



## IMPLEMENTASI METODE *FUZZY POSSIBILISTIC C-MEANS* PADA PENGELOMPOKAN PROVINSI DI INDONESIA BERDASARKAN DATA JUMLAH KEJADIAN DAN DAMPAK BENCANA BANJIR

Khairun Nida<sup>1</sup>, Memi Nor Hayati<sup>2</sup>, Rito Goejantoro<sup>3</sup>

Corresponding author : Khairun Nida

<sup>1</sup>Universitas Mulawarman, khairunida21@gmail.com

<sup>2</sup>Universitas Mulawarman, meminorhayati@fmipa.unmul.ac.id

<sup>3</sup>Universitas Mulawarman, rito.goejantoro@fmipa.unmul.ac.id

Received : 23 Juli 2023, Revised : 31 Maret 2024, Accepted : 31 Maret 2024

### Abstract

Cluster analysis is one of the techniques in data mining used to find similar groups of data objects. The Fuzzy Possibilistic C-Means (FPCM) method is a method of clustering that is a development of the Fuzzy C-Means (FCM) and Possibilistic C-Means (PCM) algorithms using the advantages of fuzzy and possibilistic modeling. This study aims to find out the optimal number of clusters based on the validity index of the Modified Partition Coefficient (MPC) as well as the results of the optimal grouping of 34 Provinces in Indonesia based on data on the number of occurrences and effects of flood disasters in 2017-2021. According to the National Disaster Management Agency (BNPB), from 2017 to 2021 the number of natural disasters that occurred in Indonesia reached 18,658 incidents where flood disasters were included in the major disaster category with a total percentage of incidents of 28% from 2017 to 2021. Therefore, it is necessary to grouping provinces in Indonesia based on the impact of flood disasters as a mitigation effort in recognizing the risk of flood disasters. The optimal number of clusters using the FPCM method based on the MPC validity index is 2 clusters, the first cluster consisting of 19 Provinces in Indonesia and the second cluster consisting of 15 Provinces in Indonesia. The first cluster was dominated by provinces in the Sumatra Islands whose area consisted mostly of highlands and mountains, as well as provinces located in the islands of Papua and Maluku which had a smaller population than other provinces. Meanwhile, in the second cluster it is dominated by provinces with a fairly high number of riverbank settlements.

*Keywords* : Cluster Analysis, Flood Disaster, FPCM, Modified Partition Coefficient

### Abstrak

Analisis *cluster* merupakan salah satu teknik dalam *data mining* yang digunakan untuk menemukan kelompok objek data yang serupa. Metode *Fuzzy Possibilistic C-Means* (FPCM) adalah salah satu metode *clustering* yang merupakan pengembangan dari algoritma *Fuzzy C-Means* (FCM) dan *Possibilistic C-Means* (PCM) dengan menggunakan kelebihan dari pemodelan *fuzzy* dan *possibilistic*. Penelitian ini bertujuan untuk mengetahui jumlah *cluster* optimal berdasarkan indeks validitas *Modified Partition Coefficient* (MPC) serta mengetahui hasil pengelompokan optimal 34 Provinsi di Indonesia berdasarkan data jumlah kejadian dan dampak bencana banjir pada tahun 2017-2021. Menurut Badan Nasional Penanggulangan Bencana (BNPB) sejak tahun 2017 hingga 2021 jumlah bencana alam yang terjadi di Indonesia mencapai 18.658 kejadian di mana bencana banjir termasuk kategori bencana yang besar dengan persentase total kejadian 28% sejak tahun 2017 hingga 2021. Oleh sebab itu, perlu dilakukan pengelompokan Provinsi di Indonesia berdasarkan dampak bencana banjir sebagai upaya mitigasi dalam mengenali risiko bencana banjir. Jumlah *cluster* optimal dengan menggunakan metode FPCM berdasarkan indeks validitas MPC adalah sebanyak 2 *cluster* yaitu *cluster* pertama beranggotakan 19 Provinsi di Indonesia dan *cluster* kedua beranggotakan 15 Provinsi di Indonesia. *Cluster* pertama didominasi oleh provinsi di Kepulauan Sumatera yang sebagian besar kawasannya terdiri dari dataran tinggi dan pegunungan, serta provinsi yang terletak di Kepulauan Papua dan Maluku yang memiliki jumlah penduduk lebih kecil dibandingkan dengan provinsi lain. Sementara pada *cluster* kedua didominasi oleh provinsi dengan jumlah pemukiman bantaran sungai yang cukup tinggi.

