

## IMPLEMENTASI ALGORITMA K-MEDOIDS UNTUK CLUSTERING WILAYAH TERINFEKSI KASUS COVID19 DI DKI JAKARTA

Muh. Arifandi<sup>1</sup>, Arief Hermawan<sup>2</sup>, Donny Avianto<sup>3</sup>

<sup>1</sup> Magister Teknologi Informasi, Program Pascasarjana, Universitas Teknologi Yogyakarta  
<sup>2,3</sup> Prodi Studi Informatika, Fakultas Sains & Teknologi, Universitas Teknologi Yogyakarta  
E-mail: <sup>1</sup>muh.arifandi@student.uty.ac.id, <sup>2</sup>ariefdb@uty.ac.id, <sup>3</sup>donny@uty.ac.id

### Abstrak

Pada awal Maret tahun 2019 Indonesia dilanda dengan wabah Covid-19 (*Corona*). Peningkatan jumlah pasien terinfeksi virus Covid-19 semakin melonjak hari demi hari dan sudah sulit dikendalikan. Untuk mencegah bertambahnya kasus Covid-19 adalah membuat *cluster* atau pengelompokan wilayah tertentu (kelurahan) berdasarkan jumlah positif, dirawat, sembuh, meninggal dan isolasi. Pengelompokan ini akan membantu pemerintah DKI Jakarta dalam memberikan penanganan yang sesuai dengan pola kelurahan tersebut. Data yang akan menjadi kajian penelitian adalah mengenai data persebaran status terinfeksi kasus Covid-19 di Provinsi DKI Jakarta pada tanggal 20 Mei 2021. Algoritma *K-Medoids* merupakan metode yang dapat menentukan suatu set *cluster* di antara sekelompok data yang mendekati suatu objek. Berdasarkan kajian penelitian yang sudah dilakukan dapat diambil kesimpulan bahwa pada teknik data mining dalam pengelompokkan total kasus terinfeksi Covid-19 berdasarkan kelurahan di Provinsi DKI Jakarta menggunakan algoritma *K-Medoids* dengan tiga *Cluster*. *Cluster 0*, *Cluster 1*, *Cluster 2*. Kasus terinfeksi Covid-19 tertinggi pada Provinsi DKI Jakarta ditunjukkan pada *Cluster 0* dengan 31 wilayah. Hasil penelitian Pengelompokkan ini akan membantu pemerintah DKI Jakarta dalam memberikan penanganan yang sesuai dengan pola kelurahan. *K-Medoids* dapat diterapkan menggunakan data yang berjumlah besar dengan atribut yang kompleks.

**Kata Kunci:** Covid-19, K-Medoids, *Data Mining*, *Clustering*.

### Abstract

*In early March 2019, Indonesia was hit by the Covid-19 (Corona) outbreak. The increase in the number of patients infected with the Covid-19 virus is increasing day by day and is already difficult to control. Jakarta is no exception. To prevent the increase in cases of COVID-19, it is necessary to create a cluster or grouping of certain areas (Urban village) based on the number of positive, treated, recovered, died and isolated. This grouping will assist the DKI Jakarta government in providing appropriate handling according to the Urban village pattern. The data that will be used as a research study is the data on the distribution of the status of infected cases of Covid-19 in DKI Jakarta Province on May 20, 2021. The K-Medoids algorithm is a method that can determine a set of clusters among a group of data that is close to an object. Based on the research studies that have been carried out, it can be concluded that in the data mining technique, the total grouping of Covid-19 infected cases based on urban areas in DKI Jakarta Province uses the k-medoids algorithm with three clusters. Cluster 0, cluster 1, cluster 2. The highest Covid-19 infected cases in DKI Jakarta Province are shown in cluster 3 with 31 regions. The results of this grouping research will assist the DKI Jakarta government in providing appropriate handling according to the Urban village pattern. K-Medoids can be implemented using large amounts of data with complex attributes.*

**Keywords:** Covid-19, K-Medoids, *Data Mining*, *Clustering*.

## I. PENDAHULUAN

Pada awal Maret tahun 2019 Indonesia dilanda dengan wabah virus Covid-19 (*Corona*). Peningkatan total pasien terinfeksi virus Covid-19 semakin melonjak dari hari kehari dan sudah sulit dikendalikan, maka dari itu diperlukannya tindakan pemerintah setempat untuk mengatasi kasus ini. Virus *Corona* berasal dari kelompok *Orthocoronavirinae* pada famili *Coronaviridae* dan ordo *Nidovirales* (Yunus NR & A, 2020). Tidak terkecuali di DKI Jakarta. Masyarakat di minta untuk melakukan pembatasan jarak yang tujuannya untuk mencegah bertambahnya kasus terinfeksi virus corona. Covid-19 dapat tertular kepada semua orang melalui hidung atau mulut dari orang yang terjangkit virus *corona*. Pada setiap wilayah yang terkena pandemi Covid-19 di DKI Jakarta memiliki penyebaran berbeda-beda, baik dilihat dari jumlah yang positif, isolasi mandiri, yang sembuh, ataupun yang dirawat serta meninggal. Kondisi yang berbeda tersebut memerlukan penanganan yang berbeda pula pada setiap daerah (kelurahan). Untuk mencegah bertambahnya kasus covid-19 adalah dengan membuat cluster atau pengelompokan wilayah tertentu (kelurahan) berdasarkan jumlah positif, dirawat, sembuh, meninggal dan isolasi. Upaya memberikan info yang sesuai dengan kondisi daerah maka diperlukan suatu sistem yang bisa menempatkan daerah sesuai dengan *cluster*-nya. Algoritma *K-medoids* merupakan metode yang dapat menentukan sebuah set *cluster* dari sekelompok data terdekat dari sebuah objek. Pengelompokan ini akan membantu pemerintah DKI Jakarta dalam memberikan penanganan yang sesuai dengan pola kelurahan tersebut.

