



Clustering of Data Covid-19 Cases in the World Using DBSCAN Algorithms

Pengelompokan Data Kasus Covid-19 di Dunia Menggunakan Algoritma DBSCAN

Nana Nurhaliza^{1*}, Mustakim²

^{1,2}Information System, State Islamic University of Sultan Syarif Kasim Riau, Indonesia

E-Mail: ¹11850324804@students.uin-suska.ac.id, ²mustakim@uin-suska.ac.id

Makalah: Diterima 01 Februari 2021; Diperbaiki 10 Februari 2021; Disetujui 19 Februari 2021
Corresponding Author: Nana Nurhaliza

Abstrak

World Health Organization (WHO) telah mendeklarasikan Virus Corona sebagai pandemi global pada tanggal 11 Maret 2020. Covid-19 sejak awal tahun 2020 hingga akhir April 2020 telah menyebar ke 210 negara dan menghasikan tiga juta kasus positif setiap harinya. Berbagai upaya dilakukan pada setiap negara dalam penanggulangan pandemi ini. Penelitian ini menggunakan teknik *clustering* untuk mengelompokkan negara-negara dengan pola kasus serupa yang dapat dijadikan rekomendasi untuk acuan penanganan pada suatu negara dengan mengamati negara lainnya yang berada pada satu kelompok. Algoritma DBSCAN diterapkan pada penelitian ini untuk mendapatkan hasil klasterisasi dan validitas *cluster* diuji dengan Silhouette Index. Dilakukan sebanyak 22 percobaan dengan rentang nilai Eps 0,1 hingga 0,2 dan nilai Minpts yaitu 3 dan 4. Pada penelitian ini diperoleh percobaan dengan nilai Eps 0,2 dan MinPts 3 yang mengasilkan sejumlah 3 *cluster* memperoleh hasil validitas *cluster* terbaik dengan nilai Silhouette Index sebesar 0,3624. Manfaat dari penelitian ini adalah didapatkannya nilai validitas *cluster* sehingga dapat diketahui *cluster* manakah yang merupakan *cluster* paling optimal.

Keyword: Covid-19, DBSCAN, *Clustering*, *Silhouette Index*.

Abstract

The World Health Organization (WHO) declared the Corona Virus a global pandemic on March 11, 2020. Covid-19 from the beginning of 2020 to the end of April 2020 has spread to 210 countries and resulted in three million positive cases every day. Various efforts were made in each country in dealing with this pandemic. This study uses a clustering technique to group countries with a similar pattern of cases which can be used as recommendations for reference in handling a country by observing other countries in one group. The DBSCAN algorithm was applied in this study to obtain clustering results and the cluster validity was tested using the Silhouette Index. A total of 22 experiments were conducted with an Eps value range of 0.1 to 0.2 and a Minpts value of 3 and 4. In this study, an experiment with an Eps value of 0.2 and a MinPts 3 resulted in 3 clusters obtaining the best cluster validity results with the Silhouette value. Index of 0.3624. The benefit of this research is getting the value of cluster validity so that it can be seen which cluster is the most optimal cluster.

Keyword: Covid-19, DBSCAN, *Clustering*, *Silhouette Index*.

1. Pendahuluan

Pada Akhir Desember 2019, penyakit Covid-19 pertama kali teridentifikasi di Wuhan, Ibukota Provinsi Hubei, China. Menurut Huang, gejala klinis yang khas dari penyakit ini yaitu demam, batuk kering, mialgia, pneumonia dan dapat menyebabkan kegagalan pernapasan progresif karena kerusakan pada alveolar dan dapat berujung pada kematian. Komisi Kesehatan Nasional China melaporkan bahwa tetesan cairan yang berasal dari saluran pernapasan serta kontak langsung dari orang ke orang merupakan rute utama dari penularan virus corona [1]. World Health Organization (WHO) telah mendeklarasikan Virus Corona sebagai pandemi global pada tanggal 11 Maret 2020 [2]. Pada tanggal 23 Maret 2020, dilaporkan oleh Komisi Kesehatan Nasional China total 81.603 kasus yang dikonfirmasi di China (61,28% di Wuhan) dan 3276 kematian telah dilaporkan

