

# Implementasi Metode K-Medoids Untuk Mengumpulkan Data Pasien SARSCoV2

Enggar Wijaya Putra<sup>1\*</sup>, Praba Milantino A.R.K<sup>2</sup>, Maskhul Ryan Ibrahim<sup>3</sup>, Dwi Hartanti<sup>4</sup>

<sup>1,2,3,4</sup>Sistem Informasi

Univesitas Duta Bangsa Surakarta

<sup>1</sup>\*202030350@mhs.udb.ac.id, <sup>2</sup>202030240@mhs.udb.ac.id, <sup>3</sup>202020088@mhs.udb.ac.id, <sup>4</sup>dwhartanti@udb.ac.id

**Abstrak**—Wuhan—ibu kota Provinsi Hubei, China—virus Covid-19 pertama kali diketahui muncul pada tahun 2019. Penyakit corona virus, juga dikenal sebagai SARSCoV2 adalah penyakit sangat menular yang ditandai pertama gejala hidung tersumbat, tidak bisa merasakan, dan juga panas di seluruh badan. Korban COVID-19 biasanya mengalami kelelahan, demam, dan batuk kering. Pengelompokan menggunakan metode K-medoids untuk mengelompokkan Pasien Penyebaran Covid-19 di Provinsi DKI Jakarta. Pengidentifikasi ini berisi jumlah rencana Cluster yang diproses dengan metode K-Medoids. Pengenal Persyaratan antara lain atribut table dan juga data-data pasien SARSCoV2 diproses dengan metode K-Medoids. dan juga cara menggunakan Rapidminer untuk melakukan Clustering Pada aplikasi rapidminer kita import data berupa spreetsheet, lalu lalu berlanjut ke proses seperti gambar 3 dan melakukan pencocokan atribut data, data yang di masukan ke dalam rapidminer semua berupa numerik Setelah melakukan preprocessing data kita melakukan pengolahan dengan metode K-Medoids yang tersedia di tab desain Berdasarkan hasil pengujian data SARSCoV2 yang dilakukan di daerah DKI Jakarta, Metode K-Medoids yang digunakan dengan menggunakan Rapidminer berhasil menghasilkan dua cluster. Proses ini dapat diperpanjang dengan metode pengelompokan lainnya agar meningkatkan akurasi penelitian dengan menambahkan informasi dan fitur yang lebih mudah.

**Kata kunci**— Data Mining, Cluster, K-Medoids, Covid-19, Pemerintah DKI Jakarta

**Abstract**— Wuhan – the capital city of Hubei Province, China – the Covid-19 virus was first known to appear in 2019. Coronavirus disease, also known as SARSCoV2, is a highly contagious disease characterized by the first symptoms of a congested nose, can not feel, and also heat throughout the body. COVID-19 victims typically experience fatigue, fever, and dry cough grouping using K-medoids methods to group Covid-19 Spread Patients in Jakarta DKI Province. This identifier contains the number of Cluster plans processed using the K-Medoids method. Identifying Requirements among other attributes table and also patient data SARSCoV2 is processed with the method of K-Medoids. and also how to use Rapidminer to do Clustering On the application rapidminer we import data in the form of spreetsheet, then continue to the process like image 3 and do matching attribute data, data that is input into the quickminer all in numerical form After preprocessing data we do processing with K-Medoids method available in the design tab Based on the results of data testing SARSCOV2 carried out in the area of DKI Jakarta, the K-medoids method used using RapidMiner succeeded in producing two clusters. This process can be extended with other grouping methods to improve research accuracy by adding information and features that are easier to use.

