

Menentukan Tingkat Penyebaran Pandemi Covid-19 di Indonesia dengan Menggunakan Algoritma K-Means

Sardo Pardingotan Sipayung¹, Tulus Pramita Sihalo², Ade Purba³, Sakaria Efrata Ginting⁴

^{1,2,3,4} Universitas Mandiri Bina Prestasi

Jl. Letjend. Djamin Gitning No. 285-287, Padang Bulan, Medan Baru, Kota Medan, Sumatera Utara, Indonesia - 20155

¹pinarsiphom@gmail.com, ²sihalohotulus@gmail.com, ³adelinhar31@gmail.com, ⁴sakariaginting1983@gmail.com

DOI: xx.xxxx/j.ccs.xxxx.xx.xxx

Abstrak

Corona virus (COVID-19) merupakan jenis virus baru yang ditemukan pada manusia di propinsi Wuhan, Cina pada bulan Desember 2019. Virus ini dapat menular dari manusia ke manusia melalui tetesan kecil (droplet) dari hidung atau mulut pada saat batuk, bersin, atau berbicara. Oleh karena itu, di masa pandemi ini sangat penting untuk menjaga jarak dengan orang lain dan menghindari wilayah dengan persebaran COVID-19 yang tinggi. Penyebaran kasus COVID-19 yang merata di seluruh provinsi di Indonesia, merupakan penyebaran yang cukup cepat dan berdampak negatif pada seluruh bidang. Luasnya wilayah Indonesia memungkinkan diperlukannya pengelompokan bagian bagian berdasarkan wilayah di Indonesia. Pengelompokan ini akan menghasilkan titik – titik pusat penyebaran kasus COVID -19. Salah satu algoritma Clustering adalah K-Means yang menggunakan beberapa kelompok untuk penempatan beberapa data dengan sistem partisi. Data-data tanpa label kelas diterima oleh Algoritma ini. Dikarenakan pandemi global yang terjadi banyak pihak berupaya ikut berperan serta dalam mengatasi. Penelitian ini dilakukan untuk penerapan dalam penyebaran pandemi covid-19 di Indonesia. Dalam penelitian ini menggunakan algoritma K-Means untuk menentukan bagaimana tingkat penyebaran covid-19 di daerah-daerah yang ada di Indonesia.

Kata Kunci: Algoritma, K-Means , COVID-19, Penyebaran, Pandemi.

1. Pendahuluan

COVID-19 adalah penyakit yang menular, dan ditandai oleh gejala pada bagian pernapasan akut sindrom pernafasan coronavirus 2 (SARS-CoV-2). Virus ini merupakan keluarga besar Coronavirus yang dapat menyerang hewan. Penularan virus ini terjadi jika adanya kontak antar sesama manusia. Ketika menyerang manusia, Penyakit yang disebabkan oleh virus ini adalah penyakit infeksi saluran pernafasan, seperti flu, MERS (Middle East Respiratory Syndrome), dan SARS (Severe Acute Respiratory Syndrome).[1]

Pada tanggal 2 Maret 2020 di Indonesia sebanyak dua kasus positif COVID-19 pertama kali dilaporkan. Terkonfirmasi berjumlah 1.528 kasus dan 136 kasus kematian pada 31 Maret 2020. Tingkat kematian COVID-19 di Indonesia sebesar 8,9%, adalah jumlah tertinggi di Asia Tenggara.[1]

Virus corona dapat menular dari manusia ke manusia melalui droplet (tetesan super kecil) yang dihasilkan oleh penderita saat batuk, bersin, atau berbicara. Beberapa penelitian juga membuktikan bahwa virus corona dapat bertahan beberapa jam di

benda-benda yang disentuh oleh penderita. Jika terdapat seseorang menyentuh benda yang sudah terkontaminasi dengan droplet dan menyentuh mata, hidung atau mulut, maka orang itu dapat tertular COVID-19. Oleh karena itu, salah satu pencegahan penularan virus ini adalah dengan menjaga jarak fisik, menggunakan masker dan senantiasa menjaga kesehatan fisik dan non-fisik. Walaupun sejumlah penelitian telah dilakukan untuk mendeteksi virus Covid-19 [2], namun hingga saat ini belum ditemukan cara menghentikan penyebaran virus ini. Pencegahan dapat dilakukan dengan membatasi pergerakan manusia, menjaga jarak fisik, menggunakan masker, dan menerapkan pola hidup sehat.

