

Komparasi Algoritma *K-Means* dan *K-Medoids* Untuk Pengelompokan Penyebaran Covid-19 di Indonesia

Anisa Fira^{1*}, Chaerur Rozikin^{2*}, Garno^{3*}

* Teknik Informatika, Universitas Singaperbangsa Karawang

anisa.fira17058@student.unsika.ac.id¹, chaerur.rozikin@staff.unsika.ac.id², garno@staff.unsika.ac.id³

Article Info

Article history:

Received 2021-08-19

Revised 2021-10-06

Accepted 2021-10-22

Keyword:

Clustering,
Covid-19,
Data Mining,
K-Medoids,
K-Means.

ABSTRACT

COVID-19 merupakan bagian dari keluarga virus penyebab *Severe Acute Respiratory Syndrome (SARS)* dan *Middle East Respiratory Syndrome (MERS)*, beberapa gejala yang dialami apabila terinfeksi virus ini antara lain batuk, demam, letih, sesak nafas, dan mengalami penurunan nafsu makan. Pada Penelitian ini data yang digunakan adalah sebanyak 34 data Provinsi pada tahun 2019 - Februari 2021 yang bersumber pada website resmi www.covid19.go.id. Dalam upaya menemukan daerah yang memiliki kasus penyakit Covid-19 dapat menggunakan *Data Mining*. Negara Indonesia merupakan salah satu dari negara di dunia yang cukup tinggi terkena virus covid-19. Tujuan penelitian ini yaitu untuk mengelompokan provinsi yang memiliki penyakit covid-19 dengan tingkat tinggi dan rendah di Indonesia dan melakukan perbandingan dengan metode algoritma yang digunakan yaitu *K-Means* dan *K-Medoids*. Hasil yang didapatkan pada penelitian ini adalah memiliki *cluster* optimal sebanyak 2 *cluster*, dengan menggunakan algoritma *K-Means* dimana *cluster* 1 beranggotakan 2 wilayah dan dikategorikan tinggi, sedangkan untuk *cluster* 2 sebanyak 32 wilayah dan dikategorikan rendah. Sedangkan menggunakan algoritma *K-Medoids* yaitu untuk *cluster* 1 beranggotakan 4 wilayah dan dikategorikan tinggi, sedangkan untuk *cluster* 2 sebanyak 30 wilayah dan dikategorikan rendah. Dari kedua perbandingan tersebut menghasilkan nilai *Silhouette Coefficient* dengan metode *K-Means* adalah sebesar 0,207. Sedangkan Nilai *Silhouette Coefficient* dengan metode *K-Medoids* adalah sebesar 0,347.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

COVID-19 merupakan bagian dari keluarga virus penyebab *Severe Acute Respiratory Syndrome (SARS)* dan *Middle East Respiratory Syndrome (MERS)* [1]. Kasus virus ini ditemukan pertama kali di provinsi Wuhan dan beberapa gejala yang dialami apabila terinfeksi virus ini antara lain batuk, demam, letih, sesak nafas, dan mengalami penurunan nafsu makan [2]. Secara umum virus ini dapat menular melalui droplet atau cairan tubuh yang dikeluarkan selama bersin dan batuk [3-5]. Negara Indonesia adalah salah satu dari berbagai negara di dunia yang mengalami dampak COVID-19. Berdasarkan Grafik perkembangan COVID-19 Indonesia jumlah kasus terkonfirmasi telah mencapai 1.263.299 kasus per 19 Februari 2021.



Gambar 1 Data Grafik Penyakit Covid-19 tahun 2019– 2020
(Sumber: www.covid19.go.id)

Negara Indonesia adalah salah satu dari berbagai negara di dunia yang mengalami dampak COVID-19. Berdasarkan Grafik perkembangan COVID-19 Indonesia jumlah kasus terkonfirmasi telah mencapai 1.263.299 kasus per 19 Februari 2021 [2]. Berdasarkan pada gambar 1 ini yakni informasi permasalahan penyakit Covid-19 di Indonesia pada tahun 2019 hingga Februari 2021 mengalami peningkatan setiap harinya.

