

IMPLEMENTASI FUZZY C-MEANS UNTUK PENGELOMPOKAN DAERAH BERDASARKAN PERSEBARAN PENULARAN COVID-19

Gibran Satya Nugraha*¹, Ramaditia Dwiyanaputra², Fitri Bimantoro³, Arik Aranta⁴

^{1,2,3,4}Universitas Mataram, Mataram

Email: ¹gibransn@unram.ac.id, ²rama@unram.ac.id, ³bimo@unram.ac.id, ⁴arikaaranta@unram.ac.id

*Penulis Korespondensi

(Naskah masuk: 04 November 2021, diterima untuk diterbitkan: 27 Februari 2023)

Abstrak

Peningkatan kasus Covid-19 di Indonesia memberikan rasa khawatir bagi hampir seluruh masyarakat, Dilihat dari persebaran tiap provinsi untuk kasus positif, sembuh, dan meninggal tidak menunjukkan sebuah grafik yang linier. Seperti pada data harian kasus per provinsi di akhir bulan April 2021 dimana kasus positif dan sembuh terbanyak terdapat pada Provinsi DKI Jakarta, untuk kasus meninggal Provinsi Jawa Timur berada di posisi pertama, dan di posisi empat untuk kasus positif dan meninggal. Data persebaran yang abstrak ini membuat pengelompokan persebaran Covid-19 di Indonesia menjadi sukar untuk dilakukan. Penelitian ini mengelompokkan provinsi-provinsi berdasarkan persebaran Covid-19 di Indonesia dengan cara mengimplementasikan metode Fuzzy C-means serta metode Elbow. Fuzzy C-means adalah metode pengelompokan berbasis *fuzzy* yang dapat melakukan persebaran data pada seluruh *cluster* berdasarkan derajat keanggotaan yang dimilikinya. Sedangkan untuk menentukan jumlah *cluster* terbaik akan diimplementasikan metode Elbow. Metode Elbow membandingkan perbandingan hasil *sum square error* (SSE) dari setiap *cluster* dan mendapatkan jumlah *cluster* terbaik dari perubahan nilai SSE yang signifikan atau membentuk siku (*elbow*). Penggunaan Fuzzy c-means sebagai metode pengelompokan untuk mencari tahu seberapa besar pengaruh yang dimiliki setiap data terhadap masing-masing *cluster*. Karera metode-metode sebelumnya yang digunakan pada objek yang sama hanya melakukan pengelompokan saja secara tegas, tanpa memperhatikan besarnya pengaruh sebuah data terhadap seluruh *cluster*. Pengelompokan dilakukan menjadi tiga buah *cluster* atau kelompok berdasarkan parameter kasus positif, sembuh, dan meninggal Covid-19 per 27 April 2021. *Cluster* 1 hanya terdiri tiga provinsi yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur. *Cluster* 2 DKI Jakarta, dan sisanya masuk ke *cluster* 3.

Kata kunci: *fuzzy c-means, covid-19, cluster, kelompok*

IMPLEMENTATION OF FUZZY C-MEANS FOR GROUP PROVINCE BASED ON THE SPREAD OF COVID-19

Abstract

The increase in Covid-19 cases in Indonesia raises concerns for all parties, When viewed for the distribution of each province, positive, recovered and dead cases do not show a linear graph. As in the daily data of cases per province at the end of April 2021 where the most positive and recovered cases were in DKI Jakarta Province, while for dead cases, East Java Province was in first position, and in fourth position for positive and dead cases. This abstract distribution data makes it difficult to classify the distribution of Covid-19 in Indonesia. This study will group provinces based on the spread of Covid-19 in Indonesia using the Fuzzy C-means method and the Elbow method. Fuzzy C-means is a fuzzy-based grouping method that allows all data to be members of all clusters formed with their respective degrees of membership. Meanwhile, to determine the best number of clusters, the Elbow method will be implemented. The Elbow method compares the sum square error (SSE) results from each cluster and gets the best number of clusters from a significant change in the SSE value or forms an elbow. The use of Fuzzy c-means as a grouping method to find out how much influence each data has on each cluster. Because the previous methods used on the same object only grouped it explicitly, without paying attention to the effect of one data on the entire cluster. The grouping was carried out into three clusters or groups based on the parameters of positive cases, recovered, and died of Covid-19 as of 27 April 2021. Cluster 1 only consisted of three provinces, namely West Java, Central Java, and East Java. Cluster 2 DKI Jakarta, and the rest go to cluster 3. It takes a grouping test to determine how accurate the results are.

