

Analisis Kluster Menggunakan Metode *Fuzzy C-Means* pada Data COVID-19 di Provinsi Lampung

Suciati, I.¹, Herawati, N.¹, Subian, S.¹, Widiarti.¹

¹Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung
Jl. Prof. Dr. Ir. Sumantri Brojonegoro, Gedong Meneng, Bandar Lampung, Indonesia

E-mail: indhahsuciati222@gmail.com

Dikirim: 25-03-2022, Diterima: 25-04-2022, Diterbitkan: 02-05-2022

Abstrak

Analisis kluster adalah suatu alat untuk mengelompokkan sejumlah n objek berdasarkan p variabel yang mempunyai kesamaan karakteristik diantara objek-objek tersebut, sehingga keragaman dalam suatu kluster tersebut lebih kecil dibandingkan dengan keragaman antar kluster. Tujuan penelitian yaitu melakukan pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh pada data COVID-19 di Provinsi Lampung kurun waktu 6 bulan (Maret 2020-Agustus 2020 dan September 2020-Februari 2021) serta kurun waktu 12 bulan (Maret 2020-Februari 2021) menggunakan metode *fuzzy c-means*. Dalam penelitian ini digunakan *partition coefficient index* untuk mengevaluasi jumlah kluster optimal, sehingga akan diperoleh hasil kluster terbaik untuk data COVID-19 di Provinsi Lampung, kemudian melihat apakah terjadi perpindahan hasil kluster dalam kurun waktu 6 bulan pertama dan 6 bulan terakhir. Hasil penelitian ini yaitu jumlah kluster 2 merupakan jumlah kluster yang optimal untuk pengklasteran setiap data COVID-19 di Provinsi Lampung dengan hasil kluster terbaik menghasilkan klusterisasi yang sama, dimana Kota Bandar Lampung merupakan kabupaten/kota di Provinsi Lampung yang masuk ke dalam kluster tinggi, sedangkan untuk kabupaten/kota lainnya masuk ke dalam kluster rendah. Selain itu, diperoleh hasil bahwa tidak terdapat perpindahan hasil kluster pada data COVID-19 di Provinsi Lampung kurun waktu 6 bulan pertama dan 6 bulan terakhir.

Kata Kunci: *analisis kluster, fuzzy c-means, partition coefficient index*

1. Pendahuluan

Coronavirus Disease (COVID-19) adalah penyakit menular yang ditandai dengan gejala pada bagian pernapasan akut (*SARS-CoV-2*) [3]. Saat ini, COVID-19 telah tersebar ke berbagai negara salah satunya yaitu Negara Indonesia, dimana Provinsi Lampung merupakan salah satu provinsi di Negara Indonesia yang terdampak COVID-19. Penyebaran COVID-19 yang merata di kabupaten/kota Provinsi Lampung merupakan penyebaran yang cukup cepat. Luasnya wilayah Provinsi Lampung memungkinkan diperlukannya pengklasteran kabupaten/kota di Provinsi Lampung. Pengklasteran ini akan menghasilkan titik-titik pusat penyebaran kasus COVID-19 yang nantinya dapat dievaluasi dan dijadikan suatu informasi. Analisis kluster adalah suatu alat untuk mengelompokkan sejumlah n objek berdasarkan p variabel yang mempunyai kesamaan karakteristik diantara objek-objek tersebut, sehingga keragaman dalam suatu kluster tersebut lebih kecil dibandingkan dengan keragaman antar kluster [10]. Analisis kluster dapat diklasifikasikan sebagai *soft clustering* dan *hard clustering* [1]. Metode *soft clustering* yang sangat populer adalah metode *Fuzzy C-Means* (FCM), dimana kelebihan metode ini terletak pada penempatan pusat kluster yang lebih tepat dibandingkan dengan metode kluster lainnya [9]. Namun, metode ini terdapat kelemahan dalam penentuan jumlah kluster yang optimal. Untuk meminimalisir hal tersebut serta mengoptimalkan hasil pengklasteran, perlu dilakukan validasi dengan menggunakan indeks validitas [11]. Indeks validitas yang sering digunakan dalam metode FCM yaitu *Partition Coefficient Index* (PCI).

Dalam penelitian ini, penulis tertarik untuk melakukan penelitian mengenai analisis kluster menggunakan metode *Fuzzy C-Means* (FCM) pada data COVID-19 di Provinsi Lampung, penulis menggunakan *Partition Coefficient Index* (PCI) untuk mengevaluasi jumlah kluster optimal, sehingga akan diperoleh hasil kluster terbaik untuk data COVID-19 di Provinsi Lampung.

