

Analisis Kluster dengan Metode *K-Means* Pada Penyebaran Kasus Covid-19 Berdasarkan Kabupaten/Kota di Sumatera Utara

Siti Ramadhani Nasution^{1✉}, Rina Filia Sari², Rina Widyasari³

^{1,2,3} Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Indonesia

Informasi Artikel

Riwayat Artikel

Diserahkan : 13-07-2023

Direvisi : 22-07-2023

Diterima : 28-07-2023

Kata Kunci:

Analisis Kluster, Metode K-Means, RLE, Covid 19

Keywords :

Cluster Analysis, K-Means Method, Covid-19

ABSTRAK

Tujuan dasar dari pendekatan multivariat yang dikenal sebagai analisis kluster adalah untuk mengkategorikan objek atau contoh menurut fiturnya. Pendekatan K-Means merupakan salah satu teknik analisis cluster. Algoritme *K-Means* membagi data yang ada menjadi beberapa kelompok, yang masing-masing mempunyai fitur berbeda dari yang lain sambil berbagi kesamaan tertentu dengan yang lain. Kasus Covid-19 di Indonesia mengalami peningkatan terkhususnya Sumatera Utara maka digunakan *K-Means clustering* untuk mengelompokkan daerah-daerah dengan tingkat penyebaran kasus Covid-19 dengan tingkat penyebaran tinggi, sedang dan rendah. Adapun variabel yang diteliti dalam penelitian ini adalah X_1 = Jumlah Penduduk, X_2 = Kepadatan Penduduk, X_3 = positif Covid-19, X_4 = pasien yang sembuh, X_5 = pasien yang meninggal berdasarkan Kabupaten/Kota yang ada di Sumatera Utara. Hasil penelitian menyatakan semua variabel yang dipakai memiliki perbedaan yang signifikan pada ketiga kluster yang dihasilkan, sesuai dengan hasil kluster berdasarkan uji ANOVA, serta variabel yang mempunyai perbedaan paling besar pada ketiga kluster tersebut ialah variabel jumlah penduduk. Variabel ini paling menonjol di antara ketiga kluster karena memiliki nilai F terbesar (182,681) dan nilai signifikansi (0,000).

ABSTRACT

The basic goal of the multivariate approach known as cluster analysis is to categorise objects or instances according to their features. The K-Means approach is one of the cluster analysis techniques. The K-Means algorithm divides existing data into multiple groups, each of which has distinct features from the others while sharing certain commonalities with the others. The spread of Covid-19 cases in Indonesia, especially North Sumatra, K Means clustering was used to group areas with a high, medium and low distribution rate of Covid-19 cases. The variables studied in this study were X_1 = Total Population, X_2 = Population Density, X_3 = Covid-19 positive, X_4 = patients who recovered, X_5 = patients who died based on districts/cities in North Sumatra. The results of the study stated that all the variables used had significant differences in the three resulting clusters, in accordance with the results of the clusters based on the ANOVA test, and the variable that had the greatest difference in the three clusters was the population variable. This variable is the most prominent among the three clusters because it has the largest F value (182.681) and a significance value (0.000).

Corresponding Author :