Keywords: *covid-19, fuzzy c-means, grouping, cluster*

1. PENDAHULUAN

Awal Maret 2020 adalah waktu pertama kali seorang pasien dilaporkan positif tertular Covid-19 di Indonesia (Sindi et al., 2020), hingga saat ini 11 Mei 2021, jumlah pasien yang positif sebanyak 1.718.575 orang, pasien sembuh mencapai 1.574.615 orang, dan pasien meninggal dunia sebanyak 47.218 (Komite Penanganan COVID-19 dan Pemulihan Ekonomi Nasional, 2021).

Peningkatan kasus yang diakibatkan oleh virus yang menyerang sistem pernapasan manusia ini memberikan rasa khawatir bagi hampir seluruh lapisan masyarakat di Indonesia, karena hampir seluruh sendi kehidupan terdampak oleh virus yang pertama kali ditemukan di Wuhan, Tiongkok ini (Sindi et al., 2020).

Apabila dilihat untuk persebaran tiap provinsi kasus positif, sembuh, dan meninggal tidak menunjukkan sebuah grafik yang linier. Seperti pada data harian kasus per provinsi tanggal 27 April 2021, kasus positif dan sembuh terbanyak terdapat pada Provinsi DKI Jakarta, sedangkan untuk kasus meninggal Provinsi Jawa Timur berada di posisi pertama, dan di posisi empat untuk kasus positif dan meninggal (Komite Penanganan COVID-19 dan Pemulihan Ekonomi Nasional, 2021). Data persebaran yang abstrak ini membuat pengelompokan persebaran Covid-19 di Indonesia menjadi sukar untuk dilakukan, hal itu tentunya akan membuat pihak-pihak terkait menjadi kesulitan dalam menerapkan kebijakan bagi provinsi-provinsi yang sebenarnya memiliki kemiripan dalam hal jumlah pasien positif, sembuh, dan meninggal. Apabila pengelompokan dilakukan maka tentunya perumusan kebijakan yang akan diterapkan akan lebih mudah, serta, efektif, dan efisien (Putri, Senjaliana dan Riadi, 2020).

Pengelompokan data dalam bidang *machine learning* ada yang bersifat *supervised learning* dan *unsupervised learning*. Pengelompokan yang bersifat *supervised learning* bercirikan bahwa kelas-kelas yang akan menampung hasil klasifikasi sudah diberikan label terlebih dahulu, sedangkan pada *unsupervised learning* kelas tidak diberikan label karena tidak ada parameter atau definisi yang tepat dan akurat dalam memberikan label untuk kelas tersebut. Perbedaan lainnya antara *supervised learning* dan *unsupervised learning* adalah pada *supervised learning* perlu dilakukan tahapan pembelajaran atau pelatihan sedangkan pada *unsupervised learning* tidak diperlukan (Abijono, Santoso dan Anggreini, 2021)

Pengelompokan provinsi seperti pada penelitian ini akan masuk ke dalam kategori *unsupervised learning*, karena disini definisi masing-masing kelas tidak dijelaskan secara detail sehingga pengelompokan masing-masing provinsi tersebut didasarkan pada kemiripan data yang mereka miliki, serta untuk kelas yang menampung data hasil pengelompokannya juga akan ditentukan jumlahnya

dengan menggunakan metode Elbow (Hardianti dan Agushinta R, 2020).

Ada beberapa penelitian yang sebelumnya sudah pernah dilakukan dengan tema pengelompokan provinsi berdasarkan data Covid 19. Salah satunya dilakukan oleh (Putri, Senjaliana dan Riadi, 2020), pengelompokan provinsi dilakukan dengan menggunakan metode K-means Clustering, variabel yang digunakan dalam pengelompokan adalah kasus positif, sembuh, meninggal, ODP, PDP, dan proyeksi jumlah penduduk. Jumlah *cluster* yang digunakan sebanyak lima buah *cluster*. Kelemahan metode K-means Clustering adalah metode tersebut tidak memiliki fleksibilitas dalam menempatkan sebuah data ke *cluster* yang lainnya, sebagaimana yang kita ketahui setiap data pasti memiliki representasi dari kelas-kelas yang akan dituju (Pickens dan Sengupta, 2021)

Penelitian lainnya adalah (Sindi et al., 2020). Metode yang digunakan adalah K-medoids Clustering, jumlah parameter yang digunakan sebanyak 3 buah yaitu jumlah kasus positif, kasus sembuh, dan kasus meninggal. Penentuan jumlah *cluster* dilakukan secara manual, dan dipilih sebanyak 3 buah *cluster*. Kelemahan dalam penelitian ini adalah penentuan jumlah *cluster* yang dilakukan secara manual berpotensi membuat hasil pengelompokan menjadi kurang akurat, karena idealnya jumlah *cluster* didapatkan berdasarkan data yang akan dikelompokkan (Putri, Senjaliana dan Riadi, 2020).

