



SNESTIK

Seminar Nasional Teknik Elektro, Sistem Informasi,
dan Teknik Informatika

<https://ejurnal.itats.ac.id/snestik> dan <https://snestik.itats.ac.id>



Informasi Pelaksanaan :

SNESTIK IV - Surabaya, 27 April 2024

Ruang Seminar Gedung A, Kampus Institut Teknologi Adhi Tama Surabaya

Informasi Artikel:

DOI : 10.31284/p.snestik.2024.5938

Prosiding ISSN 2775-5126

Fakultas Teknik Elektro dan Teknologi Informasi-Institut Teknologi Adhi Tama Surabaya
Gedung A-ITATS, Jl. Arief Rachman Hakim 100 Surabaya 60117 Telp. (031) 5945043
Email : snestik@itats.ac.id

Pengelompokan Penyebaran Covid-19 di Jawa Timur dengan Metode K-Medoids Clustering

Ghofur, Tutuk Indriyani*
Institut Teknologi Adhi Tama Surabaya
tutuk@itats.ac.id

ABSTRACT

Covid-19 transmission can only be prevented by being disciplined by staying at home. The risk of transmission can also be avoided and the Covid-19 problem in Indonesia can be immediately resolved with the best policies. The total spread of Covid-19 in Indonesia has now reached 14,265, and if the spread of Covid-19 recovers in Indonesia it will reach 2,881 with a death rate of 991. The Ministry of Health of the Republic of Indonesia (Kemenkes RI) classifies them as people under monitoring, patients under supervision, people without symptoms and confirmed patients if a positive RTPCR COVID-19 result is obtained with any symptoms. The algorithm used to cluster the Covid-19 virus is K-Medoids clustering. The dataset in this research is the Covid-19 dataset. The research result show that the spread of COVID-19 includes 3 level, namely high, medium and low. the amount of data in the 0th cluster has a total of 1032 data and a centroid point of 487 for New Cases 0.15 and with New Deaths -0.26 while the 1st cluster has a total of 1032 data and also a centroid point of 267 for New Cases -0.08 and on New Deaths -0.66. The 2nd cluster has a total of 1032 data and a centroid point of 757 with New Cases -0.11 and with New Deaths -0.45, from this dataset with 3 clusters it is obtained that cluster 2 is the best cluster for handling Covid-19 in East Java.

Keywords: Covid-19; Clustering; K-medoids; silhouette coefficient

ABSTRAK

Penularan covid-19 hanya bisa dicegah dengan disiplin tinggal di rumah. Resiko penularan juga dapat dihindari dan masalah covid-19 di Indonesia dapat segera diatasi dengan kebijakan terbaik. Total penyebaran covid-19 di Indonesia kini telah mencapai 14.265, dan jika penyebaran covid-19 pulih di Indonesia akan mencapai 2.881 dengan tingkat kematian 991. Kementerian Kesehatan Republik Indonesia (Kemenkes RI) mengklasifikasikan menjadi orang dalam pemantauan (ODP), pasien dalam pengawasan (PDP), orang tanpa gejala (OTG) dan pasien terkonfirmasi bila didapatkan hasil RTPCR COVID- 19 positif dengan gejala apapun. Algoritma yang digunakan untuk mengclustering virus Covid-19 adalah K-Medoids clustering. Dataset dalam penelitian ini adalah dataset Covid-19. Hasil dari penelitian ini, penyebaran Covid-19 dikelompokkan kedalam 3 tingkat yaitu tinggi, sedang, rendah. Jumlah penyebaran pada cluster ke-0 memiliki jumlah data sebanyak 1032 dan titik centroid 487 pada New Cases 0,15 dan dengan New Deaths -0,26 sedangkan pada kluster ke-1 memiliki jumlah 1032 data dan juga titik centroid 267 untuk New Cases -0,08 dan pada New Deaths -0,66. Pada cluster ke-2 memiliki jumlah data 1032 dan titik centroid 757 pada New Cases -0,11 dan dengan New Deaths -0,45, dari dataset tersebut dengan 3 cluster diperoleh cluster 2 merupakan kluster yang terbaik yang perlu dilakukan penanganan covid-19 di Jawa Timur.