## 1. Pendahuluan

Salah satu bidang ilmu yang digunakan untuk menangani masalah penggalian data berupa informasi dari suatu *database* yang besar disebut dengan *data mining*. Beberapa teknik yang sering digunakan dalam *data mining* diantaranya teknik prediksi, asosiasi, klasifikasi, dan analisis *cluster*. Analisis *cluster* merupakan teknik untuk menemukan dan mengelompokkan objek yang serupa pada suatu *dataset* [1]. Dalam analisis *cluster* terdapat dua cara yang dapat digunakan dalam mengelompokkan objek pengamatan, yaitu secara hierarki dan atau non hierarki [2]. Beberapa jenis metode non hierarki adalah *parallel threshold procedure*, *sequential threshold procedure*, dan metode partisi. Metode partisi merupakan sebuah metode di mana penempatan objek pada tiap *cluster* menggunakan pertimbangan optimasi [3]. Beberapa jenis metode pengelompokan yang termasuk dalam metode partisi ini diantaranya yaitu metode *K-Means*, *K-Medoids*, dan *fuzzy clustering* [4].

*Fuzzy C-Means* (FCM) termasuk salah satu algoritma *fuzzy clustering* yang menerapkan metode pengelompokan non hierarki di mana letak objek data dalam tiap-tiap *cluster* ditentukan oleh nilai keanggotaan. Dalam *fuzzy clustering* terdapat partisi possibilistik. Partisi possibilistik akan menunjukkan possibilistik dan probabilistik tiap objek data yang diinterpretasikan oleh *typicality value* (nilai keanggotaan kekhasan). Semakin tinggi nilai keanggotaan kekhasan suatu objek maka akan menyebabkan semakin besar pula kemungkinan objek tersebut masuk pada *cluster* tertentu. Metode *clustering* yang menggunakan partisi possibilistik dalam pengelompokannya adalah metode *Possibilistic C-Means* (PCM). Adapun pengembangan dari metode PCM dan FCM ialah metode *Fuzzy Possibilistic C-Means* (FPCM) (Apsari et al., 2020).

Penerapan metode FPCM telah dibahas pada penelitian sebelumnya. Penelitian [6] mengimplementasikan algoritma FCM dan FPCM untuk melakukan klasterisasi data *tweets* pada akun Twitter Tokopedia. Hasil perbandingan dengan menggunakan indeks

validitas *Modified Partition Coefficient* (MPC) ialah metode FCM merupakan metode yang lebih baik dibandingkan dengan metode FPCM. Kemudian penelitian [7] membandingkan metode FCM, FPCM, dan PFCM pada *noisy* data dengan studi kasus fasilitas tempat tinggal setiap provinsi tahun 2017 di mana menggunakan jumlah *cluster* awal sebanyak 2, 3, 4, dan 5 *cluster*. Berdasarkan indeks validitas MPC didapatkan hasil bahwa metode FPCM merupakan metode terbaik dengan jumlah *cluster* optimal sebanyak 5 *cluster*. Selain itu penelitian menurut [8] yang membahas mengenai penerapan *clustering* data titik gempa dengan menggunakan metode FPCM berhasil melakukan pengelompokan data titik gempa bumi di wilayah Pulau Sumatera menjadi 3 *cluster*. Berdasarkan beberapa penelitian yang telah dilakukan sebelumnya, selain dapat diterapkan pada bidang teknologi dan fasilitas tempat tinggal, FPCM juga dapat diterapkan dalam bidang mitigasi bencana gempa maupun bencana lain seperti banjir.

