

# Penerapan Algoritma K-Means dan K-Medoid untuk Pengelompokan Data Pasien Covid-19

Umairah Rizky Gurning<sup>1</sup>, Mustakim<sup>2</sup>

<sup>1,2</sup> Puzzle Research Data Technology (Predatech), Faculty of Science and Technology

<sup>1,2</sup> Department of Information System, Faculty of Science and Technology

Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: <sup>1,\*</sup>umairahrizkygurning@gmail.com, <sup>2</sup>mustakim@uin-suska.ac.id

Submitted: 24/06/2021; Accepted: 29/06/2021; Published: 30/06/2021

**Abstrak**—Pada akhir 2019 di Wuhan, Cina telah ditemukan virus baru yaitu Corona Virus Disease 2019. Virus ini menyebabkan masalah kesehatan yang serius dan ditetapkan sebagai pandemi sejak 11 Maret 2020. Menyebabkan kematian dan telah merenggut ribuan nyawa. Virus ini menyebar sangat cepat dan dengan berbagai cara seperti kontak langsung dengan pasien, pelaku perjalanan, pemilik penyakit bawaan dan masih banyak penularan lainnya. Untuk mengurangi penyebaran virus ini, pemerintah telah melakukan berbagai cara seperti social distancing dan scrining atau swab dadakan di pusat-pusat keramaian. Dikarenakan banyaknya jenis penularan dari virus ini, penelitian dilakukan untuk mengelompokkan kasus Covid-19 di Kota Dumai. Harapannya, hasil penelitian ini dapat dijadikan sebagai gambaran pengelompokan data pasien Covid-19 dengan menerapkan K-Means dan K-Medoid sebagai algoritma pengelompokan berdasarkan jenis penularan, usia, jenis kelamin, paskes dan kecamatan. Berdasarkan dari penelitian ini, algoritma K-Means lebih optimal dibanding K-Medoid pada mengklasterisasi data pasien Covid-19 khususnya di Kota Dumai. Dibuktikan pada nilai terbaik DBI K-Means sebesar 0,139 dengan percobaan  $k = 4$ .

**Kata Kunci:** Klustering; Perbandingan; Covid-19; Davies Bouldin Index (DBI); K-Means; K-Medoid

**Abstract**—At the end of 2019 in Wuhan, China, a new virus was discovered, that is Corona Virus Disease 2019. This virus causes serious health problems and has been declared as pandemic since March 11, 2020. It has caused death and claimed thousands of lives. This virus spreads very quickly and in various ways such as direct contact with patients, travelers, owners of congenital diseases and many other transmissions. To suppress the spread of this virus, the government has carried out various ways such as social distancing and screening or impromptu swabs in crowded centers. Due to the many types of transmission from this virus, this research was conducted to classify Covid-19 cases in Dumai City. It is hoped that the results of this study can be used as an illustration of the grouping of Covid-19 patient data by applying K-Means and K-Medoid as a grouping algorithm based on the type of transmission, age, gender, health services and district. Based on this research, the K-Means algorithm is more optimal than K-Medoid in classifying Covid-19 patient data, especially in Dumai City. It is proven that the best DBI K-Means value is 0.139 with  $k = 4$ .

**Keywords:** Clustering; Comparison; Covid-19; Davies Bouldin Index (DBI); K-Means; K-Medoid

## 1. PENDAHULUAN

Akhir 2019 di Wuhan, Cina telah ditemukan virus baru yaitu Corona Virus Disease 2019 atau Covid-19, kemudian menyebar ke seluruh dunia pada tahun 2020 dan menjadi masalah kesehatan yang serius [1][2][3]. World Health Organization (WHO) secara resmi menandai Covid-19 sebagai pandemi pada 11 Maret 2020. Ini dikarenakan penyebarannya yang cepat ke seluruh dunia [4][5][6]. Covid-19 adalah virus mematikan, virus ini telah merenggut ribuan nyawa [6][7], menjadi pandemi global pada waktu kurang dari empat bulan dan menyerang negara lain bahkan Indonesia [2][6][7][8].