Ada beberapa penelitian tentang *clustering* covid antara lain yang di lakukan oleh (Untoro, Meida Cahyo, 2021), dalam penelitiannya untuk *clustering* wilayah Provinsi Dki Jakarta. Penelitian ini serupa dengan penelitian yang telah di lakukan peneliti saat ini namun metode yang di gunakan berbeda. Metode yang di gunakan dalam penelitiannya adalah *K-Means*. Adapun hasil penelitian yang di laporkan, jumlah *cluster* yang di gunakan ada 2 *cluster*. Pada *cluster* 0 diperoleh 173 data, sedangkan pada *cluster* 1 diperoleh 18 data. Hasil penelitian yang diperoleh akan di laporkan ke pemerintah, sehingga pemerintah dapat melakukan penanganan segera sesuai wilayah yg sudah di *cluster*.

Penelitian yang serupa di laksanakan oleh (Bu'ulolo & Purba, 2021). Tujuan dari penelitiannya adalah untuk *clustering* kasus covid pada Provinsi Sumatera Selatan sehingga pemerintah setempat dapat melakukan penanganan

Covid-19 dan menanggulangnya. Dalam penelitian yang di laksanakan diketahui menggunakan metode *K-Medoids* dengan jumlah *cluster* 3. Dari ketiga *cluster* tersebut diketahui *cluster* tertinggi adalah *cluster* 1, sedangkan *cluster* 2 merupakan *cluster* yang sedang dan *cluster* 3 merupakan *cluster* yang terendah.

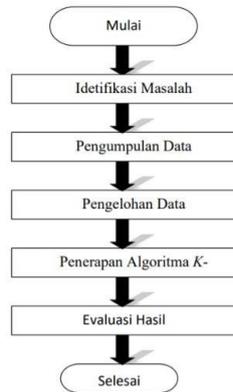
Penelitian lainnya di laksanakan oleh (Yustianti et al., 2021). Tujuan dari penelitian yang dilaksanakan adalah untuk membantu Provinsi Jawa Timur dengan cara mengelompokkan kasus penyebaran Covid-19 dengan 5 jumlah *cluster* karakteristik yakni Zona Hijau Mudah, Zona Hijau dan Zona Merah. Karakteristik klaster pada tiap *cluster* dapat diketahui secara umum adalah klaster dengan kasus PDP dan *Confirm* yang sangat tinggi (red zone) , klaster dengan kasus jumlah *Confirm* Tinggi ( *Orange Zone*), klaster dengan jumlah PDP Tinggi ( *Yellow Zone*), klaster dengan jumlah kasus PDP dan *Confirmed* Sedang ( *Green Zone*) dan klaster dengan PDP dan *Confirm* rendah ( *Light Green Zone*). Adapun metode yang di gunakan dalam penelitiannya adalah *K-Means*.

Penelitian selanjutnya di lakukan oleh (Indraputra & Fitriana, 2020) menggunakan metode *K-Means* sebagai algoritma pengelompokkan untuk mengetahui daerah daerah yang memerlukan penanganan Covid-19 darurat. Dalam penelitian yang dilaksanakan diketahui memiliki 2 *cluster*. Hasil *clustering* yang di lakukan dengan 2 jumlah *cluster* di ketahui *cluster* 2 merupakan jumlah kasus terinfeksi di bandingkan dengan *cluster* 1. Penelitian lainnya di lakukan oleh (Virgantari & Faridhan, 2020) menggunakan *K-Means* sebagai metode *clustering* kasus Covid-19 di wilayah-wilayah Indonesia. Penelitiannya merupakan upaya awal dalam rangka menginformasikan kepada masyarakat dan meningkatkan kesadaran akan penyebaran penyakit. Dalam penelitian yang telah dilaksanakan dengan jumlah 7 *cluster*. Data yang digunakan dalam penelitiannya diperoleh dari Kementerian Kesehatan Republik Indonesia per 8 November 2021.

Berdasarkan situasi yang telah di bahas di atas, maka peneliti tertarik melakukan penelitian serupa menggunakan algoritma *K-Medoids* untuk *clustering* kasus Covid-19 di wilayah Provinsi DKI Jakarta. Tujuan dari kajian ini adalah untuk membantu pemerintah memanalisis kasus terinfeksi Covid-19 serta menyajikan cara penerapan algoritma *K-Medoids* dalam mengelompokkan data kasus Covid-19 dengan jumlah 3 *cluster*. Adapun langkah-langkahnya dan bahan untuk penerapannya akan di bahas pada bagian selanjutnya.

## II. METODE

Berdasarkan masalah yang melatar belakangi penelitian yang telah di paparkan, penelitian ini di lakukan beberapa tahapan, pada Gambar 1 merupakan tahapan kerangka kerja dalam penyusunan penelitian.



Gambar 1. Tahapan Penelitian

### a) Identifikasi Masalah

Pada penelitian ini, langkah pertama yang dilakukan adalah menggunakan teknik *clustering K Medoids* untuk mendefinisikan masalah dengan tujuan untuk mengetahui berapa banyak *cluster* yang terdapat pada masalah pada kasus terinfeksi.