Siti Ramadhani Nasution

Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara, Indonesia

Jl. Lap. Golf, Kp. Tengah, Kec. Pancur Batu, Kab. Deli Serdang

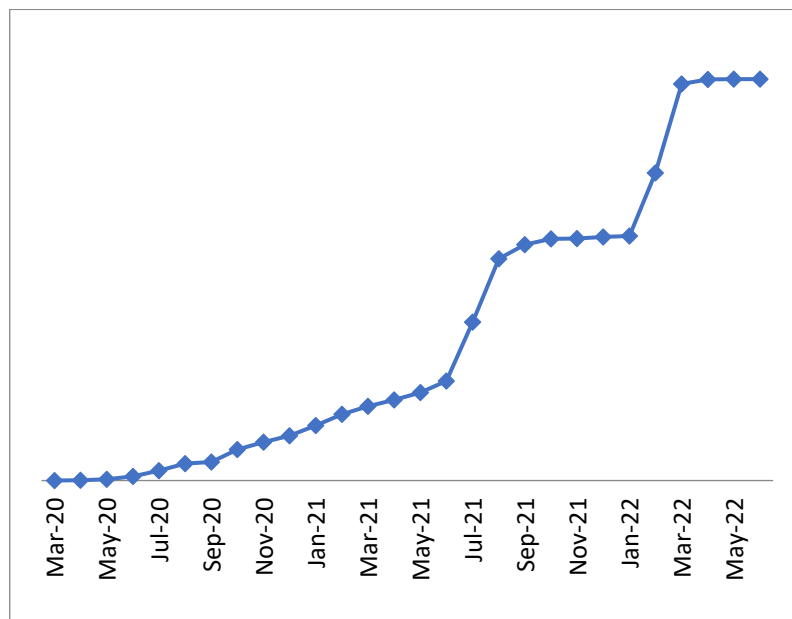
Email: sitiramdhani584@gmail.com

PENDAHULUAN

Kesehatan setiap orang sangat penting untuk kualitas hidup mereka. Hal ini penting karena akan sulit bagi seseorang untuk melakukan tugas sehari-hari jika kesehatannya buruk. Kesehatan adalah keadaan dan pertumbuhan alamiah tubuh, masyarakat, dan jiwa manusia yang dapat bergerak tanpa gangguan yang besar pada saat terjadi kemajuan di antara manusia, menurut Kementerian Kesehatan sebagaimana disinggung dalam undang-undang no. 23 Tahun 1992. Kondisi yang berkaitan dengan tubuh, jiwa, dan masyarakat, serta interaksi dengan lingkungan (Helma & et al, 2019).

Penyakit pernafasan parah coronavirus 2 (Sars-CoV-2) adalah penyebab epidemi yang dikenal sebagai coronavirus 2019 (Covid-19). Penyakit yang menyebabkan pandemi virus korona 2019–2020 ini awalnya ditemukan pada Desember 2019 di Wuhan, ibu kota Provinsi Hubei di Tiongkok. Epidemi virus corona 2019–2020 dianggap sebagai Darurat Kesehatan Masyarakat yang Menjadi Kepedulian Internasional (PHEIC) oleh Organisasi Kesehatan Dunia (WHO) pada 30 Januari 2020, dan pandemi pada 11 Maret 2020. Artinya, Covid-19 memenuhi tiga syarat: adalah virus baru yang dapat dengan cepat menginfeksi banyak orang dan sebagian besar disebarkan melalui kontak manusia (Dwitri, 2020) .

Fakta bahwa virus tersebut telah menyebar ke hampir 200 negara, termasuk Indonesia, membuat masyarakat dunia khawatir. Selain itu, pemerintah telah menerapkan sejumlah langkah untuk memutus mata rantai penularan virus ini, seperti *lockdown*, pengucilan sosial, dan vaksinasi (Solichin & Khairunnia, 2020). Indonesia telah terkena dampak Covid-19, khususnya Sumatera Utara. Covid-19 menyebar dengan cepat di wilayah Sumatera Utara. Meningkatnya kasus Covid-19 setiap bulan menjadi buktinya. Cuplikan kasus Covid -19 di Sumatera Utara dapat dilihat di bawah ini:



Gambar 1. Kasus Covid-19 Sumatera Utara

Perlunya pengelompokan lokasi yang mengarah pada titik pusat penyebaran Covid-19 ditunjukkan oleh wilayah di Sumatera Utara yang terdiri dari 33 Kabupaten/Kota. Analisis kluster merupakan salah satu pendekatan yang dapat digunakan. Dengan bantuan metode pengelompokan tidak terstruktur yang disebut analisis kluster, data dapat dibagi menjadi beberapa kelompok dan dikelompokkan menurut kesamaan sifat-sifatnya. Kemiripan antar objek adalah ukuran kesamaan. Jarak antara dua item pada akhirnya akan bertemu. Ketika dua objek berdekatan, mereka berada pada level yang sama. (Sari & Yunita, 2020)