di China daratan [1]. Untuk penyebaran di berbagai belahan dunia, penyakit Covid-19 sejak awal tahun 2020 hingga akhir April 2020 telah menyebar ke 210 negara dan menghasikan tiga juta kasus positif setiap harinya [3]. Hal yang menyebabkan pandemi ini sulit ditanggulangi adalah belum ditemukannya obat untuk menanggulangi penyebaran Covid-19 [2]. Kondisi ini memaksa organisasi pemerintahan di berbagai negara untuk mengambil berbagai langkah bijak dalam menanggulangi pandemi ini.

Berbagai langkah dilakukan oleh negara yang terdampak virus ini untuk mengurangi tingkat penularan. Langkah yang diambil sebagian besar pemerintahan di berbagai negara untuk mencegah penularan dari Covid-19 dengan melakukan *social distancing*. Sebagian negara seperti Korea Selatan dan China dengan tanggap membangun fasilitas medis yang diperlukan. Sebagai contoh Korea Selatan menunjuk rumah sakit khusus untuk penanganan medis Covid-19, membangun 120 ruang isolasi baru serta menambah jumlah peralatan medis dan ambulans khusus yang bertujuan untuk menghindari penularan di Rumah Sakit [4]. Untuk itu, diperlukan klasterisasi data 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. Klasterisasi kasus Covid-19 dapat dilakukan dengan teknik *clustering* dalam data mining.

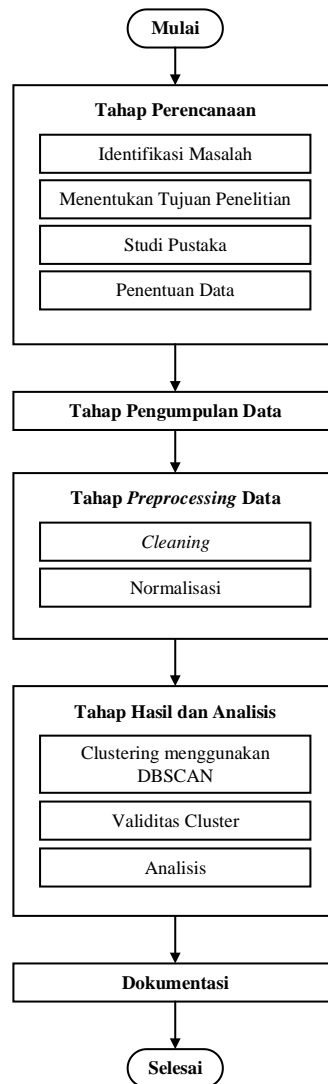
Clustering merupakan teknik dalam data mining yang bertujuan untuk mengelompokkan data-data (objek) ke dalam beberapa *cluster* atau kelompok sehingga objek yang serupa disatukan kedalam *cluster* yang sama, sedangkan yang berbeda harus menjadi bagian dari *cluster* yang berbeda. Lebih lanjut, dari perspektif optimasi, tujuan utama dari *clustering* adalah untuk memaksimalkan homogenitas (internal) dalam sebuah *cluster* dan heterogenitas (eksternal) di antara *cluster* yang berbeda [5]. Objek pengelompokan pada metode ini tidak untuk melakukan prediksi sebuah target pada variabel kelas, melainkan hanya mengelompokkan data dengan struktur serupa [6]. Menurut Suyanto (dalam Rohalidyawati dkk., 2020), pada dasarnya DBSCAN dapat menciptakan beberapa *cluster* yang berbentuk bebas dan acak (tidak bulat) dan dapat menciptakan *cluster* dengan lebih mudah jika terdapat noise pada beberapa *cluster* tersebut. Algoritma ini membentuk daerah yang memiliki kerapatan tinggi menjadi beberapa *cluster* dengan dua parameter yang diperlukan yaitu radius ketegangan, yang disimbolkan dengan ϵ , dan minimum objek yang menjadi batas kepadatan (density threshold) untuk menentukan kepadatan dari suatu wilayah tersebut, yang disimbolkan MinObj. Algoritma DBSCAN mampu menemukan setiap cluster dengan bentuk apapun dan secara efektif mengidentifikasi titik-titik noise yang ada [8].

