

Algoritma Clustering Untuk Membentuk Cluster Zona Penyebaran Covid-19

Efori Bu'ulolo¹, Bister Purba²

¹Program Studi Teknik Informatika Fakultas Ilmu Komputer & Teknologi Informasi

²Program Studi Sistem Informasi Fakultas Ilmu Komputer & Teknologi Informasi
Universitas Budi Darma

^{1,2}Jl. Sisingamangaraja XII No.338, Medan, Sumatera Utara, telp. 061 7875998

e-mail: ¹buuloloeftori21@gmail.com , ²bisterpurba36@gmail.com

Abstrak

Covid-19 yaitu suatu penyakit yang menyerang sistem pernapasan manusia dan dapat menular dengan mudah. Sumatera Utara salah satu daerah yang dilanda pandemi Covid-19. Melalui Gugus Tugas Percepatan dan Penanganan Covid-19 provinsi Sumatera Utara telah melakukan berbagai upaya untuk pencegahan penyebaran Covid-19 seperti belajar dan ibadah dirumah, himbauan pakai masker dan lain sebagainya. Untuk mempermudah identifikasi penyebaran Covid-19 Tim Gugus membagi zona penyebaran Covid-19 berdasarkan jumlah kasus positif. pembagian zona dengan menggunakan satu variabel yaitu positif menyebabkan penanganan Covid-19 tidak maksimal karena hanya terkonsentrasi pada zona dengan kasus positif yang terbanyak sedangkan potensi penyebaran bukan hanya dari kasus positif. Oleh karene itu, dibutuhkan teknik yang lain dapat mengelompokkan / cluster zona penyebaran Covid-19. Salah satu teknik yang sesuai untuk pengelompokkan / cluster yaitu algoritma clustering K-Medoids. Hasil dari implementasi algoritma Algoritma K-Medoids yaitu cluster zona penyebaran Covid-19 di Sumatera Utara dibagi dalam 3(tiga) Cluster yaitu cluster 1, cluster 2 dan cluster 3. Cluster 1 identik dengan zona merah, Cluster 2 identik dengan zona kuning dan cluster 3 identik dengan zona hijau.

Kata Kunci: Covid-19, Cluster, Zona, K-Medoids

Abstract

Covid-19 is a disease that attacks the human respiratory system and can be transmitted easily. North Sumatra is one of the areas hit by the Covid-19 pandemic. Through the Task Force for the Acceleration and Handling of Covid-19, the province of North Sumatra has made various efforts to prevent the spread of Covid-19, such as studying and worship at home, appealing to wear masks and so on. To make it easier to identify the spread of Covid-19, the Cluster Team divides the Covid-19 spread zones based on the number of positive cases. zoning by using one variable, namely positive, causes the handling of Covid-19 to be not optimal because it is only concentrated in the zone with the most positive cases, while the potential for spread is not only from positive cases. Therefore, another technique is needed to group / cluster the Covid-19 spread zones. One technique that is suitable for grouping / clustering is the K-Medoids clustering algorithm. The results of the implementation of the K-Medoids Algorithm algorithm, namely the Covid-19 spread zone cluster in North Sumatra is divided into 3 (three) clusters, namely cluster 1, cluster 2 and cluster 3. Cluster 1 is identical to the red zone, Cluster 2 is identical to the yellow zone and cluster 3 is identical to the green zone.

Keywords: Covid-19, Cluster, Zona, K-Medoids.

1. Pendahuluan

Covid-19 (*Corona Virus Disease*) yaitu suatu penyakit yang menyerang sistem pernapasan manusia yang diakibatkan oleh jenis corona virus yang baru ditemukan dan termasuk penyakit menular[1][2]. Provinsi Sumatera Utara termasuk daerah yang dilanda