2. Tinjauan Pustaka

2.1 Analisis Kluster

Analisis kluster adalah suatu alat untuk mengelompokkan sejumlah n objek berdasarkan p variabel yang mempunyai kesamaan karakteristik diantara objek-objek tersebut, sehingga keragaman dalam suatu kluster tersebut lebih kecil dibandingkan dengan keragaman antar kluster [10]. Analisis kluster dapat diklasifikasikan

sebagai *soft clustering* dan *hard clustering* [1]. Dalam *soft clustering*, himpunan *fuzzy* digunakan untuk mengelompokkan data sehingga setiap data dapat dimiliki oleh dua atau lebih kluster dengan derajat keanggotaan yang berbeda. Sedangkan pada teknik *hard clustering*, himpunan *crisp* digunakan untuk mengelompokkan data sehingga jika suatu data masuk dalam kluster tertentu maka tidak dapat masuk ke kluster lain. Dalam banyak situasi, *soft clustering* lebih alami daripada *hard clustering*.

2.2 Jarak Euclidean

Dalam pengklasteran digunakan suatu ukuran yang dapat menerangkan keserupaan atau kedekatan antar data, yaitu ukuran jarak atau similaritas. Ukuran jarak yang sering digunakan adalah ukuran jarak Euclidean [7]. Rumus jarak Euclidean adalah sebagai berikut [4]:

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

dengan:

- d_{ik} = jarak objek ke- i dengan pusat kluster ke- k
- p = jumlah variabel
- x_{ij} = data dari objek ke- i pada variabel ke- j
- c_{kj} = pusat kluster ke- k pada variabel ke- j

2.3 Metode Fuzzy C-Means (FCM)

Metode *Fuzzy C-Means* (FCM) adalah suatu teknik pengelompokkan data yang keberadaan tiap-tiap titik data suatu kluster ditentukan oleh nilai keanggotaan. Nilai keanggotaan tersebut akan mencakup bilangan real pada interval [0-1]. Metode ini pertama kali diperkenalkan oleh Jim Bezdek pada tahun 1981 dengan kelebihan metode ini yaitu penempatan pusat kluster yang lebih tepat jika dibandingkan dengan metode kluster lainnya [9].

2.4 Algoritma Metode Fuzzy C-Means (FCM)

Algoritma dari *Fuzzy C-Means* (FCM) adalah sebagai berikut [8]:

1. Memasukkan data yang akan dikluster, berupa matriks \mathbf{X} berukuran $n \times p$ (n = jumlah sampel data, p = jumlah variabel setiap data) sebagai berikut:

$$\mathbf{X} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \quad (2)$$

dengan:

- $i = 1, 2, \dots, n$
- $j = 1, 2, \dots, p$
- x_{ij} = data sampel ke- i variabel ke- j

2. Menentukan:

- a. Jumlah kluster ($1 < c < N$)
- b. Pangkat pembobot ($m > 1$)
- c. Maksimum iterasi (*MaxIter*)
- d. Galat terkecil yang diharapkan ($\epsilon > 0$)
- e. Fungsi objektif awal ($P_0 = 0$)
- f. Iterasi awal ($t = 1$)

3. Membangkitkan bilangan acak μ_{ik} sebagai elemen-elemen matriks partisi awal \mathbf{U} . Matriks partisi awal \mathbf{U} sebagai berikut:

$$\mathbf{U} = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_1) & \cdots & \mu_{1c}(x_1) \\ \mu_{21}(x_2) & \mu_{22}(x_2) & \cdots & \mu_{2c}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{n1}(x_n) & \mu_{n2}(x_n) & \cdots & \mu_{nc}(x_n) \end{bmatrix} \quad (3)$$

dengan:

- $i = 1, 2, \dots, n$
- $k = 1, 2, \dots, c$
- μ_{ik} = elemen matriks partisi awal \mathbf{U} ke- i kluster ke- k

dengan aturan,

$$\sum_{k=1}^c \mu_{ik} = 1 \quad (4)$$

4. Menghitung pusat kluster ke- k pada variabel ke- j (v_{kj}) dengan rumus sebagai berikut:

$$v_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^m x_{ij})}{\sum_{i=1}^n (\mu_{ik})^m} \quad (5)$$

dengan:

$$i = 1, 2, \dots, n$$

$$j = 1, 2, \dots, p$$

$$k = 1, 2, \dots, c$$

m = pangkat pembobot

μ_{ik} = elemen matriks partisi awal U ke- i kluster ke- k

x_{ij} = data sampel ke- i variabel ke- j

sehingga diperoleh matriks pusat kluster V sebagai berikut:

$$V = \begin{bmatrix} v_{11} & \dots & v_{1p} \\ \vdots & \ddots & \vdots \\ v_{c1} & \dots & v_{cp} \end{bmatrix} \quad (6)$$

