

KLASTERISASI PERSEBARAN COVID-19 DI KOTA TANGERANG SELATAN DENGAN METODE CLUSTERING K-MEANS DAN DENSITY BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE

¹Tukiyat, ²Makhsun, ³Yohanes, ⁴Siti Hotijah

^{1,2}Badan Riset dan Inovasi Nasional (BRIN)

^{1,2,3,4}Program Magister Teknik Informatika, Universitas Pamulang, Indonesia

E-mail: ¹dosen02711@unpam.ac.id

ABSTRACT

The Covid-19 pandemic that spread in Indonesia in early 2020 has had a significant impact on Indonesia, especially on the health sector. The impact of the pandemic has been responded by the government by anticipating through efforts to limit mobility and policies in all sectors and lines. South Tangerang City as a buffer area in DKI Jakarta certainly needs to be more vigilant in dealing with the new symptoms of Covid-19. The purpose of writing this paper is to analyze the pattern of the spread of the COVID-19 virus in South Tangerang City. The research was conducted by qualitative methods using secondary data obtained from the official website of <https://lawancovid19.tangerangselatankota.go.id/>. The Covid-19 data sample used is data from January 2021 to March 2022. The results of data collection were analyzed using the K-Means and Density Based Spatial Clustering of Applications With Noise (DBSCAN) methods. To test the accuracy level of the clustering model measured by the Davies Bouldin Index method. The results showed that the clustering formed there were 6 clusters. The clusters formed by the K-Mean method are the clusters with the highest number of members in cluster 0 as many as 12 or 22.22%. With the DBSCAN method, the most clusters in cluster 4 with a percentage of 29.63% with 16 members. The data validation test of the modeling results measured by the Davies Bouldin Index (DBI) method can be concluded that the K-Means method is more suitable for constructing Covid-19 clustering in South Tangerang City because it has a smaller DBI value than the DBSCAN method. Keywords : Virus, Covid-19, K-Means, Clustering, Density Based Clustering.

Keywords : Virus, Covid-19, K-Means, Clustering, Density Based Clustering.

ABSTRAK

Pandemi Covid-19 yang menyebar di Indonesia pada awal tahun 2020 telah memberikan dampak yang signifikan bagi Indonesia terutama pada sektor kesehatan. Dampak dari pandemi tersebut telah direspon oleh pemerintah dengan mengantisipasi melalui upaya kebijakan pembatasan mobilitas dan kebijakan pada semua sektor dan lini. Kota Tangerang Selatan sebagai daerah penyangga di DKI Jakarta tentunya perlu lebih waspada dalam menangani gejala baru covid-19. Tujuan penulisan paper ini berusaha melakukan analisis pada pola penyebaran virus covid-19 di Kota Tangerang Selatan. Penelitian dilakukan dengan metode kualitatif dengan menggunakan data sekunder yang diperoleh dari web resmi <https://lawancovid19.tangerangselatankota.go.id/>. Sampel data Covid-19 yang digunakan adalah data dari Januari 2021 hingga Maret 2022. Hasil pengumpulan data dianalisis dengan metode K-Means dan Density Based Spatial Clustering of Applications With Noise (DBSCAN). Untuk menguji tingkat akurasi model klasterisasi diukur dengan metode Davies Bouldin Index. Hasil penelitian menunjukkan bahwa klasterisasi yang terbentuk ada 6 kluster. Kluster yang terbentuk dengan metode K-Mean kluster dengan jumlah anggota terbanyak pada kluster 0 sebanyak 12 atau sebesar 22.22%. Dengan metode DBSCAN kluster terbanyak pada kluster 4 dengan persentase sebesar 29,63% dengan jumlah anggota sebanyak 16 orang. Uji validasi data hasil pemodelan diukur dengan metode Davies Bouldin Index (DBI) dapat disimpulkan bahwa metode K-Means lebih cocok digunakan untuk mengkonstruksikan clustering Covid-19 di Kota Tangerang Selatan karena mempunyai nilai DBI yang lebih kecil dibanding dengan metode DBSCAN.