### b) Pengumpulan Data

Sumber data pada penelitian ini ialah data persebaran status terinfeksi virus *corona* (Covid) di Provinsi DKI Jakarta. Data diperoleh dari situs riwayat covid-19 dengan url: <https://riwayat-file-covid-19-dkijakartajakartagis.hub.arcgis.com/>.

### c) Pengolahan Data

Pengolahan Data dengan melakukan penentuan atribut, normalisasi data dan transformasi pada data yang terpilih untuk melakukan *Clustering*.

### d) Penerapan Algoritma *K-Medoids*

Tahap ini akan membahas bagaimana cara mengelompokkan data kasus terinfeksi kasus Covid-19 menggunakan Algoritma *K-Medoids*.

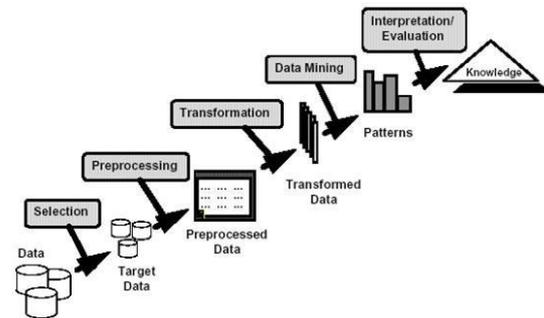
## Data Mining

Data mining merupakan sebuah proses untuk menemukan korelasi, pola dan tren dengan menganalisis sejumlah data dengan menggunakan teknik pengenalan pola (Nasari & Sianturi, 2016).

Data mining merupakan suatu teknik analisa pada data dengan tujuan untuk mendapatkan relasi yang pasti dan juga memberikan hasil yang bermanfaat kepada pemilik data karena memberikan hasil yang tidak diketahui. Data mining juga disebut metoda yang dapat di gunakan mengekstraksi informasi *predictive* yang tidak diketahui pada database. Data mining juga sebuah teknologi

potensial pada perusahaan dalam memperdayakan data *warehouse*. Data mining dibagi menjadi 2 kelompok utama, antara lain:

- Descriptive* merupakan metode untuk mencari tahu ciri-ciri penting dari data di dalam sekumpulan data. Strategi data mining yang terkategori sebagai deskriptif terdiri dari pengelompokkan, asosiasi, dan penambangan berurutan.
- Prediktif merupakan metode untuk mencari tahu pola dari data menggunakan variabel yang berbeda di masa depan. Klasifikasi adalah metode yang terletak di penambangan prediktif. Data mining juga dapat dikenal sebagai teknik penyaringan atau penambangan data dari sekumpulan basis data yang sangat besar. Data mining juga bisa disebut sebagai *Knowledge Discovery in Database* atau disingkat KDD (Darmi & Setiawan, 2016)(Agus, 2017).



Gambar 2. KDD (*Knowledge Discovery in Database*)

## Clustering

*Clustering* adalah proses membagi sekelompok data menjadi beberapa kelompok, dimana kemiripan antara data dengan satu kelompok data lebih besar daripada kemiripan antara data dengan kelompok data lainnya (Sadewo et al., 2018).

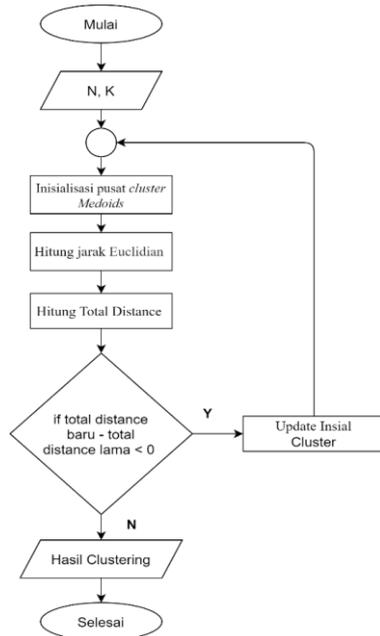
*Clustering* juga di sebut sebagai pengelompokkan, dapat dipertimbangkan sebagai masalah kritis maksimum dari pembelajaran tanpa pengawasan. Jadi, seperti masalah yang berbeda dari jenis ini, ini terkait dengan menemukan bentuk set data yang tidak berlabell (Jain dan Ambassador, 1988; Jain et al., 1999). Definisi bebas dari *Clustering* mungkin ingin menjadi "proses pengorganisasian objek menjadi kelompok yang anggota serupa dalam beberapa hal". Maka dari itu sekelompok objek yang "sebanding" di antara yang lain dan "berbeda" dari objek milik pengelompokkan yang berbeda (Velmurugan & Santhanam, 2010).

### *K-Medoids*

Algoritma *K-Medoids*, atau dikenal juga sebagai PAM (*Partitioning Around Medoid*),

dikembangkan melalui cara Leonard Kaufman dan Peter J. Rousseeu. Mirip dengan algoritma *K-Means* di mana kedua algoritma dipisahkan dan kumpulan data dibagi menjadi beberapa kelompok. Algoritma *K-Medoids* adalah untuk menentukan pusat *cluster*. Algoritma *K-Medoids* menggunakan nilai mean masing-masing kelompok sebagai titik fokus *cluster*, Algoritma *K-Medoids* juga menggunakan informasi objek sebagai medoid sebagai titik pusat *cluster*. Algoritma *K-Medoids* digunakan untuk mengatasi kekurangan perhitungan kmedian yang rentan pada *outlier* karena objek jauh dari sebagian besar atribut dalam data. Oleh karena itu, jika mereka berada dalam kumpulan data, mereka dapat mencondongkan rata-rata. Rata-rata *cluster* (Wira et al., 2019).

