

Analisis *Cluster* COVID-19 di Sumatera Barat dengan Metode *Non-Hirarki (K-Means)*

Afridho Afnanda^{#1}, Arnellis^{*2}

[#]*Student of Mathematics Departement Universitas Negeri Padang, Indonesia*

^{*}*Lecturers of Mathematics Departement Universitas Negeri Padang, Indonesia*

¹afridhoafnanda7@gmail.com

²Arnellis.mathunp25@gmail.com

Abstract – COVID-19 has spread and has become a pandemic in almost all of the world, including in Indonesia. In Indonesia, this virus was first announced on March 2, 2020, until March 14 to 27, 2020, a total of 77,261 cases were identified. To stop the spread of COVID-19, the Indonesian government has made several policies that are expected to suppress the spread of this virus. Cluster analysis is a statistical technique used to classify objects or cases into relatively homogeneous (same) groups. The type of data used is secondary data on COVID-19 in districts/cities of West Sumatra Province obtained from the West Sumatra Provincial Health Office and the COVID-19 task force website, where the data obtained is cumulative data from March 2020 to July 14, 2021. Optimization of the number of K clusters using the elbow method to produce K=3. The results of the K-means cluster analysis with the number of K=3 concluded that the distance of members in the cluster is low, and the distance between clusters is large.

Keywords - COVID-19, Non-Hierarchical Cluster Analysis, K-means, Elbow Method

Abstrak – COVID-19 telah menyebar dan menjadi pandemi di hampir seluruh dunia termasuk di Indonesia. Di Indonesia sendiri virus ini pertama kali diumumkan pada tanggal 2 Maret 2020, sampai pada tanggal 14 sampai 27 Maret 2020 total kasus yang teridentifikasi sebanyak 77.261 kasus. Dalam upaya untuk menghentikan penyebaran COVID-19 pemerintah Indonesia membuat beberapa kebijakan yang diharapkan dapat menekan penyebaran virus ini. Analisis kluster yaitu suatu teknik statistika yang digunakan untuk mengklasifikasikan objek atau kasus kedalam kelompok yang relatif homogen (sama). Jenis data yang digunakan adalah data sekunder COVID-19 di kabupaten/kota Provinsi Sumatera Barat yang diperoleh dari Dinas Kesehatan Provinsi Sumatera Barat dan *website* gugus tugas COVID-19, dimana data yang diperoleh adalah data kumulatif dari bulan Maret 2020 sampai 14 Juli 2021. Pengoptimalan jumlah K cluster memakai metode elbow dengan menghasilkan K=3, hasil analisis cluster K-means dengan jumlah K=3 didapatkan kesimpulan bahwa jarak anggota dalam cluster yang rendah dan jarak antar cluster yang besar.

Kata Kunci – COVID-19, Analisis Cluster Non Hirarki, K-means, Metode Elbow

PENDAHULUAN

COVID-19 telah menyebar dan menjadi pandemi di hampir seluruh dunia termasuk di Indonesia. Di Indonesia pertama kali teridentifikasi pada tanggal 2 Maret 2020, pada rentang tanggal 14 - 27 Maret 2020 total kasus yang teridentifikasi sebanyak 77.261 kasus. Dalam upaya untuk menghentikan penyebaran COVID-19 pemerintah Indonesia membuat beberapa kebijakan yang diharapkan dapat menekan penyebaran virus ini. Ada beberapa kebijakan yang disahkan oleh pemerintah diantaranya adalah pembatasan sosial bersekala besar atau disingkat PSBB.

Kebijakan PSBB (Pembatasan sosial bersekala besar) merupakan salah satu upaya pemerintah untuk menghadang laju penyebaran COVID-19 di Indonesia. Pelaksanaan PSBB telah diterapkan di Indonesia sejak

tanggal 2 Maret 2020. Kebijakan pembatasan sosial berskala besar ini diterapkan hampir keseluruhan daerah di Indonesia termasuk diprovinsi Sumatera Barat.