Kata Kunci: Virus, Covid-19, Data Mining, K-Means, Clustering, Density Based Clustering

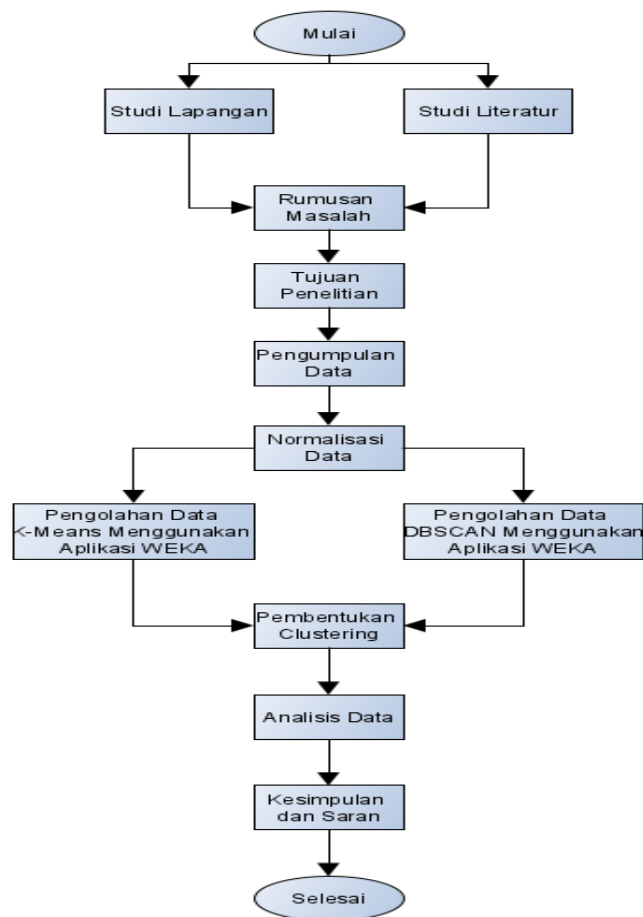
PENDAHULUAN

Tingginya trend peningkatan kasus terkonfirmasi virus covid-19 di Kota Tangerang Selatan saat ini mulai mengkhawatirkan, untuk itu diperlukan kerja keras dan kerja sama oleh semua elemen yang ada untuk meminimalisir ataupun mencegah berkembangnya virus ini secara cepat dan tepat. Kota Tangerang Selatan merupakan kota penyangga daerah Ibukota Jakarta dimana mobilitas penduduk sangat tinggi. Dari penelusuran data pada tanggal 30 Oktober 2021, diperoleh informasi penduduk yang kontak erat sebanyak 15.917 orang, kasus suspect 5.150 orang, kasus probable 11 orang dan kasus terkonfirmasi 31.113 orang. Rata-rata kasus baru pada setiap harinya sebanyak 101 kasus dengan kasus terbanyak terjadi pada tanggal 12 April 2020 (Rahmadina, 2020). Banyak masyarakat Indonesia yang tidak mematuhi himbauan dari pemerintah untuk menanggulangi pandemi virus corona ini, diakibatkan oleh salah satu konsep di dalam psikologi yang dinamakan bias kognitif. Bias kognitif adalah kesalahan sistematis dalam berpikir yang memengaruhi keputusan dan penilaian. yang dibuat seseorang. Jenis bias kognitif yang tepat untuk menjelaskan fenomena ini adalah bias optimis, bias emosional, dan efek *Dunning-Kruger* (Buana, 2017). Salah satu faktor penyebab persebaran dan perkembangan virus covid-19 di Kota Tangerang Selatan dipengaruhi oleh faktor tingginya pergerakan aktivitas penduduk yang dinamis dari dan ke Jakarta maupun kota-kota lainnya, pola komuter, urbanisasi dan migrasi penduduk. Dengan demikian, maka mobilitas penduduk tersebut akan berpengaruh pada penyebaran virus covid-19 di Tangerang Selatan.