Kata kunci: Covid-19; Clustering; K-medoids; silhouette coefficient

PENDAHULUAN

Pada awal Maret 2020, Indonesia dilanda pandemi covid-19. Virus ini pertama kali muncul di wilayah Wuhan China pada akhir Desember 2019. pemerintah dan warga melakukan Langkah social distancing untuk memutus mata rantai penyebaran covid-19 yang terus bertambah setiap harinya. Penularan covid-19 hanya bisa dicegah dengan disiplin tinggal di rumah. Resiko penularan juga dapat dihindari dan masalah covid-19 di Indonesia dapat segera diatasi dengan kebijakan terbaik. Total penyebaran covid-19 di Indonesia kini telah mencapai 14.265, dan jika penyebaran covid-19 pulih di Indonesia akan mencapai 2.881 dengan tingkat kematian 991. Kementerian Kesehatan Republik Indonesia (Kemenkes RI) mengklasifikasikan menjadi orang dalam pemantauan (ODP), pasien dalam pengawasan (PDP), orang tanpa gejala (OTG) dan pasien terkonfirmasi bila didapatkan hasil RTPCR COVID- 19 positif dengan gejala apapun (Perhimpunan Dokter Paru Indonesia, 2020). Dua permasalahan COVID- 19 awal pertama kali dilaporkan di Indonesia pada Maret 2020. Informasi per 31 Maret 2020 menampilkan 1. 528 permasalahan terkonfirmasi serta 136 kematian. Tingkatan kematian COVID- 19 Indonesia merupakan 8.9% paling tinggi di Asia Tenggara. Dan per 14 September 2020, kasus di Indonesia dilaporkan meningkat menjadi 221.523 kasus, masih menjadi angka kematian tertinggi kedua di Asia Tenggara setelah Filipina. Permasalahan bagaimana mengelompokkan tingkat penyebaran COVID - 19 di Indonesia bersumber pada informasi COVID- 19 tiap- tiap Provinsi, dengan tujuan untuk mengetahui tingkat penyebaran di setiap daerah daerah dalam satu propinsi. Daerah Indonesia yang luas membolehkan bagian-bagian dapat dikelompokkan berdasarkan wilayah Indonesia. Oleh karena itu, mengelompokkan tingkat distribusi ini memberitahu kita wilayah mana yang memiliki lebih banyak kasus dan sebaliknya. Hal ini hendak menolong pemerintah dan serta warga dengan membagikan data sepanjang masa penyebaran COVID- 19 di tiap-tiap wilayah.

Dalam penelitian ini K-Medoids adalah metode yang tersedia untuk pengelompokan. Metode ini merupakan clustering terdistribusi yang tujuannya adalah untuk menemukan sekumpulan K-cluster dalam data yang paling menggambarkan objek dalam kumpulan data

tersebut[1]. Kelebihan metode ini adalah dapat mengatasi kelemahan metode K- Means yang sensitif terhadap outlier.

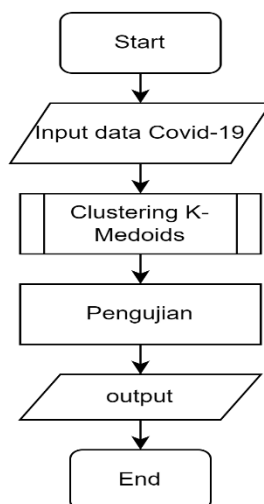
Ada dua metode yang digunakan dalam Teknik clustering yaitu K-Means adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan unsupervised learning dan menggunakan metode yang mengelompokkan data berbagi partisi[2]. Sedangkan K-Medoids merupakan metode pengelompokan positif yang tujuannya adalah untuk menemukan sekumpulan k cluster dari data yang paling menggambarkan objek dalam kumpulan data[3]. Metode K-Means dan K-Medoids sama-sama memiliki kelebihan yaitu memiliki waktu komputasi yang cepat namun Metode K-Means memiliki kelemahan sensitif terhadap data outlier. Berbeda dengan Metode K-Medoids yang hampir dapat bekerja pada setiap jenis data matriks dan mampu mengatasi outlier[4].