Kasus Covid pertama di Indonesia dikonfirmasi pada Senin, 2 Maret 2020 yang dipublikasikan secara resmi oleh pemerintah melalui website resmi Kementerian Kesehatan Republik Indonesia [2]. Sejak saat itu pemerintah Indonesia terus mengumumkan kasus yang dikonfirmasi, pulih serta meninggal setiap harinya [1][9]. Di Indonesia, pertanggal 15 September 2021 telah mencapai 4,18 juta kasus dengan kasus kematian sebesar 140 ribu kasus [10], dan sudah tersebar ke semua provinsi di Indonesia. Provinsi Riau pertama kalinya mengkonfirmasi kasus Covid-19 pada Hari Rabu, 18 Maret 2020 dan menjadi satu diantara sepuluh provinsi dengan tingkat tertinggi. Pekanbaru dan Dumai sebagai kota dengan kasus Covid tertinggi di Riau. Jumlah kasus Covid terkonfirmasi di kota Dumai mencapai 10.189 [11]. Penderita yang terkena virus Covid akan mendapatkan gangguan pernafasan ringan maupun berat dengan gejala seperti, batuk kering, mudah Lelah, demam, sesak nafas, letih dan penurunan nafsu makan [12][13]. Adapun penyebaran dari virus ini umumnya melalui kontak erat, terkena droplet pasien yang terinfeksi maupun menyentuh barang yang telah terkena droplet pasien [2][14].

Untuk menghindari penyebaran Covid-19, pemerintah telah melakukan berbagai cara seperti social distancing dan scrining atau swab dadakan di pusat-pusat keramaian [2][7][15]. Selain itu, masyarakat juga diharuskan untuk menaati protokol kesehatan seperti menjaga jarak, menggunakan masker, mencuci tangan, menghindari kerumunan dan menerapkan hidup sehat. Dikarenakan banyaknya jenis penularan virus Covid-19 maka diperlukan pengelompokan pasien berdasarkan jenis penularannya. Untuk melakukan pengelompokan dan menganalisis data yang berskala besar diperlukan sebuah algoritma data mining pengelompokan [16].

Data mining merupakan sebuah metode eksplorasi data berskala besar yang sebelumnya tidak ditemukan [17]. Data mining dapat digunakan untuk pengambilan keputusan, prediksi masa depan, pengelompokan dan

peramalan. Metode pada data mining sangat beragam antara lain Classification, Estimation, Clustering, Association dan Prediction. Clustering merupakan proses pengelompokan satu set pola yang diberikan pada suatu kelompok data. Teknik pengelompokan ini dilakukan dengan menentukan pola yang sama untuk jenis cluster yang sama. Metode ini menghasilkan cluster-cluster yang sesuai dengan pola kelompoknya. Clustering mengelompokkan menjadi beberapa kategori berdasarkan kesamaan atribut. Tujuan dari metode ini adalah untuk mengetahui perbedaan antar kategori [18]. Pada metode Clustering data atau nilai yang ada belum memiliki target sehingga memerlukan prediksi ke kategori mana data itu akan masuk [19]. Pada saat ini, metode Clustering sudah sangat banyak seperti K-Means Clustering dan K-Medoids.

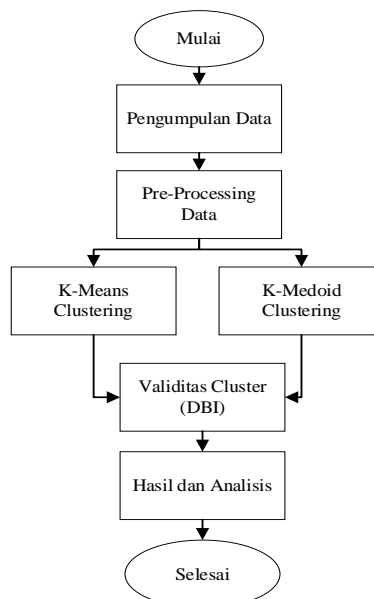
K-Medoids pernah digunakan pada penelitian terdahulu pada pengelompokan penyebaran Covid-19 di Indonesia menjadi 3 cluster, hasil cluster 1 sebanyak 1 wilayah, cluster 2 terdiri dari 2 wilayah dan 31 wilayah lainnya di cluster ke-3 [15]. Penelitian lainnya pernah dilakukan pengelompokan kasus kematian Covid-19 di Benua Asia dengan K-Means, yang menghasilkan empat negara dengan tingkat tinggi, 4 negara tingkat sedang dan 41 lainnya tingkat rendah [5]. Dan masih banyak penelitian lainnya.