Pada penelitian ini, seluruh data kasus COVID-19 di semua provinsi di Indonesia dilakukan pengelompokan berdasarkan konsep *data mining* sehingga dapat menemukan pola hubungan pada keseluruhan data. Data mining mampu mengonversi data mentah menjadi sebuah pengetahuan yang lebih bermanfaat dengan cara memetakan data menjadi lebih ringkas dan abstrak [6]. Salah satu teknik yang terdapat dalam konsep data mining adalah teknik klusterisasi. Klusterisasi diartikan sebagai suatu metode yang digunakan untuk mengklasifikasi data dan untuk melakukan segmentasi data [6]. Metode klusterisasi yang digunakan dalam penelitian ini adalah non hierarki yaitu Algoritma *K-Means* dan *K-Medoids*. Algoritma *K-Means* digunakan untuk menunjukkan jumlah anggota dan mengklasifikasikan masing-masing data di setiap kluster [7]. Keunggulan dari metode ini yaitu memiliki tingkat ketelitian yang cukup tinggi pada pengukuran dan pengolahan objek dengan jumlah besar sehingga algoritmanya lebih relatif terukur dan efisien [8-11]. Selain itu, terdapat pengembangan varian dari metode *K-Means Clustering* yaitu *K-Medoids*, bertujuan mengurangi sensitivitas dari partisi yang dihasilkan sehubungan dengan nilai-nilai ekstrim yang terdapat dalam dataset, penggunaan medoids tidak berdasarkan dari pengamatan mean yang dimiliki oleh setiap cluster.

II. METODE PENELITIAN

A. Silhouette Coefficient

Silhouette Coefficient adalah suatu metode penilaian *Cluster* yang bertujuan untuk membandingkan mutu objek dalam sesuatu *Cluster* [6]. Ada pula tahapan perhitungan *Silhouette Coefficient*, antara lain :

1. Hitung jarak rata-rata dari suatu objek ke- i dengan seluruh objek lain yang terletak dalam satu *Cluster*.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (1)$$

2. Hitung rata-rata dari objek ke- i tersebut dengan seluruh objek pada *Cluster* yang lain, serta ambil nilai terkecilnya.

$$b(i) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \quad (2)$$

3. Nilai *Silhouette Coefficient* nya merupakan: Jumlah $s(i)$ diperoleh dengan mengkombinasikan $a(i)$ serta $b(i)$:

$$s = \begin{cases} 1 - \frac{a(i)}{b(i)}, & \text{if } a(i) < b(i) \\ 0 & \text{if } a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1, & \text{if } a(i) > b(i) \end{cases} \quad (3)$$

Sehingga bisa dirumuskan :

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

Dimana :

$s(i)$ = Nilai *Silhouette*.

$a(i)$ = Rata-rata jarak antara i dengan seluruh objek pada *Cluster*-nya.

$b(i)$ = Rata-rata jarak antara data i pada seluruh objek pada *Cluster* lain.

$d(i, j)$ = Jarak antara objek i dengan j .

B. Metode Analisis Data

Metode analisis data yang digunakan pada penelitian ini adalah *Knowledge Discovery Database* (KDD). Adapun tahapan dari metode *Knowledge Discovery in Database* yang digunakan pada penelitian ini adalah sebagai berikut.

1) Data Selection

Data selection merupakan salah satu proses dimana pengambilan data/pengumpulan data, pemilihan data dan pengecekan data.

2) Pre-Processing

Tahap *pre-processing* merupakan salah satu tahap pembersihan data dengan membuang duplikasi data, mengatasi *missing value*, memeriksa data yang tidak konsisten, dan memperbaiki data seperti kesalahan pada penulisan (*tipografi*).

3) Data Transformation

Tahap ini data akan dilakukan transformasi sebelum dilakukannya *data mining*. Hal tersebut bertujuan supaya data dapat menyesuaikan pada saat diolah berdasarkan algoritma dan *tools* yang akan digunakan dalam pengolahan data.