Penelitian ini dilakukan dari review penelitian penelitian sebelumnya meliputi: K-Medoids and K-Means algorithms for grouping region of disability distribution on children [3], Algoritma K-Means dan K-Medoids dalam pengelompokan komoditas peternakan di Jawa Tengah, Algoritma K-Means & K-Medoids untuk Big Data [4] dan penerapan metode K-Medoids pada kondisi jalan cluster di kota Semarang nilai koefisien siluet rata-rata sebesar 0,5743 teramati pada lokasi cluster dan mendekati 1 pada kedua lokasi cluster[5]. Berdasarkan uraian diatas, penulis memutuskan untuk mempelajari penyebaran COVID- 19 menggunakan algoritma K-Medoids atau Partitioning Around Medoids (PAM).

Dalam penelitian ini menerapkan algoritma K-Medoids untuk pengelompokan data COVID-19 dari tahun 2020 sampai tahun 2021, dengan memiliki delapan atribut yaitu New Cases adalah kasus positif harian baru , New Deaths adalah kematian harian baru, New Recovered adalah harian baru pulih, New Active Cases adalah kasus aktif harian baru, Total Cases adalah total kasus positif hingga tanggal terkait, Total Deaths adalah total kematian hingga tanggal terkait, Total Recovered adalah total pulih hingga tanggal terkait, Total Active Cases adalah total aktif hingga tanggal terkait. Tujuan dari penelitian ini adalah mengetahui tingkat penyebaran COVID-19 di Provinsi Jawa Timur pada periode data berdasarkan hasil clustering menggunakan k-medoids. Diharapkan dapat memberikan informasi kepada pemerintah dan masyarakat tentang penyebaran COVID- 19 di Indonesia dan dapat memberikan gambaran penerapan algoritma K-Medoids untuk menghasilkan tingkat penyebaran COVID- 19 di Jawa Timur.

METODE

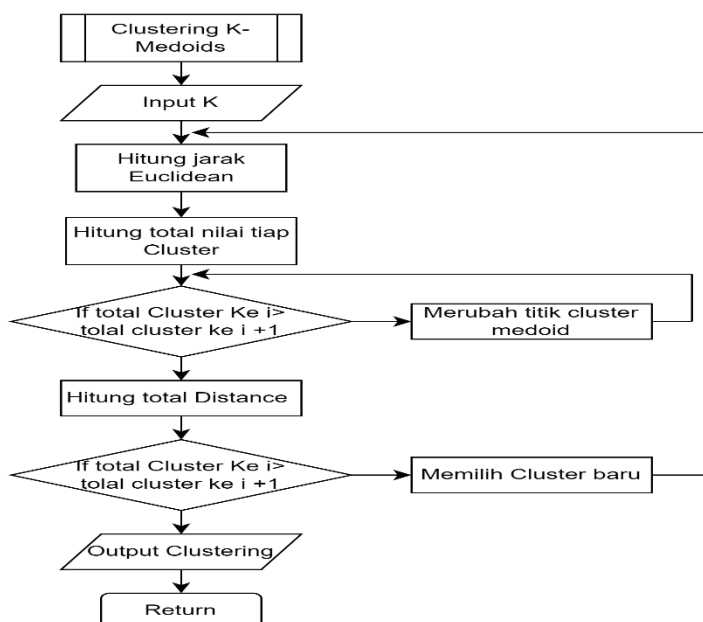
Dalam penelitian ini akan dikembangkan sistem yang menerapkan clustering data Covid-19 Di Indonesia. Metode *K-Medoids* untuk mengelompokkan Data Covid-19 Di Indonesia. Dalam penelitian ini nantinya dapat mengetahui pengelompokkan penyebaran Covid-19 di Indonesia. Untuk melakukan Langkah awal yang harus dilakukan adalah melakukan inisialisasi pusat cluster sebanyak 2 cluster dari data sampel, menentukan jumlah maksimal iterasi dan menentukan jumlah ambang batas perubahan nilai fungsi objektif (jika dibutuhkan maka menentukan nilai centroid). Langkah selanjutnya menghitung jarak terdekat dengan persamaan Euclidean distance. Setelah mendapat hasil pusat cluster selanjutnya dilakukan perhitungan fungsi objektif dengan nilai awal fungsi objektif yaitu 0, jika objektif masih diatas nilai ambang batas yang telah ditentukan maka iterasi belum tercapai dan dilanjutkan dengan iterasi berikutnya. Jika iterasi sudah maximum atau fungsi objektif sudah memenuhi maka proses cluster dihentikan.