K-Means merupakan salah satu algoritma Clustering yang masuk dalam kelompok Unsupervised learning yang digunakan untuk mengelompokkan data kedalam beberapa kelompok dengan sistem partisi. Algoritma ini menerima masukan berupa data tanpa label kelas. Pada algoritma K-Means, komputer menerima data-data yang tidak diketahui kelasnya terkebih dahulu lalu mengelompokkannya. Input yang diterima ialah data dan jumlah kelompok (cluster) yang diinginkan.[2] Algoritma ini akan memasukan data kedalam beberapa kelompok tersebut.

2. Metode Penelitian

2.1. Tahap Pengumpulan Data

Data yang digunakan untuk penelitian ini adalah data jumlah penyebaran virus COVID-19 di Indonesia yang diperoleh dari Kementerian Kesehatan Republik Indonesia. Data penyebaran virus COVID-19 yang digunakan pada penelitian ini adalah data penyebaran pada tanggal 12 Januari 2021. Template memformat teks Anda dengan menggunakan fitur Word yang disebut 'Styles'. Styles menentukan format (atau tampilan) paragraf teks sehubungan dengan ukuran huruf, lekukan, spasi baris, dll. Jika Anda tidak terbiasa menggunakan gaya, jangan khawatir; template mengatur segalanya untuk Anda dengan cara yang mudah digunakan.

2.2. COVID-19

COVID-19 adalah penyakit yang menular, dan ditandai oleh gejala pada bagian pernapasan akut (SARS-CoV-2). Virus ini merupakan keluarga besar Coronavirus yang dapat menyerang hewan. Penularan virus ini terjadi jika adanya kontak antar sesama manusia. Ketika menyerang manusia, Virus ini dapat menyebabkan penyakit infeksi saluran pernafasan, seperti MERS (Middle East Respiratory Syndrome), dan SARS (Severe Acute Respiratory Syndrome).[1] Virus ini telah tersebar lebih dari 190 negara lainnya. Oleh karena itu COVID-19 ditetapkan sebagai pandemi oleh WHO.

Sampai tanggal 11 Januari 2021, telah terjadi 90.687.987 kasus dan 1.943.090 kasus kematian di dunia. Sementara di Indonesia terdapat 836.718 kasus positif COVID-19 dan 24.343 kasus kematian. Dikarenakan pandemi global yang terjadi banyak pihak berupaya ikut berperan serta dalam mengatasi. Para dokter umum dan spesialis angkat bicara bersama guna memberi penjelasan singkat kepada masyarakat maupun imbauan agar menjaga kebersihan diri dan lingkungan sekaligus tak banyak keluar rumah.

2.3. Algoritma K-Means

Salah satu algoritma Clustering adalah K-Means yang digunakan untuk mengelompokkan data menjadi beberapa kelompok dengan beberapa Cluster.[2] Data-data dipilih menjadi beberapa kelompok dengan kriteria yang telah ditentukan lalu dikumpulkan menjadi satu dalam sebuah Cluster. Dimana setiap Cluster memiliki titik pusat yang disebut Centroid.