Berdasarkan penelitian sebelumnya, yang membedakan penelitian ini adalah membandingkan dua algoritma pengelompokan yaitu K-Means dan K-Medoids serta dataset yang menjadi dasar pengelompokan adalah jenis penularan virus, usia, jenis kelamin, asal paskes dan kecamatan. Dengan menggunakan kedua algoritma ini, data yang dikelompokkan akan ditentukan pusat data dari setiap kelompoknya kemudian dilakukan perhitungan menggunakan rumus K-Means dan K-Medoids agar menghasilkan akurasi masing-masing algoritma yang nantinya dapat dibandingkan dengan uji validitas cluster serta didapatkan cluster-cluster yang dianalisis berdasarkan atribut pendukungnya. Penelitian ini menggunakan data yang berasal dari Dinas Kesehatan Kota Dumai yang diambil secara langsung pada bulan September 2021. Dengan demikian, penelitian ini bertujuan untuk membandingkan algoritma K-Means Clustering dan K-Medoids Clustering pada klusterisasi data pasien Covid-19 di Kota Dumai. Adanya kesulitan pada pemodelan dan klusterisasi data dilapangan terkait dengan jenis penularan, jenis kelamin, usia, paskes dan kecamatan menjadi permasalahan utama sehingga diperlukan sebuah algoritma K-Means dan K-Medoid untuk melakukan klusterisasi tersebut. Penelitian ini diharapkan dapat memperoleh model algoritma terbaik pada pengelompokan pasien Covid-19 dan dapat dijadikan sebagai saran atau bahan pertimbangan kepada Dinas Kesehatan untuk menanggulangi virus Covid-19 di Kota Dumai berdasarkan dari hasil analisis yang diperoleh.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini menerapkan lima tahapan utama yang tertera pada gambar 1 yaitu pengumpulan data, pre-processing data, proses klusterisasi menggunakan algoritma K-Means dan K-Medoid serta tahap hasil dan analisis.



**Gambar 1.** Tahapan Penelitian

Tahap pengumpulan data dilakukan di dinas kesehatan kota dumai dengan menggunakan data pasien Covid-19 kota dumai selama 2 bulan yaitu sejak 1 juli hingga 4 september 2021 dengan menghasilkan sebanyak 3125 data. Data tersebut kemudian di preprocessing hingga menyisakan 3002 data. Kemudian klusterisasi

menggunakan algoritma K-Means dan K-Medoid. Tahapan terakhir sebelum didapatkan hasil dan dilakukannya analisis adalah melakukan validitas cluster yang mana pada penelitian ini menerapkan teknik Davies Bouldin Index (DBI).

### 2.1 K-Means

Algoritma partitioned clustering yang mudah digunakan serta sangat cepat mengklusterisasi data besar dan outlier adalah K-Means Clustering. Algoritma ini, mengklusterisasi berdasarkan tingkat kesesuaian dan ketidaksesuaian objek dengan cara membagi suatu objek menjadi k-cluster yang nantinya posisi objek tersebut dikelompokkan dari nilai mean terdekat [20]. Proses klusterisasi K-Means bergantung pada data yang ada serta kesimpulan yang ingin diambil. Data yang sudah diatur di tiap-tiap pengamatan cluster berdasarkan ketetanggaan pengamatan nilai mean kemudian diterapkan perhitungan berulang [21]. Untuk menghitung jarak ke- $i$  ( $x_i$ ) pada pusat cluster ke- $k$  ( $c_k$ ), diberi nama ( $d_{ik}$ ), dapat menggunakan persamaan 1 yaitu persamaan Euclidean.

$$d_{ik} = \sqrt{\sum_{j=1}^m (c_{ij} - x_{ik})^2} \quad (1)$$

K-Means akan menguji setiap atribut pada kumpulan data yang ada dan mengelompokkan atribut tersebut kedalam suatu cluster dengan jarak terdekat. Untuk menghitung hal ini dapat menerapkan persamaan 2.

$$m \sum_k^k d_{ik} = \sqrt{\sum_{j=1}^m (c_{ij} - x_{ik})^2} \quad (2)$$

### 2.2 K-Medoids

Partitioning Around Medoids (PAM) atau yang dikenal dengan K-Medoids adalah varian algoritma K-Means yang dilandaskan pada penerapan medoids yang bertujuan mengurangi sensitivitas dari partisi terkait nilai-nilai ekstrim pada dataset [21][22]. Setiap data dialokasikan ke cluster yang paling dekat dengan persamaan Euclidian pada persamaan 3.

$$d_{ik} = \sqrt{\sum_{j=1}^m (c_{ij} - x_{ik})^2} \quad (3)$$

### 2.3 Davies-Bouldin Index (DBI)

Tahun 1979, David L. Davies dan Donald W. Bouldin pertama kali memperkenalkan DBI. DBI berfungsi untuk memaksimalkan jarak inter-cluster, dan juga mencoba meminimalkan jarak antar-titik pada cluster pada saat yang bersamaan [23][24][25]. Nilai rata-rata yang similar antara setiap cluster dan paling mirip merupakan ukuran dari DBI. Semakin kecil nilai DBI atau bernilai positif terkecil yang mendekati 0 maka menunjukkan skema cluster yang paling optimal [26]. Untuk Menghitung nilai DBI dapat menggunakan persamaan 4.