Penelitian mengenai pengelompokan provinsi atau sejenisnya berdasarkan persebaran Covid-19 juga dilakukan oleh (Solichin dan Khairunnisa, 2020). Peneliti mengkhususkan data persebaran hanya di Provinsi DKI Jakarta saja. Parameter yang digunakan adalah jumlah ODP, PDP, kasus positif, pasien sembuh dan pasien meninggal. Pada penelitian ini, untuk melakukan klasterisasi data digunakan metode K-means dan metode pengukuran jarak Euclidean. Sama seperti penelitian yang lainnya metode K-means memiliki kekurangan yaitu metode tersebut secara tegas mengelompokkan sebuah data ke suatu *cluster* tanpa mempertimbangkan pengaruh data tersebut terhadap *cluster* lainnya

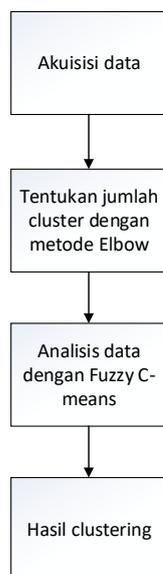
Penelitian mengenai pengelompokan provinsi berdasarkan persebaran penularan Covid-19 yang dilakukan oleh (Abdullah et al., 2021). Penelitian tersebut menggunakan algoritma K-means dalam pengelompokannya, menggunakan parameter kasus terkonfirmasi, sembuh, dan kematian. Pengelompokan data dimasukkan ke dalam 3 kelas. Metode K-means juga dibandingkan dengan pengelompokan berbasis *cluster* lainnya seperti k-medoids (Utomo, 2021), dan hasil yang didapatkan algoritma k-means lebih baik dari k-medoids. Penelitian yang lain adalah mengenai analisis jumlah *cluster* dengan metode K-means (Sari dan Sukestiyarno, 2021), jumlah *cluster* terbaik adalah sebanyak 3 *cluster*.

Berdasarkan beberapa penelitian sebelumnya tersebut, maka dalam penelitian ini penulis akan mengelompokkan provinsi-provinsi berdasarkan persebaran Covid-19 di Indonesia menggunakan metode Fuzzy C-means dan metode Elbow dalam menentukan jumlah *cluster* yang efektif. Fuzzy C-means adalah metode pengelompokan berbasis *fuzzy* yang dapat melakukan persebaran data pada seluruh *cluster* berdasarkan derajat keanggotaan yang dimilikinya (Syarif, Furqon dan Adinugroho, 2018). Sehingga dari beberapa penelitian sebelumnya yang menggunakan K-means Clustering kelemahan penelitian tersebut dapat diatasi. Sedangkan untuk menentukan jumlah *cluster* terbaik akan diimplementasikan metode Elbow. Metode Elbow membandingkan perbandingan hasil *sum square error* (SSE) dari setiap *cluster* dan mendapatkan jumlah *cluster* terbaik dari perubahan nilai SSE yang signifikan atau membentuk siku (*elbow*).

2. METODE PENELITIAN

2.1 Akuisisi Data

Data Covid-19 yang dijadikan *sample dataset* dalam penelitian ini adalah data persebaran Covid-19 di seluruh provinsi Indonesia per tanggal 27 April 2021. Parameter yang digunakan adalah jumlah kasus positif, meninggal, dan sembuh. Data diperoleh dari website resmi Komite Penanganan COVID-19 dan Pemulihan Ekonomi Nasional, yaitu sebuah lembaga atau komite yang bertanggung jawab pada penanganan Covid-19 di Indonesia (Komite Penanganan COVID-19 dan Pemulihan Ekonomi Nasional, 2021). Data persebaran Covid-19 di beberapa provinsi di Indonesia dapat dilihat pada Tabel 1.