Penelitian menggunakan algoritma DBSCAN telah banyak dilakukan. Hermanto dan Sunandar pada tahun 2020 melakukan penelitian tentang analisis data sebaran penyakit menggunakan algoritma *Density Based Spatial Clustering of Applications with Noise*. Penelitian ini melakukan pengolahan *dataset* melalui beberapa skenario dengan menggunakan beberapa nilai *epsilon* dan nilai minPts. Pada skenario ditentukan nilai *epsilon* sebesar 0,01, 0,011, 0,012 dan 0,013, minPts sebesar 1, 2, 3, 4, 5, dan diterapkan pada 4 penyakit yaitu ISPA, TBC, Asma, dan Pneumonia. Diperoleh hasil berupa 5 cluster penyakit ISPA dengan daerah sebaran penyakit paling tinggi terletak pada cluster 1, 4 Cluster Penyakit TBC dengan daerah sebaran penyakit tertinggi pada cluster 4, 4 Cluster penyakit asma dengan daerah sebaran penyakit tertinggi pada cluster 4, serta 4 Cluster penyakit pneumonia dengan daerah sebaran penyakit tertinggi pada cluster 4. Selanjutnya Yuwono dkk. tahun 2009 telah melakukan implementasi metode *Density Based Spatial Clustering of Applications with Noise* untuk mencari arah penyebaran wabah demam berdarah. Pada penelitian ini algoritma DBSCAN dapat menjawab kebutuhan informasi tentang daerah yang memiliki tingkat kepadatan penderita penyakit demam berdarah yang tinggi ataupun rendah yang ditunjukkan oleh adanya *noise*.

Berdasarkan penelitian sebelumnya, maka pada penelitian ini akan dilakukan klasterisasi kasus Covid-19 di dunia dengan menerapkan algoritma DBSCAN. Penelitian ini menggunakan data total kasus dan total kematian dari kasus Covid-19 setiap harinya pada tiap negara di dunia untuk menemukan jumlah cluster yang paling optimal.

2. Metodologi Penelitian

Pada Gambar 1 dapat dilihat gambaran dari metodologi penelitian yang dilakukan.



Gambar 1. Metodologi Penelitian

Metode riset pada penelitian ini dimulai dari tahap perencanaan. Pada tahapan ini dilakukan identifikasi masalah terlebih dahulu yang dilanjutkan dengan menentukan tujuan penelitian. Dilakukannya penentuan dari tujuan penelitian ini agar penelitian dapat menjadi lebih jelas dan terarah. Berikutnya dilakukan studi pustaka agar peneliti memperoleh dasar-dasar referensi yang kuat dalam menerapkan metode yang digunakan dalam penelitian. Teori-teori yang dipelajari yaitu Covid-19, Data Mining, dan DBSCAN yang diperoleh dari buku, jurnal, artikel, serta sumber lainnya. Setelah dilakukan studi pustaka, peneliti melakukan penentuan data yang akan digunakan pada penelitian. *Dataset* yang digunakan pada penelitian ini merupakan data kasus Covid-19 di dunia.