Berikut adalah tahapan-tahapan untuk melakukan optimasi menggunakan algoritma K-Means:[2]

- Pilihlah jumlah Cluster (k) yang diinginkan pada dataset
- Tentukan titik pusat (Centroid) secara acak/random pada tahap awal
- Hitunglah jarak terdekat setiap data dengan Centroid. Untuk menghitung jarak terdekat dengan Centroid adalah Euclidean distance dapat menggunakan rumus di bawah ini:

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

$d_{(x,y)}$ = jarak data ke i ke pusat cluster j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut ke k

- Hitung kembali pusat Cluster dengan keanggotaan Cluster yang sekarang. Pusat Cluster adalah rata-rata dari semua data dalam sebuah Cluster. Dapat dihitung menggunakan rumus:

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \dots\dots\dots(2)$$

Keterangan:

V_{ij} = Centroid rata-rata pada Cluster ke - i untuk variabel ke -j

N_i = Jumlah anggota Cluster ke -i

i,k = Indeks dari Cluster j = Indeks variabel

X_{kj} = Nilai data ke -k variabel ke -j untuk Cluster tersebut

- Hitung kembali setiap objek menggunakan pusat Cluster baru (Centroid baru), ini merupakan tahap awal pembukaan iterasi baru. Jika anggota Cluster tidak mengalami perpindahan Cluster lagi, maka proses Clustering dinyatakan selesai. Tetapi, jika anggota Cluster mengalami perpindahan, maka kembali lagi pada langkah ke-c sampai anggota Cluster tidak mengalami perpindahan lagi.

3. Hasil dan Pembahasan

3.1. Perhitungan dengan Algoritma K-Means

Saat dilakukan perhitungan manual dengan data sampel sebanyak 34 provinsi yang terkena dampak penyebaran COVID-19 di Indonesia. Berikut adalah sampel data penyebaran COVID-19 di Indonesia per tanggal 12 Januari 2021:

Tabel 1.

Sampel Data Penyebaran

| No | Provinsi | Kasus Positif | Kasus Meninggal |
|----|---------------------|---------------|-----------------|
| 1 | Aceh | 8967 | 368 |
| 2 | Bali | 19835 | 569 |
| 3 | Bangka Belitung | 3132 | 53 |
| 4 | Banten | 20864 | 468 |
| 5 | Bengkulu | 4096 | 133 |
| 6 | DI Yogyakarta | 15482 | 336 |
| 7 | Dki Jakarta | 211252 | 3550 |
| 8 | Gorontalo | 4011 | 106 |
| 9 | Jambi | 3649 | 63 |
| 10 | Jawa Barat | 100585 | 1245 |
| 11 | Jawa Tengah | 94087 | 4161 |
| 12 | Jawa Timur | 94249 | 6576 |
| 13 | Kalimantan Barat | 3383 | 28 |
| 14 | Kalimantan Selatan | 16227 | 601 |
| 15 | Kalimantan Tengah | 10574 | 286 |
| 16 | Kalimantan Timur | 31076 | 827 |
| 17 | Kalimantan Utara | 5139 | 69 |
| 18 | Kepulauan Riau | 7357 | 180 |
| 19 | Lampung | 7541 | 368 |
| 20 | Maluku | 5951 | 90 |
| 21 | Maluku Utara | 2972 | 94 |
| 22 | Nusa Tenggara Barat | 6237 | 289 |
| 23 | Nusa Tenggara Timur | 2575 | 69 |
| 24 | Papua | 13767 | 155 |
| 25 | Papua Barat | 6246 | 104 |
| 26 | Riau | 26686 | 622 |
| 27 | Sulawesi Barat | 2368 | 53 |
| 28 | Sulawesi Selatan | 37766 | 652 |
| 29 | Sulawesi Tengah | 5017 | 139 |
| 30 | Sulawesi Tenggara | 8499 | 164 |
| 31 | Sulawesi Utara | 10699 | 343 |
| 32 | Sumatera Barat | 24620 | 548 |
| 33 | Sumatera Selatan | 12644 | 635 |
| 34 | Sumatera Utara | 19212 | 701 |

Pada tahap ini dapat diketahui bahwa:

Jumlah Cluster: 3

Jumlah Data: 34

Jumlah Atribut: 2

(Kasus Positif dan Kasus Meninggal)

- Menentukan centroid awal secara acak. C1 (DKI Jakarta) = (21152,3550), C2 (Jawa Barat) = (100585,1245), C3 (DI Yogyakarta) = (15482,336)
 Pada Cluster penyebaran COVID-19 yang menjadi penyebaran dalam jangka besar di ambil pada data ke 7 yaitu DKI Jakarta, lalu untuk Cluster dengan penyebaran jangka menengah adalah data ke 10 yaitu Jawa Barat, dan untuk Cluster dengan penyebaran jangka kecil adalah data ke 6 yaitu DI Yogyakarta.