Pertanggal 20 April 2020 Kasus COVID-19 di Sumatera Barat teridentifikasi sebanyak 74 kasus dan menempatkan Sumatera Barat menjadi peringkat 15 besar jumlah kasus penyebaran COVID-19 di Indonesia. Kasus ini menyebar di 19 kabupaten/kota di Sumatera Barat, dan menunjukkan kurva peningkatan kasus positif setiap harinya. Semenjak April untuk pertama kalinya salah satu daerah di Sumatera Barat yaitu kabupaten 50 kota berubah status menjadi zona merah, sekitar 9 daerah yang berada di zona *orange* dan 9 daerah lainnya berada di zona kuning. Melihat hal tersebut perlu adanya kebijakan yang tepat untuk menghambat laju penyebaran COVID-19.

Penerapan PSBB dimasing-masing daerah yang ada di Indonesia tentu berbeda-beda tergantung kebijakan pada daerah masing-masing dan sesuai syarat yang harus dipenuhi oleh suatu daerah seperti: jumlah kasus dan/atau jumlah kematian akibat penyakit meningkat dan menyebar secara signifikan dan cepat ke beberapa wilayah, dan terdapat kaitan epidemiologis dengan kebijakan serupa di wilayah atau negara lain.

Dengan adanya perbedaan kebijakan disetiap daerah dalam menerapkan kebijakan PSBB, maka akan ada perbedaan penanganan yang dilakukan dalam upaya penerapan kebijakan. Hal ini akan mempengaruhi cepat dan lambatnya penyebaran terjadi dari daerah tingkat penyebaran tinggi ke daerah yang tingkat penyebaran rendah. Kebijakan yang berlaku dilakukan sesuai dengan keadaan epidemiologis daerah masing-masing maka kebijakan di daerah dengan penyebaran yang relatif tinggi akan berbeda dengan kebijakan dan penerapannya pada daerah yang penyebaran yang relatif rendah, sehingga perlu dikelompokkan daerah mana saja tingkat penyebaran COVID-19 yang tinggi dan yang relatif rendah, untuk itu perlu dilakukan pengelompokan.

Penelitian ini merupakan penelitian pengembangan yang relevan dengan penelitian terdahulu yaitu penelitian dari Rizkiana Prima R, Yashintia Arien E, dan Sutikno yang berjudul Analisis Cluster virus Corona (COVID-19) di Indonesia pada Maret 2020-14 Juli 2020 dengan Metode *K-Means Clustering* metode ini juga digunakan dalam pengelompokan COVID-19 di Sumatera Barat dimana data yang digunakan, cakupan dan pemilihan jarak *clustering* yang berbeda.

Untuk mengenali keadaan daerah yang terdampak COVID-19 disetiap kabupaten/kota yang ada di Sumatera Barat, maka dilakukan pengklasifikasian berdasarkan kondisi tingkat penyebaran COVID-19 disetiap kabupaten/kota yang ada di Sumatera Barat, pengelompokan pada penelitian ini berdasarkan persentase penyebaran COVID-19 dapat dilakukan dengan menggunakan analisis *clustering*.

Dengan melihat penyebaran COVID-19 di Sumatera Barat akan diperoleh data yang cenderung lebih banyak sehingga peneliti menggunakan metode *K-Means* yang bisa mengolah data yang lebih banyak dan juga mempercepat mendapatkan hasil analisisnya, juga peneliti dapat menentukan jumlah cluster yang diinginkan.

METODE

Penelitian ini merupakan penelitian terapan. Penelitian ini menggunakan jenis data sekunder mengenai COVID-19 di kabupaten/kota Provinsi Sumatera Barat yang diperoleh dari Dinas Kesehatan Provinsi Sumatera Barat dan *website* gugus tugas COVID-19, dimana data yang diperoleh adalah data kumulatif dari bulan Maret 2020 sampai 14 Juli 2021. Objek pengamatan pada penelitian ini yaitu 19 kabupaten/kota di Sumatera Barat dengan variabel yang digunakan adalah seperti Tabel I.