$$DB = \sum_{i=1}^p \left( \frac{\sigma_i + \sigma_j}{p} \right) \quad (4)$$

### 2.4 Corona Virus Disease 2019 (Covid-19)

Covid-19 adalah jenis penyakit menular berupa sindrom pernapasan akut yang disebabkan oleh SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) [27]. Virus ini tergolong ke dalam keluarga coronaviridae yang ketika menyerang manusia dapat mengakibatkan infeksi saluran pernapasan, seperti flu, MERS (Middle East Respiratory Syndrome), dan SARS (severe acute respiratory syndrome). Pada akhir 2019, COVID-19 ditemukan sebagai coronavirus jenis baru di wilayah Wuhan, Hubei, China [28]. Covid-2019 kemudian menyebar ke seluruh dunia pada tahun 2020 dan menjadi masalah kesehatan yang serius [1][2][3].

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pengumpulan dan Preprocessing Data

Sumber data pada penelitian ini diambil langsung pada Dinas Kesehatan Kota Dumai. Data yang diambil berupa data kasus Covid-19 pada 2 bulan terakhir yaitu sejak 1 Juli hingga 4 September 2021. Data yang diperoleh pada penumpulan ada adalah 3125 data. Atribut yang terdapat pada data antaralain Jenis Kelamin, Umur, Alamat atau Kecamatan, Paskes, Status Isolasi, Lokasi pengambilan Swab dan Status Penyebab Tertularnya Virus Covid-19. Pada tahap preprocessing data, dilakukan cleaning, transformasi dan normalisasi terhadap data-data di tiap atribut karena memiliki banyak noise dan data kecamatan, status isolasi, paaskes, lokasi pengambilan swab dan status penyebab tertularnya pasien masih berupa data kualitatif sehingga harus ditransformasi ke data numerik terlebih dahulu. Data yang digunakan setelah proses preprocessing ini menjadi 3002 data. Pada tabel 1 dapat terlihat data hasil cleaning penelitian ini.

**Tabel 1.** Data Covid-19 hasil Cleaning

Id	Jenis Penularan	Jenis kelamin	Usia (th)	Kecamatan	Puskesmas	Status Isolasi	Lokasi Pengambilan Sampel
1	Suspek	L	26	Dumai Barat	Dumai Barat	Isolasi Mandiri	RSPD
2	Suspek	L	45	Dumai Barat	Dumai Barat	Isolasi Mandiri	RSPD
3	Kontak Erat	L	58	Dumai Barat	Purnama	Isolasi Mandiri	P. Purnama
4	Kontak Erat	P	53	Dumai Barat	Purnama	Isolasi Mandiri	P. Purnama
5	Kontak Erat	L	11	Dumai Barat	Purnama	Isolasi Mandiri	P. Purnama
6	Kontak Erat	P	42	Dumai Kota	Dumai Kota	Isolasi Mandiri	P. Dumai Kota
7	Suspek	P	46	Dumai Selatan	Bumi Ayu	Isolasi Mandiri	P. Bumi Ayu
...	...	...	...	...	...	...	...
3001	Suspek	P	41	Dumai Timur	Jaya Mukti	Isolasi Mandiri	P. Jaya Mukti
3002	Suspek	L	36	Dumai Timur	Jaya Mukti	Isolasi Mandiri	P. Jaya Mukti

**Transformasi Data**

Transformasi diterapkan pada data kasus Covid-19 untuk atribut Jenis Kelamin, Umur, Alamat atau Kecamatan, Paskes, Status Isolasi, Lokasi pengambilan Swab dan Status Penyebab Tertularnya Virus Covid-19 yang tertera pada tabel 2.

**Tabel 2.** Data Covid-19 hasil Transformasi

Id	Jenis Penularan	Jenis kelamin	Usia (th)	Kecamatan	Puskesmas	Status Isolasi	Lokasi Pengambilan Sampel
1	1	1	26	1	1	1	2
2	1	1	45	1	1	1	2
3	2	1	58	1	2	1	3
4	2	2	53	1	2	1	3
5	2	1	11	1	2	1	3
6	2	2	42	5	3	1	4
7	1	2	46	3	4	1	5
...	...	...	...	...	...	...	...
3001	1	2	41	2	5	1	6
3002	1	1	36	2	5	1	6

**Normalisasi Data**

Tabel 3 merupakan data hasil normalisasi dengan menggunakan teknik min-max normalization pada data pasien Covid-19 di Kota Dumai.