*K-Medoids* bisa menjadi strategi pembagian karena menggunakan objek yang dikumpulkan (*medoids*) pada *cluster* untuk membentuk titik pusat *cluster* berdasarkan hasil rata-rata dari objek dalam set *cluster*. Adapun alur dalam melakukan proses *clustering* data menggunakan algoritma *K-Medoids* dapat dilihat pada Gambar 3 (Sindi et al., 2020).



Gambar 3. Proses *Clustering K-Medoids*

Keterangan :

1. Inisialisasi sebagai banyak pusat *cluster* K (jumlah *cluster*).
2. Gunakan persamaan *Euclidean Distance* untuk menetapkan pada tiap data (Objek) pada *cluster* yang dekat. Contoh persamaan sebaga berikut :

$$d_{ij} = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2} = \sqrt{(X_i - X_j)'(X_i - X_j)} \quad (1)$$

dimana  $i=1, \dots, n$ ;  $j=1, \dots, n$  dan  $p$  adalah banyak variable, serta  $V$  adalah matrik varian kovarian.

3. Memilih objek secara random pada tiap *cluster* sebagai wakil medoid terbaru.
4. Gunakan wakil medoid baru untuk menghitung jarak tiap objek di tiap anggota *cluster*.
5. Deviasi total (S) dihitung dengan menghitung total jarak baru dari total jarak sebelumnya. Jika  $S < 0$ , maka objek tersebut ditukar dengan data *cluster* untuk membentuk himpunan baru k objek sebagai medoid.
6. Mengulangi langkah 3 hingga 5 sampai tidak mengubah medoid untuk mendapatkan sekelompok *cluster* dan kelompok *cluster* yang sesuai.

Algoritma *K-Medoids* secara komputasi lebih rentan dibandingkan dengan *K-Means* karena perhitungan medoid berdasarkan frekuensi yang telah terjadi. *K-Medoids* memiliki karakteristik yang berpotensi penting, yang pusatnya termasuk dalam data yang diindeks sendiri. Algoritma *K-Medoids* baru diusulkan, yang bekerja mirip dengan algoritma *K-Means*, dan metode yang berbeda diuji untuk memilih sumber medoid (Arbin et al., 2016).

### Rapid Miner

Rapid Miner dapat menjadi solusi pemeriksaan penambangan informasi, penambangan teks, dan analitik terkini. Rapid Miner menggunakan teknik deskriptif dan prediktif untuk menyajikan informasi kepada pemakai sehingga mereka dapat menarik kesimpulan tebaik. Rapid Miner memiliki sekitar 500 operator penambangan informasi, termasuk *inputs*, *Reasoning instructions*, data *preprocessing*, and *display instructions*. Rapid Miner dikembangkan menggunakan bahasa pemograman Java, sehingga dapat berjalan disemua versi perangkat lunak.

### III. HASIL DAN PEMBAHASAN

#### Dataset

Data yang digunakan pada penelitian merupakan data kasus riwayat Covid-19. Jumlah Data status terinfeksi pada tanggal 20 Mei 2021 adalah 842.708 kasus, yang terdiri dari 269 Wilayah, Status Positif 421.354 kasus, Dirawat 4.674 kasus, Isolasi 3.797 kasus, Meninggal 7.169 kasus dan Sembuh 405.714 kasus. Data tersebut akan di normalisasi sebelum dilakukan proses *clustering*. Agar lebih jelas data dapat di lihat pada Tabel 1.

Tabel 1. Contoh Dataset Covid-19 Provinsi DKI Jakarta

| No. | ID_KEL     | kelurahan   | Positif | Dirawat | Sembuh | Meninggal | Isolasi |
|-----|------------|-------------|---------|---------|--------|-----------|---------|
| 1   | 3172051003 | Ancol       | 986     | 13      | 951    | 12        | 10      |
| 2   | 3173041007 | Angke       | 789     | 8       | 759    | 14        | 8       |
| 3   | 3175041005 | Bal Kambang | 1011    | 4       | 979    | 20        | 8       |
| 4   | 3175031003 | Bali Mester | 397     | 0       | 386    | 8         | 3       |
| 5   | 3175101006 | Bambu Apus  | 1141    | 23      | 23     | 1089      | 17      |

**Pengolahan Data**

Pengolahan data ini dilakukan dengan menghilangkan data dan *noise* yang tidak konsisten, menyalin data, mengoreksi kesalahan data, dan dapat dilengkapi dengan data eksternal yang sesuai. *Preprocessing* ini digunakan untuk *Clustering* data dan memberikan hasil yang sama dengan tujuan kajian penelitian. Adapun hal-hal yang dilaksanakan dalam proses pengolahan data kajian penelitian antara lain:

a) Penentuan Atribut

Penentuan Atribut, dalam data penelitian terdapat banyak atribut. Agar didapatkan hasil *cluster* yang sesuai dengan tujuan penelitian maka dilakukan penentuan atribut yang penting. Adapun atribut yang terdapat pada data penelitian adalah sebagai berikut.