5. Menghitung nilai fungsi objektif pada iterasi ke- t (P_t) dengan rumus sebagai berikut:

$$P_t = \sum_{i=1}^n \sum_{k=1}^c ((\mu_{ik})^m d_{ik}) \quad (7)$$

dengan:

$$i = 1, 2, \dots, n$$

$$k = 1, 2, \dots, c$$

m = pangkat pembobot

μ_{ik} = elemen matriks partisi awal U ke- i kluster ke- k

d_{ik} = jarak Euclidean dari objek ke- i dengan pusat kluster ke- k

6. Menghitung perubahan matriks keanggotaan μ_{ik} dengan rumus sebagai berikut:

$$\mu_{ik} = \left[\frac{[\sum_{j=1}^p d_{ik}]^{\frac{1}{m-1}}}{\sum_{k=1}^c [\sum_{j=1}^p d_{ik}]^{\frac{1}{m-1}}} \right]^{-1} \quad (8)$$

dengan:

$$i = 1, 2, \dots, n$$

$$j = 1, 2, \dots, p$$

$$k = 1, 2, \dots, c$$

m = pangkat pembobot

μ_{ik} = elemen matriks partisi awal U ke- i kluster ke- k

d_{ik} = jarak Euclidean dari objek ke- i dengan pusat kluster ke- k

7. Memeriksa kondisi berhenti:

- Jika $(|P_t - P_{t-1}| < \epsilon)$ atau $(t > MaxIter)$ maka iterasi berhenti
- Jika tidak, maka $t = t + 1$, kembali ke langkah 4.

2.5 Partition Coefficient Index (PCI)

Partition Coefficient Index (PCI) adalah indeks validitas yang hanya mengevaluasi nilai derajat keanggotaan, tanpa memandang nilai vektor (data) yang biasanya mengandung informasi sebaran data [12]. Rumus yang digunakan untuk menghitung nilai PCI yaitu:

$$PCI = \frac{1}{n} \sum_{k=1}^c \sum_{i=1}^n (\mu_{ik})^2 \quad (10)$$

dimana n adalah banyak objek penelitian, c adalah banyak kluster, dan μ_{ik} adalah nilai keanggotaan objek ke- i dengan pusat kluster ke- k . PCI memiliki rentang nilai antara 0 sampai 1 dengan jumlah kluster yang optimal ditunjukkan oleh nilai PCI yang paling besar.

3. Hasil dan Pembahasan

3.1 Analisis Kluster Menggunakan Metode Fuzzy C-Means (FCM)

Dalam penelitian ini, akan dilakukan pengklasteran kabupaten/kota di Provinsi Lampung berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung menggunakan data COVID-19 di Provinsi Lampung kurun waktu 6 bulan (Maret 2020-Agustus 2020 dan September 2020-Februari 2021) dan kurun waktu 12 bulan (Maret 2020-Februari 2021) dengan menggunakan metode Fuzzy C-Means (FCM). Pengklasteran dengan metode FCM, dilakukan dengan langkah-langkah sebagai berikut.

- Membangun matriks berukuran 15×3 yang merupakan nilai standarisasi data penelitian. Standarisasi data ini dilakukan untuk menghindari masalah yang dihasilkan dari penggunaan nilai skala yang berbeda antar

variabel pengklasteran. Dalam penelitian ini, data distandarisasi dengan menggunakan *z-score*. Adapun matriks-matriks berukuran 15×3 yang merupakan nilai standarisasi dari data penelitian sebagai berikut.