pandemi Covid-19. Sistem zona penyebaran Covid-19 yang dibuat Gugus Tugas Percepatan Penanganan Covid-19 dan Gugus daerah hanya menggunakan satu variabel yaitu jumlah positif Covid-19 dan tidak berdasarkan variabel lain seperti PDP (pasien dalam pemantauan) dan ODP (orang dalam pemantauan), sehingga muncul masalah yaitu pencegahan Covid-19 kurang optimal karena penanganan dan pencengahan hanya terkonsentrasi pada zona merah, sedangkan potensi peningkatan kasus bukan hanya dari jumlah yang positif tetapi juga dari yang PDP, ODP bahkan yang meninggal dunia karena Covid-19. Selain itu, karena konsentrasi penanganan dan pencegahan hanya pada zona merah maka potensi penambahan jumlah kasus sulit diprediksikan, disebabkan oleh pola penyebaran Covid-19 yang sangat mudah, juga disebabkan karena sebagian masyarakat kurang peduli dan kurang taat terhadap protokol kesehatan yang telah diterbitkan oleh pemerintah Provinsi Sumatera Utara.



Gambar 1: Pembagian Zona Penyebaran Covid-19 Berdasarkan Kasus Positif

Sejak diumumkan kasus pertama Covid-19 di Provinsi Sumatera Utara, maka data-data yang berkaitan dengan jumlah kasus Covid-19 sudah mulai dikumpulkan (rekap) seperti jumlah kasus positif, PDP, ODP dan meninggal dunia termasuk penyebarannya di seluruh daerah di Sumatera Utara. Data tersebut dapat dimanfaatkan untuk membentuk cluster zona dan memprediksikan jumlah kasus Covid-19 di Sumatera Utara. Salah satu fungsi data mining adalah fungsi clustering / pengelompokan atau algoritma clustering[3][4]. Data mining atau KDD (*knowlegde discovery in database*) merupakan aktivitas yang berkaitan dengan pengumpulan dan pemakaian data historis untuk penggalian / pencarian / penemuan informasi / pengetahuan / pola hubungan / kelompok data yang berukuran besar (*big data*)[5][6][7]. Salah satu algoritma yang termasuk dalam clustering adalah K-Medoids. K- Medoids merupakan teknik pengelompokan data ke dalam beberapa *cluster* / kelompok berdasarkan jarak terpendek, kemiripan, atau kesamaan karakteristik. Data dalam satu *cluster* memiliki jarak minimum dan antar *cluster* memiliki jarak maksimum[8][9][10]. Data jumlah kasus Covid-19 di Sumatera Utara dapat dimanfaatkan untuk membentuk *cluster* zona terbaru berdasarkan jumlah kasus positif Covid-19, PDP, dan ODP dengan menggunakan algoritma clustering K-Medoids.

Penelitian yang dilakukan oleh Ahmad Husain Ardiansyah dkk dengan judul Penerapan data mining menggunakan clustering untuk menentukan status provinsi di Indonesia 2020 menyimpulkan dengan menggunakan kriteria positif, sembuh dan meninggal dunia status provinsi di Indonesia dibagi 3 cluster yaitu zona merah yaitu DKI Jakarta dan Jawa Timur, zona kuning yaitu Banten, Jawa Barat, Jawa Tengah, Kalimantan Selatan, Sumatera Selatan dan Sulawesi Selatan sedangkan zona hijau yaitu Aceh dan propinsi lainnya[11]. Penelitian yang dilakukan oleh Dyang Falila Pramesti, M. Tanzil Furqon, dan Candra Dewi tahun 2017 dengan topik K-Medoids clustering pengelompokan data potensi kebakaran hutan / lahan. Hasil penelitian menyimpulkan cluster 1 menghasilkan nilai rata-rata brightness temperature lebih dari 3300K dan nilai rata-rata confidence lebih dari 80% sehingga memiliki potensi tinggi kebakaran hutan / lahan. Cluster 2 menghasilkan nilai rata-rata brightness temperature kurang dari 3300K dan nilai rata-rata confidence dibawah 80% sehingga potensi sedang kebakaran hutan / lahan[9].