4) Data Mining

Data mining merupakan salah satu proses pengolahan data berdasarkan algoritma sesuai dengan teknik data mining. Algoritma yang digunakan pada penelitian ini adalah algoritma *k-means* dan *k-medoids* yang merupakan salah satu algoritma *clustering*.

5) Knowledge Interpretation/Evaluation.

Tahap ini merupakan proses dari penginterpretasian dan evaluasi hasil dari *data mining*. Sehingga hasil yang didapat menjadi informasi.

III. HASIL DAN PEMBAHASAN

A. Hasi Penelitian

1) Data Selection

Dataset yang digunakan dalam penelitian ini menggunakan dataset penyakit Covid-19 di Indonesia tahun Desember 2019 - Februari 2021 dengan atribut Provinsi, jumlah kasus sembuh, dan jumlah jumlah kasus meninggal di Indonesia. Dataset diperoleh dari website *www.covid19.go.id*. Berikut merupakan tabel kasus penyakit covid-19 yang ada di Indonesia.

Berdasarkan Tabel I dibawah ini, kasus penyakit Covid-19 di Indonesia tahun 2019-2021 terdiri dari 4 variabel dan 34 objek data.

TABEL I
DATASET PENYAKIT COVID-19 DI INDONESIA TAHUN 2019-2021

| No | Provinsi | Sembuh | Meninggal |
|----|---------------------|--------|-----------|
| 1 | Aceh | 7776 | 382 |
| 2 | Sumatera Utara | 20239 | 803 |
| 3 | Sumatera Barat | 26730 | 639 |
| 4 | Riau | 28802 | 737 |
| 5 | Jambi | 3927 | 80 |
| 6 | Sumatera Selatan | 13283 | 744 |
| 7 | Bengkulu | 4510 | 146 |
| 8 | Lampung | 9946 | 612 |
| 9 | Bangka Belitung | 5737 | 103 |
| 10 | Kepulauan Riau | 8102 | 212 |
| 11 | Dki Jakarta | 304799 | 5095 |
| 12 | Jawa Barat | 159519 | 2169 |
| 13 | Jawa Tengah | 92458 | 6113 |
| 14 | DI Yogyakarta | 19717 | 624 |
| 15 | Jawa Timur | 111822 | 8798 |
| 16 | Banten | 20553 | 605 |
| 17 | Bali | 28340 | 856 |
| 18 | Nusa Tenggara Barat | 6673 | 336 |
| 19 | Nusa Tenggara Timur | 5162 | 210 |
| 20 | Kalimantan Barat | 3991 | 32 |
| 21 | Kalimantan Tengah | 11713 | 345 |
| 22 | Kalimantan Selatan | 18006 | 701 |
| 23 | Kalimantan Timur | 41868 | 1206 |
| 24 | Kalimantan Utara | 6661 | 136 |
| 25 | Sulawesi Utara | 11275 | 488 |
| 26 | Sulawesi Tengah | 7085 | 242 |
| 27 | Sulawesi Selatan | 49051 | 819 |
| 28 | Sulawesi Tenggara | 9076 | 191 |
| 29 | Gorontalo | 4238 | 126 |
| 30 | Sulawesi Barat | 3345 | 96 |
| 31 | Maluku | 6035 | 104 |

| | | | |
|----|--------------|------|-----|
| 32 | Maluku Utara | 3202 | 111 |
| 33 | Papua | 8928 | 171 |
| 34 | Papua Barat | 6436 | 120 |

2) Data Pre-Processing

Setelah dilakukan data selection atau pemilihan data kemudian dilakukan penanganan missing value, duplikasi data dan data yang tidak sesuai atau inkonsisten data. Dari dataset dilakukannya proses perbersihan data atau pemilihan data, dimana hanya data-data tertentu yang akan digunakan pada proses data mining.

berikut merupakan data akan diolah dari hasil seleksi dataset penyakit Covid-19 tahun 2019-2021 dapat dilihat pada Tabel II sebagai berikut.