Tahapan yang dilakukan berikutnya adalah pengumpulan data. Pengumpulan data dilakukan secara online pada situs web resmi World Health Organization (WHO) pada alamat <https://covid19.who.int>. Dilakukan proses *preprocessing* pada data yang telah diperoleh dengan melakukan *cleaning* dan normalisasi pada data. Adapun normalisasi yang diterapkan pada data yaitu normalisasi min-max. Pada tahapan Hasil dan Analisis dilakukan *clustering* dengan algoritma DBSCAN. Setelah diperoleh hasil *cluster* melalui beberapa percobaan, selanjutnya dihitung nilai dari *Silhouette Index* pada setiap percobaan. Jika nilai validitas *cluster* sudah diperoleh, maka dapat dilakukan analisis.

2.1 Clustering

Clustering merupakan teknik dalam data mining yang bertujuan untuk mengelompokkan data-data (objek) ke dalam beberapa *cluster* atau kelompok sehingga objek dengan pola serupa disatukan kedalam *cluster* yang sama, sedangkan objek dengan pola berbeda harus menjadi bagian dari *cluster* yang berbeda. Adapun tujuan utama dari *clustering* adalah untuk memaksimalkan tingkat kesamaan pada sebuah *cluster* dan perbedaan pada

cluster yang berbeda [5]. Data yang dikelompokkan menggunakan metode ini tidak untuk diprediksi, melainkan hanya mengelompokkan data dengan struktur serupa [6].

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

Algoritma DBSCAN (Density-Based Spatial Clustering of Applications with Noise) adalah salah satu metode clustering data spasial berdasarkan kepadatan yang dikemukakan oleh Ester Martin. Algoritma ini dapat menemukan cluster dengan bentuk apa pun pada satu kondisi kepadatan [8]. Adapun urutan algoritma DBSCAN sebagai berikut [11]:

- 1) Inisialisasi parameter input MinPts dan Eps.
- 2) Tentukan titik awal atau p secara acak.
- 3) Lakukan perhitungan Eps atau semua jarak titik yang density reachable terhadap p menggunakan rumus jarak euclidean.

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

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) Jika titik yang memenuhi Eps lebih dari MinPts maka titik p merupakan core point dan terbentuk sebuah cluster.
- 5) Ulangi 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.

2.3 Silhouette Index (SI)

Menurut Aini dkk, metode validasi *Silhouette Index* merupakan salah satu ukuran dari validasi yang berbasis kriteria internal [12]. *Silhouette Index* dapat menganalisa posisi setiap objek dalam *cluster* dan dapat memperoleh jumlah *cluster* yang optimal dengan perbedaan antara jarak rata-rata di dalam *cluster* dan jarak minimum antara *cluster* [13]. Untuk menghitung nilai *Silhouette Index* dapat menggunakan persamaan berikut:

$$SI = \frac{1}{n} \sum_{i=1}^n \left(\frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \right) \quad (2)$$

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

3. Hasil dan Analisis

Adapun hasil dan analisis penelitian ini dijabarkan sebagai berikut.:

3.1 Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data kasus Covid-19 pada 197 negara di Dunia. Atribut yang digunakan adalah total kasus (TC) serta total kematian (TD) setiap harinya. Adapun sumber data pada penelitian ini berasal dari halaman *web* resmi *World Health Organization* (WHO) pada alamat <https://covid19.who.int>. Berikut hasil *cleaning* data pada Tabel 1 sebagai berikut:

Tabel 1. Data *Cleaning*

Lokasi	Tanggal								
	01/05/2020		02/05/2020		03/05/2020		...	31/08/2020	
	TC	TD	TC	TD	TC	TD	...	TC	TD
Afghanistan	2171	64	2335	68	2469	72	...	38162	1402
Albania	773	31	782	31	789	31	...	9380	280
Algeria	4006	450	4154	453	4295	459	...	44146	1501
Andorra	745	42	746	43	747	44	...	1124	53
Angola	27	2	29	2	35	2	...	2624	107
Antigua and Barbuda	24	3	25	3	25	3	...	94	3
Argentina	4415	218	4519	225	4668	237	...	401226	8401
Armenia	2148	33	2273	33	2386	35	...	43750	877
Aruba	100	2	100	2	100	2	...	1997	10
Australia	6762	92	6767	93	6783	93	...	25670	611
Azerbaijan	1804	24	1854	25	1894	25	...	36309	531
Bahamas	81	11	82	11	83	11	...	2167	43
...