Penelitian yang berkaitan dengan clustering zona penyebaran Covid-19 sudah dilakukan beberapa penelitian sebelumnya seperti penelitian yang dilakukan oleh Achmad Solichin dan Khansa Khairunnisa yaitu klasterisasi Covid-19 di DKI Jakarta dengan algoritma K-Means hasil penelitiannya klaster penyebaran Covid-19 terbagi menjadi 3(tiga) klaster berdasarkan kasus pada tingkat kecamatan ada di DKI Jakarta[12]. Penelitian yang dilakukan oleh R. A. Indraputra dan R. Fitriana Clustering Data Covid-19 yaitu mengelompokan penyebaran Covid-19 di beberapa propinsi yang ada di beberapa negara dan hasilnya adalah berdasarkan data yang ada dibagi menjadi 2 cluster, dimana cluster ke-2 dengan jumlah terjangkit dan meninggal dunia lebih banyak dibandingkan dengan cluster ke-1[13]. Dari kedua penelitian tersebut berdasarkan zona penyebaran dan data yang digunakan belum ada penelitian yang berkaitan dengan clustering zona penyebaran Covid-19 khususnya di Sumatera Utara.

Berdasarkan uraian diatas bahwa algoritma clustering K-Medoids dapat digunakan untuk membentuk *cluster* zona penyebaran Covid-19 di Sumatera Utara menggunakan data jumlah kasus Covid-19 berdasarkan variabel jumlah kasus positif, PDP, ODP dan meninggal dunia. Hasil dari *clustering* berupa *cluster* zona terbaru penyebaran Covid-19 mulai pada tingkat kecamatan hingga kabupaten / kota di Sumatera Utara. Setiap kecamatan dan kabupaten / kota yang memiliki jumlah kasus yang sama atau mirip berdasarkan variabel yang digunakan maka dikelompokkan dalam satu *cluster* zona, sedangkan antar *cluster* zona memiliki jumlah kasus yang berbeda / tingkat kemiripan jauh[14][15]. Dengan penerapan pemodelan *clustering* untuk membentuk *cluster* zona penyebaran dan prediksi jumlah kasus Covid-19 diharapkan dapat membantu Gugus Tugas Percepatan Penanganan Covid 19 Sumatera Utara dan Gugus daerah dalam mengidentifikasi wilayah yang menjadi prioritas penanganan, serta mendukung proses perencanaan penanganan dan penanggulangan Covid-19 ke depan.

2. Metode Penelitian

Langkah paling awal yang dilakukan pada penelitian ini yaitu

2.1 Menyiapkan dataset

Mengumpulkan data, data yang digunakan penulis sebagai dataset adalah data jumlah kasus Covid-19 yang bersumber dari Gugus Tugas Percepatan Penanganan Covid-19 Provinsi Sumatera Utara yang diakses langsung melalui situs covid19.sumutprov.go.id dan variabel dataset digunakan adalah jumlah kasus positif, PDP, ODP, dan meninggal dunia.

2.2 Implementasi metode

Pembentukan cluster zona penyebaran Covid-19 dengan algoritma K-Means pada gambar 2

1) Sample data

Sample data yang digunakan diperoleh dari situs Gugus Tugas Percepatan Penanganan Covid-19 provinsi Sumatera Utara dengan kriteria data yaitu jumlah kasus positif, PDP, ODP, dan meninggal dunia. Data yang digunakan adalah data kasus Covid-19 perkabupaten / kota yang ada di Sumatera Utara dan jumlah data yang digunakan sebanyak 33 objek data.

2) Iterasi I

a. Jumlah Cluster

Jumlah cluster yang digunakan adalah 3(tiga) klaster sesuai dengan pembagian zona penyebaran Covid-19 yaitu zona merah, zona kuning dan zona hijau

b. Menentukan Nilai Centroid Awal

Penentuan nilai centroid awal / titik pusat awal dipilih secara acak dengan menggunakan 3(tiga) objek dataset

c. Menghitung Jarak

Tahap selanjutnya, data jumlah kasus Covid-19 dibentuk *cluster* zona dengan algoritma *clustering* K- Medoids. Rumus perhitungan yang digunakan dalam proses pembentukan *cluster* adalah persamaan Euclidean Distance (d_{ij}) dengan rumus sebagai berikut[16]:

$$d_{ij} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \dots + (x_{ki} - x_{kj})^2}$$