- Menghitung centroid terdekat.

Misalnya: Sumatera Utara (19212,652)

$$C1 = \sqrt{(19212 - 211252)^2 + (652 - 3550)^2} = 192061.132$$

Sumatera Utara

$$C2 = \sqrt{(19212 - 100585)^2 + (652 - 1245)^2} = 81374.81837$$

Sumatera Utara

$$C3 = \sqrt{(19212 - 15482)^2 + (652 - 336)^2}$$

Sumatera Utara
 = 3747.816031

Langkah selanjutnya ialah menghitung jarak setiap data dengan Centroid dengan persamaan Euclidean Distance. Pada tahap ini jarak terdekat antara data dengan Cluster akan menentukan suatu data masuk ke dalam Cluster mana. Berikut adalah hasil perhitungan jarak pada Iterasi ke 1.

Tabel 2.

Hasil Perhitungan Iterasi 1

| Provinsi | Jarak Ke Centroid | | |
|---------------------|-------------------|----------|----------|
| | C1 | C2 | C3 |
| Aceh | 202310.025 | 91622.2 | 6515.079 |
| Bali | 191440.211 | 80752.83 | 4359.231 |
| Bangka Belitung | 208149.378 | 97460.29 | 12353.24 |
| Banten | 190412.944 | 79724.79 | 5383.618 |
| Bengkulu | 207184.179 | 96495.41 | 11387.81 |
| DI Yogyakarta | 195796.381 | 85107.85 | 0 |
| Dki Jakarta | 0.000 | 110691 | 195796.4 |
| Gorontalo | 207269.615 | 96580.72 | 11473.31 |
| Jambi | 207632.283 | 96943.21 | 11836.15 |
| Jawa Barat | 110691.002 | 0 | 85107.85 |
| Jawa Tengah | 117166.593 | 7122.293 | 78698.01 |
| Jawa Timur | 117042.124 | 8280.366 | 79013.78 |
| Kalimantan Barat | 207898.835 | 97209.62 | 12102.92 |
| Kalimantan Selatan | 195047.295 | 84360.46 | 790.7275 |
| Kalimantan Tengah | 200704.542 | 90016.11 | 4908.255 |
| Kalimantan Timur | 180196.575 | 69510.26 | 15601.73 |
| Kalimantan Utara | 206142.393 | 95453.24 | 10346.45 |
| Kepulauan Riau | 203922.848 | 93234.08 | 8126.497 |
| Lampung | 203735.850 | 93048.13 | 7941.064 |
| Maluku | 205330.154 | 94641.05 | 9534.174 |
| Maluku Utara | 208308.671 | 97619.79 | 12512.34 |
| Nusa Tenggara Barat | 205040.933 | 94352.84 | 9245.119 |
| Nusa Tenggara Timur | 208706.032 | 98017.06 | 12909.76 |
| Papua | 197514.180 | 86824.84 | 1724.525 |
| Papua Barat | 205034.960 | 94345.9 | 9238.913 |
| Riau | 184589.224 | 73901.63 | 11207.65 |
| Sulawesi Barat | 208913.270 | 98224.23 | 13117.05 |
| Sulawesi Selatan | 173510.203 | 62821.8 | 22286.24 |
| Sulawesi Tengah | 206263.206 | 95574.4 | 10466.85 |
| Sulawesi Tenggara | 202781.271 | 92092.34 | 6985.118 |
| Sulawesi Utara | 200578.640 | 89890.53 | 4783.005 |
| Sumatera Barat | 186656.142 | 75968.2 | 9140.459 |
| Sumatera Selatan | 198629.391 | 87943.12 | 2853.707 |
| Sumatera Utara | 192061.132 | 81374.82 | 3747.816 |

- Melakukan pengelompokkan berdasarkan cluster. Berikut adalah hasil pengelompokkan cluster pada Iterasi 1:

Tabel 3.