Gambar 1. Flowchart sistem penelitian

Penjelasan tahapan setiap penelitian sebagai berikut:

1. Data yang digunakan adalah dari dataset Kaggle oleh akun yang Bernama Hedratno pada tanggal 03 Februari 2020. Data ini berformat csv. Informasi ini berisi informasi berbeda tentang COVID-19 di setiap provinsi mulai 01 Maret 2020 hingga 30 Juni 2021 Pengelompokan. Data yang digunakan dalam penelitian ini adalah 16284 data. meliputi variabel new cases, new deaths, new recovered, new active cases, total cases, total deaths, total recovered dan total active cases.
2. Proses *K-Medoids*, tahapan pada metode ini dijelaskan pada Gambar 2.



Gambar 2. Flowchart Metode K-Medoids

Langkah-langkah dalam penyelesaian perhitungan manual data mining menggunakan K-medoids sebagai berikut :

- Inisialisasi pusat *cluster* sebanyak 2 *cluster* dari data sampel.
- Untuk pemilihan sampel setiap medoid dipilih secara acak dan diasumsikan Total Cases dan Total Recovered sebagai medoid awal.
- Menghitung nilai jarak terdekat dengan persamaan *Euclidean distance*. Untuk menghitung jarak antara titik objek menggunakan *Euclidean distance*. rumus untuk menghitung jarak menggunakan Persamaan 2.1.
- Setelah didapatkan hasil jarak dari setiap objek pada iterasi ke-I, maka lanjut ke iterasi ke-i+1. Kandidat medoid baru (non-medoid) pada iterasi ke-i+1. Hitung ulang jarak setiap objek pada iterasi ke-i+1 dengan menggunakan medoid. Maka mendapatkan hasil keseluruhan dari iterasi ke-i+1.
- Hitung Total Simpangan (S) setelah didapatkan hasil nilai jarak iterasi ke-I dan iterasi ke-i+1 hitung total simpangan (S) dengan mencari selisih dari total cost baru dan cost lama. Dengan ketentuan jika $S < 0$, maka tukar nilai objek dengan menentukan medoid baru.
- Ulangi langkah sebelumnya hingga $S > 0$, sehingga didapatkan cluster beserta anggota cluster masing-masing. Setelah mendapatkan hasil nilai jarak iterasi ke-i+1 hitung total simpangan (S) dengan mencari selisih dari nilai total *cost* baru dengan nilai Total *cost* lama. Dengan ketentuan jika $S < 0$, maka tukar nilai objek dengan menentukan medoids baru.

$$S = \text{Total cost baru} - \text{Total cost lama}$$

Dengan nilai $S > 0$, maka proses cluster di hentikan. Sehingga diperoleh anggota tiap cluster. Dibawah ini merupakan contoh perhitungan manual metode *K-Medoids* dari data Covid-19 dengan menggunakan medoid awal sebagai berikut[5]:

Menghitung nilai jarak terdekat dengan persamaan *Euclidean Distance*.

$$\text{Rumus} = d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Dimana : X1 = variabel total cases

X2 = centroid 1

Y1 = variabel total recovered

Y2 = centroid 2

3. Pengujian Menggunakan Silhouette Coefficient

Melakukan pengujian menggunakan Silhouette Coefficient. Adapun Langkah untuk menghitung Silhouette Coefficient diawali dengan mencari jarak rata-rata data ke-I dengan semua data dalam kluster yang sama[2], [6], disini anggap saja data ke-I terletak pada cluster A. rumus dari a(i) terdapat di persamaan berikut:

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

Dimana,

A = total data di kluster A

Kemudian tentukan nilai b(i) yang merupakan nilai minimal dari rata-rata jarak data ke-I antara semua data pada setiap kluster. Sekarang mari kita perhatikan kluster lain selain A dan kluster C. untuk menghitung jarak rata-rata data ke-I dengan semua data di cluster C. menggunakan persamaan berikut

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

Dimana,

C = total data di cluster C

Setelah menghitung $d(i,C)$ untuk semua kluster, Langkah selanjutnya untuk memilih jarak paling minimum sebagai nilai $b(i)$, menggunakan persamaan berikut.

$$b(i) = \min_{C \neq A} d(i, C)$$

Cluster terbaik kedua untuk data ke- i , setelah kluster A , adalah cluster B , yang disebut sebagai tetangga dari data ke- i jika kluster B memiliki nilai jarak minimal, yaitu $d(i,B)$. Silhouette Coefficient ditentukan pada Langkah terakhir setelah $a(i)$ dan $b(i)$ diketahui.

Jika nilai indeks Silhouette mendekati 1 berarti proses clustering berhasil, dan jika mendekati 0, berarti prosedur clustering tidak berhasil. Berikut rumus Silhouette Coefficient, pada persamaan berikut[1].

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

$S(i)$ = nilai Silhouette Coefficient pada object ke- i

(i) = object yang akan diteliti

$a(i)$ = rata-rata i jarak i antar anggota i dalam kluster yang sama

$b(i)$ = nilai minimal jarak rata-rata antara item ke- i dengan objek pada cluster lain.

Nilai $s(i)$ terletak diantara -1 dan 1, di mana setiap nilai ditafsirkan sebagai berikut:

$s(i) = 1 \Rightarrow$ data ke- i digolongkan dengan baik (dalam A)

$s(i) = 0 \Rightarrow$ data ke- i terletak di Tengah antara dua cluster (A dan B)

$s(i) = -1 \Rightarrow$ data ke- i tergolong lemah (dekat ke cluster B daripada A)

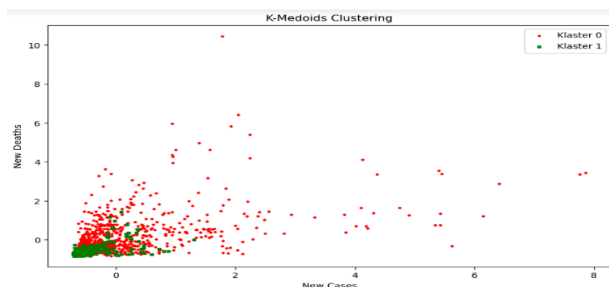
HASIL DAN PEMBAHASAN

Tahapan uji coba berdasarkan jumlah Cluster = 2

Pada gambar 4.5 dibawah merupakan hasil uji coba pada dataset covid-19 di jawa timur dan dapat ditarik kesimpulan bahwa jumlah data pada cluster ke-0 memiliki jumlah data sebanyak 1032 data dan titik centroid 115 pada New Cases 1,43 dan dengan New Deaths 0,14 sedangkan pada kluster ke-1 memiliki jumlah 1032 data dan juga titik centroid 187 untuk New Cases 1,65 dan pada New Deaths 0,42. Maka hasil yang dapat disimpulkan dari dataset tersebut dengan dua cluster diperoleh kluster 0 merupakan cluster yang terbaik yang perlu dilakukan penanganan covid-19 di jawa timur dengan segera, serta diperoleh titik di setiap penyebaran covid-19 tiap cluster yang sesuai dengan titik centroid yang telah didapatkan

Tabel 1 Hasil uji coba data covid-19 kluster = 2

| Data | C1 | C2 | Kluster |
|------|----------|----------|---------|
| 0 | 4,670880 | 5,365578 | 0 |
| 1 | 2,164707 | 2,844248 | 0 |
| 2 | 4,753694 | 5,448770 | 0 |
| ... | ... | ... | ... |
| 1030 | 6,284592 | 6,028297 | 1 |
| 1031 | 5,407748 | 5,407298 | 1 |