Banjir adalah peristiwa tergenangnya tanah yang diakibatkan oleh luapan air sungai. Banjir dapat diakibatkan oleh curah hujan maupun aliran air yang turun dari daerah lain yang lebih tinggi [9]. Penelitian [10] menyatakan bahwa bencana banjir menduduki posisi ketiga sebagai penyebab kerugian ekonomi dari semua bencana alam di seluruh dunia. Menurut Badan Nasional Penanggulangan Bencana (BNPB) sejak tahun 2017 hingga 2021 jumlah bencana alam yang terjadi di Indonesia mencapai 18.658 kejadian di mana bencana banjir termasuk kategori bencana yang besar dengan persentase total kejadian 28% sejak tahun 2017 hingga 2021. Adapun jumlah kejadian banjir dari tahun 2017 hingga 2021 berturut-turut adalah 992, 883, 796, 1.522, dan 1.175 kejadian.

Terjadinya bencana banjir tidak bisa terus dihindari akan tetapi dapat dilakukan berbagai upaya mitigasi bencana. Adapun upaya mitigasi ini tidak dapat disusun secara seragam pada setiap daerah karena perlu memperhatikan kondisi geografis setiap wilayah yang berbeda. Oleh sebab itu, perlu

dilakukan pengelompokan Provinsi di Indonesia berdasarkan dampak bencana banjir sebagai upaya mitigasi khususnya dalam mengenali risiko bencana banjir [11].

Berdasarkan latar belakang tersebut, maka akan dilakukan penelitian ilmiah dengan judul “Implementasi Metode *Fuzzy Possibilistic C-Means* Pada Pengelompokan Provinsi di Indonesia Berdasarkan Data Jumlah Kejadian dan Dampak Bencana Banjir”.

## 2. Metode

Metode analisis yang digunakan ialah metode *clustering* FPCM. Adapun tahapan analisis pada penelitian ini dengan menggunakan bantuan *software R* adalah sebagai berikut :

- Melakukan penginputan data ( $X_{ij}$ ) ke dalam matriks  $\mathbf{X}$  yang memiliki ukuran  $n \times m$  dengan  $i = 1, 2, \dots, 34$ ; dan  $j = 1, 2, \dots, 10$ .
- Mendeskripsikan data penelitian.
- Melakukan standarisasi data dengan metode standarisasi z-score menggunakan persamaan (1).

$$X'_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j} \quad (1)$$

$$\bar{X}_j = \frac{1}{n} \times \sum_{i=1}^n X_{ij} \quad (2)$$

$$S_j = \sqrt{\frac{1}{n-1} \times \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2} \quad (3)$$

di mana :

$X'_{ij}$  = standarisasi data ke- $i$  variabel ke- $j$

$n$  = banyaknya data

$X_{ij}$  = data ke- $i$  variabel ke- $j$

$\bar{X}_j$  = rata-rata variabel ke- $j$

$S_j$  = deviasi standar variabel ke- $j$  [12]

- Mendeteksi multikolinearitas data dengan menghitung nilai VIF (*Variance Inflation Factor*) setiap variabel menggunakan persamaan (4).

$$VIF_j = \frac{1}{1 - R_j^2} \quad (4)$$

di mana  $R_j^2$  merupakan koefisien determinasi antara variabel  $X_j$  yang

diregresikan dengan variabel lainnya di mana jika ditemukan variabel yang memiliki nilai VIF lebih besar dari 10 maka dapat diindikasikan adanya multikolinearitas pada data pengamatan [13].