**Keywords**— Data Mining, Cluster, K-Medoids, Covid-19, DKI Jakarta Government

## I. PENDAHULUAN

Wuhan—ibu kota Provinsi Hubei, China— virus Covid-19 pertama kali diketahui muncul pada tahun 2019. Penyakit corona virus, juga dikenal sebagai SARSCoV2 adalah penyakit sangat menular yang ditandai pertama gejala hidung tersumbat, tidak bisa merasakan, dan juga panas di seluruh badan. Korban COVID-19 biasanya mengalami kelelahan, demam, dan batuk kering. Beberapa pasien menunjukkan gejala yang lebih ringan, seperti nyeri di badan, sakit kepala berlebihan, tenggorokan terasa aneh, hilangnya indra perasa dan indra penciuman, dan perubahan warna pada semua jari. rentan suhu ruangan 26-27 derajat Celcius. Penyebaran terjadi melalui droplet, yang sudah terkena virus yang menyebabkan penyakit, yang dilepaskan oleh pasien saat kita bicara dan batuk maka dari itu kita saat sakit

batuk sebisa mungkin menggunakan masker agar tidak mengeluarkan virus ke udara dan menulari orang terdekat kita [1].

Pada Senin, awal Maret 2020, virus ini masuk ke Indonesia dan sejak saat itu pemerintah mulai memberitahu agar memakai masker saat keluar agar tidak terjangkit virus tersebut, pemulihan, dan kematian. Per 15 September 2021, jumlah kasus di Indonesia kira-kira ada 4,18 juta dan 140 ribu orang meninggal karena virus ini. Kasus ini tersebar di semua provinsi Indonesia. Pada pertengahan Maret 2020, Riau adalah salah satu dari sepuluh provinsi dengan sebagian besar ada di Kota-kota di Riau terjangkit virus COVID tertinggi adalah Dumai dan pekanbaru, masing-masing dengan 10.189. Virus COVID-19 [2].

Menurut laporan dari 34 daerah provinsi di Indonesia, daerah DKI Jakarta mempunyai kasus

tertinggi sebesar 25% ,Jawa Timur sebesar 11%, dan di Jawa Barat memiliki 9%. [3]. Metode klusterisasi parsial digunakan oleh algoritma K-Medoids dalam pengolahannya. Untuk menemukan kesesuaian klaster, yang memungkinkannya untuk menghasilkan klusterisasi yang efektif.[4]. Klusterisasi, juga dikenal sebagai clustering, adalah alat tambahan untuk pengolahan data. Tujuan clustering adalah metode dengan cara mengelompokkan objek ke satu atau lebih cluster hingga semua objek sangat mirip satu sama lain[5].

## II. LANDASAN TEORI

### A. Data Mining

Data mining adalah pengumpulan dan pengolahan yang dilakukan untuk mengekstrak informasi penting dari data. Proses ini dapat dilakukan dengan menggunakan perangkat lunak yang memanfaatkan perhitungan matematika, statistika, atau AI[4]. Tahapan antara lain :

#### 1. Cleaning and Integration

Pembersihan data, juga dikenal sebagai "Pembersihan data yaitu proses yang dilakukan untuk menghapus data yang tidak konsisten dan tidak relevan dengan data. Menggabungkan objek data dari beberapa database ke satu database baru bisa di sebut integrasi data[6].

#### 2. Selection and Transformation

Karena di database seringkali tidak digunakan sepenuhnya, maka diperlukan hanyalah data yang tepat untuk diproses dari database. Data di ganti ke dalam bentuk mining data yang disebut sebagai transformasi data mining[6].

#### 3. Data Mining

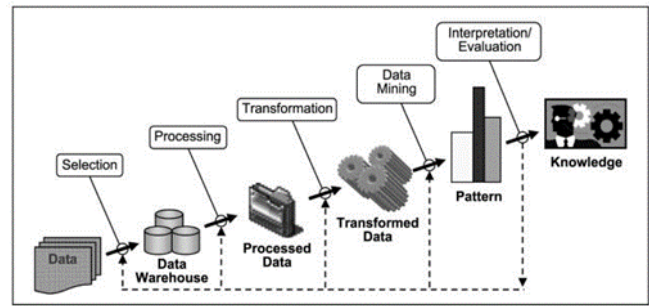
Metode yang untuk mengetahui informasi penting dan rahasia dari sebuah data[6].

#### 4. Evaluation and Presentation

Evaluasi pola data mining yang tepat dan menunjukkan apakah output yang ada telah dipenuhi[6].