METODE PENELITIAN

Data sekunder untuk penelitian ini, yang memakan waktu sekitar satu tahun, disediakan oleh Dinas Kesehatan Provinsi Sumatera Utara, yang dapat ditemukan di Jl. Prof.H.M. Yamin SH. No.41AA, Perintis, Kec. Medan Timur, Kota Medan, Sumatera Utara 20232. Jenis penelitian ini adalah kuantitatif. Metode penelitian induktif, obyektif, dan ilmiah seperti penelitian kuantitatif mengumpulkan data berupa angka atau nilai melalui proses perhitungan dan pengukuran. Untuk mengumpulkan data kuantitatif dari populasi dan sampel untuk penelitian ini, kerangka matematika dan teori terkait digunakan. Studi ini menggunakan populasi sebagai full item yang akan dievaluasi, meliputi jumlah populasi, kepadatan populasi, jumlah pasien positif Covid-19, pasien yang selamat, dan pasien yang meninggal dunia.

HASIL DAN PEMBAHASAN

Pemilihan cluster adalah langkah awal. Distribusi kasus Covid-19 tinggi, distribusi kasus Covid-19 sedang, dan distribusi kasus Covid-19 rendah digunakan untuk mengkategorikan klaster dalam penelitian ini. Titik pusat dipilih secara acak setelah cluster diidentifikasi. Centroid tiap cluster ditentukan secara acak, dengan centroid cluster 1 pada data ke-28 (Medan), centroid cluster 2 pada data ke-21 (Simalungun), dan centroid cluster 3 pada data ketiga (From). Setelah menetapkan centroid awal, jarak dihitung untuk setiap potongan informasi untuk masing-masing dari tiga cluster menggunakan jarak Euclidean. Cluster dengan interval terkecil adalah cluster yang memiliki data jika ada jarak terkecil antara data dan setiap cluster. Data tersebut kemudian dikelompokkan.

Data rata-rata untuk setiap cluster digunakan untuk menghitung *centroid* baru. Clustering selesai jika *centroid* baru sama persis dengan centroid lama. Ulangi proses pengelompokan cluster dan perhitungan nilai jarak untuk menentukan nilai centroid baru jika berbeda dengan *centroid* lama. Ulangi proses ini sampai *centroid* baru cocok dengan centroid lama. (Suprawoto, 2016). *Centroid* yang diperoleh pada penelitian ini adalah sebagai berikut:

Tabel 1. Centroid Pada Tiap Iterasi

<i>Centroid Awal</i>			
Klaster	K1	K2	K3
Penduduk	2435252	990246	30874
Kepadatan Penduduk	9,189.63	226.65	160.16
Positif	72961	0	2963
Sembuh	71929	0	2856
Meninggal	1019	0	107
<i>Centroid Baru Iterasi 1</i>			
Klaster	K1	K2	K3
Penduduk	2183346.5	861974.5	248402.963
Kepadatan Penduduk	5,025.62	236.2875	895.3762963
Positif	40883.5	1579	2150.740741
Sembuh	40,260.50	1492	2100.592593
Meninggal	616.5	83.75	37.55555556
<i>Centroid Baru Iterasi 2</i>			
Klaster	K1	K2	K3
Penduduk	2183346.5	861974.5	248402.963
Kepadatan Penduduk	5,025.62	236.2875	895.3762963

Positif	40883.5	1579	2150.740741
Sembuh	40,260.50	1492	2100.592593
Meninggal	616.5	83.75	37.55555556
Centroid Baru Iterasi 3			

Hasil pengelompokan terbaik dicapai pada iterasi terbaru.