$$\begin{matrix}
 X = \begin{bmatrix} -0.44105576 & 0.03305898 & -0.54315906 \\ -0.53728611 & 0.03305898 & -0.54315906 \\ 0.25661426 & 0.03305898 & 0.28879232 \\ -0.53728611 & -0.46282572 & -0.60053502 \\ 0.37690220 & 0.03305898 & 0.37485625 \\ 0.06415357 & 0.03305898 & -0.05546343 \\ -0.15236472 & 0.03305898 & -0.14152736 \\ -0.60945888 & -0.46282572 & -0.60053502 \\ -0.34482542 & -0.46282572 & -0.28496725 \\ -0.51322853 & -0.46282572 & -0.51447108 \\ -0.60945888 & -0.46282572 & -0.60053502 \\ -0.36888300 & -0.46282572 & -0.37103119 \\ 0.49719014 & -0.46282572 & 0.71911199 \\ 3.36004301 & 3.50425191 & 3.30103006 \\ -0.44105576 & -0.46282572 & -0.42840715 \\ -0.37860960 & -0.39976238 & -0.37634174 \\ -0.25715802 & -0.21444207 & -0.22745505 \\ -0.07070420 & -0.08207042 & -0.07566796 \\ 0.08752496 & 0.01058973 & 0.06548437 \\ 0.93255421 & 0.65921081 & 0.87952722 \\ 0.17048836 & -0.25415356 & 0.12735936 \\ -0.61809157 & -0.50565970 & -0.60063859 \\ -0.58473515 & -0.43947387 & -0.58516984 \\ -0.25202627 & -0.24091640 & -0.24679098 \\ -0.38716252 & -0.33357655 & -0.38987690 \\ -0.63519743 & -0.51889686 & -0.61514054 \\ -0.56506342 & -0.37328805 & -0.55133195 \\ -0.55223403 & -0.54537119 & -0.55133195 \\ 3.29829399 & 3.45225259 & 3.32745657 \\ -0.18787931 & -0.21444207 & -0.18008201 \end{bmatrix} \\
 \\
 Y = \begin{bmatrix} -0.37510615 & -0.41105443 & -0.36913277 \\ -0.24604371 & -0.22091912 & -0.21563201 \\ -0.08250567 & -0.08510819 & -0.08804697 \\ 0.11020401 & 0.02354056 & 0.08837922 \\ 0.94999390 & 0.67543305 & 0.89375980 \\ 0.17385124 & -0.26166240 & 0.13323334 \\ -0.63323105 & -0.51970318 & -0.61433527 \\ -0.58195966 & -0.43821662 & -0.58243901 \\ -0.24781168 & -0.23450021 & -0.24453800 \\ -0.38129408 & -0.32956787 & -0.38408414 \\ -0.63411503 & -0.51970318 & -0.61333851 \\ -0.57046780 & -0.37031115 & -0.55552654 \\ -0.58903158 & -0.54686536 & -0.59340335 \\ 3.28549379 & 3.44597611 & 3.31588209 \\ -0.17797652 & -0.20733803 & -0.17077790 \end{bmatrix} \\
 \\
 Z = \begin{bmatrix} -0.44105576 & 0.03305898 & -0.54315906 \\ -0.53728611 & 0.03305898 & -0.54315906 \\ 0.25661426 & 0.03305898 & 0.28879232 \\ -0.53728611 & -0.46282572 & -0.60053502 \\ 0.37690220 & 0.03305898 & 0.37485625 \\ 0.06415357 & 0.03305898 & -0.05546343 \\ -0.15236472 & 0.03305898 & -0.14152736 \\ -0.60945888 & -0.46282572 & -0.60053502 \\ -0.34482542 & -0.46282572 & -0.28496725 \\ -0.51322853 & -0.46282572 & -0.51447108 \\ -0.60945888 & -0.46282572 & -0.60053502 \\ -0.36888300 & -0.46282572 & -0.37103119 \\ 0.49719014 & -0.46282572 & 0.71911199 \\ 3.36004301 & 3.50425191 & 3.30103006 \\ -0.44105576 & -0.46282572 & -0.42840715 \\ -0.37860960 & -0.39976238 & -0.37634174 \\ -0.25715802 & -0.21444207 & -0.22745505 \\ -0.07070420 & -0.08207042 & -0.07566796 \\ 0.08752496 & 0.01058973 & 0.06548437 \\ 0.93255421 & 0.65921081 & 0.87952722 \\ 0.17048836 & -0.25415356 & 0.12735936 \\ -0.61809157 & -0.50565970 & -0.60063859 \\ -0.58473515 & -0.43947387 & -0.58516984 \\ -0.25202627 & -0.24091640 & -0.24679098 \\ -0.38716252 & -0.33357655 & -0.38987690 \\ -0.63519743 & -0.51889686 & -0.61514054 \\ -0.56506342 & -0.37328805 & -0.55133195 \\ -0.55223403 & -0.54537119 & -0.55133195 \\ 3.29829399 & 3.45225259 & 3.32745657 \\ -0.18787931 & -0.21444207 & -0.18008201 \end{bmatrix}
 \end{matrix}$$

Matriks **X** merupakan matriks standarisasi data COVID-19 di Provinsi Lampung kurun waktu 6 bulan (Maret 2020-Agustus 2020), matriks **Y** merupakan matriks standarisasi data COVID-19 di Provinsi Lampung kurun waktu 6 bulan (September 2020-Februari 2021), dan matriks **Z** merupakan matriks standarisasi data COVID-19 di Provinsi Lampung kurun waktu 12 bulan (Maret 2020-Februari 2021).

- Setelah membangun matriks yang merupakan nilai standarisasi data penelitian, maka langkah selanjutnya yaitu menentukan jumlah klaster, pangkat pembobot, maksimum iterasi, galat terkecil, fungsi objektif awal, dan iterasi awal. Dalam penelitian ini, jumlah klaster dan pangkat pembobot yang digunakan yaitu 2 sampai 5, maksimum iterasi yang digunakan yaitu 100, galat terkecil yaitu 10^{-9} , fungsi objektif awal yaitu 0, dan iterasi awal yaitu 1.
- Membangkitkan bilangan acak μ_{ik} sebagai elemen matriks partisi awal **U** dengan aturan $\sum_{k=1}^c \mu_{ik} = 1$ dimana $i = 1, 2, \dots, 15$ dan $k = 1, 2, \dots, 5$. Dalam penelitian ini, untuk *output* langkah 3 tidak ditampilkan pada software RStudio, maka diberikan contoh untuk membangun matriks partisi awal dengan menggunakan jumlah klaster 3 sebagai berikut.