Tabel 2. Atribut data penelitian

| ATRIBUT        | INISIALISASI |
|----------------|--------------|
| Kelurahan      | X1           |
| Positif        | X2           |
| Dirawat        | X3           |
| Sembuh         | X4           |
| Meninggal      | X5           |
| Self Isolation | X6           |

b) Normalisasi Data

Proses normalisasi disebut juga dengan proses binning, yaitu proses transformasi untuk merubah nilai data. Normalisasi digunakan untuk menyamakan skala atribut data kedalam *range* yang spesifik yang lebih kecil seperti -1 sampai 1 atau 0 sampai 1. *Min-Max Normalization* merupakan teknik normalisasi dengan melakukan transformasi linear pada atribut data asli untuk menghasilkan range nilai yang sama. *Min-Max Normalization* memetakan sebuah *value* V dari atribut A menjadi *v'* kedalam range [new\_minA, new\_MaxA] dengan persamaan berikut ini.

$$V^1 = \frac{V - \min_a}{\max_a - \min_a} (new\_max_A = new\_min_A) + new\_min_A$$

Keterangan:

- v'* = Hasil yang normalisasi
- v* = Yang akan dinormalisasi
- min<sub>a</sub>* = Nilai terendah (minimal)
- max<sub>a</sub>* = Nilai tertinggi (maksimal)
- new\_min<sub>A</sub>* = Nilai minimal baru, 0
- new\_max<sub>A</sub>* = Nilai maksimal baru, 1

Tabel 3. Hasil Normalisasi

| Kelurahan    | New Positif | New Dirawat | New Sembuh | New Wafat | New Isolasi |
|--------------|-------------|-------------|------------|-----------|-------------|
| Ancol        | 0.022       | 0.029       | 0.021      | 0.019     | 0.028       |
| Angke        | 0.017       | 0.018       | 0.017      | 0.023     | 0.022       |
| Bale Kambang | 0.022       | 0.009       | 0.022      | 0.032     | 0.022       |

c) Transformasi Data

Mengubah data yang terpilih, sehingga data sesuai untuk melakukan proses penambangan data. Transformasi data dengan mengubah data menjadi format CSV agar bisa dibaca oleh *software* yang digunakan.

**Penerapan Algoritma K-Medoids**

Pada tahapan ini pengolahan data kasus terinfeksi Covid-19 menurut jenis kasus di Provinsi DKI Jakarta menggunakan algoritma *K-Medoids*. Langkah-langkah dalam menyelesaikan perhitungan manual data mining menggunakan *K-Medoids* clustering sebagai berikut.

- a) Inisialisasi pusat *cluster* sebanyak 3 dari jumlah data yang di gunakan. Medoid yang telah dipilih secara acak di asumsikan kelurahan Roa Malaka, Gunung Sahari Utara dan kelurahan

Ceger. Contoh inialisasi pusat *cluster* disajikan pada tabel berikut.

Tabel 4. Contoh inialisasi pusat *cluster*

| Inisial (C=kluster) | Positif | Dirawat | Sembuh | Wafat | Isolasi |
|---------------------|---------|---------|--------|-------|---------|
| BALE KAMBAN G (C1)  | 0.022   | 0.009   | 0.022  | 0.032 | 0.022   |
| CIKOKO (C2)         | 0.011   | 0.013   | 0.011  | 0.011 | 0.006   |
| DUREN SAWIT (C3)    | 0.063   | 0.068   | 0.062  | 0.088 | 0.064   |

- b) Menentukan jarak terdekat menggunakan persamaan *Euclidean Distance*. Sebagai contoh perhitungan Ancol ke titik pusat *cluster* pertama dengan persamaan (1) seperti yang di sajikan di bawah ini :

$$d(1,C1) = \sqrt{(0.022 - 0.022)^2 + (0.009 - 0.029)^2} + \sqrt{(0.022 - 0.021)^2 + (0.032 - 0.019)^2} + \sqrt{(0.022 - 0.028)^2} = 0.24$$

$$d(1,C2) = \sqrt{(0.011 - 0.022)^2 + (0.013 - 0.029)^2} + \sqrt{(0.011 - 0.021)^2 + (0.001 - 0.019)^2} + \sqrt{(0.006 - 0.028)^2} = 0.032$$

$$d(1,C3) = \sqrt{(0.036 - 0.022)^2 + (0.068 - 0.029)^2} + \sqrt{(0.062 - 0.021)^2 + (0.088 - 0.019)^2} + \sqrt{(0.064 - 0.028)^2} = 0.104$$

Berdasarkan perhitungan persamaan yang telah di lakukan, maka diperoleh jarak yang terdekat adalah *cluster* 2, sehingga kelurahan Ancol dikategorikan dalam *cluster* kedua. Hasil perhitungan pada iterasi pertama dapat dilihat pada tabel berikut

Tabel 5. Hasil perhitungan jarak terdekat.

| Id                   | Jarak Ke Medoid |       |       | Terdekat      | Cluster |
|----------------------|-----------------|-------|-------|---------------|---------|
|                      | C1              | C2    | C3    |               |         |
| 1                    | 0.024           | 0.032 | 0.104 | 0.024         | 1       |
| 2                    | 0.015           | 0.022 | 0.112 | 0.015         | 1       |
| 3                    | 0.000           | 0.031 | 0.107 | 0.000         | 1       |
| .....                |                 |       |       |               |         |
| Total Jarak Terdekat |                 |       |       | <b>11.985</b> |         |