Keterangan:

d_{ij} = Jarak dari data ke i ke pusat cluster j

x_{ki} = Data dari ke – i pada attribute data ke – k

x_{kj} = Data dari ke – j pada attribute data ke – k

d. Membandingkan Hasil

Dari hasil perhitungan jarak dilakukan cluster/ pengelompokan dengan membandingkan nilah hasil perhitungan objek

e. Hitung Cost

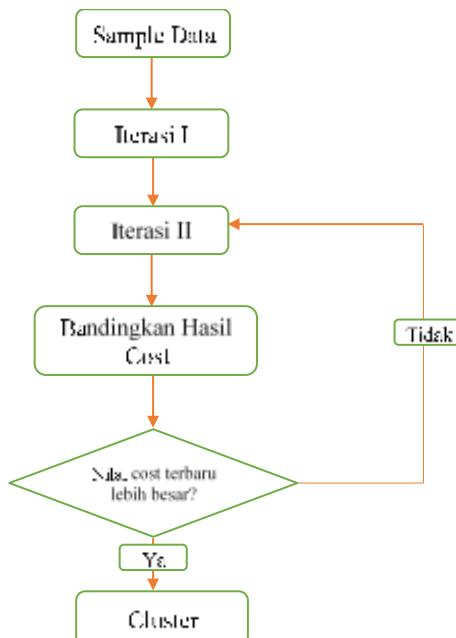
Dari hasil perhitungan jarak hitunglah nilai cost berdasarkan nilai terkecil

3) Iterasi II

Selanjutnya adalah menentukan nilai centroid terbaru dan lakukan tahapan seperti pada iterasi I sebelumnya

4) Membandingkan Hasil Cost

Dari perhitungan cost lama dan yang baru dibandingkan hasilnya, jika cost terbaru lebih besar dari cost yang lama maka proses berhenti dan jika tidak ulangi iterasi II hingga nilai cost terbaru lebih besar dari pada yang lama.



Gambar 2: Implementasi Algoritma K-Medoids

3. Hasil Dan Pembahasan

Data yang digunakan bersumber dari Gugus Tugas Percepat Penangan Covid-19 Sumatera Utara.

Tabel 1. Data Kasus Covid-19 di Sumatera Utara

No	Kode	Kabupaten/Kota	OPD	PDP	Positif	Meninggal
1	0201	Kota Medan	17	75	414	28
2	0202	Kota Pematang Siantar	0	4	34	3
3	0203	Kota Tanjung Balai	2	1	1	1
4	0204	Kota Binjai	48	0	6	0

No	Kode	Kabupaten/Kota	OPD	PDP	Positif	Meninggal
5	0205	Kota Tebing Tinggi	101	1	2	0
6	0206	Kota Sibolga	0	0	0	0
7	0207	Kota Padang Sidempuan	0	0	1	1
8	0208	Deli Serdang	10	18	82	10
9	0209	Langkat	2	0	5	1
10	0210	Karo	0	0	7	0
11	0211	Simalungun	29	14	22	0
12	0212	Asahan	0	1	6	1
13	0213	Labuhan Batu	0	1	1	0
14	0214	Tapanuli Utara	2	1	4	0
15	0215	Tapanuli Tengah	0	0	0	0
16	0216	Tapanuli Selatan	0	0	1	0
17	0217	Nias	0	0	0	0
18	0218	Dairi	1	0	1	0
19	0219	Toba	2	2	2	0
20	0220	Mandailing Natal	5	0	1	0
21	0221	Nias Selatan	104	0	0	0
22	0222	Pakpat Barat	0	0	0	0
23	0223	Humbang Hasundutan	0	0	0	0
24	0224	Samosir	0	0	0	0
25	0225	Serdang Bedagai	10	2	8	1
26	0226	Batu Bara	9	0	0	0
27	0227	Padang Lawas	0	0	0	0
28	0228	Padang Lawas Utara	0	2	0	0
29	0229	Labuhan Batu Utara	0	2	2	0
30	0230	Labuhan Batu Selatan	0	0	0	0
31	0231	Kota GunungSitoli	70	0	0	0
32	0232	Nias Utara	0	0	0	0
33	0233	Nias Barat	0	0	0	0