Gambar 1. Alur penelitian

Tabel 1. Data persebaran Covid-19 di Indonesia

Provinsi	Kasus Positif	Kasus Sembuh	Kasus Meninggal
Aceh	10814	9367	432
Sumatera Utara	29198	25952	965
Sumatera Barat	36268	33058	783
Riau	42698	37374	1050
Jambi	7407	5987	108
Sumatera Selatan	20068	17712	979
Bengkulu	6630	5743	172
Kep. Bangka Belitung	12822	11272	201
Lampung	10838	9581	253
Kepulauan Riau	9381	8801	230
DKI Jakarta	406205	393166	6625
Jawa Barat	277553	243650	3678
Jawa Tengah	183027	162823	8001
D.I Yogyakarta	38703	33636	948
Jawa Timur	146808	134525	10600
Banten	47101	43892	1200
Bali	44236	41608	1317
NTB	10291	7750	373
NTT	14200	12468	366
Kalimantan Barat	7503	6548	43
Kalimantan Tengah	19780	16577	442
Kalimantan Selatan	32612	29518	937
Kalimantan Timur	68136	64718	1631
Kalimantan Utara	11702	10978	186
Sulawesi Utara	15638	12961	504
Sulawesi Tengah	12125	11109	325
Sulawesi Selatan	61419	60164	931
Sulawesi Tenggara	10396	9818	209
Gorontalo	5355	4990	161
Sulawesi Barat	5443	5251	118
Maluku	7515	7056	116
Maluku Utara	4414	4117	120
Papua	20308	11223	207
Papua Barat	8879	8371	147

2.2 Penentuan Jumlah Cluster dengan Metode Elbow

Metode Elbow ialah metode yang mampu merekomendasikan jumlah total *cluster* yang paling bagus, sehingga dapat memberikan hasil yang terbaik pada saat melakukan proses *clustering*. Dasar ide dari metode Elbow adalah dengan cara membandingkan hasil *Sum Square of Error* (SSE) yang diperoleh dari setiap percobaan, tergantung dari jumlah *cluster* yang diinginkan (Ishak dan Bengnga, 2022). Algoritma dari metode Elbow ialah sebagai berikut (Fuadah, Arifin dan Juwita, 2021)

1. Jumlah *cluster* (k) ditentukan dengan memberikan sebuah label nilai
2. Menentukan pusat *cluster*
3. Nilai k terus bertambah hingga mencapai titik optimal, perhitungan nilai SSE juga tetap dilakukan pada langkah ini
4. Hitung selisih nilai SSE pada setiap nilai k
5. Selisih nilai SSE dari satu k ke k lainnya yang paling besar, dijadikan sebagai jumlah total klaster yang paling baik

Perhitungan SSE dapat dilihat pada Persamaan (1) (Winarta dan Kurniawan, 2021):

$$SSE(A, B) = \sum_{k=1}^m \sum_{i \in S_k} \square \quad \square \quad (1)$$

Dengan:

- SSE : nilai SSE yang dihitung
- A : jumlah *k-cluster* yang berhasil dibentuk
- y_i : Data pada y sesuai dengan indeks nilai ke- i
- B_k : Rerata pada *k-cluster* sesuai nilai k yang telah ditentukan
- m : batas jumlah *cluster*

2.3 Analisis Data dengan Fuzzy C-means

Fuzzy C-means (FCM) adalah salah satu algoritma pengelompokan data berbasis pada logika *fuzzy*. FCM merupakan algoritma *fuzzy clustering* yang paling banyak digunakan. Dasar algoritma yang digunakan dalam FCM adalah pengelompokan suatu data bergantung pada derajat keanggotaan yang dimiliki data tersebut terhadap suatu *cluster*. Derajat keanggotaan memiliki nilai antara 0 dan 1. Jika nilai derajat keanggotaan data pada suatu kelompok semakin besar, maka semakin tepat pengelompokan data tersebut (Askari, 2021).

Algoritma FCM memiliki beberapa tahapan, yang pertama yaitu menentukan *centroid*. *Centroid* atau yang lebih dikenal dengan nama pusat *cluster* titik rata-rata bagi seluruh *cluster* secara individu. Pada tahap awal, nilai pada *centroid* tersebut masih belum akurat. Hal itu juga berlaku untuk derajat keanggotaan yang dimiliki oleh seluruh *cluster*. Oleh karena itu, perlu dilakukan modifikasi nilai *centroid* dan derajat keanggotaan yang dilakukan secara iterative. Hasil dari modifikasi ini membuat *centroid* bergerak menuju titik yang optimal (Kusumadewi dan Purnomo, 2004).