**Tabel 3.** Data Covid-19 hasil Normalisasi

Id	Jenis Penularan	Jenis kelamin	Usia (th)	Kecamatan	Puskesmas	Status Isolasi	Lokasi Pengambilan Sampel
1	0,28571	0,66667	0,27957	0,22222	0,16667	0,50000	0,17647
2	0,28571	0,66667	0,48387	0,22222	0,16667	0,50000	0,17647
3	0,42857	0,66667	0,62366	0,22222	0,25000	0,50000	0,23529
4	0,42857	1,00000	0,56989	0,22222	0,25000	0,50000	0,23529
5	0,42857	0,66667	0,11828	0,22222	0,25000	0,50000	0,23529
6	0,42857	1,00000	0,45161	0,66667	0,33333	0,50000	0,29412
7	0,28571	0,66667	0,27957	0,22222	0,16667	0,50000	0,17647
...	...	...	...	...	...	...	...
3001	0,28571	1,00000	0,44086	0,33333	0,50000	0,50000	0,41176
3002	0,28571	0,66667	0,38710	0,33333	0,50000	0,50000	0,41176

**3.1.1 Proses Clustering dan Validitas Cluster**

Proses Clustering pada penelitian ini dilakukan dengan menggunakan dua algoritma yaitu K-Means dan K-Medoid. Algoritma tersebut mengolah data Covid kota dumai sebanyak 3002 data yang sebelumnya sudah dipreprocessing.

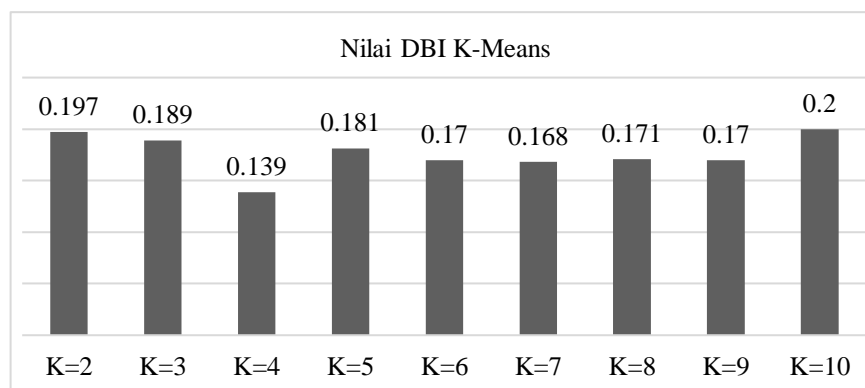
**K-Means**

Proses Clustering pada penelitian ini menggunakan algoritma K-Means. Berdasarkan penelitian yang dilakukan oleh Insanul Kamila, Ulya Khairunnisa, dan Mustakim (2019) menggunakan percobaan cluster 2, 3, 4, 5, 6, 7, 8, 9, dan 10. Hasil cluster tersebut dapat dilihat pada tabel 4.

**Tabel 4.** Hasil Klasterisasi K-Means

PERCOBAAN	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
NILAI K = 2	2142	860	-	-	-	-	-	-	-	-
NILAI K = 3	1562	651	789	-	-	-	-	-	-	-
NILAI K = 4	652	789	1560	1	-	-	-	-	-	-
NILAI K = 5	751	598	639	1	1013	-	-	-	-	-
NILAI K = 6	746	597	58	587	1013	1	-	-	-	-
NILAI K = 7	584	739	58	379	569	672	1	-	-	-
NILAI K = 8	58	1	680	565	519	154	354	671	-	-
NILAI K = 9	291	1	58	518	455	154	541	678	306	-
NILAI K = 10	150	193	284	1	681	304	554	58	376	401

Setelah dilakukan proses pengelompokan menggunakan algoritma k-means kemudian dilakukan validitas cluster dengan teknik DBI. Nilai DBI berdasarkan hasil Clustering data menggunakan algoritma K-Means tertera pada gambar 2.



**Gambar 2.** Nilai DBI K-Means

Berdasarkan nilai DBI terbaik, cluster terbaik pada algoritma k-means terletak pada percobaan dengan k=4 dengan nilai DBI yang diperoleh sebesar 0,139 yang membagi data menjadi 4 cluster dengan 653 anggota pada cluster 1, 789 anggota pada cluster 2, 1560 anggota pada cluster 3 dan 1 anggota pada cluster 4.