- c) Setelah proses perhitungan pada literasi satu maka di dilanjutkan dengan perhitungan medoid baru atau *non medoids*. Objek yang akan

digunakan sebagai medoid baru (*non mdoids*) adalah data yang di pilih secara acak yakni kelurahan Tugu Selatan, Jati Padang dan Pasar Minggu disajikan pada tabel berikut :

Tabel 6. Tabel medoid baru (*Non Medoids*)

| inisial (C=kluster) | Positif | Dirawat | Sembuh | Wafat | Isolasi |
|---------------------|---------|---------|--------|-------|---------|
| Tugu Selatan (C1)   | 0.025   | 0.011   | 0.025  | 0.032 | 0.011   |
| Jati Padang (C2)    | 0.038   | 0.029   | 0.038  | 0.055 | 0.030   |
| Pasar Minggu (C3)   | 0.033   | 0.029   | 0.033  | 0.052 | 0.022   |

- d) Setelah data objek sudah ditentukan, dilanjutkan ltersi kedua yaitu perhitungan *non medoids* atau medoid baru dengan persamaan yang telah di gunakan pada langkah sebelumnya. tabel dibawah ini diperlihatkan hasil perhitungan jarak pada iterasi kedua.

Tabel 7. Hasil perhitungan iterasi kedua

| Id                    | Jarak Ke Medoid |       |       | Terdekat      | Cluster |
|-----------------------|-----------------|-------|-------|---------------|---------|
|                       | C1              | C2    | C3    |               |         |
| 1                     | 0.024           | 0.032 | 0.104 | 0.024         | 1       |
| 2                     | 0.015           | 0.022 | 0.112 | 0.015         | 1       |
| 3                     | 0.000           | 0.031 | 0.107 | 0.000         | 1       |
| .....                 |                 |       |       |               |         |
| 269                   | 0.043           | 0.055 | 0.083 | 0.043         | 1       |
| Jumlah Jarat Terdekat |                 |       |       | <b>11.882</b> |         |

- e) Menghitung total jarak baru dari total jarak sebelumnya. Setelah diperoleh hasil perhitungan jarak antara iterasi pertama dan kedua, di lanjutkan penjumlahan total simpangan dengan mencari selisi pada total *cost* baru-nilai *cost* lama. Dengan syarat apabila  $S < 0$ , maka gantikan nilai objek dengan medoid baru. Berikut ini merupakan cara perhitungannya:

$$S = \text{Total Baru} - \text{Total Lama} = 11.985 - 11.882 = 0.103$$

- f) Karena  $S < 0$  maka di lanjutkan ke iterasi ketiga hingga  $S > 0$  dengan *non medoids* pada iterasi kedua. Objek yang di pilih secara acak untuk iterasi ketiga dapat dilihat pada Tabel 8.

Setelah objek ditentukan maka di lakukan perhitungan menggunakan persamaan yang telah di gunakan sebelumnya. Hasil

perhitungan iterasi ketiga dapat dilihat pada Tabel 9.

Tabel 8. Medoid iterasi ketiga

| inisial (c=klu ster) | Positif | Dirawat | Sembuh | Wafat | Isolasi |
|----------------------|---------|---------|--------|-------|---------|
| Suntr Agung          | 0.068   | 0.106   | 0.066  | 0.100 | 0.064   |
| Kayu Putih           | 0.037   | 0.110   | 0.035  | 0.049 | 0.033   |
| Cilanda k Barat      | 0.059   | 0.066   | 0.058  | 0.058 | 0.097   |

Tabel 9. Hasil perhitungan iterasi ketiga

| Id                    | Jarak Ke Medoid |       |       | Terdekat      | Cluster |
|-----------------------|-----------------|-------|-------|---------------|---------|
|                       | C1              | C2    | C3    |               |         |
| 1                     | 0.134           | 0.089 | 0.102 | 0.089         | 2       |
| 2                     | 0.143           | 0.101 | 0.113 | 0.101         | 2       |
| 3                     | 0.141           | 0.105 | 0.111 | 0.105         | 2       |
| .....                 |                 |       |       |               |         |
| 269                   | 0.108           | 0.065 | 0.094 | 0.065         | 2       |
| Jumlah Jarat Terdekat |                 |       |       | <b>24.928</b> |         |

Berdasarkan tabel yang di sajikan di atas jumlah total jarak terdekat dari seluruh kluster diperoleh 24.928 kemudian akan di jumlah dengan jumlah jarak terdekat pada literasi .

$$\begin{aligned}
 S &= \text{Total Baru} - \text{Total Lama} \\
 &= 24.928 - 11.882 \\
 &= 12,046
 \end{aligned}$$

Karena hasil selisih yang diperoleh lebih dari 0, maka tidak ada lagi perubahan medoid pada iterasi ketiga dan proses klastering di berhentikan. Kemudian telah di hasilkan total jarak minimum 12,046 pada iterasi ketiga. Hasil klastering dapat dilihat pada tabel berikut :

Tabel 10. Hasil klastering iterasi terakhir

| Cluster 0        | Cluster 1     | Cluster 2    |
|------------------|---------------|--------------|
| Cengkareng Barat | Ancol         | Bangka       |
| Cengkareng Timur | Angke         | Baru         |
| Ciracas          | Babel Kambang | Batu Ampar   |
| Duren Sawir      | Bali Master   | Bidadar Cina |
| .....            |               |              |