#### 5. Knowledge

Pengetahuan, yang merupakan komponen terakhir dari penambahan data, yang diperoleh melalui berbagai proses penambahan data[6].



Gambar 1. Knowledge Discovery In Database (KDD)

### B. Metode K-Medoids

PAM (Partitioning Around Medoid) adalah metode yang di gunakan oleh kelompok atau medoid. metode K-Medoids lebih baik daripada K- Means karena mereka menggunakan jumlah jarak geometris pada objek data, sedangkan K-Medoids menggunakan objek K untuk mengetahui ketidak samaan objek data[9].

### C. Clustering

Clustering digunakan untuk menemukan pola distribusi dalam sebuah set data. Sementara Kebanyakan orang menggambarkan objek data sebagai titik dalam ruang multidimensi, dan persamaan objek biasanya diketahui dari nilai kedekatan atribut data yang di ketahui dari objek-objek data itu sendiri[7]. Pada penelitian ini, berdasarkan kondisi dataset dan tujuan clustering untuk identifikasi penyebaran virus[8], maka tahapan penelitian antara lain :

1. Persiapan Pembersihan dan Integrasi yaitu bertujuan memisahkan data yang tidak konsisten. Selanjutnya Seleksi dan Transformasi data di ubah menjadi bentuk numerik agar proses mining bisa di lakukan.
2. Proses Mining, proses ekstraksi pola dari data yang ada.
3. Eksperimen beberapa kombinasi atribut data yang digunakan.
4. Presentasi pengetahuan dengan teknik visualisasi.
5. Proses interpretasi pola yang dihasilkan dan visualisasnyai dengan Teknik filtering dan seleksi untuk identifikasi pola penyebaran virus sehingga diperoleh pengetahuan. Pada bagian ini membahas metode yang dilaksanakan dalam penelitian dimulai dari Persiapan Data hingga Interpretasi dan Visualisasi[8].

### III. METODOLOGI PENELITIAN

#### A. Metode K-Medoids

K-Medoids termasuk dalam kategori clustering partitioning, yang memungkinkan data dengan nilai sangat besar di luar distribusi data. Metode yang digunakan dalam K-Medoids antara lain :

1. Cari nilai K untuk jumlah cluster yang diinginkan. Menggunakan indikator (C1) dan (C2).
2. Hitung jarak masing- masing objek ke medoid sementara dengan rumus Euclidean Distance:

$$d = \sqrt{(x1 - x2)^2 + (y1 - y2)^2}$$

Keterangan :

d = Jarak

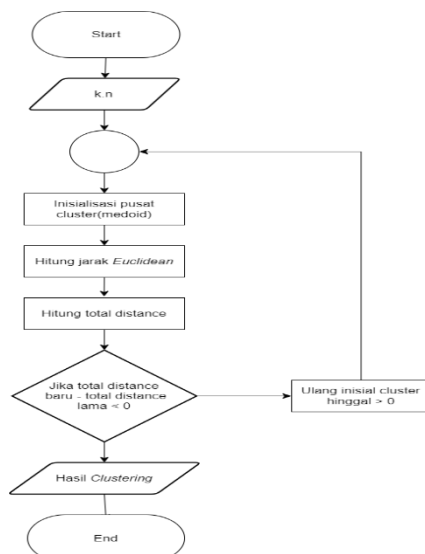
$\chi_1$  = latitude 1

$\chi_2$  = latitude 2

$\lambda_1$  =Longitude 1

$\lambda_2$  = Longitude 2

3. Pilih objek
4. Menghitung dengan medoid baru
5. Menghitung total kedekatan (S) yaitu mencari selisih jarak medoid baru dengan total jarak medoid lama. Jika  $S < 0$ , maka mencari medoid baru lagi hingga mendapatkan selisih jarak antara medoid baru dan lama  $> 0$ .
6. Ulangi pemilihan medoid hingga selisih kedekatan  $> 0$  media berhenti berganti untuk mendapatkan Maximize  $th = \sum P X p Q p V p$



Gambar 2. Alur Metode K-Medoids

#### B. Identifikasi Masalah

Pengelompokan menggunakan metode K-medoids untuk mengelompokkan Pasien Penyebaran Covid-19 di Provinsi DKI Jakarta. Pengidentifikasi ini berisi jumlah rencana Cluster yang diproses dengan metode K-Medoids. Pengenal Persyaratan antara lain atribut table dan juga data-data pasien SARSCoV2 diproses dengan metode K-Medoids. dan juga cara menggunakan Rapidminer untuk melakukan Clustering [4].