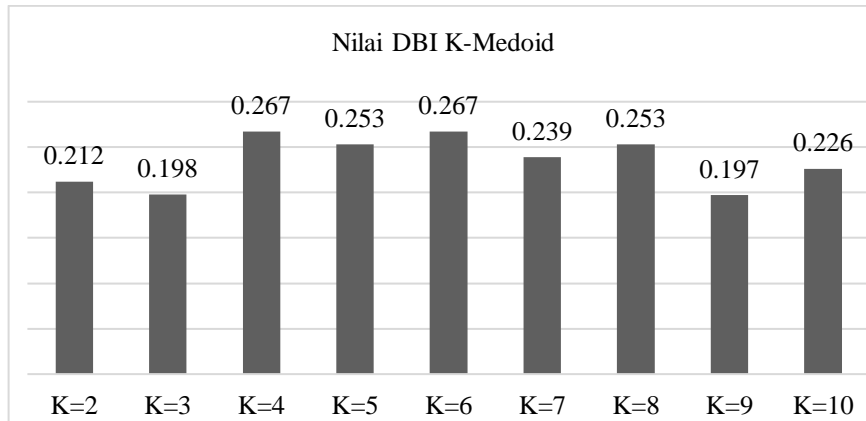
**K-Medoid**

Proses Clustering pada penelitian ini juga menggunakan algoritma K-Medoid dengan menggunakan percobaan cluster yang sama dengan algoritma sebelumnya, sehingga juga menggunakan percobaan cluster 2, 3, 4, 5, 6, 7, 8, 9, dan 10. Hasil cluster tersebut dapat dilihat pada tabel 5.

**Tabel 5.** Hasil Klasterisasi K-Medoid

PERCOBAAN	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10
NILAI K = 2	2335	667	-	-	-	-	-	-	-	-
NILAI K = 3	621	509	1872	-	-	-	-	-	-	-
NILAI K = 4	587	446	896	1073	-	-	-	-	-	-
NILAI K = 5	520	922	880	391	289	-	-	-	-	-
NILAI K = 6	222	839	732	235	587	387	-	-	-	-
NILAI K = 7	230	520	229	380	731	117	795	-	-	-
NILAI K = 8	384	325	191	285	52	374	745	646	-	-
NILAI K = 9	713	135	228	116	492	51	519	365	383	-
NILAI K = 10	619	349	205	377	216	123	634	230	199	50

Kemudian juga dilakukan validitas cluster dengan menggunakan Davies Bouldin Index (DBI). Nilai DBI berdasarkan hasil Clustering data menggunakan algoritma K-Medoid dapat dilihat pada gambar 3.

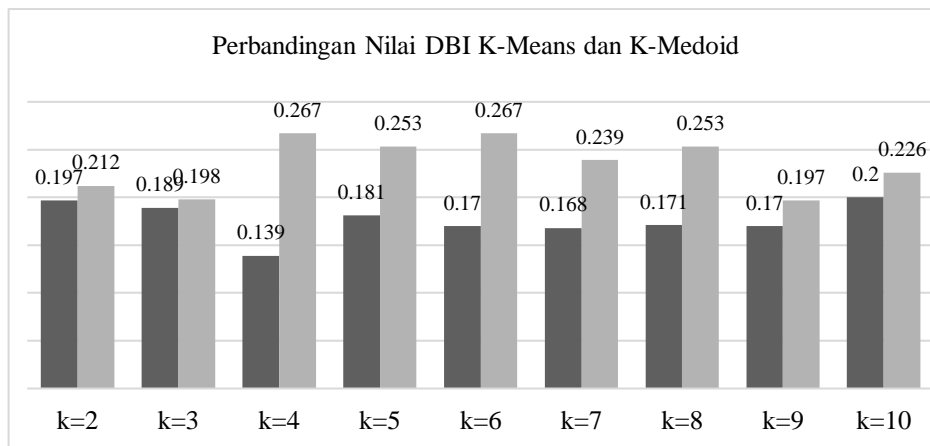


**Gambar 3.** Nilai DBI K-Medoid

Berdasarkan nilai DBI terbaik, cluster terbaik pada algoritma k-medoid terletak pada percobaan dengan k=9 dengan nilai DBI yang diperoleh adalah 0,197 disusul percobaan dengan nilai k=2 diperoleh nilai DBI sebesar 0,198. Pada percobaan nilai k=9 membagi data menjadi 9 cluster dengan 713 anggota pada cluster 1, 135 anggota pada cluster 2, 228 anggota pada cluster 3, 116 anggota pada cluster 4, 492 anggota pada cluster 5, 51 anggota pada cluster 6, 519 anggota pada cluster 7, 365 anggota pada cluster 8, dan 383 anggota pada cluster 9.

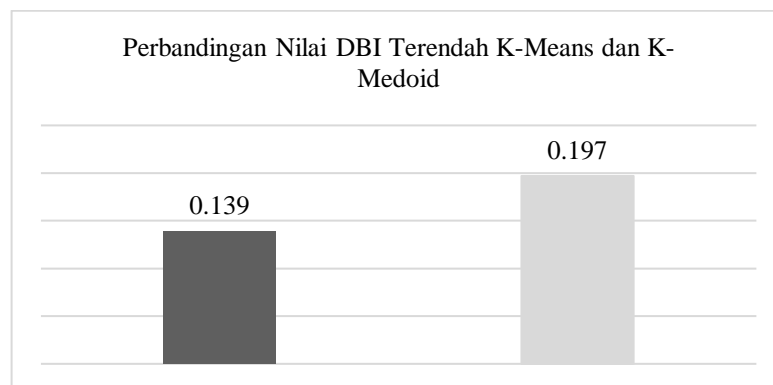
**Perbandingan Hasil DBI Algoritma K-Means dan K-Medoid**

Adapun perbandingan algoritma K-Means dan K-Medoid dalam mengelompokkan data pasien Covid-19 dapat dilihat pada gambar 4.