Tabel 1 Membangun Matriks Partisi Awal untuk Jumlah Klaster 3

Kabupaten/Kota	Matriks Partisi Awal			Jumlah
Lampung Barat	0.31	0.15	0.54	1
Tanggamus	0.32	0.16	0.52	1
Lampung Selatan	0.11	0.33	0.56	1
Lampung Timur	0.65	0.22	0.13	1
Lampung Tengah	0.44	0.21	0.35	1
Lampung Utara	0.14	0.30	0.56	1
Way Kanan	0.45	0.41	0.14	1
Tulang Bawang	0.34	0.32	0.34	1
Pesawaran	0.66	0.11	0.23	1
Pringsewu	0.32	0.33	0.35	1
Mesuji	0.41	0.35	0.24	1
Tulang Bawang Barat	0.28	0.51	0.21	1
Pesisir Barat	0.61	0.22	0.17	1
Bandar Lampung	0.39	0.37	0.24	1
Metro	0.42	0.32	0.26	1

Dalam membangkitkan bilangan acak μ_{ik} , bilangan-bilangan tersebut dibangkitkan dengan melihat aturan sebagai berikut.

$$\sum_{k=1}^c \mu_{ik} = 1$$

Sebagai contoh untuk perhitungan, digunakan matriks partisi awal pada Tabel 1 dengan jumlah kluster 3 untuk Kabupaten Lampung Barat sebagai berikut.

$$\sum_{k=1}^3 \mu_{1k} = 0.31 + 0.15 + 0.54 = 1$$

4. Menghitung pusat kluster ke- k pada variabel ke- j (v_{kj}) dimana $k = 1, 2, \dots, 5$ dan $j = 1, 2, 3$, dengan menggunakan persamaan sebagai berikut.

$$v_{kj} = \frac{\sum_{i=1}^{15} ((\mu_{ik})^m x_{ij})}{\sum_{i=1}^{15} (\mu_{ik})^m}$$

Dalam menentukan kategori kluster, maka dapat dilihat dari nilai rata-rata pusat klusternya.

5. Menghitung nilai fungsi objektif yang berfungsi sebagai syarat perulangan untuk mendapatkan pusat kluster yang tepat, sehingga akan diperoleh kecenderungan data untuk masuk ke suatu kluster. Apabila nilai mutlak pengurangan nilai fungsi objektif ke- t dengan fungsi objektif ke- $(t - 1)$ telah mencapai kondisi konvergen, maka iterasi dihentikan. Sebaliknya, apabila belum mencapai kondisi konvergen, maka iterasi akan dilanjutkan. Untuk menghitung nilai fungsi objektif persamaan yang digunakan yaitu sebagai berikut.

$$P_t = \sum_{i=1}^{15} \sum_{k=1}^c ((\mu_{ik})^m d_{ik})$$

Untuk perhitungan nilai akhir fungsi objektif pada masing-masing kurun waktu data penelitian dapat dilihat pada Tabel 2 sebagai berikut.

Tabel 2 Nilai Akhir Perubahan Fungsi Objektif Data COVID-19 di Provinsi Lampung

k	m	Nilai Fungsi Objektif		
		Maret 2020- Agustus 2020	September 2020- Februari 2021	Maret 2020- Februari 2021
2	2	0.3302258595	0.3549726334	0.3432506453
	3	0.2595653652	0.2442393086	0.2368642545
	4	0.1673963235	0.1496120259	0.1453065396
	5	0.0977273992	0.0857488112	0.0833759989
3	2	0.0805845225	0.1250621001	0.1197303404
	3	0.0448155224	0.0577453403	0.0554314312
	4	0.0199177264	0.0229018839	0.0220350900
	5	0.0079388309	0.0085207341	0.0082313084
4	2	0.0443637758	0.0302838381	0.0286630850
	3	0.0183208446	0.0142798175	0.0229248395
	4	0.0058839211	0.0067672328	0.0066362649
	5	0.0017244969	0.0018589388	0.0014869678
5	2	0.0259338758	0.0119090635	0.0111310237
	3	0.0090604925	0.0049574337	0.0047005736
	4	0.0024499941	0.0014927360	0.0014298316
	5	0.0005710944	0.0003835573	0.0036543234

6. Melakukan perhitungan untuk mengetahui perubahan nilai keanggotaan setiap data pada setiap kluster menggunakan persamaan sebagai berikut.

$$\mu_{ik} = \left[\frac{[\sum_{j=1}^p d_{ik}]^{\frac{1}{m-1}}}{\sum_{k=1}^c [\sum_{j=1}^p d_{ik}]^{\frac{1}{m-1}}} \right]^{-1}$$

Apabila telah mencapai nilai konvergen, maka perubahan akhir nilai keanggotaan pada iterasi terakhirlah yang digunakan sebagai penentu suatu data masuk ke dalam kluster tertentu. Dalam penelitian ini, digunakan perubahan akhir nilai keanggotaan untuk data penelitian dengan jumlah kluster 2 dan pangkat pembobot 2. Untuk perubahan akhir nilai keanggotaan data penelitian dengan jumlah kluster 2 dan pangkat pembobot 2 dapat dilihat pada Tabel 3 sebagai berikut.