#### C. Pengumpulan Data

Distribusi pasien di Provinsi DKI Jakarta pada tahun 2021. Data Informasi tentang distribusi kasus diperlukan untuk data sampel untuk tujuan referensi Saat menerapkan metode K-Medoids.

#### D. PreProcessing Data

Untuk menyesuaikan data dengan kebutuhan mendapatkan Maximize  $th = \sum P X p Q p V p$ , preprocessing data biasanya dikerjakan sebelum proses data mining. Mengidentifikasi ID tabel yang akan digunakan untuk klasterisasi, yang dilakukan dengan mengganti atribut menjadi tipe numerik serta beberapa atribut yang akan digunakan untuk referensi untuk penyusunan hasil cluster, yang dilakukan dengan menggunakan tipe data integer untuk angka, adalah bagian dari preprocessing penelitian ini[4].

### IV. HASIL DAN PEMBAHASAN

#### A. Data Awal

Sumber dikumpulkan langsung dari website <https://data.jakarta.go.id/> pada Cluster ini terlebih dahulu dihitung berdasarkan data pasien provinsi DKI Jakarta.

Tabel 1. Cleaning Data

Id	Kategori	Sub Kategori	Jumlah
1	Suspek	Perawatan RS	2
2	Pelaku Perjalanan Probable	Isolasi di Rumah	40
3		Meninggal	9
4	Suspek	Selesai Isolasi	857
5	Suspek	Perawatan RS	4
6	Suspek	Isolasi di Rumah	101
7	Suspek	Meninggal	2
8	Pelaku Perjalanan	Selesai Isolasi	596

9	Pelaku Perjalanan	Selesai Isolasi	447
10	Pelaku Perjalanan	Selesai Isolasi	251
...	...	...	...
2000	Pelaku Perjalanan	Selesai Isolasi	590
2001	Pelaku Perjalanan	Selesai Isolasi	377

Tabel 2. Transformasi Data

Id	Kategori	Sub Kategori	Jumlah
1	1	1	2
2	2	2	40
3	3	3	9
4	1	4	857
5	1	1	4
6	1	2	101
7	1	3	2
8	2	4	596
9	2	4	447
10	2	4	251
...	...	...	...
2000	2	4	590
2001	2	4	377

**B. Centroid Data**

Dengan menggunakan metode K-Medoid, diambil pengacakan data saat dikumpulkan. Ini dilakukan asumsi bahwa Jumlah kelompokan yang diharapkan adalah dua., dan bahwa pengelompokan dibagi menjadi dua bagian (C1)&(C2) centroid, atau Nilai cluster ini diambil secara tidak sengaja. Dengan centroid awal yang telah ditentukan sebelumnya, kita melakukan proses menggunakan rumus jarak Euclidian Distance. C1 digunakan dalam skenario ini:[10]

Tabel 3. Data Covid Iterasi ke - 1

Id	Kategori	Sub Kategori	Jumlah
1	0.000	-0.500	0.001
2	0.000	-0.500	0.001

Tabel 4. Data Covid Iterasi ke - 2

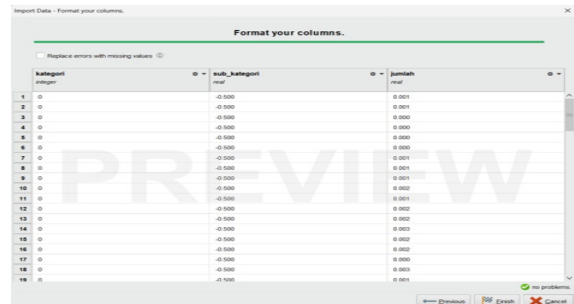
Id	Kategori	Sub Kategori	Jumlah
2000	1.000	1.000	0.000
2001	0.000	1.000	0.352