Sumber : Gugus Tugas Percepatan Penanganan Covid-19 Sumut (08 Juni 2020)[17]

Untuk pembentukan *cluster* zona penyebaran Covid-19 di Sumatera Utara menggunakan algoritma K-Medoids yaitu

Iterasi I

Pada iterasi I ditentukan terlebih dahulu nilai *centroid* awal / titik pusat awal. Dalam kasus ini, nilai *centroid* awal yang digunakan adalah 3(tiga) *record* data yang dipilih secara acak dari Tabel 1.

Tabel 2. Nilai centroid awal Iterasi I

No	Kode	Kabupaten/Kota	OPD	PDP	Positif	Meninggal
1	0201	Medan	C1	17	75	414
11	0211	Simalungun	C2	29	14	22
32	0232	Nias Utara	C3	0	0	0

Untuk hasil perhitungan jarak dengan menggunakan model Euclidean pada Iterasi I adalah

Tabel 3. Hasil Perhitungan Jarak Iterasi I

No	Kode	Kabupaten/Kota	dc1	dc2	dc3	Cluster
1	0201	Kota Medan	0	397,8857	422,0118	1
		Kota Pematang Siantar	387,7564	33,07567	34,36568	2
2	0202	Kota Tanjung Balai	420,7125	36,60601	2,645751	3
3	0203	Kota Binjai	416,934	28,51315	48,37355	2
4	0204	Kota Tebing Tinggi	427,8551	75,84853	101,0247	2
5	0205	Kota Sibolga	422,0118	39	0	3
		Kota Padang				
6	0206	Sidempuan	420,9656	38,45777	1,414214	3
7	0207	Deli Serdang	337,4107	63,85139	85,13519	2
8	0208	Langkat	416,9652	34,85685	5,477226	3
9	0209	Karo	415,147	35,52464	7	3
10	0210	Simalungun	397,8857	0	39	2
11	0211	Asahan	415,8822	35,59494	6,164414	3
12	0212	Labuhan Batu	420,8539	38,09199	1,414214	3
13	0213	Tapanuli Utara	417,8337	34,95712	4,582576	3
14	0214	Tapunuli Tengah	422,0118	39	0	3
15	0215	Tapanuli Selatan	421,0309	38,44477	1	3
16	0216	Nias	422,0118	39	0	3
17	0217	Dairi	420,9917	37,69615	1,414214	3
18	0218	Toba	419,6213	35,67913	3,464102	3
19	0219	Mandailing Natal	420,8586	34,82815	5,09902	3
20	0220	Nias Selatan	430,5508	79,40403	104	2
21	0221	Pakpat Barat	422,0118	39	0	3
		Humbang				
22	0222	Hasundutan	422,0118	39	0	3
23	0223	Samosir	422,0118	39	0	3
24	0224	Serdang Bedagai	413,4525	26,49528	13	3
25	0225	Batu Bara	421,7452	32,86335	9	3
26	0226	Padang Lawas	422,0118	39	0	3
		Padang Lawas				
27	0227	Utara	421,661	38,32754	2	3
28	0228	Labuhan Batu Utara	419,6975	37,21559	2,828427	3
29	0229	Labuhan Batu				
30	0230	Selatan	422,0118	39	0	3
31	0231	Kota GunungSitoli	424,9871	48,59012	70	2
32	0232	Nias Utara	422,0118	39	0	3
33	0233	Nias Barat	422,0118	39	0	3