TABEL I
VARIABEL PENELITIAN

Variabel	Keterangan
X_1	total kasus positif
X_2	kasus meninggal dunia
X_3	Persentase kasus meninggal dunia
X_4	kasus sembuh
X_5	persentase kasus sembuh
X_6	Kasus aktif
X_7	Persentase kasus aktif

Penelitian ini menggunakan teknik analisis data dengan metode analisis *cluster non-hierarki (K-means)* dengan ukuran jarak *euclidean*. Proses analisis akan dilaksanakan dengan menggunakan bantuan *software MiniTab 16*. Output dari *software MiniTab 16* kemudian dianalisis.

Adapun tahap-tahap teknik analisis data yang digunakan:

1. Mengumpulkan data sekunder yang diperoleh dari Dinas Kesehatan Provinsi Sumatera Barat dan juga *website* resmi gugus tugas COVID-19 di Sumatera Barat.
2. Melakukan proses standarisasi dengan menggunakan analisis komponen utama untuk melihat hubungan korelasi antara variable.
3. Menentukan K optimal dengan menggunakan SSE (*Sum Square of Error*) atau disebut juga dengan metode Elbow. Persamaan yang digunakan sebagai berikut:

$$SSE = \sum_{k=1}^K \sum_{X_i \in S_k} \|X_i - C_k\|_2^2 \quad (1)$$

4. Melakukan analisis *cluster* dengan menggunakan metode *non-hierarki (K-Means)*, dengan Menentukan ukuran jarak dengan menggunakan metode *Euclidean*. Urutan langkah melakukan analisis *cluster non-hierarki K-Means* sebagai berikut:

- a. Tentukan besar K
- b. Tentukan titik pusat (*centroid*) secara acak pada awal tahap
- c. Hitung jarak terdekat pada setiap data dengan centroid. Untuk menghitung jarak *Centroid* digunakan jarak *Euclidean*.

$$d(i, j) = [\sum_{k=1}^p (\bar{x}_j - \bar{x}_i)^2]^{1/2} \quad (2)$$

- d. Hitung kembali pusat *Cluster* dengan keanggotaan *cluster* yang baru terbentuk. Jika M menyatakan jumlah data dalam satu cluster, i menyatakan data ke- i dalam sebuah *cluster* dan p menyatakan dimensi data, maka didapat persamaan sebagai berikut

$$C^1 = \frac{1}{M} \sum_{j=1}^M X_j \quad (3)$$

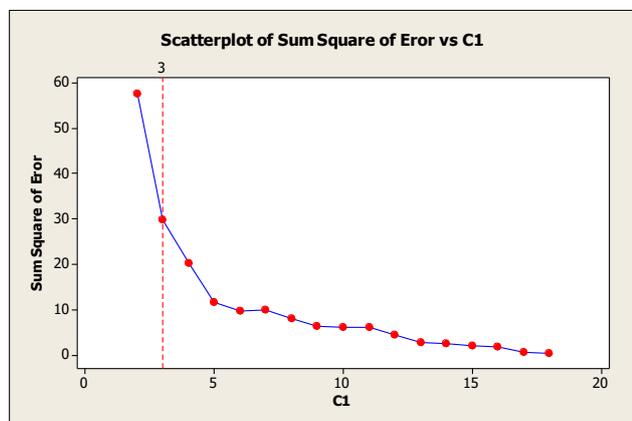
B. Analisis data

1). Optimasi K Cluster dengan metode Elbow

Sebelum menggunakan analisis cluster K-Means, ditentukan terlebih dahulu jumlah K optimal yang akan dibentuk, penentuan jumlah K dilakukan dengan menggunakan metode elbow untuk melihat nilai *sum of square error* (SSE). Hasil perhitungan SEE untuk setiap K didapatkan nilai sebagai berikut:

TABEL II
NILAI SSE SETIAP K

K	SEE	Jarak
2	57,496	
3	29,778	27,718
4	20,309	9,469
5	11,546	8,763
6	9,789	1,757
7	9,936	-0,147
8	7,983	1,953
9	6,361	1,622
10	6,142	0,219
11	6,027	0,115
12	4,486	1,541
13	2,656	1,83
14	2,645	0,011
15	1,97	0,675
16	1,705	0,265
17	0,599	1,106
18	0,452	0,147



Gambar 5. Plot Metode Elbow

Dari plot yang terbentuk dapat dilihat dengan signifikan penurunan nilai SEE terjadi pada saat K=3 dengan membentuk sudut terkecil atau elbow yang jelas. Penentuan nilai K dilakukan juga dengan menghitung jarak atau tingkat penurunan nilai SSE yang paling besar, pada tabel 2 didapatkan bahwa penurunan paling besar terjadi pada saat nilai K = 3 dengan jarak penurunan sebesar 27.718. Maka didapat nilai K optimum dengan jumlah K sebesar 3 untuk penentuan jumlah cluster yang akan dibentuk.

2). Analisis Cluster K-means

Tahap selanjutnya adalah melakukan analisis cluster dengan menggunakan jumlah cluster (k) sebanyak 3 kelompok. Jarak yang digunakan pada penelitian ini adalah jarak *euclidean* dan menghasilkan tabel cluster sebagai berikut

TABEL III
JUMLAH PENGAMATAN CLUSTER

cluster ke-i	Jumlah anggota	Jarak rerata dari centroid	Jarak maksimum dari centroid
Cluster1	1	0	0
Cluster2	12	1,128	2,378
Cluster3	6	1,223	2,098

Dari Tabel III pada cluster 1 memuat satu anggota pada observasi, pada cluster 2 memuat 12 anggota observasi dimana rata-rata jarak pusat pada cluster 2 adalah 1,128 dan jarak maksimum dari pusat adalah 2,378, sedangkan pada cluster 3 terdapat 6 anggota observasi yang memiliki rata-rata jarak pusat sebesar 1,223 dan jarak maksimum dari pusat centroid nya adalah 2,098. Dari data ini dapat dikelompokkan cluster untuk setiap kabupaten/kota yang ada di Sumatera Barat seperti pada Tabel VI.

TABEL VI
PENGELompokAN KABUPATEN/KOTA DI SUMATERA BARAT BERDASARKAN PENYEBARAN COVID-19

No	Kab/Kota	Cluster
1	Padang	1
2	Padang Panjang	2
3	Bukittinggi	3
4	Payakumbuh	2
5	Kota Solok	2
6	Sawahlunto	2
7	Pasaman	2
8	Padang Pariaman	3
9	Agam	3
10	Lima Puluh Kota	2
11	Kabupaten Solok	2
12	Tanah Datar	3
13	Sijunjung	2
14	Pesisir Selatan	2
15	Mentawai	2
16	Pariaman	3
17	Pasaman Barat	3
18	Dharmasraya	2
19	Solok Selatan	2

Hasil analisis cluster diperoleh cluster 1 dimana diprediksi sebagai daerah yang tingkat penyebaran COVID-19 dengan penyebaran yang tinggi dengan karakteristik cluster 1 merupakan cluster yang diprediksi memiliki tingkat penyebaran COVID-19 yang tertinggi dimana dapat dilihat dari kasus positif kota padang merupakan kasus positif terbesar dengan selisih yang

sangat jauh dari daerah lainnya. Pada kasus aktif sampai tanggal 14 Juli 2021 hasil data akumulasi menunjukkan jumlah rata-rata kasus aktif terbesar dibandingkan dengan cluster 2 dan 3. Dilihat dari kasus meninggal kota padang juga menempati daerah yang kasus meninggal terdampak COVID-19 yang tertinggi di Sumatera Barat. Hasil dari cluster 2 menunjukkan tingkat penyebaran COVID-19 dengan taraf sedang, dimana persentase kasus sembuh memiliki rata-rata yang sedang dan merata disetiap anggota clusternya. Pada persentase kasus meninggal cluster 2 menunjukkan angka rata-rata yang lebih tinggi dibandingkan dengan cluster 3. Hasil karakteristik dari cluster 3 menunjukkan jumlah persentase kasus meninggal tertinggi dari cluster 1 dan 2, pada persentase kasus aktif juga menunjukkan tingkat yang tertinggi, sedangkan pada kasus persentase sembuh menunjukkan angka rata-rata terkecil dari cluster yang lainnya.