Tempat-tempat di mana kasus Covid-19 lazim dikelompokkan, dan hasilnya adalah sebagai berikut:

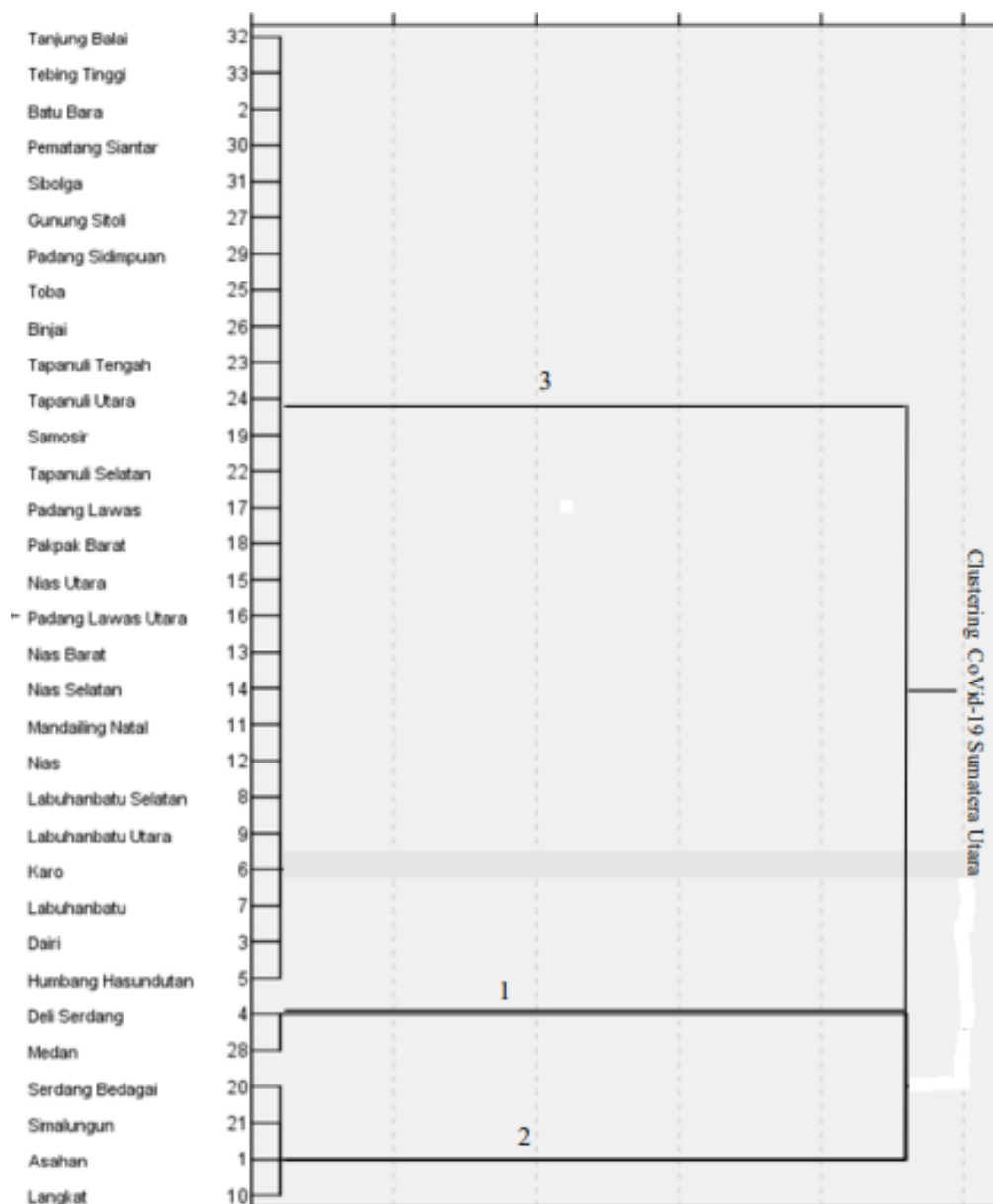
Tabel 2 Pengelompokan Penyebaran Kasus Covid-19 Menggunakan *K-Means*