- Mengelompokkan data dengan menggunakan metode FPCM. Adapun langkah-langkahnya adalah sebagai berikut :

a. Menentukan :

a) Jumlah *cluster* =  $c$  ( $\geq 2$ )

b) Bobot eksponen untuk partisi *fuzzy* = ( $w > 1$ )

c) Bobot eksponen untuk partisi *possibilistik* = ( $\eta > 1$ )

d) Maksimum iterasi (*MaxIter*)

e) *Error* terkecil yang diharapkan ( $\epsilon$ )

f) Fungsi objektif awal ( $P_0$ ) = 0

- Memanggil matriks kekhasan relatif ( $\mathbf{U}$ ) dan pusat *cluster* ( $V_{kj}$ ) pada hasil akhir algoritma FCM untuk menghitung matriks kekhasan absolut berukuran  $n \times c$  di mana  $i = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, c$ ; dan  $j = 1, 2, \dots, m$ ; sebagai berikut :

$$\mathbf{T} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1c} \\ t_{21} & t_{22} & \dots & t_{2c} \\ \vdots & \vdots & \dots & \vdots \\ t_{n1} & t_{n2} & \dots & t_{nc} \end{bmatrix} \quad (5)$$

dengan elemen matriks sebagai berikut :

$$t_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1}}{\sum_{i=1}^n [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1}} \quad (6)$$

- Memperbaiki pusat *cluster* di mana  $i = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, c$ ; dan  $j = 1, 2, \dots, m$  dengan persamaan berikut :

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik}^w + t_{ik}^\eta) X_{ij}}{\sum_{i=1}^n (\mu_{ik}^w + t_{ik}^\eta)} \quad (7)$$

- Menghitung fungsi objektif pada iterasi ke- $t$  dengan menggunakan persamaan :

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik}^w + t_{ik}^\eta) \right) \quad (8)$$

- Memperbaiki matriks kekhasan relatif ( $\mathbf{U}$ ) dengan elemen ( $\mu_{ik}$ ) di mana  $i = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, c$ ; dan  $j = 1, 2, \dots, m$  dengan persamaan berikut :

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-1}} \quad (9)$$

- Memperbaiki matriks kekhasan absolut ( $\mathbf{T}$ ) dengan elemen ( $t_{ik}$ ) dengan persamaan :

$$t_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{\eta-1}}}{\sum_{i=1}^n [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{\eta-1}}} \quad (10)$$

- g. Mengulangi langkah ke-c sampai langkah ke-f hingga  $(|P_t - P_{t-1}| < \epsilon)$  atau  $t > MaxIter$  terpenuhi [6].
- h. Mengulangi langkah ke-b sampai langkah ke-g untuk *cluster* yang berbeda.
6. Menghitung indeks validitas untuk setiap *cluster* menggunakan metode validitas *Modified Partition Coefficient* (MPC) dengan persamaan (11) dan (12).