Lokasi	Tanggal								
	01/05/2020		02/05/2020		03/05/2020		...	31/08/2020	
	TC	TD	TC	TD	TC	TD	...	TC	TD
Zimbabwe	34	4	34	4	34	4	...	6412	196

3.2 Normalisasi

Normalisasi yang diterapkan pada data adalah *min-max normalization*. Hasil dari normalisasi data dapat dilihat pada Tabel 2 sebagai berikut:

Tabel 2. Normalisasi

Lokasi	Tanggal								
	01/05/2020		02/05/2020		03/05/2020		...	31/08/2020	
	TC	TD	TC	TD	TC	TD	...	TC	TD
Afghanistan	0,2141	0,0808	0,2209	0,0850	0,2273	0,0866	...	0,0610	0,0724
Albania	0,0759	0,0391	0,0736	0,0388	0,0723	0,0373	...	0,0150	0,0145
Algeria	0,3956	0,5682	0,3934	0,5663	0,3958	0,5523	...	0,0706	0,0775
Andorra	0,0731	0,0530	0,0702	0,0538	0,0684	0,0529	...	0,0018	0,0027
Angola	0,0021	0,0025	0,0022	0,0025	0,0027	0,0024	...	0,0042	0,0055
Antigua and Barbuda	0,0018	0,0038	0,0018	0,0038	0,0018	0,0036	...	0,0001	0,0002
Argentina	0,4360	0,2753	0,4280	0,2813	0,4302	0,2852	...	0,6419	0,4338
Armenia	0,2118	0,0417	0,2150	0,0413	0,2196	0,0421	...	0,0700	0,0453
Aruba	0,0093	0,0025	0,0089	0,0025	0,0087	0,0024	...	0,0032	0,0005
Australia	0,6681	0,1162	0,6412	0,1163	0,6254	0,1119	...	0,0410	0,0316
Azerbaijan	0,1778	0,0303	0,1752	0,0313	0,1742	0,0301	...	0,0581	0,0274
Bahamas	0,0074	0,0139	0,0072	0,0138	0,0071	0,0132	...	0,0034	0,0022
...
Zimbabwe	0,0028	0,0051	0,0027	0,0050	0,0026	0,0048	...	0,0102	0,0101

3.3 Klasterisasi dengan DBSCAN

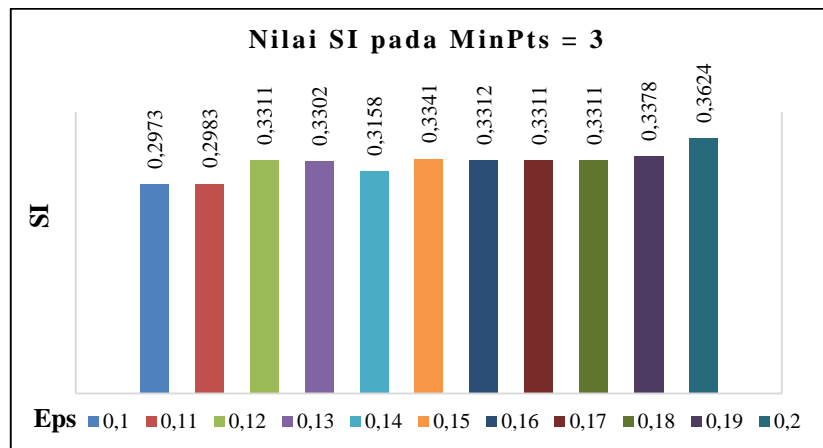
Dilakukan beberapa percobaan clustering menggunakan algoritma DBSCAN untuk menemukan jumlah cluster optimal. Setiap percobaan menggunakan nilai Eps dan MinPts yang berbeda-beda. Nilai Eps yang digunakan pada penelitian ini pada rentang 0,1 hingga 0,2 serta nilai MinPts 3 dan 4. Hasil klasterisasi dapat dilihat pada Tabel 3 berikut:

Tabel 3. Hasil Cluster Data

Percobaan	Eps	MinPts	Jumlah Cluster	Noise
1	0,10	3	5	76
2	0,10	4	3	83
3	0,11	3	4	73
4	0,11	4	4	77
5	0,12	3	4	69
6	0,12	4	3	75
7	0,13	3	4	65
8	0,13	4	3	70
9	0,14	3	4	64
10	0,14	4	3	68
11	0,15	3	3	62
12	0,15	4	2	66
13	0,16	3	3	59
14	0,16	4	2	64
15	0,17	3	3	57
16	0,17	4	2	62
17	0,18	3	3	57
18	0,18	4	2	61
19	0,19	3	4	53
20	0,19	4	3	57
21	0,20	3	3	53
22	0,20	4	2	56

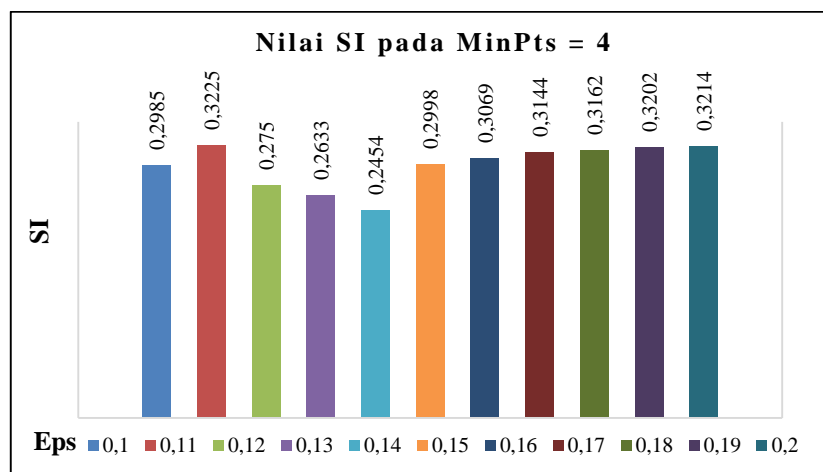
3.4 Validitas Cluster

Validitas cluster dilakukan setelah dilakukannya tahap clustering untuk menemukan jumlah *cluster* yang paling optimal. Untuk menemukan nilai validitas cluster menggunakan *Silhouette Index* (SI). Adapun hasil validitas cluster dapat dilihat pada Gambar 2 dan Gambar 3 berikut:



Gambar 2. Validitas Cluster pada MinPts 3

Pada Gambar 2 terlihat bahwa nilai SI pada tiap percobaan mengalami pola naik turun. Seluruh nilai SI pada percobaan ini bernilai positif yang artinya objek ke-*i* sudah berada pada kelompok yang tepat. Nilai SI tertinggi terletak pada nilai Eps 0,2. Percobaan yang menggunakan nilai Eps 0,2 dan MinPts 3 merupakan percobaan ke-21. Pada percobaan tersebut dihasilkan sebanyak 3 *cluster* dengan *noise* sebanyak 53.



Gambar 3. Validitas Cluster pada MinPts 4

Gambar 3 memperlihatkan grafik naik turun nilai SI pada tiap percobaan yang menggunakan nilai MinPts 4. Pada percobaan menggunakan nilai MinPts 4, nilai SI mengalami pola kenaikan dari nilai Eps 0,14 hingga 0,2, namun nilai SI tertinggi terletak pada percobaan menggunakan nilai Eps 0,11. Seluruh nilai SI pada percobaan menggunakan MinPts 4 juga bernilai positif yang artinya objek ke-*i* pada percobaan ini sudah berada pada kelompok yang tepat. *Cluster* paling optimal ditentukan dengan nilai SI terbaik yaitu yang paling mendekati 1. Percobaan dengan nilai Eps 0,2 dan MinPts 3 memiliki nilai SI yang paling mendekati 1 yaitu sebesar 0,3624 sehingga dapat dikatakan sebagai *cluster* paling optimal.