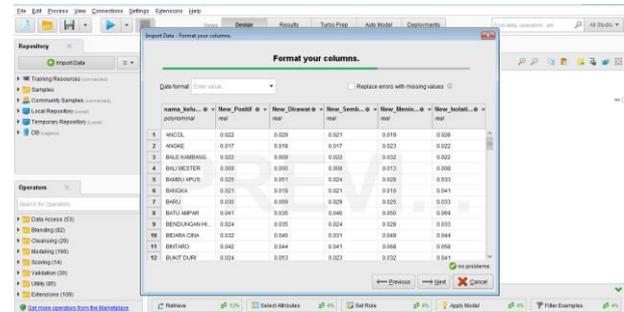
### Implementasi pada Rapid Miner

Setelah di lakukan perhitungan manual dilanjutkan dengan ipelentasi pada aplikasi Rapid Miner yang bertujuan untuk mlakukan validasi apakah data yang digunakan sama dengan

perhitungan aplikasi Rapid Miner dengan perhitungan manual. Berikut ini disajikan langkah-langkah penerapan pada aplikasi Rapid Miner.

#### 1) Import data

Pertama kita yang harus kita adalah import data yang sudah di lakukan normalisasi pada tahapan sebelumnya agar bisa diolah pada aplikasi Rapid Miner.

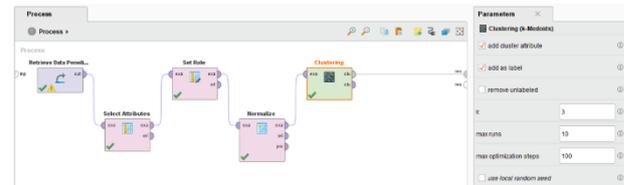


Gambar 4. Import data

Setelah di import data tersebut disimpan dalam Local repository aplikasi Rapid Miner.

#### 2) Membuat Simulasi

Setelah melakukan proses import data, selanjutnya adalah membuat proses clustering data. Pada bagian operator Clustering K-Medoids



parameter K=3 di tentukan jumlah cluster sesuai perhitungan manual yang sudah di lakukan pada tahap sebelumnya. dengan rancangan teknik clustering seperti berikut.

Gambar 5. Desain proses Clustering

#### 3) Hasil Cluster dengan Rapid Miner

Dari proses pengujian desain dengan cluster berjumlah 3, didapatkan anggota masing-masing cluster sebagai berikut.

#### Cluster Model

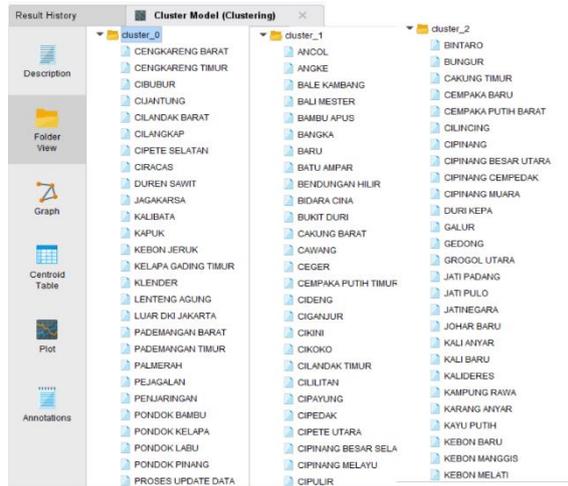
```

Cluster 0: 31 items
Cluster 1: 172 items
Cluster 2: 66 items
Total number of items: 269
    
```

Gambar 6. Hasil Cluster

Berdasarkan hasil tersebut menunjukkan bahwa pada cluster pertama atau 0 didapatkan anggota cluster berjumlah 31 kelurahan. Pada cluster kedua atau cluster 1 didapatkan jumlah anggota cluster sebanyak 177 cluster dan pada cluster ketiga atau cluster 2 didapatkan jumlah anggota cluster 66 kelurahan. Adapun masing-masing anggota pada

setiap *cluster* ditunjukkan pada Gambar 7 berikut ini. Berikut ini merupakan isi pada masing-masing *cluster*.



Gambar 7. Isi tiap *cluster*

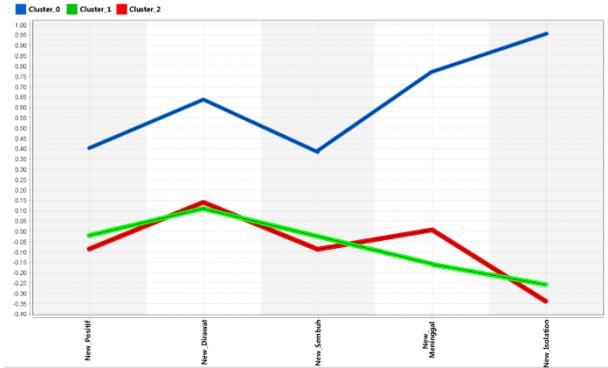
Pada Tabel 10 yang telah di sajikan nilai centroid pada masing-masing *cluster* berdasarkan kasus. Seperti yang terlihat pada Tabel 10 *Cluster 0* atau pertama memiliki kasus terinfeksi covid-19 dengan centroid/rata-rata paling tinggi di antara kasus lain, dengan jenis status yang menonjol yaitu status positif, dirawat, isolasi meninggal. Akan tetapi dalam klaster ini status isolasi memiliki centroid/rata-rata paling rendah diantara status lain.