Fuzzy C-means memiliki langkah-langkah perhitungan:

1. Pengelompokan data dituangkan dalam bentuk matriks ukuran $n \times m$, dengan n mengacu pada jumlah sampel data dan m adalah atribut setiap data
2. Menentukan jumlah *cluster* (c), *weight* (w), maksimum iterasi yang digunakan, batas ambang *error*, serta nilai fungsi objektif awal yang digunakan sebelum melakukan pengolahan data
3. Menentukan bilangan acak untuk mengisi elemen matriks awal U . Menghitung jumlah tiap kolom atribut Q dengan Persamaan (2) dan (3), Q_j adalah nilai matriks partisi awal:

$$Q_j = \sum_{k=1}^c \mu_{ik} \quad (2)$$

Hitung

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j} \quad (3)$$

4. Menentukan pusat *cluster* menggunakan Persamaan (4), sebagai acuan dalam menghitung jarak dari sebuah data ke pusat *cluster*. X adalah nilai dari parameter data

$$v_{ij} = \frac{\sum_{k=1}^n ((\mu_{ik})^w * X_{kj})}{\sum_{k=1}^n (\mu_{ik})^w} \quad (4)$$

5. Menghitung nilai fungsi objektif (P) menggunakan Persamaan (5) sebagai dasar dalam menentukan jarak titik data menuju pusat *cluster* yang memiliki bobot sesuai dengan derajat keanggotaan yang dimilikinya.

$$Per_t = \sum_{k=1}^N \sum_{l=1}^c \left(\left[\sum_{j=1}^m (X_{lk} - V_{kj})^2 \right] (\mu_{lk})^2 \right) \quad (5)$$

6. Hitung perubahan matriks partisi

$$\mu_{lk} = \frac{\left[\sum_{j=1}^m (X_{lj} - V_{kj})^2 \right]^{-\frac{1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{lj} - V_{kj})^2 \right]^{-\frac{1}{w-1}}} \quad (6)$$

Dengan: $l=1,3,5,\dots,n$; dan $k=1,2,\dots,c$

7. Cek bagaimana kondisi akan berhenti
 - a. Jika: $(|Per_t - Per_{t-1}| < \xi)$ atau $(t > IterMax)$ selanjutnya iterasi akan dihentikan
 - b. Apabila $t=t+1$, langkah 4 diulangi kembali

2.4 Hasil Clustering

Hasil *clustering* berupa pengelompokan data provinsi menjadi beberapa kelompok, tergantung pada hasil jumlah *cluster* terbaik dengan metode Elbow

3. HASIL DAN PEMBAHASAN

Hasil dari pengelompokan provinsi berdasarkan kasus positif, kasus meninggal, dan kasus sembuh Covid-19 akan dibahas pada bab ini.

3.1. Akuisisi Data Provinsi

Data yang digunakan dalam pengelompokan provinsi berdasarkan kasus positif, kasus meninggal, dan kasus sembuh Covid-19 diambil dari (Komite Penanganan COVID-19 dan Pemulihan Ekonomi Nasional, 2021), data tersebut dapat dilihat pada Tabel 1.

3.2. Penentuan Jumlah Cluster dengan Metode Elbow

Berikut adalah tahapan dalam menentukan jumlah *cluster* menggunakan metode Elbow

1. Jumlah *cluster* yang dicoba dalam penelitian ini berurutan dari 2, 3, 4, dan 5. Jumlah *cluster* yang dicoba pada saat menggunakan metode Elbow ini hingga 5 *cluster*, dan perubahan nilai SSE yang signifikan berada pada jumlah *cluster* 2 dan 3. Berdasarkan penelitian sebelumnya, metode Elbow mampu menentukan jumlah *cluster* yang optimal secara akurat, oleh karena itu dalam penelitian ini akan digunakan jumlah *cluster* sebanyak 3.