Hasil pembentukan *cluster* berdasarkan aturan jika $dc1 < dc2$ dan $dc1 < dc3$, maka *cluster* 1, jika $dc2 < dc1$ dan $dc2 < dc3$, maka *cluster* 2, dan jika $dc3 < dc1$ dan $dc3 < dc2$ maka termasuk *cluster* 3. Perhitungan nilai total *cost* iterasi I = $0 + 33,07567 + 2,645751 + 28,51315 + 75,84853 + 0 + 1,414214 + 63,85139 + 5,477226 + 7 + 0 + 6,164414 + 1,414214 + 4,582576 + 0 + 1 + 0 + 1,414214 + 3,464102 + 5,09902 + 79,40403 + 0 + 0 + 0 + 13 + 9 + 0 + 2 + 2,828427 + 0 + 48,59012 + 0 + 0 = \text{395,7871}$

Iterasi II

Pada iterasi II sama dengan Iterasi I ditentukan terlebih dahulu nilai *centroid* awal / atau titik pusat awal. Dalam kasus ini, nilai *centroid* awal yang digunakan adalah 3(tiga) *record* data yang dipilih secara acak dari Tabel 1 dan nilai *centroid* awal pada Iterasi II berbeda dengan nilai *centroid* awal yang digunakan pada Iterasi I.

Tabel 4. Nilai centroid awal Iterasi II

No	Kode	Kabupaten/Kota	OPD	PDP	Positif	Meninggal
8	0208	Deli Serdang	C1	10	18	82
19	0219	Toba	C2	2	2	0
33	0233	Nias Barat	C3	0	0	0

Tabel 5. Hasil Perhitungan Jarak Iterasi II

No	Kode	Kabupaten/Kota	dc1	dc2	dc3	Cluster
1	0201	Kota Medan	337,4107	419,6213	422,0118	1
2	0202	Kota Pematang Siantar	51,46844	32,26453	34,36568	2
3	0203	Kota Tanjung Balai	83,63612	1,732051	2,645751	2
4	0204	Kota Binjai	87,42997	46,21688	48,37355	2
5	0205	Kota Tebing Tinggi	122,7599	99,00505	101,0247	2
6	0206	Kota Sibolga	85,13519	3,464102	0	3
7	0207	Kota Padang Sidempuan	84,0595	3,162278	1,414214	3
8	0208	Deli Serdang	0	82,58329	85,13519	1
9	0209	Langkat	79,9875	3,741657	5,477226	2
10	0210	Karo	78,41556	5,744563	7	2
11	0211	Simalungun	63,85139	35,67913	39	2
12	0212	Asahan	79,03164	4,690416	6,164414	2
13	0213	Labuhan Batu	83,96428	2,44949	1,414214	3
14	0214	Tapanuli Utara	80,85172	2,236068	4,582576	2
15	0215	Tapanuli Tengah	85,13519	3,464102	0	3
16	0216	Tapanuli Selatan	84,17244	3	1	3
17	0217	Nias	85,13519	3,464102	0	3
18	0218	Dairi	84,0595	2,44949	1,414214	3
19	0219	Toba	82,58329	0	3,464102	2
20	0220	Mandailing Natal	83,72574	3,741657	5,09902	2
21	0221	Nias Selatan	126,4278	102,0392	104	2
22	0222	Pakpat Barat	85,13519	3,464102	0	3
23	0223	Humbang Hasundutan	85,13519	3,464102	0	3
24	0224	Samosir	85,13519	3,464102	0	3
25	0225	Serdang Bedagai	76,24303	10,04988	13	2
26	0226	Batu Bara	84,55176	7,549834	9	2
27	0227	Padang Lawas	85,13519	3,464102	0	3
28	0228	Padang Lawas Utara	84,73488	2,828427	2	3
29	0229	Labuhan Batu Utara	82,80097	2	2,828427	2
30	0230	Labuhan Batu Selatan	85,13519	3,464102	0	3
31	0231	Kota GunungSitoli	103,6726	68,0588	70	2
32	0232	Nias Utara	85,13519	3,464102	0	3
33	0233	Nias Barat	85,13519	3,464102	0	3

Tabel 5 menampilkan hasil perhitungan jarak dengan model Euclidean pada Iterasi II. Hasil pembentukan *cluster* berdasarkan aturan jika $dc1 < dc2$ dan $dc1 < dc3$, maka *cluster* 1, jika $dc2 < dc1$ dan $dc2 < dc3$, maka *cluster* 2, dan jika $dc3 < dc1$ dan $dc3 < dc2$ maka *cluster* 3.