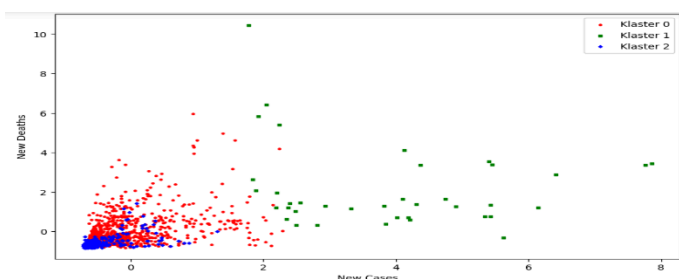
Gambar 3. Hasil uji coba pada data covid-19 dengan jumlah cluster=2

Tahapan uji coba dengan jumlah Cluster = 3

Pada gambar 4 dibawah ini merupakan hasil dari uji coba pada dataset covid-19 di jawa timur dan dapat ditarik kesimpulan bahwa jumlah data pada cluster ke-0 memiliki jumlah data sebanyak 1032 dan titik centroid 487 pada New Cases 0,15 dan dengan New Deaths -0,26 sedangkan pada kluster ke-1 memiliki jumlah 1032 data dan juga titik centroid 267 untuk New Cases -0,08 dan pada New Deaths -0,66. Pada cluster ke-2 memiliki jumlah data 1032 dan titik centroid 757 pada New Cases -0,11 dan dengan New Deaths -0,45. Maka hasil yang dapat disimpulkan dari dataset tersebut dengan 3 cluster diperoleh cluster 2 merupakan cluster yang terbaik yang perlu dilakukan penanganan covid-19 di jawa timur dengan segera serta diperoleh titik setiap lokasi covid-19 di jawa timur yang sesuai dengan titik centroid yang telah didapatkan.

Tabel 2. Hasil uji coba pada data covid-19 dengan K=3

| Data | C1 | C2 | C3 | Kluster |
|------|----------|----------|----------|---------|
| 0 | 4,463794 | 2,949624 | 4,860164 | 1 |
| 1 | 2,569108 | 1,540308 | 3,198401 | 1 |
| 2 | 4,567511 | 2,988677 | 4,952308 | 1 |
| ... | ... | ... | ... | ... |
| 548 | 4,265344 | 2,654992 | 4,659114 | 1 |
| 549 | 2,251122 | 2,053758 | 2,514028 | 1 |



Gambar 4. Hasil uji coba terhadap dataset dengan jumlah cluster = 3

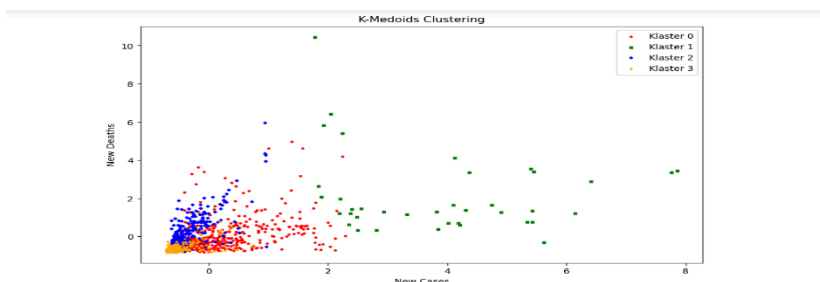
Tahapan uji coba dengan jumlah Cluster=4

Pada gambar 5 dibawah ini adalah hasil uji coba pada dataset covid-19 di jawa timur dan dapat ditarik kesimpulan bahwa jumlah data pada ke-0 memiliki jumlah data sebanyak 1032 dan titik centroid 483 pada New Cases -0,08 dan dengan New Deaths -0,63 sedangkan pada kluster ke-1 memiliki jumlah data 1032 data juga titik centroid 621 untuk New Cases 0,55 dan pada New Deaths -0,32. Pada cluster ke-2 memiliki jumlah data 1032 dan titik centroid 211 pada New Cases 1,33 dan dengan New Deaths 0,51. Kemudian pada kluster ke-3 mempunyai jumlah

data sebanyak 1032 dengan titik centroid 811 pada New Cases -0,25 dan dengan New Deaths -0,23. Maka hasil yang dapat disimpulkan dari dataset tersebut dengan 4 cluster diperoleh cluster 2 merupakan cluster yang terbaik yang perlu dilakukan penanganan covid-19 di Jawa Timur dengan segera, serta diperoleh titik setiap lokasi covid-19 di Jawa Timur yang sesuai dengan titik centroid yang telah didapatkan.