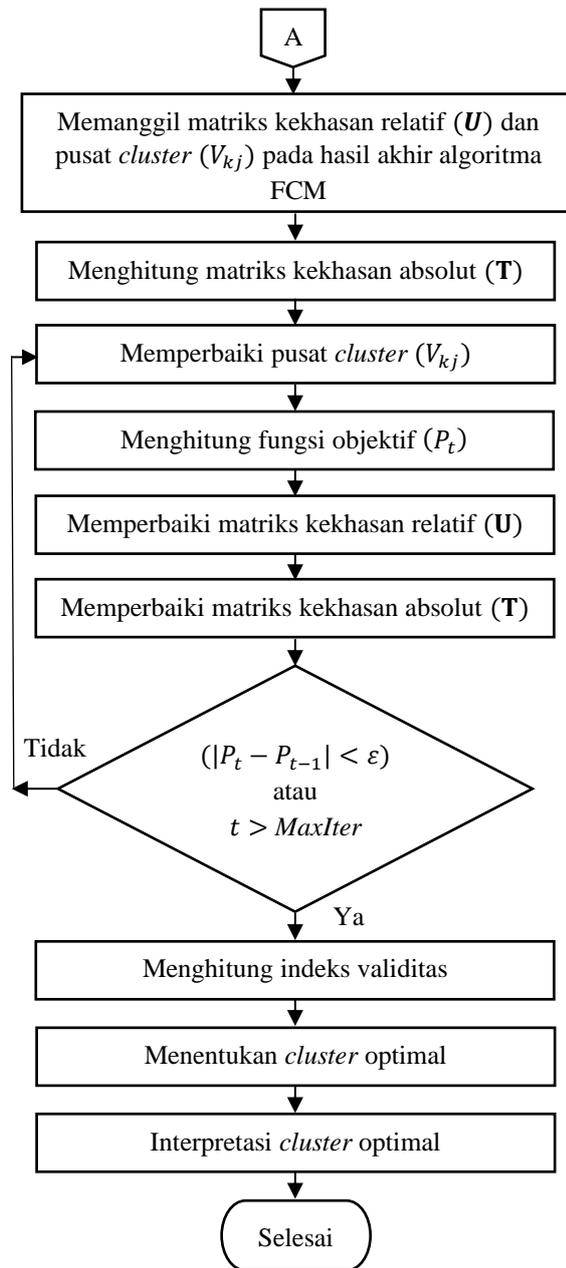
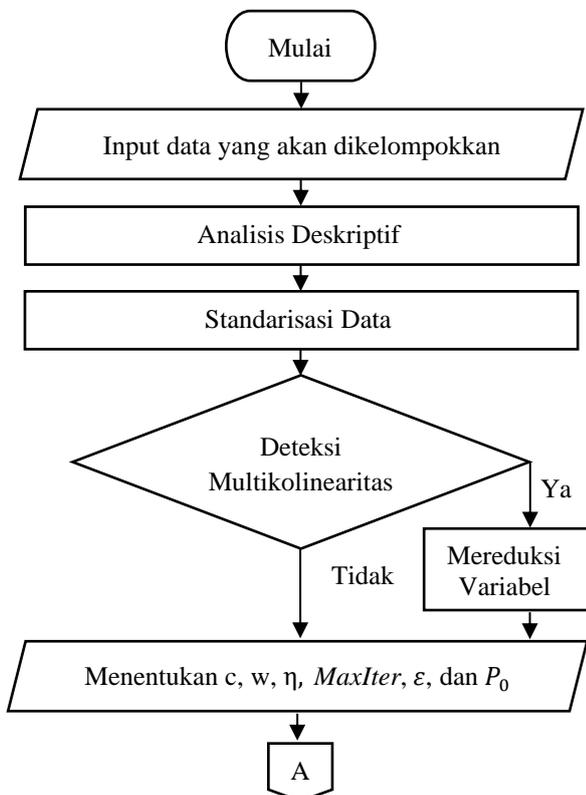
$$MPC(c) = 1 - \frac{c}{c-1} (1 - PC) \quad (11)$$

$$PC(c) = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^c (\mu_{ik})^2 \quad (12)$$

[14]

Perhitungan indeks validitas digunakan untuk mengukur kinerja hasil pengelompokan. Indeks MPC memiliki rentang antara  $0 < MPC(c) \leq 1$ , semakin besar nilai indeks MPC (mendekati 1) maka akan semakin baik pula *cluster* yang dihasilkan [14].

7. Menginterpretasikan hasil pengelompokan berdasarkan indeks validitas. Tahapan metode FPCM dapat dilihat pada Gambar 1. berikut ini :



Gambar 1. Tahapan metode FPCM

### 3. Pembahasan

#### 3.1 Data Penelitian

Data penelitian yang digunakan adalah data sekunder yaitu data jumlah kejadian dan dampak bencana banjir menurut 34 Provinsi di Indonesia tahun 2017-2021. Data diperoleh melalui website <https://dibi.bnpb.go.id/> Badan Nasional Penanggulangan Bencana (BNPB) Indonesia. Adapun variabel penelitian adalah sebanyak 10 variabel sebagai berikut :

- Jumlah Kejadian Bencana
- Korban Meninggal
- Korban Hilang
- Korban Terluka

- e. Korban Menderita
- f. Korban Mengungsi
- g. Kerusakan Rumah
- h. Kerusakan Fasilitas Pendidikan
- i. Kerusakan Fasilitas Kesehatan
- j. Kerusakan Fasilitas Peribadatan

### 3.2 Analisis Statistika Deskriptif

Analisis statistika deskriptif pada data jumlah kejadian dan dampak bencana banjir menurut 34 Provinsi di Indonesia tahun 2017-2021 menggunakan nilai minimum, nilai maksimum, dan rata-rata. Adapun hasil analisis statistika deskriptif disajikan pada Tabel 1.