Tabel 11. Centroid setiap *cluster*

| Kasus           | Cluster 1 | Cluster 2 | Cluster 3 |
|-----------------|-----------|-----------|-----------|
| New -Sembuh     | 0.285     | -0.027    | -0.086    |
| New_Positif     | 0.389     | -0.020    | -0.085    |
| New_Dirawat     | 0.635     | 0.112     | 0.142     |
| Nw_Mninggal     | 0.772     | -0.159    | 0.008     |
| New_Isolasi     | 0.958     | 0.256     | -0.34     |
| Standar Deviasi | 0.274     | 0.157     | 0.176     |

Dari hasil pengelompokan yang di peroleh dari aplikasi Rapid Miner, mula dari anggota *cluster* terkecil, menengah sampai *cluster* dengan jumlah tertinggi. Adapun perbandingan hasil *cluster* berdasarkan atribut pada data penelitian adalah sebagai berikut:

Seperti yang terlihat pada Gambar 8, *Cluster 0* memiliki jumlah kasus paling tinggi diantara kasus *cluster* lain yang terdiri dari 36 kelurahan. Sedangkan *Cluster 2* memiliki jumlah kasus paling rendah di antara kasus pada *cluster* lain yang terdiri dari 66 kelurahan.



Gambar 8. Grafik perbandingan *Cluster*

#### IV. Kesimpulan

Berdasarkan kajian penelitian yang telah di lakukan dapat diambil kesimpulan bahwa pada teknik data mining dalam pengelompokan kasus terinfeksi Covid-19 berdasarkan kelurahan di Provinsi DKI Jakarta menggunakan algoritma K-Medoids dengan perhitungan manual dan perhitungan mnggunakan Rapid Miner memiliki hasil yang valid dengan jumlah tiga *cluster* antara lain *Cluster 0*, *Cluster 1* dan *Cluster 2*. Diketahui kasus paling tinggi yakni *Cluster 0* yang terdiri dari 31 kelurahan sedangkan *cluster* paling rendah diketahui *Cluster 2* yang terdiri 66 kelurahan. Berdasarkan hasil *cluster* tersebut pemerintah dapat melakukan penanganan berdasarkan kasus yang telah di kelompokkan terutamanya pada *cluster* yang memiliki kasus yang paling tinggi.

#### V. Daftar Pustaka

Agus, P. (2017). Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method. *International Journal Of Artificial Intelligence Research*, 1(2), 26–33.

Arbin, N., Suhaimi, N. S., Mokhtar, N. Z., & Othman, Z. (2016). Comparative analysis between k-means and k-medoids for statistical clustering. *Proceedings - AIMS 2015, 3rd International Conference on Artificial Intelligence, Modelling and Simulation*, 117–121. <https://doi.org/10.1109/AIMS.2015.82>

Bu’ulolo, E., & Purba, B. (2021). Algoritma Clustering Untuk Membentuk Cluster Zona Penyebaran Covid-19. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 12(1), 59–67. <https://doi.org/10.31849/digitalzone.v12i1.6572>

Darmi, Y. D., & Setiawan, A. (2016). Penerapan Metode Clustering K-Means Dalam

- Pengelompokan Penjualan Produk. *Jurnal Media Infotama*, 12(2).  
<https://doi.org/10.37676/jmi.v12i2.418>
- Indraputra, R. A., & Fitriana, R. (2020). *K-Means Clustering Data COVID-19*. 10(3), 275–282.
- Nasari, F., & Sianturi, C. J. M. (2016). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat. *CogITO Smart Journal*, 2(2), 108.  
<https://doi.org/10.31154/cogito.v2i2.19.108-119>
- Sadewo, M. G., Windarto, A. P., & Wanto, A. (2018). Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means. *KOMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, 2(1).  
<https://doi.org/10.30865/komik.v2i1.943>
- Sindi, S., Ningse, W. R. O., Sihombing, I. A., Ilmi R.H.Zer, F., & Hartama, D. (2020). Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia. *Jti (Jurnal Teknologi Informasi)*, 4(1), 166–173.
- Untoro, Meida Cahyo, A. (2021). Penerapan metode k-means clustering data COVID-19 di Provinsi Jakarta. *Teknologi: Jurnal Ilmiah Sistem Informasi* 11, 11(2), 59–68.  
<https://doi.org/10.26594/teknologi.v11i2.2323>
- Velmurugan, T., & Santhanam, T. (2010). Computational complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points. *Journal of Computer Science*, 6(3), 363–368.  
<https://doi.org/10.3844/jcssp.2010.363.368>
- Virgantari, F., & Faridhan, Y. E. (2020). *K-Means Clustering of COVID-19 Cases in Indonesia's Provinces*. 5(2), 1–7.
- Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. *RAINSTEK : Jurnal Terapan Sains & Teknologi*, 1(3), 53–68. <https://doi.org/10.21067/jtst.v1i3.3046>
- Yunus NR, & A, R. (2020). Kebijakan Pemberlakuan Lock Down Sebagai Antisipasi Penyebaran Corona Virus Covid-19. *Salam: Jurnal Sosial Dan Budaya Syar-I*, 7(3), 227–238.  
<http://journal.uinjkt.ac.id/index.php/salam/article/view/15083>
- Yustianti, W., Rahmawati, N., & Yamasari, Y. (2021). Klastering Wilayah Kota/Kabupaten Berdasarkan Data Persebaran Covid-19 di Propinsi Jawa Timur dengan Metode K-Means. *Journal Information Engineering and Educational Technology*, 04(01), 20.