Kabupaten/Kota	Penduduk	Kepadatan Penduduk	Positif	Sembuh	Meninggal	Klaster
Asahan	769960	207.97	1718	1688	30	Sedang
Batu Bara	410678	445.32	931	923	8	Rendah
Dairi	30874	160.16	2963	2856	107	Rendah
Deli Serdang	1931441	861.6	8806	8592	214	Tinggi
Humbang Hasundutan	197751	84.68	3615	3234	33	Rendah
Karo	404998	190.41	1504	1466	58	Rendah
Labuhanbatu	493899	229.08	2321	2237	84	Rendah
Labuhanbatu Selatan	314094	87.35	1438	1367	71	Rendah
Labuhanbatu Utara	381994	106.97	1773	1694	79	Rendah
Langkat	1030202	164.52	1573	1511	49	Sedang
Mandailing Natal	472886	77.09	1164	1119	45	Rendah
Nias	146672	79.6	12	8	0	Rendah
Nias Barat	89994	189.97	0	0	0	Rendah
Nias Selatan	360531	197.53	0	0	0	Rendah
Nias Utara	147274	122.44	37	33	0	Rendah
Padang Lawas Utara	260720	66.54	513	484	29	Rendah
Padang Lawas	261011	67.05	121	118	3	Rendah
Pakpak Barat	52351	42.97	557	524	33	Rendah
Samosir	136441	65.94	1849	1815	34	Rendah
Serdang Bedagai	657490	346.01	3025	2769	256	Sedang
Simalungun	990246	226.65	0	0	0	Sedang
Tapanuli Selatan	300911	49.9	1935	1890	45	Rendah
Tapanuli Tengah	365177	166.9	1179	1161	18	Rendah
Tapanuli Utara	312758	82.49	5001	4921	80	Rendah
Toba	206199	88.54	53	48	1	Rendah
Binjai	291842	4930.6	25733	25590	143	Rendah
Gunung Sitoli	136017	484.43	115	112	3	Rendah
Medan	2435252	9189.63	72961	71929	1019	Tinggi

Padang Sidempuan	225105	1963.24	519	516	3	Rendah
Pematang Siantar	268254	4819.51	2550	2517	33	Rendah
Sibolga	89584	2168.58	0	0	0	Rendah
Tanjung Balai	176027	1632.45	614	585	29	Rendah
Tebing Tinggi	172838	5575.42	1573	1498	75	Rendah

Berdasarkan tabel tersebut di atas, terdapat total 2 wilayah yang termasuk dalam kluster 1 (wilayah penyebaran kasus Covid-19 di Medan), 4 wilayah yang termasuk dalam kluster 2 (wilayah penyebaran kasus Covid-19 sedang), dan 27 wilayah masuk dalam kluster 3 (wilayah dengan kasus Covid-19 rendah) setelah dilakukan pengelompokan menggunakan *K-Means Clustering* dengan jumlah penduduk 14490597 dan kepadatan penduduk 35011,38 jiwa/km².

Dalam bentuk dendrogram dapat dilihat sebagai berikut:



Gambar 2. Dendrogram *Clustering Covid-19 Sumatera Utara*

Temuan cluster tes ANOVA mengungkapkan bahwa semua faktor yang digunakan memiliki perbedaan yang signifikan. Variabel populasi memiliki perbedaan terbesar diantara ketiga cluster yang dibuat. Variabel yang memiliki nilai F terbesar dari variabel lainnya yaitu 182,681 dan nilai signifikan 0,000 merupakan variabel yang paling membedakan individu-individu dalam ketiga cluster tersebut. Karena *klaster* yang dihasilkan pada populasi yang bervariasi dan tingkat kasus yang paling berbeda dalam penularan Covid-19, Kota Medan, provinsi terpadat di Sumatera Utara, termasuk dalam *klaster* 1 dengan tingkat kasus penularan yang tinggi.