**Gambar 4.** Perbandingan Nilai DBI K-Means dan K-Medoid

Grafik perbandingan nilai DBI tertinggi pada algoritma k-means yaitu pada k=4 dengan nilai DBI sebesar 0,139 dan pada algoritma k-medoids yaitu pada k=9 dengan nilai DBI sebesar 0,197 dapat dilihat pada gambar 5.



**Gambar 5.** Perbandingan Nilai DBI Terendah K-Means dan K-Medoid

Dari hasil yang telah ditampilkan pada grafik 4 maka dapat kita ketahui bahwa pada penelitian ini, algoritma K-Means memiliki nilai validitas cluster lebih baik dibandingkan algoritma K-Medoid. Dengan demikian, pada penelitian ini diperoleh cluster paling optimal menggunakan algoritma K-Means dengan percobaan nilai  $k = 4$ .

#### 4. KESIMPULAN

Berdasarkan hasil analisis dan pengolahan data menggunakan algoritma k-means dan k-medoid. Maka diperoleh cluster terbaik pada algoritma k-means dengan nilai  $k=4$ . Sementara itu, pada algoritma K-Medoid diperoleh cluster terbaik pada  $k=9$ . Hasil klusterisasi tersebut sudah di validasi menggunakan Davies Bouldin Index (DBI). Maka pada penelitian ini K-Means lebih unggul daripada K-medoid pada pengelompokan kasus Covid-19 di Kota Dumai berdasarkan jenis tertularnya dengan nilai DBI terbaik yaitu 0,139 dengan nilai  $k=4$ . Percobaan k-means dengan nilai  $k=4$  membagi data menjadi 4 cluster dengan 653 anggota pada cluster 1, 789 anggota pada cluster 2, 1560 anggota pada cluster 3 dan 1 anggota pada cluster 4. Pola yang diperoleh dari penelitian ini dapat dijadikan sebagai acuan gambarkan model clustering Covid-19.

#### UCAPAN TERIMAKASIH

Ucapan terimakasih kepada Dinas Kesehatan Kota Dumai yang telah memberikan data untuk penelitian ini. kepada Organisasi Research Puzzle Research Data Teknologi atau Predatech yang memberikan dukungan, menjadi wadah pembelajaran dan memberikan motivasi terhadap penelitian ini sehingga dapat berjalan dengan baik sesuai dengan yang di harapkan.

