] I. D. Iskandar, M. W. Pertiwi, M. Kusmira, and I. Amirulloh, "Komparasi Algoritma Clustering Data Media Online," *J. IKRA-ITH Inform.*, vol. 2, no. 4, pp. 1–8, 2018.
- [16] M. A. Nahdliyah, T. Widiharih, and A. Prahutama, "Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index," vol. 8, no. 2, pp. 161–170, 2019.
- [17] X. Wang and Y. Xu, "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 569, no. 5, 2019, doi: 10.1088/1757-899X/569/5/052024.