Perhitungan total nilai *cost* pada Iterasi II = $337,4107 + 32,26453 + 1,732051 + 46,21688 + 99,00505 + 0 + 1,414214 + 0 + 3,741657 + ,744563 + 35,67913 + 4,690416 + 1,414214 + 2,236068 + 0 + 1 + 0 + 1,414214 + 0 + 1,414214 + 0 + 3,741657 + 102,0392 + 0 + 0 + 10,04988 + 7,549834 + 3,464102 + 2 + 0 + 68,0588 + 0 + 0 = \textbf{769,4031}$

Perbandingan total nilai *cost* Iterasi I (*cost* lama) dengan total nilai *cost* Iterasi II (*cost* terbaru) adalah total nilai *cost* Iterasi I yaitu **395,7871** dan total nilai *cost* Iterasi II yaitu **769,4031**, dimana total nilai *cost* terbaru lebih tinggi dari pada total *cost* lama maka proses perhitungan jarak dengan iterasi selanjutnya dihentikan. Maka berdasarkan proses perhitungan pada Iterasi II maka hasil pembentukan *cluster* zona penyebaran Covid-19 khususnya di Sumatera adalah sebagai berikut:

Tabel 6. *Cluster* zona penyebaran Covid-19 di Sumatera Utara

<i>Cluster</i> 1	<i>Cluster</i> 2	<i>Cluster</i> 3
Kota Medan	Kota Pematang Siantar	Kota Sibolga
Kabupaten Deli Serdang	Kota Tanjung Balai Kota Binjai Kota Tebing Tinggi Langkat Karo Simalungun Asahan Tapanuli Utara Toba Mandailing Natal Nias Selatan Serdang Bedagai Batu Bara Labuhan Batu Utara Kota Gunung Sitoli	Kota Padang Sidempuan Labuhan Batu Tapunuli Tengah Tapanuli Selatan Nias Dairi Pakpat Barat Humbang Hasundutan Samosir Padang Lawas Padang Lawas Utara Labuhan Batu Selatan Nias Utara Nias Barat

Berdasarkan Tabel 6. *Cluster* zona penyebaran Covid-19 di Sumatera Utara dan dihubungkan dengan data Tabel 1. Data Kasus Covid-19 di Sumatera Utara dengan menggunakan variabel OPD, PDP, Positif dan meninggal dunia bahwa *cluster* 1 menunjukkan sebagai daerah di Sumatera Utara dengan jumlah kasus Covid-19 yang paling tinggi / *cluster* merah, *cluster* 2 merupakan daerah dengan jumlah kasus Covid-19 sedang/ *cluster* kuning dan *cluster* 3 merupakan daerah dengan jumlah kasus sedikit / *cluster* hijau.

Hasil penelitian *cluster* zona penyebaran Covid-19 di Sumatera Utara terbagi menjadi 3(tiga) *cluster* yaitu zona merah, zona kuning dan zona hijau menggunakan variabel OPD, DPD, Positif dan meninggal dunia. Hasil tersebut menggambarkan kondisi nyata yang terjadi dilapangan. Berbeda dengan penelitian-penelitian sebelumnya yang berkaitan dengan clustering Covid-19 sebagian penelitian tersebut dalam proses perhitungannya hanya menggunakan sebagian variabel / variabelnya terbatas hanya variabel positif dan meninggal dunia sedangkan potensi penyebaran Covid-19 dipengaruhi oleh variabel jumlah kasus positif, OPD, PDP dan meninggal dunia.