3.5 Analisis Cluster

Berdasarkan beberapa percobaan yang telah dilakukan, diperoleh percobaan ke-21 sebagai cluster yang paling optimal. Percobaan tersebut menghasilkan sebanyak 3 cluster. Adapun jumlah anggota pada setiap cluster dapat dilihat pada Tabel 4 berikut:

Tabel 4. Hasil Cluster Paling Optimal

cluster 0	cluster 1	cluster 2
Afghanistan	Albania	Bosnia and Herzegovina
Algeria	Andorra	Bulgaria
Argentina	Angola	Macedonia
Armenia	Guyana	
Australia	Aruba	
Azerbaijan	Bahamas	
Bahrain	Barbados	
...	...	
Bolivia	Belize	

Berdasarkan Tabel 4, terdapat 158 negara yang berhasil di klaster dengan kasus covid-19 yang beragam. Adapun masing-masing karakteristik cluster dijelaskan pada analisis di bawah ini.

- 1) Pada cluster 0 memiliki 53 anggota, merupakan data *outlier* yang juga dianggap sebagai data noise. Indonesia dan Philipina menjadi negara dengan kasus Covid-19 tertinggi di Asia Tenggara termasuk ke dalam cluster ini. Negara-negara dengan total kasus akhir sebanyak ratusan ribu termasuk pada *cluster* ini. Akan tetapi, titik pada anggota *cluster* ini memiliki jarak yang jauh dengan medoid. Sehingga *cluster* 0 tidak digunakan untuk keperluan analisis *cluster*.
- 2) Pada cluster 1 memiliki 103 anggota, negara di dalam kelompok ini tidak di dominasi oleh negara dengan kasus tertinggi di benuanya masing-masing. Akan tetapi, pada kelompok ini mulai mengalami lonjakan kasus Covid-19. Pada kelompok ini lebih cenderung diisi oleh negara kecil atau negara kepulauan.
- 3) Pada cluster 2 memiliki tiga anggota yaitu diisi oleh negara-negara Balkan atau negara Eropa Tenggara. Kasus Covid-19 ketiga negara tersebut mengalami kenaikan yang signifikan, namun memiliki total kasus akhir tidak lebih dari 20.000 kasus. Negara pada *cluster* ini seperti Bosnia and Herzegovina kesulitan dalam impor dan pendistribusian APD serta peralatan medis lainnya sehingga memengaruhi kualitas pelayanan medis [16].

4. Kesimpulan

Berdasarkan penelitian yang dilakukan, diperoleh beberapa kesimpulan yaitu Hasil validitas cluster tertinggi menggunakan Silhouette Index (SI) terletak pada percobaan ke-21 dengan nilai Eps 0,2 dan Minpts 3. Pada percobaan tersebut diperoleh nilai SI sebesar 0,3624. Percobaan yang menghasilkan Cluster paling optimal pada penelitian ini menghasilkan 3 cluster yaitu cluster 0 yang merupakan data outlier, cluster 1 yang diisi oleh negara-negara yang mulai mengalami lonjakan kasus Covid-19, dan cluster 2 yang beranggotakan negara-negara dengan kenaikan kasus yang signifikan.