2. Menentukan pusat *cluster* secara acak sesuai dengan jumlah *cluster* yang akan diujikan. Banyaknya nilai pada pusat *cluster* mengacu pada jumlah parameter yang digunakan dalam pengelompokan. Dalam hal ini ada tiga parameter maka banyaknya nilainya ada tiga. Nilai pusat *cluster* ini diambil dalam *range* data masing-masing parameter. Pusat *cluster* dapat dilihat pada Gambar 2. Pusat *cluster* atau *centroid* ini sangat berperan dalam menentukan sejauh dan secepat apa sebuah data akan mengelompok ke dalam sebuah *cluster* secara optimal pada proses algoritma Fuzzy c-means
3. Nilai *k* atau jumlah *cluster* akan terus bertambah untuk melihat selisih nilai SSE yang signifikan
4. Perhitungan nilai SSE dapat dilakukan menggunakan Persamaan (1)
5. Selisih nilai SSE tiap jumlah *cluster* dihitung untuk mencari jumlah *cluster* yang paling maksimal

Tabel 2. Perbandingan nilai SSE untuk setiap *k*

K	SSE	SELISIH
2	9065720	
3	5247550	3818170
4	7318462	2070912
5	7048492	269969.7

6. Selisih nilai SSE dari satu *k* ke *k* lainnya yang paling besar, dijadikan sebagai jumlah total klaster yang paling baik. Grafiknya dapat dilihat pada Tabel 2 dan Gambar 3. Tabel 2 dan Gambar 3 menunjukkan bahwa perubahan nilai SSE yang paling besar terjadi pada saat penambahan jumlah *cluster* dari 2 menjadi 3, sehingga membentuk sudut siku yang paling tajam. Nilai SSE yang paling kecil serta yang turun secara signifikan, akan mengotimalkan jumlah *cluster* secara akurat. Hal itu nantinya akan berpengaruh pada hasil pengelompokan. Oleh karena itu jumlah *cluster* sebanyak 3 akan digunakan dalam pengelompokan provinsi berdasarkan kasus positif, kasus meninggal, dan kasus sembuh Covid-19.

3.3. Analisis Data dengan Fuzzy C-means

1. Tabel 1 berisi data yang akan dianalisis
2. Inisialisasi beberapa nilai:
 Total *cluster* = 3;
 Bobot pangkat = 2;
 Iterasi maksimum= *IterMax*=100
 Batas toleransi *error*= 0.1
 Iterasi awal = *t*=1;
3. Menentukan nilai-nilai yang digunakan untuk mengisi elemen matriks partisi yang dibangkitkan secara acak dengan rentang nilai 0-1, jumlahnya didapatkan dengan mengalikan jumlah data

dengan jumlah total *cluster*. Tabel 3 menunjukkan sepuluh data pertama nilai-nilai elemen matriks.

Tabel 3. nilai-nilai pada elemen matriks

Nilai Matrik pada Partisi Awal		
0.801	0.451	0.965
0.453	0.124	0.177
0.761	0.235	0.833
0.218	0.144	0.087
0.989	0.547	0.464
0.319	0.334	0.259
0.394	0.444	0.373
0.910	0.044	0.303
0.704	0.532	0.095
0.529	0.760	0.066
...

4. Persamaan (4) digunakan untuk perhitungan pusat *cluster*. Tabel 4 berisikan nilai pusat *cluster*.

Tabel 4. Nilai centroid

Nilai centroid		
37,618.706	34,750.281	728.967
52,972.680	49,160.709	1,184.125
30,284.049	27,284.361	731.481

5. Menghitung nilai fungsi objektif dengan menggunakan Persamaan (5). Didapatkan nilai fungsi objektif pada iterasi ke-0 sebesar 343,284,205,880.44. Apabila nilai fungsi objektif kurang dari 0.1, maka iterasi berhenti
6. Matriks partisi mengalami perubahan, dapat dihitung dengan Persamaan (6). Sembilan data pertama nilai-nilai matriks partisi ditunjukkan pada Tabel 4. Untuk iterasi selanjutnya, nilai inilah yang akan dijadikan sebagai nilai derajat keanggotaan

Tabel 4. Nilai matriks partisi

Nilai matriks partisi		
0.30	0.12	0.58
0.02	0.00	0.98
0.93	0.01	0.06
0.79	0.11	0.10
0.31	0.14	0.55
0.23	0.07	0.70
0.31	0.14	0.55
0.29	0.11	0.60
0.30	0.12	0.58
...