4. Kesimpulan

Berdasarkan uraian diatas, mulai dari bagian pendahuluan sampai pada hasil dan pembahasan maka kesimpulan dari penelitian yaitu Algoritma *clustering* K-Medoids dapat diterapkan dalam pembentukan *cluster* zona penyebaran Covid-19 khususnya di Sumatera Utara, Berdasarkan data yang digunakan, penyebaran Covid-19 dapat di *cluster* menjadi 3(tiga) kelompok, Hasil dari pembentukan *cluster* zona penyebaran Covid-19 adalah *cluster* 1(satu)

merupakan zona dengan kasus tinggi, *cluster 2*(dua) merupakan zona dengan kasus sedang dan *cluster 3* merupakan zona dengan kasus rendah, dan pembentukan cluster zona penyebaran Covid-19 dapat mempermudah Gugus Tugas Percepatan Penanganan Covid-19 Sumatera Utara dalam menangani dan menanunggali penyebaran Covid-19 berdasarkan tingkat *cluster*.

Daftar Pustaka

- [1] I. Wahidah, M. A. Septiadi, M. C. A. Rafqie, N. Fitria, S. Hartono, and R. Athallah, “Pandemik Covid-19 : Analisis Perencanaan Pemerintah dan Masyarakat dalam Berbagai Upaya Pencegahan COVID-19,” vol. 11, no. 3, pp. 179–188, 2020.
- [2] H. A. Diah Handayani, Dwi Rendra Hadi, Fathiyah Isbaniah, Erlina Burhan, “Penyakit Virus Corona 2019,” vol. 40, no. 2, 2019.
- [3] E. Buulolo, *Data Mining Untuk Perguruan Tinggi*. Yogyakarta: deepublish, 2020.
- [4] S. Masripah, “Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit,” vol. 3, no. 1, pp. 187–193, 2016.
- [5] A. E. Wicaksono, “Implementasi Data Mining Dalam Pengelompokan Data Peserta Didik Di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means (Studi Kasus SMAN 16 Bekasi),” *J. Ilm. Teknol. dan Rekayasa*, vol. 21, no. 3, 2017
- [6] P. N. Harahap and S. Sulindawaty, “Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT.Arma Anugerah Abadi Cabang Sei Rampah),” *Matics*, vol. 11, no. 2, p. 46, 2019, doi: 10.18860/mat.v11i2.7821.
- [7] I. Zulfa, R. Rayuwati, and K. Koko, “Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode apriori,” *Tek. J. Sains dan Teknol.*, vol. 16, no. 1, p. 69, 2020, doi: 10.36055/tjst.v16i1.7601.
- [8] P. Arora, Deepali, and S. Varshney, “Analysis of K-Means and K-Medoids Algorithm for Big Data,” 2016, doi: 10.1016/j.procs.2016.02.095.
- [9] D. F. Pramesti, Lahan, M. Tanzil Furqon, and C. Dewi, “Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 9, pp. 723–732, 2017, doi: 10.1109/EUMC.2008.4751704.
- [10] E. Buulolo and R. Syahputra, “Implementasi Algoritma Clustering K-Means Untuk Mengelompokkan Mahasiswa Baru Yang Berpotensi (Studi Kasus : Stmik Budi Darma),” vol. 2, no. September, pp. 17–24, 2019.
- [11] A. H. Ardiansyah, W. Nugroho, N. H. Alfiyah, R. A. Handoko, and M. A. Bakhtiar, “Penerapan Data Mining Menggunakan Metode Clustering untuk Menentukan Status Provinsi di Indonesia 2020,” *Semin. Nas. Inov. Teknol.*, vol. 4, no. 3, pp. 329–333, 2020.
- [12] M. M. K-means, A. Solichin, and K. Khairunnisa, “Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta,” vol. 5, no. 2, 2020.
- [13] R. A. Indraputra and R. Fitriana, “K-Means Clustering Data COVID-19,” vol. 10, no. 3, pp. 275–282, 2020.
- [14] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [15] A. Sani, “Penerapan Metode K-Means Clustering Pada Perusahaan,” *J. Ilm. Teknol. Inf.*, 2018.
- [16] I. Kamila *et al.*, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau,” vol. 5, no. 1, pp. 119–125, 2019.
- [17] covid19.sumutprov.go.id, “Situasi COVID-19 di Sumatera Utara,” *covid19.sumutprov.go.id*, 2020.