Penelitian ini bertujuan untuk menemukan model cluster tentang pola penyebaran virus covid-19 di kota Tangerang Selatan. Metode analisis yang digunakan dengan Algoritma *K-Means Clustering* dan *Density Based Spatial Clustering Of Applications With Noise (DBSCAN)*. Metode klusterisasi dengan *K-Means* menggunakan beberapa komponen yang diperoleh sebagai pusat *cluster* awal, pusat *cluster* ditentukan dengan memilih dari dataset secara acak. *DBSCAN* akan mendeteksi *cluster* serta menentukan parameter *Epsilon* secara otomatis dengan cara yang akurat dalam menemukan parameter input dan menemukan *cluster* dengan kepadatan yang berbeda-beda. Algoritma *DBSCAN* dapat dilakukan untuk memperoleh hasil *clustering* yang lebih baik dan memberikan pengaruh langsung terhadap jumlah *cluster* yang dihasilkan (Izhari, 2020) metode *k-means* digunakan untuk mengolah data menjadi *cluster-cluster* yang terbentuk.

METODE

Penelitian menggunakan pendekatan kualitatif dengan pendekatan rancangan pembelajaran mesin untuk membangun model klusterisasi. Clustering merupakan suatu metode data mining untuk mengelompokkan data dengan karakteristik yang sama ke satu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain. Model clusterisasi ini dapat menggambarkan sebaran data virus covid-19 di Kota Tangerang Selatan. Dalam konteks data mining model clusterisasi termasuk dalam kelompok data mining *unsupervised learning* dimana dalam *unsupervised learning* merupakan model analisis data yang tidak menggunakan label sebagai dasar untuk membuat prediksi variabel melainkan menggunakan kesamaan atribut dari atribut yang dimiliki. Secara konseptual rancangan penelitian ini dapat ditunjukkan seperti pada gambar 1.



Gambar 1. Rancangan Penelitian

Penelitian diawali dengan studi literatur dan studi lapangan untuk menetapkan rumusan masalah dan tujuan penelitian yang ingin dicapai. Penelitian dengan menggunakan data sekunder yang diperoleh dari sumber *website* penanganan covid-19 pemerintah daerah Kota Tangerang Selatan dengan link <https://lawancovid19.tangerangselatankota.go.id>. Setting data atribut yang diteliti adalah nama kecamatan, nama kelurahan, kontak erat, *suspek*, kemungkinan terpapar, positif dan meninggal. Pengumpulan data dengan studi literatur. Model analisis data digunakan dengan *K-Means* dan *DBSCAN*. Untuk mengetahui tingkat akurasi dari model dilakukan divalidasi dengan menggunakan metode *Davies Bouldin Index (DBI)*.

HASIL DAN PEMBAHASAN

Pada tahapan awal data diolah melalui proses normalisasi data, proses normalisasi data merupakan proses mereduksi perhitungan komputasi yang terlalu besar maka dilakukan proses normalisasi data kedalam range yang lebih kecil yaitu 0.0 s/d 0,2. Tahap proses selanjutnya dilakukan pada proses *K-Means Clustering* yaitu dengan menentukan jumlah *K* terbaik. Untuk menentukan jumlah *K* terbaik menggunakan metode *Elbow*. Pada metode *Elbow* dalam menentukan jumlah *K* terbaik dengan melihat pada fungsi dari nilai *cluster* pada dataset yang ada yaitu dengan melihat nilai *Sum Of Square Error (SSE)*. Hasil

pengolahan data penelitian untuk penentuan jumlah *cluster* terbaik menggunakan metode *Elbow* seperti pada tabel 1 berikut ini.