TABEL II
DATASET HASIL SELEKSI COVID-19 DI INDONESIA TAHUN 2019-2021

| Provinsi | Sembuh | Meninggal |
|---------------------|--------|-----------|
| Aceh | 7776 | 382 |
| Sumatera Utara | 20239 | 803 |
| Sumatera Barat | 26730 | 639 |
| Riau | 28802 | 737 |
| Jambi | 3927 | 80 |
| Sumatera Selatan | 13283 | 744 |
| Bengkulu | 4510 | 146 |
| Lampung | 9946 | 612 |
| Bangka Belitung | 5737 | 103 |
| Kepulauan Riau | 8102 | 212 |
| Dki Jakarta | 304799 | 5095 |
| Jawa Barat | 159519 | 2169 |
| Jawa Tengah | 92458 | 6113 |
| Di Yogyakarta | 19717 | 624 |
| Jawa Timur | 111822 | 8798 |
| Banten | 20553 | 605 |
| Bali | 28340 | 856 |
| Nusa Tenggara Barat | 6673 | 336 |
| Nusa Tenggara Timur | 5162 | 210 |
| Kalimantan Barat | 3991 | 32 |
| Kalimantan Tengah | 11713 | 345 |
| Kalimantan Selatan | 18006 | 701 |
| Kalimantan Timur | 41868 | 1206 |
| Kalimantan Utara | 6661 | 136 |
| Sulawesi Utara | 11275 | 488 |
| Sulawesi Tengah | 7085 | 242 |
| Sulawesi Selatan | 49051 | 819 |
| Sulawesi Tenggara | 9076 | 191 |
| Gorontalo | 4238 | 126 |

| | | |
|----------------|------|-----|
| Sulawesi Barat | 3345 | 96 |
| Maluku | 6035 | 104 |
| Maluku Utara | 3202 | 111 |
| Papua | 8928 | 171 |
| Papua Barat | 6436 | 120 |

3) Data Transformation

Pada tahap ini data ditransformasikan dan atribut-atribut yang akan digunakan Pada proses pengolahan data menggunakan bantuan tools Rapidminer sehingga didapatkan atribut yang digunakan seperti pada tabel di bawah ini. Berikut merupakan tabel atribut yang digunakan pada dataset penyakit Covid-19 di indonesia tahun 2019-2021.

TABEL III
ATRIBUT DATASET YANG AKAN DIGUNAKAN

| Atribut | Tipe Data | Keterangan |
|-----------------|------------|------------------------|
| Provinsi | Polynomial | Nama Provinsi |
| Kasus Sembuh | Intejer | Jumlah Kasus Sembuh |
| Kasus Meninggal | Intejer | Jumlah Kasus Meninggal |

4) Data Mining

Pada tahap data mining, algoritma yang digunakan yaitu algoritma K-Medoids dan K-Means. Dalam pemilihan jumlah cluster yang optimal dengan menggunakan tools Rapidminer, Setelah hasil cluster dilakukan maka akan diproses tahap evaluasi berdasarkan nilai *Silouette Coefficient*.

Pada Tabel IV dapat diketahui terdapat 2 cluster pada dataset Penyakit Covid-19 di indonesia Tahun 2019-2021, yaitu cluster 1 memiliki jumlah 2 Provinsi, dengan tingkatan cluster provinsi yang memiliki kasus penyakit Covid-19 tinggi di indonesia. Pada cluster 2 memiliki jumlah 32 Provinsi, dengan tingkatan cluster yang memiliki daerah penyakit Covid-19 Rendah di indonesia. sehingga hasilnya dapat dilihat pada tabel IV berikut.