Referensi

- [1] H. Qi *dkk.*, "COVID-19 Transmission in Mainland China is Associated with Temperature and Humidity: A Time-series Analysis," *Sci. Total Environ.*, vol. 728, no. 138778, hal. 1–6, 2020, doi: 10.1016/j.scitotenv.2020.138778.
- [2] F. S. D. Arianto dan N. P., "Prediksi Kasus Covid-19 di Indonesia Menggunakan Metode Backpropagation dan Fuzzy Tsukamoto," *J. Teknol. Inf.*, vol. 4, no. 1, hal. 120–127, 2020.
- [3] M. Mandal, S. Jana, S. K. Nandi, A. Khatua, S. Adak, dan T. K. Kar, "A Model Based Study on the Dynamics of COVID-19: Prediction and Control," *Chaos, Solitons & Fractals*, vol. 136, no. 1, hal. 1–12, 2020, doi: 10.1016/j.chaos.2020.109889.
- [4] M. Her, "Repurposing and Reshaping of Hospitals During the COVID-19 Outbreak in South Korea," *One Heal.*, vol. 10, no. 1, hal. 1–3, 2020, doi: 10.1016/j.onehlt.2020.100137.
- [5] A. C. Benabdellah, A. BENGHABRIT, dan I. Bouhaddou, "A survey of clustering algorithms for an industrial context," *Procedia Comput. Sci.*, vol. 148, hal. 291–302, 2019, doi: 10.1016/j.procs.2019.01.022.
- [6] I. D. Iskandar, M. W. Pertiwi, M. Kusmira, dan I. Amirulloh, "Komparasi Algoritma Clustering Data Media Online," *J. IKRA-ITH Inform.*, vol. 2, no. 4, hal. 1–8, 2018.
- [7] W. Rohalidyawati, R. Rahmawati, dan M. Mustafid, "Segmentasi Pelanggan E-Money Dengan Menggunakan Algoritma Dbscan (Density Based Spatial Clustering Applications With Noise) Di Provinsi Dki Jakarta," *J. Gaussian*, vol. 9, no. 2, hal. 162–169, 2020, doi: 10.14710/j.gauss.v9i2.27818.
- [8] W. Jing, C. Zhao, dan C. Jiang, "An improvement method of DBSCAN algorithm on cloud computing," *Procedia Comput. Sci.*, vol. 147, hal. 596–604, 2019, doi: 10.1016/j.procs.2019.01.208.
- [9] T. I. Hermanto dan M. A. Sunandar, "Analisis Data Sebaran Penyakit Menggunakan Algoritma Density

- Based Spatial Clustering of Applications With Noise,” *J. Sains Komput. dan Teknol. Inf.*, vol. 3, no. 1, hal. 104–110, 2020.
- [10] A. Yuwono, Y. Oslan, S. Kom, dan D. D. Dwijono, “Implementasi Metode Density Based Spatial Clustering of Applications With Noise Untuk Mencari Arah Penyebaran Wabah Demam Berdarah,” *J. EKSIS*, vol. 2, no. 1, hal. 11–21, 2009.
- [11] A. S. Devi, I. K. G. D. Putra, dan I. M. Sukarsa, “Implementasi Metode Clustering DBSCAN pada Proses Pengambilan Keputusan,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 6, no. 3, hal. 185, 2015, doi: 10.24843/lkjiti.2015.v06.i03.p05.
- [12] M. A. Nahdliyah, T. Widiharih, dan A. Prahutama, “Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index,” vol. 8, no. 2, hal. 161–170, 2019.
- [13] X. Wang dan 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.
- [14] Nicolaus, E. Sulistianingsih, dan H. Perdana, “Penentuan Jumlah Cluster Optimal pada Median Linkage dengan Indeks Validitas Silhouette,” *Bul. Ilm. Math. Stat. dan Ter.*, vol. 05, no. 2, hal. 97–102, 2016.
- [15] S. Chakraborty dan N. K. Nagwani, “Analysis and Study of Incremental DBSCAN Clustering Algorithm,” 2014, [Daring]. Tersedia pada: <http://arxiv.org/abs/1406.4754>.
- [16] E. Puca *dkk.*, “Short epidemiological overview of the current situation on COVID-19 pandemic in Southeast European (SEE) countries,” *J. Infect. Dev. Ctries.*, vol. 14, no. 5, hal. 433–437, 2020, doi: 10.3855/jidc.12814.