Tabel 1. Hasil *Sum Of Square Error* (SSE)

Cluster	Jumlah Data	Sum Square Error (SSE)	Selisih
K4	54	78.022	0
K5	54	69.898	6.803
K6	54	63.094	7.072
K7	54	56.022	1.365
K8	54	54.657	1.320
K9	54	53.336	2.213
K10	54	51.123	2.203

Dari hasil pengolahan data seperti terlihat pada Tabel 1, bahwa nilai *SSE* terbesar *cluster* pada K6 yang mempunyai selisih terbesar yaitu sebesar 7.072. Hal ini berarti bahwa *cluster* 6 memiliki selisih yang paling besar dibanding dengan *cluster* lainnya. Dengan demikian maka dapat diindikasikan bahwa jumlah *cluster* terbaik yang digunakan sebanyak 6 *cluster*. Dengan demikian, maka jumlah *cluster* adalah 6 sebagai dasar untuk melakukan analisis clusterisasi dengan metode metode *K-Means* dan DBSCAN.

A. Analisis *Cluster K-Means*

Berdasar pada hasil analisis data dengan proses *clustering* dengan penggunaan aplikasi *WEKA* (*Waikato Environment for Knowledge Analysis*) telah ditemukan 6 *cluster* dengan jumlah anggota dari masing-masing *cluster* berikut ini:

Cluster 0 dengan jumlah anggota sebanyak 12 dengan presentase 22,22%.

Cluster 1 dengan jumlah anggota sebanyak 8 dengan presentase 14.81%.

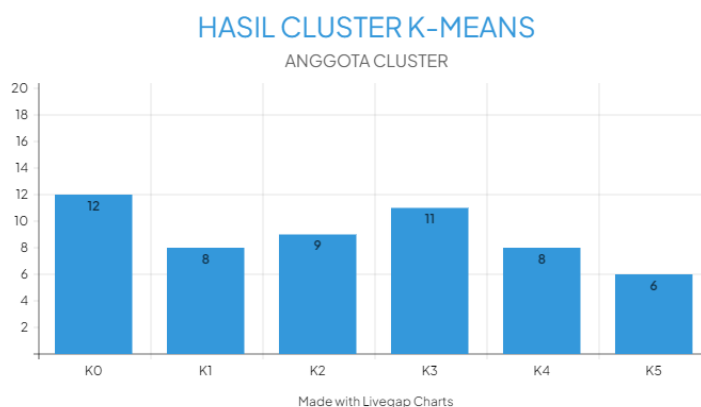
Cluster 2 dengan jumlah anggota sebanyak 9 dengan presentase 16.67 %.

Cluster 3 dengan jumlah anggota sebanyak 11 dengan presentase 20.37%.

Cluster 4 dengan jumlah anggota sebanyak 8 dengan presentase 14.81%.

Cluster 5 dengan jumlah anggota sebanyak 6 dengan presentase 11,11%.

Secara visual distribusi pengelompokan *cluster* covid-19 di Kota Tangerang Selatan dengan metode *K-Means* dapat ditunjukkan pada gambar 2.



Gambar 2. Grafik Hasil *Cluster K-Means*

B. Analisis *Density Based Spatial Clustering of Application With Noise (DBSCAN)*

Hasil pengolahan data dan analisis *clustering* dengan menggunakan metode *DBSCAN* untuk 6 cluster persebaran covid-19 di Kota Tangerang Selatan dapat diperoleh komposisi klasifikasi *cluster* sebagai berikut:

Cluster 0 dengan jumlah anggota sebanyak 6 dengan presentase 11.11%.

Cluster 1 dengan jumlah anggota sebanyak 5 dengan presentase 9.26%.