Pengelompokkan Cluster Iterasi 1

| Provinsi | Jarak Terdekat |
|---------------------|----------------|
| Aceh | C3 |
| Bali | C3 |
| Bangka Belitung | C3 |
| Banten | C3 |
| Bengkulu | C3 |
| DI Yogyakarta | C3 |
| Dki Jakarta | C1 |
| Gorontalo | C3 |
| Jambi | C3 |
| Jawa Barat | C2 |
| Jawa Tengah | C2 |
| Jawa Timur | C2 |
| Kalimantan Barat | C3 |
| Kalimantan Selatan | C3 |
| Kalimantan Tengah | C3 |
| Kalimantan Timur | C3 |
| Kalimantan Utara | C3 |
| Kepulauan Riau | C3 |
| Lampung | C3 |
| Maluku | C3 |
| Maluku Utara | C3 |
| Nusa Tenggara Barat | C3 |
| Nusa Tenggara Timur | C3 |
| Papua | C3 |
| Papua Barat | C3 |
| Riau | C3 |
| Sulawesi Barat | C3 |
| Sulawesi Selatan | C3 |
| Sulawesi Tengah | C3 |
| Sulawesi Tenggara | C3 |
| Sulawesi Utara | C3 |
| Sumatera Barat | C3 |
| Sumatera Selatan | C3 |
| Sumatera Utara | C3 |

| | |
|---------------------|----|
| Gorontalo | C3 |
| Jambi | C3 |
| Jawa Barat | C2 |
| Jawa Tengah | C2 |
| Jawa Timur | C2 |
| Kalimantan Barat | C3 |
| Kalimantan Selatan | C3 |
| Kalimantan Tengah | C3 |
| Kalimantan Timur | C3 |
| Kalimantan Utara | C3 |
| Kepulauan Riau | C3 |
| Lampung | C3 |
| Maluku | C3 |
| Maluku Utara | C3 |
| Nusa Tenggara Barat | C3 |
| Nusa Tenggara Timur | C3 |
| Papua | C3 |
| Papua Barat | C3 |
| Riau | C3 |
| Sulawesi Barat | C3 |
| Sulawesi Selatan | C3 |
| Sulawesi Tengah | C3 |
| Sulawesi Tenggara | C3 |
| Sulawesi Utara | C3 |
| Sumatera Barat | C3 |
| Sumatera Selatan | C3 |
| Sumatera Utara | C3 |

4. Menentukan Centroid baru untuk iterasi berikutnya dengan mencari rata-rata nilai dari setiap cluster. Berikut adalah nilai centroid baru untuk iterasi 2

| Centroid | C1 | 211252 | 3550 |
|----------|----|-----------|---------|
| | C2 | 96307 | 3994 |
| | C3 | 11553.066 | 303.767 |

Jika tahapan iterasi telah mencapai hasil yang sama tanpa ada perpindahan lagi, maka perhitungan dihentikan.

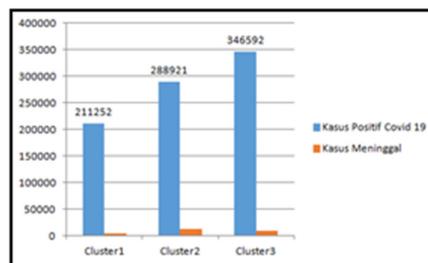
Pada penelitian ini iterasi dilakukan sampai iterasi ke 2. Hasil akhir posisi Cluster dari Centroid iterasi ke 2 dapat dilihat pada table dibawah ini:

Tabel 4.