Tabel 5. Data Covid-19 Total Kedekatan 2 Percobaan Table 4 & 5

Percobaan Ke 1	Percobaan Ke 2	Total Kedekatan
284.27224	269.35152	14.92072

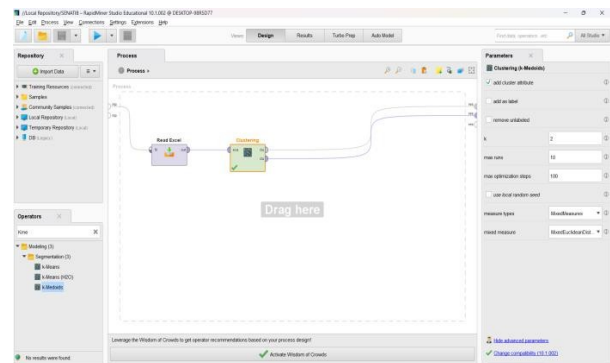
**C. Proses Analisa Pengujian**

Setelah data normalisasi di dapat maka selanjutnya melakukan Pengujian Algoritma ini menggunakan Aplikasi Rapidminer Studio versi 10.1.002 terdapat 3 sample data pada gambar 1.



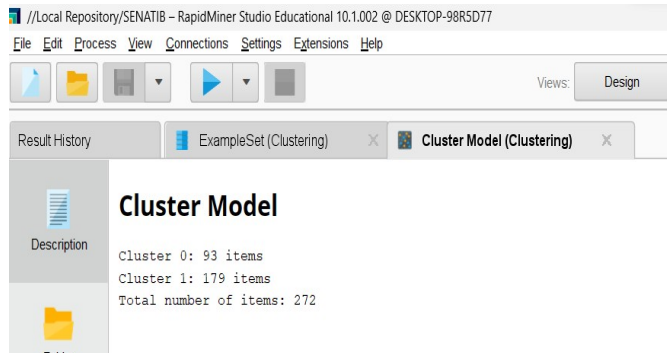
Gambar 1. Sample Data Kumulatif Covid-19 di Provinsi DKI Jakarta

Pada aplikasi rapidminer kita import data berupa spreetsheet, lalu lalu berlanjut ke proses seperti gambar 3 dan melakukan pencocokan atribut data, data yang di masukan ke dalam rapidminer semua berupa numerik. Setelah melakukan preprocessing data kita melakukan pengolahan dengan metode K-Medodis yang tersedia di tab desain.



Gambar 2. Algoritma K-Medoids untuk Clustering Data

Hasil dari pengolahan data ClusterModel menggunakan algoritma K-Medoids pada gambar 4



Gambar 5. Jumlah Data Hasil Cluster Model

Attribute	cluster_0	cluster_1
kategori	1	0
sub_kategori	1	1
jumlah	0	0.352

Gambar 6. Nilai Centroid

## V. KESIMPULAN

Berdasarkan hasil pengujian data SARSCoV2 yang dilakukan di daerah DKI Jakarta, Metode K-Medoids yang digunakan dengan menggunakan Rapidminer berhasil menghasilkan dua cluster. Proses ini dapat diperpanjang dengan metode

pengelompokan lainnya agar meningkatkan akurasi penelitian dengan menambahkan informasi dan fitur yang lebih mudah

## REFERENSI

- [1] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
- [2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
- [3] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, "A novel ultrathin elevated channel low-temperature poly-Si TFT," *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.
- [4] M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, "High resolution fiber distributed measurements with coherent OFDR," in *Proc. ECOC'00*, 2000, paper 11.3.4, p. 109.
- [5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, "High-speed digital-to-RF converter," U.S. Patent 5 668 842, Sept. 16, 1997.
- [6] (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [7] M. Shell. (2002) IEEETran homepage on CTAN. [Online]. Available: <http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEETran/>
- [8] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
- [9] "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.
- [10] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
- [11] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
- [12] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.