7. Melakukan pengecekan untuk kondisi berhenti atau tidak. Iterasi akan berakhir jika selisih nilai fungsi objektif terbaru dengan nilai fungsi objektif sebelumnya kurang dari 0.1
8. Perhitungan ini berhenti pada iterasi ke-38, dan mendapatkan *objective function* sebesar 0.057, dapat dilihat pada Tabel 5
9. Sedangkan pengelompokan provinsi menjadi tiga buah *cluster* dapat dilihat pada Tabel 6, dengan daftar provinsi seperti pada Tabel 1. Nilai derajat keanggotaan yang paling besar dari sebuah data pada masing-masing *cluster* menjadi acuan untuk menentukan data tersebut masuk ke *cluster* 1,2, atau 3

K2	C1			C2											
	36236	336192	9707	162759	261960	3329									
K3	C1			C2			C3								
	264478	171019	6576	178006	40692	2302	127639	199240	3554						
K4	C1			C2			C3			C4					
	291785	296826	6222	392134	361765	7047	246703	80758	5166	242626	254885	8096			
K5	C1			C2			C3			C4			C5		
	217627	196033	2119	238081	39557	8728	79752	234184	8379	330742	42705	4508	131703	264429	10032

K=jumlah cluster
C=cluster ke-n

Gambar 2. Nilai pusat cluster



Gambar 3. Perbandingan nilai SSE untuk setiap jumlah cluster yang diuji

Tabel 5. Pusat cluster

Pusat Cluster		
184,277.491	164,871.425	8,064.029
396,184.051	381,568.879	6,399.637
20,139.593	18,003.355	497.599

Derajat keanggotaan (m) data pada Cluster ke-		
0.007	0.001	0.991
0.001	0.000	0.999
0.004	0.001	0.995

Tabel 6. Nilai derajat keanggotaan setiap data

Derajat keanggotaan (m) data pada Cluster ke-		
1	2	3
0.003	0.001	0.996
0.003	0.001	0.996
0.012	0.002	0.986
0.024	0.004	0.973
0.005	0.001	0.994
0.000	0.000	1.000
0.006	0.001	0.993
0.002	0.000	0.998
0.003	0.001	0.997
0.004	0.001	0.996
0.002	0.997	0.001
0.634	0.286	0.081
1.000	0.000	0.000
0.015	0.002	0.983
0.911	0.017	0.071
0.040	0.006	0.955
0.031	0.005	0.964
0.004	0.001	0.996
0.001	0.000	0.999
0.005	0.001	0.994
0.000	0.000	1.000
0.007	0.001	0.992
0.157	0.018	0.825
0.002	0.000	0.997
0.001	0.000	0.999
0.002	0.000	0.998
0.116	0.014	0.870
0.003	0.001	0.996
0.007	0.001	0.992
0.007	0.001	0.992
0.005	0.001	0.994

4. KESIMPULAN

Berdasarkan hasil dan pembahasan yang sudah dijelaskan sebelumnya dapat diambil sebuah kesimpulan algoritma Fuzzy C-means yang digunakan dalam penelitian ini dapat menjadi sebuah alternatif untuk mengelompokkan data provinsi berdasarkan kasus positif, kasus meninggal, dan kasus sembuh Covid-19.

Fuzzy C-means yang memiliki kelebihan yaitu memungkinkan setiap data masuk ke dalam seluruh cluster sesuai dengan derajat keanggotaannya masing-masing. Derajat keanggotaan melambangkan sebuah nilai yang merepresentasikan sebuah data lebih ideal masuk ke dalam sebuah cluster. Nilai derajat keanggotaan yang paling tinggi dari sebuah data di dalam sebuah cluster menunjukkan data tersebut paling ideal untuk masuk ke dalam cluster tersebut. Dari definisi ini dapat diketahui sebesar besar sebuah data merepresentasikan dirinya di dalam sebuah cluster yang dilambangkan dengan derajat keanggotaan.

Penelitian-penelitian sebelumnya menggunakan metode pengelompokan yang crisp atau tegas serta penentuan jumlah cluster juga masih manual. Penentuan jumlah cluster yang dilakukan secara manual, membuat hasil pengelompokan menjadi tidak optimal, karena jumlah cluster sangat

bergantung pada karakteristik data yang akan dikelompokkan. Selanjutnya, pengelompokan yang *crisp* akan menjadi permasalahan karena setiap data yang digunakan untuk proses *learning* pasti memiliki sebuah kecenderungan atau probabilitas untuk masuk ke seluruh kelas, namun nilai probabilitas itu yang berbeda-beda. Melalui derajat keanggotaan pada Fuzzy C-means kita dapat mengetahui pengelompokan data secara akurat, dengan melihat langsung nilai derajat keanggotaan sebagai representasi sebuah data terhadap masing-masing kelas.