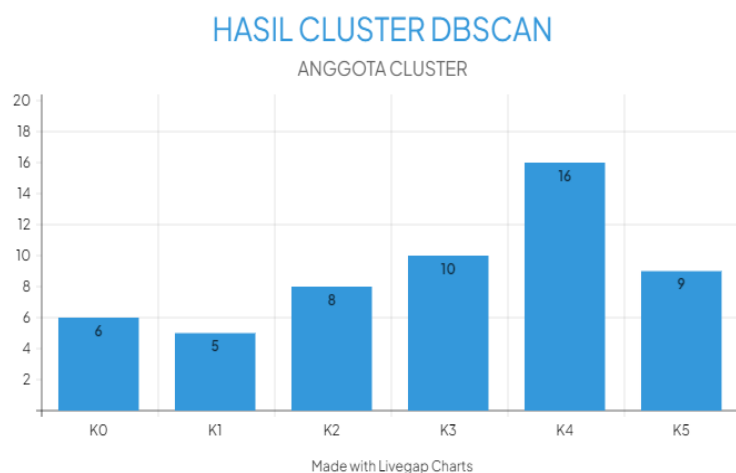
Cluster 2 dengan jumlah anggota sebanyak 8 dengan presentase 14.81%.

Cluster 3 dengan jumlah anggota sebanyak 10 dengan presentase 18.52%.

Cluster 4 dengan jumlah anggota sebanyak 16 dengan presentase 29.63%.

Cluster 5 dengan jumlah anggota sebanyak 9 dengan presentase 16.67%.

Distribusi cluster berdasar pada jumlah anggota untuk masing-masing cluster dapat ditunjukkan pada gambar 3 berikut ini



Gambar 3. Grafik Hasil Cluster DBSCAN

Berdasar pada hasil analisis *clusterisasi* dengan metode *DBSCAN* diperoleh sebaran data bahwa *cluster* terendah pada *cluster 1* dengan persentase 9,26% dengan jumlah anggota sebanyak 5. Sedang *cluster* terbesar pada kelompok *cluster 4* dengan persentase sebesar 29,63% dengan jumlah anggota sebanyak 16.

C. Pengujian Akurasi Data *Clustering*

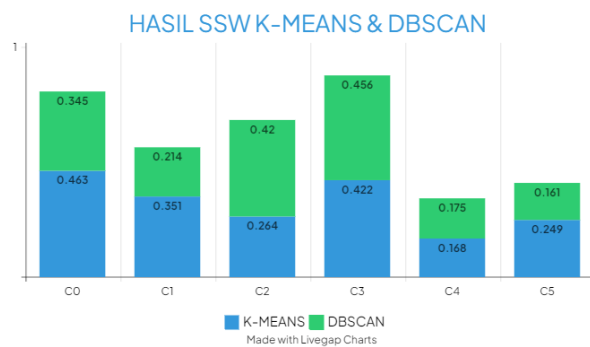
Untuk mengetahui seberapa baik kualitas dari *clustering* yang dihasilkan dari kedua metode *clustering* di atas dapat dilakukan evaluasi hasil *clustering* dengan menggunakan metode *Davies-Bouldin Index*. Untuk menguji akurasi dengan metode *Davies-Bouldin Index* perlu dilakukan langkah-langkah menghitung nilai *Sum of Square Within Cluster (SSW)* kemudian *Sum of Square Between Cluster (SSB)* dan *Ratio*.

1. *Sum of Square Within Cluster (SSW)*

Nilai dari *Sum of Square Within Cluster (SSW)* digunakan untuk mengetahui kohesi dalam sebuah *cluster ke -i*. Kohesi didefinisikan sebagai sebuah jumlah dari kedekatan data terhadap titik pusat *cluster* dari sebuah *cluster* yang diikutinya. Hasil perhitungan nilai *SSW* kedua metode tersebut terlihat pada Tabel 2 dan Gambar 4.

Tabel 2. Hasil SSW K-Means Dan DBSCAN.