SIMPULAN

Hasil dari analisis *Cluster* yang dilakukan dapat disimpulkan bahwa yang termasuk pada Cluster 1 adalah Kota Padang dengan karakteristik jumlah terkonfirmasi kasus positif terbesar dibandingkan daerah-daerah yang lainnya, juga menyumbang angka kematian terbesar di Sumatra Barat. Selanjutnya, untuk Cluster 2 adalah Kota Padang Panjang, Kota Payakumbuh, Kota Solok, Kota Sawahlunto, Kabupaten Pasaman, Kabupaten Lima Puluh Kota, Kabupaten Solok, Kabupaten Sijunjung, Kabupaten Pesisir Selatan, Kabupaten Mentawai, Kabupaten Dharmasraya, Kabupaten Solok Selatan karakteristik pada cluster 2 menunjukkan angka yang relatif menengah, dimana kasus-kasus seperti total positif, meninggal dunia, kasus sembuh, kasus aktif dan persentase kasus aktif menunjukkan angka terendah di Sumatera Barat. Terakhir

untuk Cluster 3 yaitu Kota Bukittinggi, Kabupaten Padang Pariaman, Kabupaten Agam, Kabupaten Tanah Datar, Kota Pariaman, Kabupaten Pasaman Barat, dengan karakteristik persentase kasus kesembuhan yang masih rendah jika dibandingkan dengan daerah-daerah lain yang ada di Sumatera Barat.

REFERENSI

- [1] Ariawan, P, W., Ardana, I, M., & Sugiarta, I, M, (2017), *Paket Aplikasi Statistika*, Depok: PT Raja Grafindo Persada,
- [2] Charles, E, M, (2015), *Penggunaan Analisis Two Step Clustering Untuk Data Campuran*, Jdc, 10-11,
- [3] Nalim, Y., & Salafudin, (2012), *Statistika Deskriptif*, Pekalongan: STAIN Pekalongan Press,
- [4] R, R, P., E, Y, A., & Sutikno, (2020), *Analisis Cluster Virus Corona (COVID-19) di Indonesia pada 2 Maret 2020-12 April 2020 dengan Metode K-Means Clustering*, Departemen Statistika, Fakultas Sains dan Analisis Daata, ITS,
- [5] Ristyawati, A, (2020), *Efektifitas Kebijakan Pembatasan Sosial Berskala Besar Dalam Masa Pandemi Corona Virus 2019 oleh Pemerintah Sesuai Amanat UUD NRI Tahun 1945*, *Administratif Law & Governance Journal*, 240-249,
- [6] Santoso, (2010), *Statistik Multivariat*, Jakarta: Elex Media Komputindo,
- [7] Setiati, S, (2020), *Epidemiologi Klinik dan Oat Berbasis Bukti*, Fakultas Kedokteran Universitas Indonesia, 85-89,
- [8] Singgih, S, (2002), *Buku Latihan SPSS Statistik Multivariate*, Jakarta: PT Elex Komputindo,
- [9] Supranto, (2010), *Analisis Multivariat arti dan interpretasi*, Jakarta: Rineka Cipta,
- [10] Suryanto,1998, *Metode Statistika Multivariat*, Jakarta: Proyek Pengembangan Lembaga Pendidikan Tenaga Kependidikan,
- [11] Talakua, dkk, 2017, *Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014*, Ambon, Barekeng
- [12] Velavan, T, P., & Meyer, C, G, (2020), *The COVID-19 Epidemic*, *Institute of Tropical Medicine*, 278-280,