Tabel 3. Hasil uji coba data covid-19 cluster = 4

| Dat | C1 | C2 | C3 | C4 | Kluster |
|-----|----------|----------|----------|----------|---------|
| a | | | | | |
| 0 | 3,010163 | 3,712711 | 5,012031 | 4,926377 | 0 |
| 1 | 1,443151 | 1,714469 | 2,436525 | 3,223969 | 0 |
| 2 | 3,059785 | 3,752462 | 5,096580 | 5,019358 | 0 |
| ... | ... | ... | ... | ... | ... |
| 304 | 4,556123 | 4,607184 | 4,542575 | 5,443650 | 2 |
| 305 | 2,431504 | 3,012368 | 3,981634 | 3,415773 | 0 |



Gambar 5. Hasil uji coba pada data covid-19 dengan jumlah cluster = 4

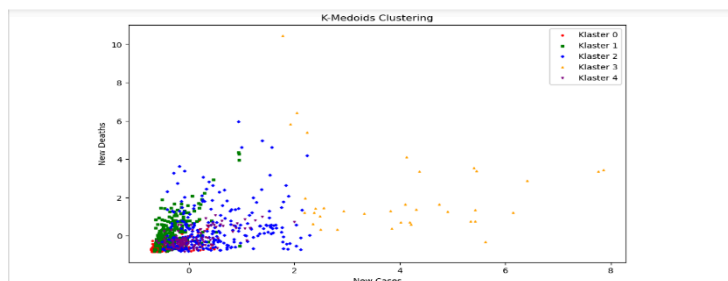
Tahapan uji coba dengan jumlah cluster = 5

Pada gambar 4.5 dibawah ini adalah hasil uji coba pada dataset covid-19 di Jawa Timur dan dapat ditarik kesimpulan bahwa jumlah data pada ke-0 memiliki jumlah data sebanyak 1032 dan titik centroid 523 pada New Cases -0,16 dan dengan New Deaths -0,54 sedangkan pada cluster ke-1 memiliki jumlah data 1032 data juga titik centroid 734 untuk New Cases -0,62 dan pada New Deaths -0,72. Pada cluster ke-2 memiliki jumlah data 1032 dan titik centroid 59 pada New Cases -0,14 dan dengan New Deaths 1,49. Kemudian pada cluster ke-3 mempunyai jumlah data sebanyak 1032 dengan titik centroid 953 pada New Cases -0,07 dan dengan New Deaths 0,72. Dan pada cluster ke-4 memiliki jumlah data sebanyak 1032 dengan titik centroid 510 pada New Cases -0,55 dan dengan New Deaths -0,82. Maka hasil yang dapat disimpulkan dari dataset tersebut dengan cluster diperoleh cluster 0 merupakan cluster yang terbaik yang perlu dilakukan penanganan covid-19 di Jawa Timur dengan segera, serta diperoleh titik setiap lokasi covid-19 di Jawa Timur yang sesuai dengan titik centroid yang telah didapatkan.