Pengelompokan dilakukan menjadi tiga buah *cluster* atau kelompok, yang dimana penentuan jumlah *cluster* menggunakan metode Elbow dengan mencari penurunan selisih nilai SSE yang paling besar untuk dapat ditentukan sebagai jumlah *cluster* yang paling tepat. *Cluster* 1 hanya terdiri tiga provinsi yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur. *Cluster* 2 DKI Jakarta, dan sisanya masuk ke *cluster* 3.

DAFTAR PUSTAKA

- ABDULLAH, D., SUSILO, S., AHMAR, A.S., RUSLI, R. dan HIDAYAT, R., 2021. The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. *Quality & Quantity*, hal.1–9.
- ABIJONO, H., SANTOSO, P. dan ANGGREINI, N.L., 2021. Algoritma Supervised Learning Dan Unsupervised Learning Dalam Pengolahan Data. *Jurnal Teknologi Terapan: G-Tech*, 4(2), hal.315–318.
- ASKARI, S., 2021. Fuzzy C-Means clustering algorithm for data with unequal cluster sizes and contaminated with noise and outliers: Review and development. *Expert Systems with Applications*, 165, hal.113856.
- FUADAH, A.W., ARIFIN, F.N. dan JUWITA, O., 2021. Optimasi K-Klasterisasi Ketahanan Pangan Kabupaten Jember Menggunakan Metode Elbow. *INFORMAL: Informatics Journal*, 6(3), hal.136–141.
- HARDIANTI, A. dan AGUSHINTA R, D., 2020. Universitas Darma Persada Menggunakan Metode Clustering K-Means Pattern Analysis of the Student Study Period in the Faculty of Engineering At Darma Persada University Using K-Means. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 7(4), hal.861–868.
- ISHAK, R. dan BENGNGA, A., 2022. Clustering Tingkat Pemahaman Mahasiswa Pada Perkuliahan Probabilitas Statistika Dengan Metode K-Means. *Jambura Journal of Electrical and Electronics Engineering*, 4(1), hal.65–69.
- KOMITE PENANGANAN COVID-19 dan PEMULIHAN EKONOMI NASIONAL, 2021. *Data Sebaran COVID-19 di Indonesia*. [daring] <https://covid19.go.id/>. Tersedia pada: <<https://covid19.go.id/>>.
- KUSUMADEWI, S. dan PURNOMO, H., 2004. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Yogyakarta: Graha Ilmu.
- PICKENS, A. dan SENGUPTA, S., 2021. Benchmarking Studies Aimed at Clustering and Classification Tasks Using K-Means, Fuzzy C-Means and Evolutionary Neural Networks. *Machine Learning and Knowledge Extraction*, 3(3), hal.695–719.
- PUTRI, A.Q., SENJALIANA, F. dan RIADI, R.A., 2020. Pengelompokan provinsi terdampak pandemi covid-19 di Indonesia. In: *Seminar Nasional Matematika dan Pendidikan Matematika (5th SENATIK)*. hal.342–351.
- SARI, D.N.P. dan SUKESTIYARNO, Y.L., 2021. Analisis cluster dengan metode K-Means pada persebaran kasus COVID-19 berdasarkan Provinsi di Indonesia. In: *PRISMA, Prosiding Seminar Nasional Matematika*. hal.602–610.
- SINDI, S., NINGSE, W.O.R., SIHOMBING, I.A., ZER, F.I.R.H. dan HARTAMA, D., 2020. Analisis algoritma k-medoids clustering dalam pengelompokan penyebaran covid-19 di indonesia. *Jurnal Teknologi Informasi*, 4(1), hal.166–173.
- SOLICHIN, A. dan KHAIRUNNISA, K., 2020. Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means. *Fountain of Informatics Journal*, 5(2), hal.52.
- SYARIF, R., FURQON, M.T. dan ADINUGROHO, S., 2018. Perbandingan Algoritme K-Means Dengan Algoritme Fuzzy C Means (FCM) Dalam Clustering Moda Transportasi Berbasis GPS. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, [daring] 2(10), hal.4107–4115. Tersedia pada: <<http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2852>>.
- UTOMO, W., 2021. The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia. *Ilk. J. Ilm*, 13(1), hal.31–35.
- WINARTA, A. dan KURNIAWAN, W.J., 2021. Optimasi cluster K-means menggunakan metode Elbow pada data pengguna narkoba dengan pemrograman Python. *JTIK (Jurnal Teknik Informatika Kaputama)*, 5(1), hal.113–119.