Cluster	Metode Analisis	
	K-Mean	DBSCAN
C0	0.463	0.345
C1	0.351	0.214
C2	0.264	0.420
C3	0.422	0.456
C4	0.168	0.175
C5	0.249	0.161



Gambar 4. Grafik Hasil SSW K-Means dan DBSCAN

2. Sum of Square Between Cluster (SSB)

Nilai dari *Sum of Square Between Cluster (SSB)* digunakan untuk mengetahui separasi antar *cluster*. Hasil dari perhitungan nilai SSB kedua metode tersebut terlihat pada Tabel 3 dan Tabel 4.

Tabel 3. Hasil SSB K-Means

SSB	C0	C1	C2	C3	C4	C5	MAX
C0	0.000	0.167	0.335	0.387	0.424	0.386	0.424
C1	0.167	0.000	0.361	0.312	0.440	0.378	0.440
C2	0.335	0.361	0.000	0.703	0.142	0.130	0.703
C3	0.387	0.312	0.703	0.000	0.703	0.678	0.703
C4	0.424	0.440	0.142	0.703	0.000	0.214	0.703
C5	0.386	0.378	0.130	0.678	0.214	0.000	0.678

Tabel 4. Hasil SSB DBSCAN

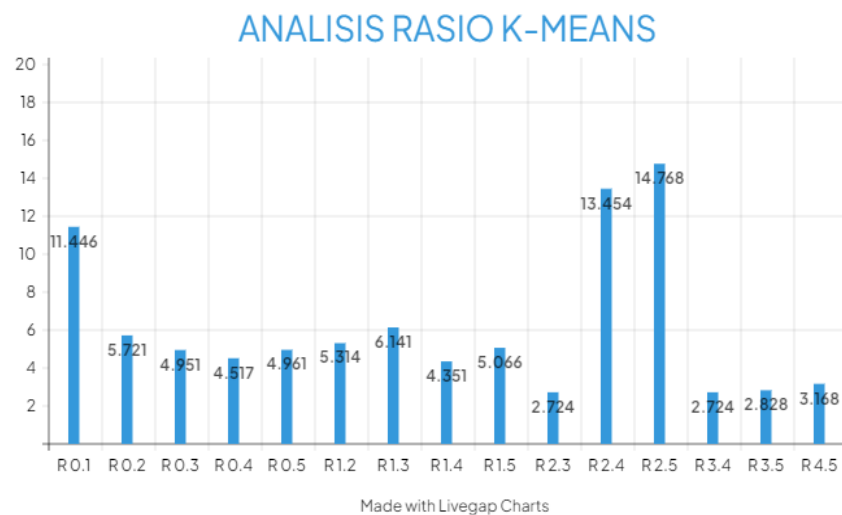
SSB	C0	C1	C2	C3	C4	C5	MAX
C0	0.000	0.259	0.382	0.301	0.462	0.496	0.496
C1	0.259	0.000	0.435	0.440	0.673	0.718	0.718
C2	0.382	0.435	0.000	0.330	0.735	0.692	0.735
C3	0.301	0.440	0.330	0.000	0.574	0.579	0.579
C4	0.462	0.673	0.735	0.574	0.000	0.172	0.735
C5	0.496	0.718	0.692	0.579	0.172	0.000	0.718

3. *Ratio*

Analisis *Ratio* (Rasio) digunakan untuk mengetahui nilai perbandingan antar *cluster* ke -*i* dan *cluster* ke -*j*. Hasil dari perhitungan nilai *Ratio* kedua metode tersebut terlihat pada Tabel 5, Tabel 6, Gambar 5 dan Gambar 6.