Tabel 4. Hasil uji coba data covid-19 cluster = 5

| Data | C1 | C2 | C3 | C4 | C5 | Kluster |
|------|----------|----------|----------|----------|----------|---------|
| 0 | 4,435954 | 0,448449 | 2,892974 | 3,858320 | 0,517660 | 1 |
| 1 | 2,800497 | 2,625830 | 2,567607 | 2,945157 | 2,649026 | 2 |
| 2 | 4,53947 | 0,557690 | 2,959471 | 3,943855 | 0,617502 | 1 |
| ... | ... | ... | ... | ... | ... | ... |
| 1030 | 7,701005 | 8,867666 | 6,344805 | 6,415778 | 8,909644 | 2 |
| 1031 | 6,003344 | 6,750725 | 3,883307 | 3,979098 | 6,795947 | 2 |

Gambar 6.
Hasil uji coba data
covid-19 cluster = 5



KESIMPULAN

Berdasarkan dari hasil pengujian dan evaluasi terhadap sistem pengelompokkan data covid-19 di jawa timur menggunakan metode K-medoids, maka ditarik kesimpulan sebagai berikut : Pada penelitian ini menggunakan Metode K-medoids yang telah digunakan untuk mengelompokkan data covid-19 berdasarkan dataset yang telah digunakan. Setelah itu dilakukan pengujian menggunakan silhouette coefficient jumlah data pada kluster ke-0 memiliki jumlah data sebanyak 1032 dan titik centroid 523 pada New Cases -0,16 dan dengan New Deaths -0,54 sedangkan pada kluster ke-1 memiliki jumlah data 1032 data juga titik centroid 734 untuk New Cases -0,62 dan pada New Deaths -0,72. Pada cluster ke-2 memiliki jumlah data 1032 dan titik centroid 59 pada New Cases -0,14 dan dengan New Deaths 1,49. Kemudian pada kluster ke-3 mempunyai jumlah data sebanyak 1032 dengan titik centroid 953 pada New Cases -0,07 dan dengan New Deaths 0,72. Dan pada cluster ke-4 memiliki jumlah data sebanyak 1032 dengan titik centroid 510 pada New Cases -0,55 dan dengan New Deaths -0,82. Maka hasil yang dapat disimpulkan dari dataset tersebut dengan kluster diperoleh kluster 0 merupakan kluster yang terbaik yang perlu dilakukan penanganan covid-19 di jawa timur.

DAFTAR PUSTAKA

- [1] A. Sobrinho Campolina Martins, L. Ramos de Araujo, and D. Rosana Ribeiro Penido, "K-Medoids clustering applications for high-dimensionality multiphase probabilistic power flow," *International Journal of Electrical Power and Energy Systems*, vol. 157, Jun. 2024, doi: 10.1016/j.ijepes.2024.109861.
- [2] T. Indriyani, M. Kurniawan, G. E. Yuliasuti, A. Rachman, C. N. Prabiantissa and R. K. Hapsari, "An Improve KNN Method for Classification of Sexually Transmitted Diseases," *2023 Sixth International Conference on Vocational Education and Electrical Engineering (ICVEE)*, Surabaya, Indonesia, 2023, pp. 315-319, doi: 10.1109/ICVEE59738.2023.10348340.
- [3] S. Akter, F. Reza, and M. Ahmed, "Convergence of Blockchain, k-medoids and homomorphic encryption for privacy preserving biomedical data classification," *Internet of Things and Cyber-Physical Systems*, vol. 2, pp. 99–110, Jan. 2022, doi: 10.1016/j.iotcps.2022.05.006.
- [4] N. Sureja, B. Chawda, and A. Vasant, "An improved K-medoids clustering approach based on the crow search algorithm," *Journal of Computational Mathematics and Data Science*, vol. 3, Jun. 2022, doi: 10.1016/j.jcmds.2022.100034.

- [5] Q. Li, X. Zhang, T. Ma, D. Liu, H. Wang, and W. Hu, "A Multi-step ahead photovoltaic power forecasting model based on TimeGAN, Soft DTW-based K-medoids clustering, and a CNN-GRU hybrid neural network," *Energy Reports*, vol. 8, pp. 10346–10362, Nov. 2022, doi: 10.1016/j.egy.2022.08.180.

- [6] T. Indriyani, S. Nurmuslimah, A. Taufiqurrahman, R. K. Hapsari, C. N. Prabiantissa, and A. Rachmad, "Steganography on Color Images Using Least Significant Bit (LSB) Method," 2023, pp. 39–48. doi: 10.2991/978-94-6463-174-6_5.