KESIMPULAN DAN SARAN

Kesimpulan

Temuan dari penelitian ini adalah mendapatkan kelompok dari setiap lokasi persebaran kasus Covid-19 di Sumatera Utara yaitu: populasi, kepadatan penduduk, positif, sembuh, dan meninggal pada tanggal 24 Juni 2022. Hal ini berdasarkan pembahasan pada bab sebelumnya. Tiga kategori penularan kasus Covid-19 yang tinggi, distribusi kasus Covid-19 yang sedang, dan distribusi kasus Covid-19 yang rendah digunakan untuk mengkategorikan klaster. Ditemukan 2 area di cluster 1, 4 area di cluster 2, dan 27 area di cluster 3, berdasarkan pengelompokan menggunakan K-Means Clustering dengan jumlah penduduk 14490597 dan kepadatan penduduk 35011,38 jiwa/km². Semua variabel yang dipakai memiliki perbedaan yang signifikan pada ketiga klaster yang dihasilkan, sesuai dengan hasil klaster berdasarkan uji ANOVA, serta variabel yang mempunyai perbedaan paling besar pada ketiga klaster tersebut ialah variabel jumlah penduduk. Variabel ini paling menonjol di antara ketiga klaster karena memiliki nilai F terbesar (182,681) dan nilai signifikansi (0,000).

Saran

Dari kesimpulan di atas, hal yang bisa diharapkan kedepannya adalah pengolahan data dengan metode yang berbeda agar hasilnya bisa dibandingkan. Penelitian ini dapat dijadikan dokumentasi pemerintah apabila terjadi pandemi yang serupa dimasa yang akan datang.

REFERENSI

- Bastian,A., Sujadi, H., Febrianto, G. 2018.Penerapan Algoritma K-Meansi Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). *Jurnal Sistem Informasi(Journal Of Information System)*, Volume14
- Dwitri, N., *et al.* 2020. Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia *Jurnal Teknologi Informasi*.Vol. 4, No. 1
- Ediyanto,*et al.*2013. Pengklasifikasian Karakteristik Dengan Metode *K-MeansCluster Analysis*. Buletin Ilmiah Mat.Stat. Dan Terapannya(Bimaster). Volume02 , No. 2
- Al Fatta, Hanif. (2007). Analisa dan Perancangan Sistem Informasi. Yogyakarta, Andi.
- Helma, S.S., *et al.* 2019. *Clustering Pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma K-Means.Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI. ISSN (Online) : 2579-5406*
- Metisen, B. M., Sari, H. L. 2015.Analisis Clustering Menggunakan Metode K-Means Dalam Pengelompokan Penjualan Produk Pada Swalayan Fadhila.*Jurnal Media Infotama*.Vol.11 No. 2
- Nasari, F Dan Darma, S. 2015. Penerapan *K-Means Clustering* Pada Data Penerimaan Mahasiswa Baru(Studi Kasus : Universitas Potensi Utama). Seminar Nasional Teknologi Informasi Dan Multimedia *ISSN : 2302-3805*
- Narimawati, U. 2008. *Metodologi Penelitian Kualitatif dan Kuantitatif, Teori dan Aplikasi*. Bandung : Agung Media

- Ningrat, D. R., Maruddani, D. A. I., Wuryandari. T. 2016. Analisis *Cluster* Dengan Algoritma K-Means Dan Fuzzy C-Means Clustering Untuk Pengelompokan Data Obligasi Korporasi *Jurnal Gaussian*, Volume5, Nomor 4
- Priyatman, H., Sajid, F., Haldivany, D. 2019. Menggunakan Algoritma *K-Means Clustering* Untuk Memprediksi Waktu Kelulusan Mahasiswa *Jepin (Jurnal Edukasi Dan Penelitian Informatika)*Issn(E): 2548-9364