Pembagian Cluster pada Iterasi ke -2

| Provinsi | Jarak Terdekat |
|-----------------|----------------|
| Aceh | C3 |
| Bali | C3 |
| Bangka Belitung | C3 |
| Banten | C3 |
| Bengkulu | C3 |
| DI Yogyakarta | C3 |
| Dki Jakarta | C1 |

Hasil Pengolahan Data secara manual dengan 34 Provinsi dapat dilihat pada gambar berikut ini:



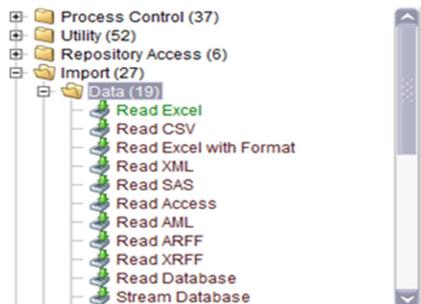
Gbr 1. Cluster tingkat penyebaran COVID-19

Hasil yang didapat dari grafik tersebut adalah Cluster 1 sebagai penyebaran dalam jangka besar yaitu sebanyak 211252 kasus pasien positif COVID-19 dengan 3550 kasus meninggal. Untuk Cluster 2 sebagai penyebaran dalam jangka menengah yaitu dengan 288921 kasus positif COVID-19 dengan 11982 kasus meninggal. Dan pada Cluster 3 sebagai penyebaran jangka kecil yaitu dengan 346592 kasus positif COVID19 dengan 9113 kasus meninggal.

3.2. Implementasi Pada Aplikasi RapidMiner

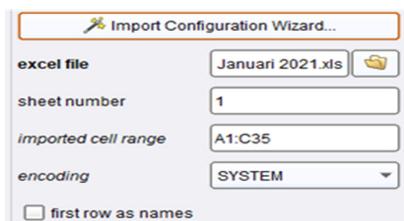
Dalam pengimplementasian pada aplikasi RapidMiner dapat dilihat sebagai berikut:

1. Import Data kedalam RapidMiner dalam bentuk sheet Excel



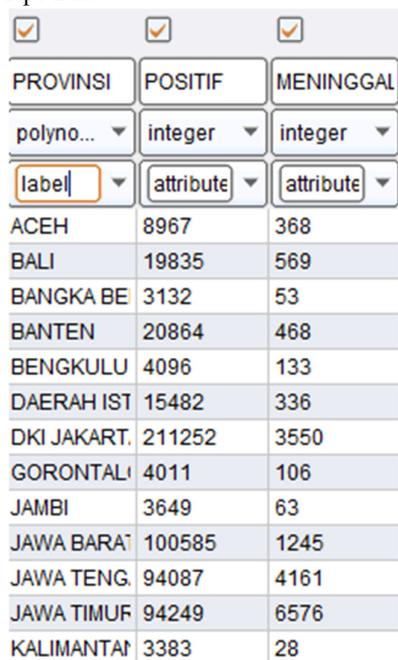
Gbr 2. Import Data

2. Import Configuration Wizard



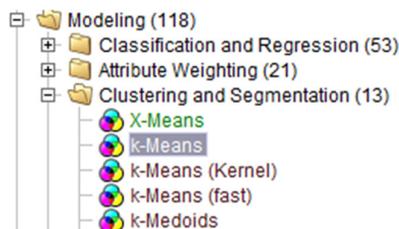
Gbr 3. Konfigurasi Wizard

3. Pilih File, sesuaikan Anotasi dan Memilih Atribut dan Tipe Data



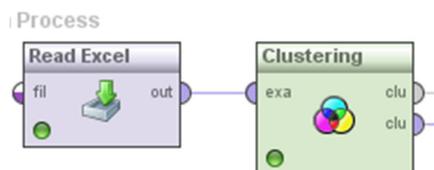
Gbr 4. Penyesuaian Atribut dan Tipe Data