Tabel 3 Perubahan Akhir Nilai Keanggotaan Data COVID-19 di Provinsi Lampung untuk Jumlah Kluster 2 dengan Pangkat Pembobot 2

Kabupaten/Kota	Nilai Keanggotaan pada Iterasi Terakhir					
	Maret 2020- Agustus 2020		September 2020- Februari 2021		Maret 2020- Februari 2021	
	(μ_1)	(μ_2)	(μ_1)	(μ_2)	(μ_1)	(μ_2)
Lampung Barat	0.9949776	0.005022369	0.9988796731	0.0011203	0.998895392	0.0011046082
Tanggamus	0.9939649	0.006035110	0.9998772299	0.0001227	0.999896819	0.0001031812
Lampung Selatan	0.9801086	0.019891418	0.9971941235	0.0028058	0.996936223	0.0030637770

Lampung Timur	0.9944618	0.005538169	0.9889060373	0.0110939	0.990402106	0.0095978940
Lampung Tengah	0.9716484	0.028351571	0.8290947044	0.1709053	0.836842124	0.1631578761
Lampung Utara	0.9937452	0.006254817	0.9893364472	0.0106635	0.989690488	0.0103095123
Way Kanan	0.9972210	0.002778954	0.9927789040	0.0072210	0.993360911	0.0066390891
Tulang Bawang	0.9935285	0.006471479	0.9946449467	0.0053550	0.994582956	0.0054170440
Pesawaran	0.9986695	0.001330503	0.9999584333	0.0000415	0.999975217	0.0000247825
Pringsewu	0.9958359	0.004164106	0.9991685098	0.0008314	0.999083717	0.0009162831
Mesuji	0.9935285	0.006471479	0.9927796995	0.0072203	0.992777408	0.0072225923
Tulang Bawang Barat	0.9982387	0.001761282	0.9955269446	0.0044730	0.995657314	0.0043426856
Pesisir Barat	0.9520756	0.047924450	0.9934128196	0.0065871	0.994459404	0.0055405964
Bandar Lampung	0.0000108	0.999989104	0.0004079183	0.9995921	0.000344957	0.9996550430
Metro	0.9973574	0.002642578	0.9994748778	0.0005251	0.999598054	0.0004019458

7. Memeriksa kondisi berhenti, dengan melihat suatu nilai jika $|P_t - P_{t-1}| < \epsilon$ atau $t > MaxIter$, maka iterasi dihentikan karena telah mencapai nilai konvergen dan jika tidak memenuhi kondisi tersebut, maka iterasi dilanjutkan dengan $t = t + 1$. Sebagai contoh perhitungan untuk memeriksa kondisi berhenti, digunakan perubahan fungsi objektif data COVID-19 di Provinsi Lampung kurun waktu 6 bulan (Maret 2020-Agustus 2020) dengan jumlah klaster 2 dan pangkat pembobot 2 sebagai berikut.

Tabel 4 Perubahan Fungsi Objektif Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (Maret 2020-Agustus 2020) dengan Jumlah Klaster 2 dan Pangkat Pembobot 2

Iterasi	Fungsi Objektif
1	1.2834328288
2	1.1864714059
3	1.1291232502
4	1.0047765623
5	0.5800165734
6	0.3320852692
7	0.3302283634
8	0.3302258641
9	0.3302258595

Dari tabel perubahan nilai fungsi objektif di atas, maka perhitungan untuk memeriksa kondisi berhentinya yaitu sebagai berikut.

$$\begin{aligned}
 |P_t - P_{t-1}| < \epsilon &\Leftrightarrow |P_9 - P_8| < \epsilon \\
 &\Leftrightarrow |0.3302258595 - 0.3302258641| < \epsilon \\
 &\Leftrightarrow |-0.0000000046| < \epsilon \\
 &\Leftrightarrow 0.0000000046 < \epsilon
 \end{aligned}$$

Dari perhitungan di atas, diperoleh nilai $|P_t - P_{t-1}| = 0.0000000046$. Karena nilai tersebut sudah kurang dari nilai galat yang diharapkan yaitu 10^{-9} atau telah mencapai nilai konvergen, maka iterasi dihentikan pada iterasi ke-9.

8. Melakukan klasterisasi setiap objek berdasarkan perubahan akhir nilai keanggotaan di setiap datanya. Dalam proses pengklasteran suatu objek, apabila nilai keanggotaan suatu objek pada suatu klaster tertentu lebih besar dibandingkan dengan nilai keanggotaan klaster lainnya, maka objek tersebut lebih cenderung untuk masuk ke dalam klaster tersebut. Dalam penelitian ini, hasil klasterisasi data COVID-19 di Provinsi Lampung untuk kurun waktu 6 bulan (Maret 2020-Agustus 2020) untuk jumlah klaster 2 dengan pangkat pembobot 2. Untuk hasil klasterisasi data penelitian dengan jumlah klaster 2 dan pangkat pembobot 2 dapat dilihat pada Tabel 5.