TABEL IV
HASIL CLUSTER K-MEANS

| Cluster | Anggota | Jumlah | Kriteria |
|-----------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|----------|
| Cluster 1 | DKI Jakarta, Jawa Barat | 2 | Tinggi |
| Cluster 2 | Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, Jawa Tengah, Di Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, | 32 | Rendah |

| | | | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|
| Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua, Papua Barat | | | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|

Berdasarkan Tabel V dibawah ini, dapat diketahui terdapat 2 cluster pada dataset Penyakit Covid-19 di indonesia Tahun 2019-2021, yaitu cluster 1 memiliki jumlah 4 Provinsi, dengan tingkatan cluster provinsi yang memiliki kasus penyakit Covid-19 tinggi di indonesia. Pada cluster 2 memiliki jumlah 30 Provinsi, dengan tingkatan cluster yang memiliki daerah penyakit Covid-19 Rendah di indonesia.

TABEL V
HASIL CLUSTER K-MEDOIDS

| Cluster | Anggota | Jumlah | Kriteria |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|----------|
| Cluster 1 | DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur | 4 | Tinggi |
| Cluster 2 | Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, Di Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua, Papua Barat | 30 | Rendah |

5) Knowledge Interpretation/Evaluation

Setelah dilakukannya analisis pada cluster dengan menggunakan algoritma *K-Means* dan *K-medoids*. Tahap selanjutnya dilakukan evaluasi terhadap model yang telah dibuat dengan menggunakan *Silouette Coefficient* yang bertujuan untuk mengetahui kualitas dari suatu objek pada suatu *cluster*. Dalam melakukan pengukuran pada kualitas cluster dengan menggunakan pengujian yang berpengaruh pada jumlah nilai cluster.

Berdasarkan Tabel VI dibawah ini, menunjukkan hasil dari pengujian pada kualitas cluster pada dataset Penyakit Covid-19 di indonesia Tahun 2019-2021 dengan data uji 34

data, pengujian kualitas cluster yang di uji yaitu nilai $k=2$ sampai $k=8$. Kualitas cluster paling baik diperoleh pada cluster 2 dengan nilai Silhouette Coefficient sebesar 0,207.

TABEL VI
HASIL UJI K-MEANS

| Data Uji | Nilai K | Silhouette Coefficient |
|----------|---------|------------------------|
| 34 | 2 | 0.207 |
| | 3 | 0.130 |
| | 4 | 0.189 |
| | 5 | 0.142 |
| | 6 | 0.132 |
| | 7 | 0.094 |
| | 8 | 0.073 |

Berdasarkan Tabel VII dibawah ini, menunjukkan hasil dari pengujian pada kualitas cluster pada dataset Penyakit Covid-19 di indonesia Tahun 2019-2021 dengan data uji 34 data, pengujian kualitas cluster yang di uji yaitu nilai $k=2$ sampai $k=8$. Kualitas cluster paling baik diperoleh pada cluster 2 dengan nilai Silhouette Coefficient sebesar 0,347.

TABEL VII
HASIL UJI K-MEDOIDS

| Data Uji | Nilai K | Silhouette Coefficient |
|----------|---------|------------------------|
| 34 | 2 | 0.347 |
| | 3 | 0.358 |
| | 4 | 1.298 |
| | 5 | 0.218 |
| | 6 | 0.346 |
| | 7 | 1.620 |
| | 8 | 0.775 |

Berdasarkan Tabel VIII dapat diketahui bahwa Nilai SC untuk validasi data hasil pengelompokan data Covid-19 di Indonesia tahun 2019-2021 dengan metode *K-Means* adalah sebesar 0,207. Sedangkan Nilai *Silhouette Coefficient* dengan metode *K-Medoids* adalah sebesar 0,347 yang menyatakan bahwa struktur cluster yang dihasilkan pada kedua metode pengelompokan ini adalah sama-sama *weak structure* atau terstruktur dengan lemah, Karena berdasarkan nilai *Silhouette Coefficient* antara $0,25 < SC \leq 0,5$ termasuk dalam kategori Weak Struktire. Berdasarkan hasil penelitian menggunakan Jumlah dataset sebesar 34 data tersebut, dapat disimpulkan bahwa metode *K-Medoids* menghasilkan nilai SC lebih besar dari pada metode *K-Means*, sehingga *K-Medoids* dapat memberikan hasil pengelompokan yang lebih baik.