4. Menambahkan Model K-Means



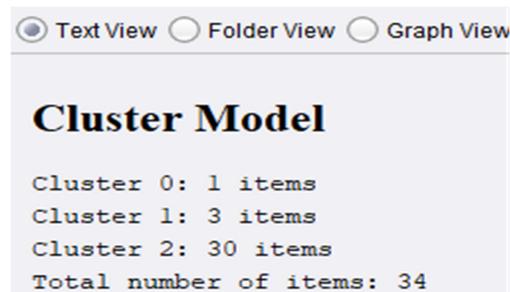
Gbr 5. Pemilihan Model K-Means

5. Menghubungkan Model K-Means

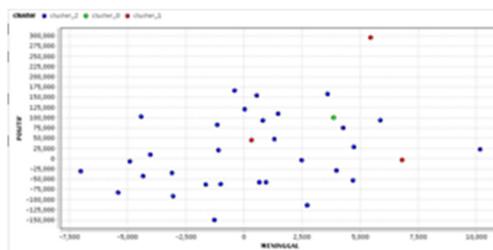


Gbr 6. Menghubungkan K-Means

6. Jalankan Project Tampilan Text View



Gbr 7. Hasil Tampilan Text View Tampilan Plot View



Gbr 8. Hasil Tampilan Plot View

Setelah melakukan implementasi menggunakan Aplikasi RapidMiner, hasil yang didapat setelah di implementasikan pada RapidMiner sama dengan hasil pada perhitungan manual.

4. Kesimpulan

Berdasarkan penelitian dan hasil perhitungan menggunakan Algoritma K-Means dalam penyebaran

COVID-19 di Indonesia, maka dapat ditarik kesimpulan sebagai berikut:

1. Pusat Cluster yang diperoleh yaitu untuk Cluster pertama berada pada provinsi DKI Jakarta. Pusat Cluster kedua berada pada provinsi Jawa Barat, dan pusat Cluster ketiga berada pada provinsi DI Yogyakarta.
2. Pusat Cluster pertama merupakan daerah penyebaran COVID-19 terbesar di Indonesia dengan kasus positif dan kasus meninggal paling besar.
3. Daerah-daerah pada Cluster kedua dan ketiga merupakan penyebaran kasus COVID-19 yang cukup potensial, sehingga harus menjadi perhatian pemerintah untuk penanganannya

Referensi

- [1] Nayuni Dwitri, et al, Data Mining : Algoritma dan Implementasi, Medan.
- [2] Yayasan Kita Menulis, 2020 A.R. Setiawan, Lembar Kegiatan Literasi Saintifik untuk Pembelajaran Jarak Jauh Topik

Penyakit Coronavirus 2019 (COVID-19), Edukatif: Jurnal Ilmu Pendidikan Vol 2, No 1, 2020.

- [3] A. Susilo, et al, Coronavirus Disease 2019: Tinjauan Literatur Terkini . Jurnal Penyakit Dalam Indonesia, Vol. 7, No. 1, 2020.
- [4] A. Bastian, et al, Penerapan Algoritma Kmeans Clustering Analysis Pada Penyakit Menular Manusia (Studi Kasus Kabupaten Majalengka). Jurnal Sistem Informasi (Journal of Information System), Volume 14, Issue 1, 2018.
- [5] M.T.I. Rahmayani, Analisis Clustering Tingkat Keperahan Penyakit Pasien Menggunakan Algoritma K-Means (Studi Kasus di Puskesmas Bandar Seikijang), JITI, Vol.1, No. 2, 2018.
- [6] Nayuni Dwitri , Penerapan Algoritma K-means dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 di Indonesia: Jurnal Teknologi Informasi) Vol.4, No.1, Juni 2020
- [7] Achmad Solichin, Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means, Fountain of Informatics Journal Volume 5, No. 2, November 2020
- [8] Wiyli Yustanti, Klastering Wilayah Kota/Kabupaten Berdasarkan Data Persebaran Covid-19 di Propinsi Jawa Timur dengan Metode K-Means, (Journal Information Engineering and Educational Technology) JIEET: Volume 04 Nomor 01, 2020