Tabel 5 Klasterisasi Data COVID-19 di Provinsi Lampung Kurun Waktu 6 Bulan (Maret 2020-Agustus 2020) untuk Jumlah Klaster 2 dengan Pangkat Pembobot 2

Kabupaten/Kota	Nilai Keanggotaan pada Iterasi Terakhir		Kecenderungan Data Masuk pada Klaster	
	(μ_1)	(μ_2)	C1	C2
Lampung Barat	0.9949776	0.005022369	*	
Tanggamus	0.9939649	0.006035110	*	
Lampung Selatan	0.9801086	0.019891418	*	
Lampung Timur	0.9944618	0.005538169	*	
Lampung Tengah	0.9716484	0.028351571	*	

Lampung Utara	0.9937452	0.006254817	*	
Way Kanan	0.9972210	0.002778954	*	
Tulang Bawang	0.9935285	0.006471479	*	
Pesawaran	0.9986695	0.001330503	*	
Pringsewu	0.9958359	0.004164106	*	
Mesuji	0.9935285	0.006471479	*	
Tulang Bawang Barat	0.9982387	0.001761282	*	
Pesisir Barat	0.9520756	0.047924450	*	
Bandar Lampung	0.0000108	0.999989104		*
Metro	0.9973574	0.002642578	*	

3.2 Pemilihan Jumlah Kluster Optimal Menggunakan *Partition Coefficient Index (PCI)*

Dalam penelitian ini, untuk memperoleh jumlah kluster yang optimal pada hasil pengklasteran, akan dilakukan pemilihan jumlah kluster dengan melihat kriteria pada nilai *Partition Coefficient Index (PCI)*. Berdasarkan kriteria nilai PCI, jumlah kluster yang optimal dalam suatu pengklasteran ditunjukkan dengan nilai PCI yang paling besar diantara nilai PCI lainnya. Apabila jumlah kluster tertentu memiliki nilai PCI paling besar, maka jumlah kluster optimal yang digunakan dalam pengklasteran adalah jumlah kluster tersebut. Adapun perbandingan nilai PCI untuk data penelitian dapat dilihat pada Tabel 6 sebagai berikut.

Tabel 6 Perbandingan Nilai PCI Data COVID-19 di Provinsi Lampung

k	m	Nilai PCI		
		Maret 2020- Agustus 2020	September 2020- Februari 2021	Maret 2020- Februari 2021
2	2	0.981213222867858	0.973370329481600	0.974509194247106
	3	0.865384105743657	0.879843548955733	0.882559200567662
	4	0.752181423419151	0.773495511950479	0.776642737272181
	5	0.676479002723250	0.695054310400800	0.697808947868284
3	2	0.871008933652634	0.826038740794927	0.831029752804562
	3	0.699189733050337	0.659782376364209	0.664878538307290
	4	0.583866694340234	0.558690751584379	0.559677756382243
	5	0.513362385994510	0.501461867426699	0.499136997022760
4	2	0.794867053032111	0.853840608951535	0.853040023757533
	3	0.598490634514269	0.670228520283933	0.609699760594083
	4	0.485235513130723	0.503494547607457	0.497704138545581
	5	0.427348328617540	0.443768957097599	0.470291351891630
5	2	0.808397986724951	0.840220059275874	0.841126070542464
	3	0.567776303810071	0.642767277114434	0.640684908237989
	4	0.455840798269207	0.512752882859468	0.504383550179010
	5	0.417971063347432	0.443417424674412	0.433725896836339

Berdasarkan Tabel 6, nilai PCI paling besar untuk setiap data penelitian terdapat pada jumlah kluster 2 dengan pangkat pembobot yang digunakan yaitu 2. Dengan melihat kriteria PCI, maka jumlah kluster yang optimal untuk pengklasteran data COVID-19 di Provinsi Lampung yaitu 2 dengan pangkat pembobot yang digunakan yaitu 2.

3.3 Hasil Kluster Terbaik Menggunakan Metode *Fuzzy C-Means (FCM)*

Setelah mendapatkan jumlah kluster optimal dan pangkat pembobot yang digunakan untuk pengklasteran data penelitian, maka diperoleh hasil kluster terbaik untuk pengklasteran kabupaten/kota di Provinsi Lampung yang dapat dilihat pada Tabel 7 sebagai berikut.