TABEL VIII
PERBANDINGAN HASIL EVALUASI

| Algoritma | Jumlah Cluster | Silhouette Coefficient |
|------------------|----------------|------------------------|
| <i>K-Means</i> | 2 | 0.207 |
| <i>K-Medoids</i> | 2 | 0.347 |

B. Pembahasan

Dalam penelitian ini untuk mengelompokan provinsi yang memiliki kasus penyakit covid-19 yang ada di indonesia menggunakan algoritma *K-Means* dan *K-Medoids* dengan menggunakan metodologi knowledge discovery in database (KDD) yang terdiri beberapa tahapan yaitu, data selection, data preprocessing, data transformation, data mining, dan knowledge interpretation atau evaluation. Serta menggunakan tools RapidMiner Studio dalam melakukan proses pengolahan data.

Data selection merupakan proses pemilihan data yang akan dilakukan dalam proses data mining, kemudian dilanjutkan pada tahap data preprocessing dimana data dilakukan pengecekan apabila terdapat data yang hilang atau terdapat typography (kesalahan input) maka data harus ditangani terlebih dahulu, kemudian dilakukan data transformation dimana data yang telah dibersihkan melalui tahap data preprocessing atribut-atribut yang digunakan diubah menjadi standarisasi data yang masih berbentuk kategori juga ditransformasikan pada bentuk integer dan polynomial, selanjutnya adalah proses data mining yaitu data tersebut diekstraksi bentuk dan polanya dengan menggunakan algoritma *K-Means* dan *K-Medoids*, dan hasil dari nilai Device Boudin [14] cluster yang tersedia di tools RapidMiner studio pengukuran jarak tersebut digunakan pada proses data mining, sedangkan untuk melakukan proses mining menggunakan algoritma *K-Means* dan *K-Medoids* [15] untuk menampilkan hasil clustering dari cluster. Selanjutnya dilakukannya evaluasi dari hasil model data mining yang telah didapatkan yaitu untuk mengukur optimalisasi dari cluster yang dihasilkan pada tools RapidMiner dengan menggunakan nilai Device Boudin untuk mengetahui nilai evaluasi berupa nilai silhouette.

Adapun cluster optimal yang dihasilkan pada penelitian ini adalah pada dataset Penyakit covid-19 di indonesia Tahun 2019-2021 menghasilkan 2 cluster dan optimal cluster pada cluster 2. Parameter untuk mengetahui kualitas suatu cluster dapat dilakukan dengan menggunakan Silhouette Coefficient yang bertujuan untuk mengukur kualitas dari cluster. Dari dataset yang digunakan yaitu dataset Penyakit covid-19 di Indonesia Tahun 2019-2021 dengan hasil perbandingan nilai Silhouette Coefficient dengan metode *K-Means* adalah sebesar 0,207. Sedangkan Nilai Silhouette Coefficient dengan metode *K-Medoids* adalah sebesar 0,347 yang menyatakan bahwa struktur cluster yang dihasilkan pada kedua metode pengelompokan ini adalah sama-sama *weak structure* atau terstruktur dengan lemah.

V. KESIMPULAN

Hasil pengelompokan yang terbentuk dengan menggunakan algoritma K-Means yaitu untuk cluster 1 beranggotakan 2 wilayah, sedangkan untuk cluster 2 sebanyak 32 wilayah. Hasil pengelompokan yang terbentuk dengan menggunakan algoritma K-Medoids yaitu untuk cluster 1 beranggotakan 4 wilayah, sedangkan untuk cluster 2 sebanyak 30 wilayah. Hasil Cluster dievaluasi menggunakan Nilai Silhouette Coefficient dengan metode K-Means adalah sebesar 0,207. Sedangkan Nilai Silhouette Coefficient dengan metode K-Medoids adalah sebesar 0,347 yang menyatakan bahwa metode K-Medoids menghasilkan nilai SC lebih besar dari pada K-Means, sehingga K-Medoids dapat memberikan hasil pengelompokan yang lebih baik.