#### REFERENCES

- [1] D. Y. Liliانا, H. Maulana, and A. Setiawan, "Data Mining untuk Prediksi Status Pasien Covid-19 dengan Pengklasifikasi Naïve Bayes," *J. Multinetics*, vol. 7, no. 1, pp. 48–53, 2021.
- [2] F. Adiba, "Penerapan Data Mining dalam Mengklasifikasikan Tingkat Kasus Covid-19 di Sulawesi Selatan Menggunakan Algoritma Naive Bayes," *Indones. J. Fundam. Sciences*, vol. 7, no. 1, pp. 18–28, 2021.
- [3] G. Mognolo and L. Nanni, "A Critic Evaluation of Methods for COVID-19 Automatic Detection from X-Ray Images," <http://arxiv.org/abs/2004.12823>, 2020.
- [4] N. Nurhalimah, "Upaya Bela Negara Melalui Sosial Distancing Dan Lockdwon," *Sekol. Tinggi Tarb. Insa. Kamil*, pp. 1–6, 2020.
- [5] Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian," *J. Inform. dan Komput.*, vol. 22, no. 2, pp. 183–188, 2020.
- [6] G. D. Rembulan, T. Wijaya, D. Palullungan, K. N. Alfina, and M. Qurthuby, "Kebijakan Pemerintah Mengenai Coronavirus Disease (COVID-19) di Setiap Provinsi di Indonesia Berdasarkan Analisis Klaster," *JIEMS (Journal Ind. Eng. Manag. Syst.)*, vol. 13, no. 2, 2020, doi: 10.30813/jiems.v13i2.2280.
- [7] D. D. Darmansah and N. W. Wardani, "Analisis Pesebaran Penularan Virus Corona di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 105–117, 2021, doi: 10.35957/jatisi.v8i1.590.
- [8] Z. Zaharah, G. I. Kirilova, and A. Windarti, "Impact of corona virus outbreak towards teaching and learning activities in Indonesia," *SALAM J. Sos. dan Budaya Syar-i*, vol. 7, no. 3, pp. 269–282, 2020.
- [9] M. Sukmana, M. Aminuddin, and D. Nopriyanto, "Indonesian government response in COVID-19 disaster prevention," *East African Sch. J. Med. Sci.*, vol. 3, no. 3, pp. 81–86, 2020.
- [10] Covid-19.go.id, "Situasi Covid-19 Di Indonesia," <http://covid19.go.id>, 2021.
- [11] Riau Tanggap Virus Corona, "Sebaran Covid-19 Provinsi Riau," <https://corona.riau.go.id>, 2021.
- [12] C. Long et al., "Diagnosis of the Coronavirus disease (COVID-19): rRT-PCR or CT?," *Eur. J. Radiol.*, vol. 126, p. 108961, 2020.
- [13] N. Mona, "Konsep isolasi dalam jaringan sosial untuk meminimalisasi efek contagious (kasus penyebaran virus corona di Indonesia)," *J. Sos. Hum. Terap.*, vol. 2, no. 2, 2020.
- [14] A. Hafeez, S. Ahmad, S. A. Siddqui, M. Ahmad, and S. Mishra, "A review of COVID-19 (Coronavirus Disease-2019) diagnosis, treatments and prevention," *EJMO*, vol. 4, no. 2, pp. 116–125, 2020.
- [15] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. Ilmi R.H.Zer, and D. Hartama, "Analisis algoritma K-Medoids clustering dalam pengelompokan penyebaran Covid-19 di Indonesia," *Jti (Jurnal Teknol. Informasi)*, vol. 4, no. 1, pp. 166–173, 2020.
- [16] C. A. Sugianto, A. H. Rahayu, and A. Gusman, "Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Cigugur Tengah," *J. Inf. Technol.*, vol. 2, no. 2, pp. 39–44, 2020.
- [17] A. F. Watratana and D. Moeis, "Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 7–14, 2020.
- [18] Y. Cheng, K. Chen, H. Sun, Y. Zhang, and F. Tao, "Data and knowledge mining with big data towards smart production," *J. Ind. Inf. Integr.*, vol. 9, no. September, pp. 1–13, 2018, doi: 10.1016/j.jii.2017.08.001.
- [19] M. Arhami, M. Kom, and S. T. Muhammad Nasir, *Data Mining-Algoritma dan Implementasi*. Penerbit Andi, 2020.
- [20] D. K. Sharma, S. K. Dhurandher, D. Agarwal, and K. Arora, "kROp: k-Means clustering based routing protocol for opportunistic networks," *J. Ambient Intell. Humaniz. Comput.*, vol. 10, no. 4, pp. 1289–1306, 2019.
- [21] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, "K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir," *IEEE Access*, vol. 8, pp. 151847–151855, 2020.



- [22] I. Kamila, U. Khairunnisa, and M. Mustakim, “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau,” *J. Ilm. Rekayasa dan Manaj. Sist. Inf.*, vol. 5, no. 1, pp. 119–125, 2019.
- [23] S. Nawrin, M. R. Rahman, and S. Akhter, “Exploreing k-means with internal validity indexes for data clustering in traffic management system,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 3, pp. 264–272, 2017.
- [24] P. Prasetyawan, I. Ahmad, R. I. Borman, Y. A. Pahlevi, and D. E. Kurniawan, “Classification of the Period Undergraduate Study Using Back-propagation Neural Network,” in *2018 International Conference on Applied Engineering (ICAE)*, 2018, pp. 1–5.
- [25] S. Pramono, I. Ahmad, and R. I. Borman, “ANALISIS POTENSI DAN STRATEGI PENEMBAAN EKOWISATA DAERAH PENYANGGA TAMAN NASIONAL WAY KAMBAS,” *J. Teknol. dan Sist. Inf.*, vol. 1, no. 1, pp. 57–67, 2020.
- [26] H. Sulistiani, I. Darwanto, and I. Ahmad, “Penerapan Metode Case Based Reasoning dan K-Nearest Neighbor untuk Diagnosa Penyakit dan Hama pada Tanaman Karet,” *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 6, no. 1, pp. 23–28, 2020.
- [27] R. Adha, N. Nurhaliza, U. Sholeha, and M. Mustakim, “Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia,” *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 18, no. 2, pp. 206–211, 2021.
- [28] H. Lu, C. W. Stratton, and Y. Tang, “Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle,” *J. Med. Virol.*, vol. 92, no. 4, p. 401, 2020.