Tabel 7 Hasil Kluster Terbaik Menggunakan Metode *Fuzzy C-Means*

Kurun Waktu	Kluster	Jumlah Anggota	Kabupaten/Kota
Maret 2020- Agustus 2020	1	14	Lampung Barat, Tanggamus, Lampung Selatan, Lampung Timur, Lampung Tengah, Lampung Utara, Way Kanan, Tulang Bawang, Pesawaran, Pringsewu, Mesuji, Tulang Bawang Barat, Pesisir Barat, Metro
	2	1	Bandar Lampung
September 2020-Februari	1	14	Lampung Barat, Tanggamus, Lampung Selatan, Lampung Timur, Lampung Tengah, Lampung Utara, Way Kanan, Tulang

2021			Bawang, Pesawaran, Pringsewu, Mesuji, Tulang Bawang Barat, Pesisir Barat, Metro
	2	1	Bandar Lampung
Maret 2020- Februari 2021	1	14	Lampung Barat, Tanggamus, Lampung Selatan, Lampung Timur, Lampung Tengah, Lampung Utara, Way Kanan, Tulang Bawang, Pesawaran, Pringsewu, Mesuji, Tulang Bawang Barat, Pesisir Barat, Metro
	2	1	Bandar Lampung

4. Kesimpulan

Berdasarkan hasil dan pembahasan pada penelitian ini, diperoleh kesimpulan bahwa jumlah kluster optimal untuk pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung menggunakan metode FCM adalah 2. Hal ini dikarenakan nilai PCI pada jumlah kluster 2 merupakan nilai PCI yang paling besar jika dibandingkan dengan jumlah kluster lainnya. Berdasarkan jumlah kluster optimal sebelumnya, maka diperoleh hasil kluster terbaik untuk pengklasteran kabupaten/kota berdasarkan kasus positif, kasus meninggal, dan kasus sembuh COVID-19 di Provinsi Lampung untuk kurun waktu 6 bulan (Maret 2020-Agustus 2020 dan September 2020-Februari 2021) serta kurun waktu 12 bulan (Maret 2020-Februari 2021) dengan menggunakan metode FCM, yaitu untuk klusterisasi data penelitian menghasilkan hasil kluster yang sama, dimana Kota Bandar Lampung merupakan kabupaten/kota di Provinsi Lampung yang masuk ke dalam kluster tinggi, sedangkan untuk kabupaten/kota lainnya masuk ke dalam kluster rendah. Selain itu, diketahui bahwa tidak terdapat perpindahan hasil kluster pada data COVID-19 di Provinsi Lampung untuk kurun waktu 6 bulan pertama (Maret 2020-Agustus 2020) dan 6 bulan terakhir (September 2020-Februari 2021).

Daftar Pustaka:

- [1] Bora, D.J. & Gupta, A.K. 2014. A Comparative Study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm. *International Journal of Computer Trends and Technology (IJCTT)*. **10**(2): 108-113.
- [2] Djarwanto. 2012. *Mengenal Beberapa Uji Statistik dalam Penelitian*. Liberty, Yogyakarta.
- [3] Dwitri, N., dkk. 2020. Penerapan Algoritma K-Means dalam Menentukan Tingkat Penyebaran Pandemi COVID-19 di Indonesia. *Jurnal Teknologi Informasi*. **4**(1): 128-132.
- [4] Fathia, A.N., Rahmawati, R., & Tarno. 2016. Analisis Kluster Kecamatan di Kabupaten Semarang Berdasarkan Potensi Desa Menggunakan Metode Ward dan Single Linkage. *Jurnal Gaussian*. **5**(4): 801-810.
- [5] Furqon. 1999. *Statistika Terapan untuk Penelitian*. Alfabeta, Bandung.
- [6] Gaspersz, V. 1989. *Statistika*. Armico, Bandung.
- [7] Johnson, R.A. & Wincern, D.W. 1982. *Applied Multivariate Statistical Analysis*. Prentice Hal, Inc., New Jersey.
- [8] Kusumadewi, S. 2010. *Aplikasi Logika Fuzzy untuk Pendukung Keputusan*. Graha Ilmu, Yogyakarta.
- [9] Megawati, N., Mukid, M.A., & Rahmawati, R. 2013. Segmentasi Pasar pada Pusat Perbelanjaan Menggunakan Fuzzy C-Means (Studi Kasus: Rita Pasaraya Cilacap). *Jurnal Gaussian*. **2**(4): 343-350.
- [10] Sitepu, R., Irmeilyana, & Gultom, B. 2011. Analisis Cluster terhadap Tingkat Pencemaran Udara pada Sektor Industri di Sumatera Selatan. *Jurnal Penelitian Sains*. **14**(3): 11-17.
- [11] Sutoyo, M.S. & Sumpala, A.T. 2015. Penerapan Fuzzy C-Means untuk Deteksi Dini Kemampuan Penalaran Matematis. *Scientific Journal of Informatics*. **2**(15): 129-136.
- [12] Wu, K.L. & Yang, M.S. 2005. A Cluster Validity Index for Crisp Clusters. *Pattern Analysis and Applications*. **26**(9): 1275-1291.