DAFTAR PUSTAKA

- [1] S. Ahmad, "A review of COVID-19 (Coronavirus disease-2019) diagnosis, treatments and prevention," *Eurasian J. Med. Oncol.*, 2020. <https://doi.org/10.14744/ejmo.2020.90853>.
- [2] 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.
- [3] A. S. Abdelhafiz et al., "Knowledge, Perceptions, and Attitude of Egyptians Towards the Novel Coronavirus Disease (COVID-19)," *J. Community Health*, vol. 45, no. 5, pp. 881–890, Oct. 2020, doi: 10.1007/s10900-020-00827-7.
- [4] A. Alqurneh, A. Mustapha, and N. M. Sharef, "A Partitioning-based Approach for Clustering COVID-19 Drugs and Co-Medication for Safe Use," *Int. J. Integr. Eng.*, vol. 12, no. 5, pp. 224–232, 2020.
- [5] M. L. F. Nascimento, "A multivariate analysis on spatiotemporal evolution of Covid-19 in Brazil," *Infect. Dis. Model.*, vol. 5, pp. 670–680, 2020, doi: 10.1016/j.idm.2020.08.012.
- [6] B. Riyanto, "Penerapan Algoritma K-Medoids Clustering Untuk Pengelompokan Penyebaran Diare Di Kota Medan (Studi Kasus: Kantor Dinas Kesehatan Kota Medan)," *KOMIK Konf. Nas. Teknol. Inf. Dan Komput.*, vol. 3, no. 1, 2019.
- [7] P. Arora, Deepali, and S. Varshney, "Analysis of K-Means and K-Medoids Algorithm For Big Data," *Procedia Comput. Sci.*, vol. 78, pp. 507–512, 2016, doi: 10.1016/j.procs.2016.02.095.
- [8] P. Kumar, B. Negi, and N. Bhoi, "Detection of Healthy and Defected Diseased Leaf of Rice Crop using K-Means Clustering Technique," *Int. J. Comput. Appl.*, vol. 157, no. 1, pp. 24–27, Jan. 2017, doi: 10.5120/ijca2017912601.
- [9] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. Dan Teknol. Inf.*, vol. 4, no. 2, p. 64, Dec. 2018, doi: 10.24014/coreit.v4i2.4498.
- [10] S. Sindi, W. R. O. Ningse, I. A. Sihombing, F. I. R. Zer, and D. Hartama, "Analisis algoritma k-medoids clustering dalam pengelompokan penyebaran covid-19 di indonesia," *JurTI J. Teknol. Inf.*, vol. 4, no. 1, pp. 166–173, 2020.
- [11] Velmurugan, "Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points," *J. Comput. Sci.*, vol. 6, no. 3, pp. 363–368, Mar. 2010, doi: 10.3844/jcssp.2010.363.368.
- [12] A. R. Mansyur, "Dampak covid-19 terhadap dinamika pembelajaran di indonesia," *Educ. Learn. J.*, vol. 1, no. 2, pp. 113–123, 2020.
- [13] R. K. Klimberg, S. Ratick, and H. Smith, "A Novel Approach to Forecasting Regression and Cluster Analysis," in *Advances in Business and Management Forecasting*, Emerald Publishing Limited, 2017.
- [14] F. Mahmuda, M. A. R. Sitorus, H. Widyastuti, and D. E. Kurniawan, "Clustering Profil Pengunjung Perpustakaan Menggunakan Algoritma K-means," *J. Appl. Informatics Comput.*, vol. 1, no. 1, pp. 14–21, 2017.
- [15] L. Purba, S. Saifullah, and R. Dewi, "Pengelompokan Kasus Penyakit Aids Berdasarkan Provinsi Dengan Data Mining K-Medoids Clustering," *KOMIK Konf. Nas. Teknol. Inf. Dan Komput.*, vol. 3, no. 1, 2019.