Tabel 5. Analisis Rasio *K-Means*

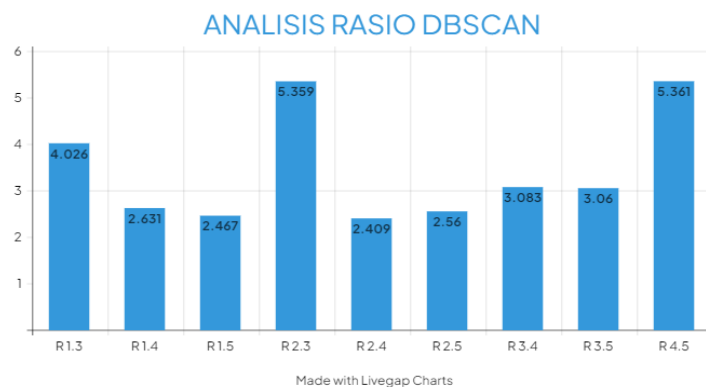
No	Ratio Metode <i>K-Mean</i>	
	Ratio	Nilai
1	R 0,1	11.446
2	R 0,2	5.721
3	R 0,3	4.951
4	R 0,4	4.517
5	R 0,5	4.961
6	R 1,2	5.314
7	R 1,3	6.141
8	R 1,4	4.351
9	R 1,5	5.066
10	R 2,3	2.724
11	R 2,4	13.454
12	R 2,5	14.768
13	R 3,4	2.724
14	R 3,5	2.828
15	R 4,5	3.168



Gambar 5. Grafik Hasil Analisa Rasio *K-Means*

Tabel 6. Analisis Rasio DBSCAN

No	Ratio Metode DBSCAN	
	Ratio	Nilai
1	R 1,3	4.026
2	R 1,4	2.631
3	R 1,5	2.467
4	R 2,3	5.359
5	R 2,4	2.409
6	R 2,5	2.560
7	R 3,4	3.083
8	R 3,5	3.060
9	R 4,5	3.361



Gambar 6. Grafik Hasil Analisa Rasio DBSCAN

Berdasar pada hasil perhitungan pada SSB, dan rasio maka dapat diperoleh nilai *DBI* tersebut yang selanjutnya akan digunakan sebagai perbandingan efektivitas *clustering* kedua metode yang digunakan yaitu *K-Means* dan *DBSCAN*. Semakin kecil nilai *Davies Bouldin Index (DBI)* yang diperoleh (*Non Negatif*), maka akan semakin baik *cluster* yang dihasilkan dari pengelompokan menggunakan kedua metode tersebut. Hasil dari perhitungan nilai *DBI* dari kedua metode tersebut terlihat pada Tabel 7.

Tabel 7 Hasil *DBI K-Means* Dan *DBSCAN*.

Nilai <i>Davies Bouldin Index</i>	
Metode	<i>DBI</i>
K-Means	0.609
DBSCAN	0.663

Berdasar pada hasil *DBI* tersebut di atas dapat ditunjukkan bahwa Algoritma *K-Means* lebih baik dibanding algoritma *DBSCAN* karena *DBI* karena nilai *DBI* pada algoritma *K-Means* lebih kecil dibanding dengan *DBI* pada *DBSCAN*. Dengan demikian, maka model clusterisasi dengan algoritma *K-Means* lebih cocok untuk mengkonstruksikan model clusterisasi pademi covid-19 di Kota Tangerang Selatan.

KESIMPULAN

Dari hasil pengolahan data menggunakan metode *K-Means* dan *DBSCAN* dalam membangun klasterisasi penyebaran *covid-19* di wilayah Kota Tangerang Selatan dapat disimpulkan beberapa hal sebagai berikut:

1. Pola penyebaran virus *covid-19* di wilayah Kota Tangerang Selatan dapat dikelompokkan menjadi 6 klaster.
2. *Clustering* dengan metode *K-Means* menghasilkan 6 *cluster* dari datasets yang dianalisis dengan komposisi: *Cluster 0* sebesar 22,22%, *Cluster 1* sebesar 14.81%, *Cluster 2* sebesar 16.67 %, *Cluster 3* sebesar 20.37%, *Cluster 4* sebesar 14.81% dan *Cluster 5* sebesar 11,11%. Hasil validasi model dengan nilai *DBI* sebesar 0,609.
3. *Clustering* dengan metode *DBSCAN* menghasilkan 6 *cluster* dari datasets yang dianalisis dengan komposisi klasifikasi: *Cluster 0* sebesar 11.11%, *Cluster 1* sebesar 9.26%, *Cluster 2* sebesar 14.81%, *Cluster 3* sebesar 18.52%, *Cluster 4* 29.63%, dan *Cluster 5* sebesar 16.67%. Validasi model diukur dengan nilai *DBI* sebesar 0.663.
4. Proses validasi hasil *cluster* menggunakan metode *Davies Bouldin Index (DBI)* menghasilkan nilai 0.609 untuk *K-Means* dan 0.663 untuk *DBSCAN*. Maka dapat disimpulkan metode yang optimal untuk klasifikasi data *covid-19* di wilayah Kota Tangerang Selatan cocok dengan metode *K-Means*.

DAFTAR PUSTAKA

- Adhitama, R., Burhanuddin, A., & Ananda, R. (2020). Penentuan Jumlah Cluster Ideal Smk Di Jawa Tengah Dengan Metode X-Means Clustering Dan K-Means Clusterin. *JIKO (Jurnal Informatika Dan Komputer)*, 3(1), 1–5. <https://doi.org/10.33387/jiko.v3i1.1635>
- Asroni, R. A. (2015). Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang. *Ilmiah Semesta Teknika*, 18(1), 76–82.
- Budiman, S., Safitri, D., & Ispriyanti, D. (2016). Perbandingan Metode K-Means Dan Metode Dbscan Pada Pengelompokan Rumah Kost Mahasiswa Di Kelurahan Tembalang Semarang. *None*, 5(4), 757–762.
- Handayani, R. T., Arradini, D., Darmayanti, A. T., Widiyanto, A., & Atmojo, J. T. (2020). Pandemi covid-19, respon imun tubuh, dan herd immunity. *Jurnal Ilmiah Stikes Kendal*, 10(3), 373–380.
- Hartanti, N. T. (2020). Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 6(2), 82–89. <https://doi.org/10.25077/teknosi.v6i2.2020.82-89>
- Izhari, F. (2020). Analisis Algoritma Dbscan Dalam Menentukan Parameter Epsilon Pada Clustering Data Numerik. *Seminar Nasional Teknologi Komputer & Sains*, 156–158.
- Novi, S. (2020). *Clustering Status Desa Menggunakan Metode DBSCAN (Density-Based Spatial Clustering of Application With Noise)*.
- Prima, R. R., & Arien, E. Y. (n.d.). Analisis Cluster Virus Corona (COVID-19) di Indonesia pada 2 Maret 2020-12 April 2020 dengan Metode K-Means

Clustering. In *researchgate.net*. Retrieved February 24, 2021, from <https://www.researchgate.net/publication/342697385>

- Rahmadina, R. P. (2020). *Analisis Cluster Virus Corona (COVID-19) di Indonesia pada 2 Maret 2020 – 12 April 2020 dengan Metode K-Means Clustering*. https://www.researchgate.net/publication/342697385_Analisis_Cluster_Virus_Corona_COVID-19_di_Indonesia_pada_2_Maret_2020_-_12_April_2020_dengan_Metode_K-Means_Clustering
- Sari, B. N., & Primajaya, A. (2019). Penerapan Clustering DbSCAN Untuk Pertanian Padi Di Kabupaten Karawang. *Jurnal Informatika Dan Komputer*, 4(1), 28–34. www.mapcoordinates.net/en.
- Yuwono, A., Oslan, Y., Kom, S., & Dwijono, D. D. (2009). Implementasi Metode Density Based Spatial Clustering of Applications With Noise Untuk Mencari Arah Penyebaran Wabah Demam Berdarah. *Jurnal EKSIS*, 2(1), 11–21.