**Tabel 1. Statistika Deskriptif**

Variabel	Minimum	Maksimum	Rata-Rata
$X_1$	12	924	157,82
$X_2$	0	250	31,15
$X_3$	0	101	6,50
$X_4$	0	1.113	85
$X_5$	770	2.702.033	399.698
$X_6$	59	538.343	48.128
$X_7$	0	54.430	4899,5
$X_8$	1	603	98,71
$X_9$	0	129	17,59
$X_{10}$	1	646	78

### 3.3 Standardisasi Data

Perhitungan standardisasi data jumlah kejadian dan dampak bencana banjir pada tiap variabel dihitung menggunakan persamaan (3) di mana perhitungan standardisasi data dilakukan pada data pengamatan ke-1 (Provinsi Aceh) hingga data pengamatan ke-34 (Provinsi Papua) untuk setiap variabel. Hasil standardisasi data terdapat pada Tabel 2.

**Tabel 2. Hasil Standardisasi Data**

No.	Provinsi	$X_1$	$X_2$	...	$X_{10}$
1.	Aceh	0,9817	-0,4658	...	-0,4174
2.	Sumatera Utara	0,4264	0,6337	...	-0,0942
⋮	⋮	⋮	⋮	⋮	⋮
34.	Papua	-0,5699	1,5017	...	-0,3366

### 3.4 Pendeteksian Multikolinearitas

Pendeteksian multikolinearitas dihitung menggunakan persamaan (4). Nilai VIF setiap variabel penelitian disajikan pada Tabel 3.

**Tabel 3. Nilai VIF**

Variabel	VIF
$X_1$	12,7012
$X_2$	9,5109
$X_3$	10,4489
$X_4$	7,7349
$X_5$	30,2803
$X_6$	10,3063
$X_7$	4,0003
$X_8$	6,5933
$X_9$	7,6938
$X_{10}$	25,6508

Berdasarkan Tabel 3, terdapat 5 variabel yang memiliki nilai VIF lebih besar dari 10, sehingga dapat diindikasikan adanya multikolinearitas antar variabel pengamatan dan belum dapat dilanjutkan ke dalam proses pengelompokan. Untuk mengatasi multikolinearitas, variabel yang memiliki nilai VIF terbesar akan dikeluarkan yaitu variabel  $X_5$ , kemudian diregresikan kembali variabel yang tersisa. Setelah mengeluarkan variabel  $X_5$  dan melakukan perhitungan regresi didapatkan nilai VIF sebagai berikut :

**Tabel 4. Nilai VIF setelah mengeluarkan variabel  $X_5$**

Variabel	VIF
$X_1$	3,0567
$X_2$	9,5091
$X_3$	8,5771
$X_4$	5,9593
$X_6$	8,5920
$X_7$	3,8158
$X_8$	6,5622
$X_9$	7,6935
$X_{10}$	18,8013

Berdasarkan Tabel 4, dapat diketahui bahwa masih terdapat 1 variabel yang memiliki nilai VIF lebih besar dari 10, sehingga variabel yang memiliki nilai VIF paling besar dikeluarkan yaitu variabel  $X_{10}$ , kemudian diregresikan kembali variabel yang tersisa. Setelah mengeluarkan variabel  $X_{10}$  dan melakukan perhitungan regresi didapatkan nilai VIF seperti pada Tabel 5. Berdasarkan Tabel 5, terlihat bahwa semua variabel memiliki nilai VIF lebih kecil dari 10, sehingga dapat disimpulkan bahwa tidak ditemukan lagi multikolinearitas antar variabel dan dapat dilanjutkan pada proses pengelompokan dengan menggunakan variabel yang tidak dikeluarkan.

**Tabel 5. Nilai VIF setelah mengeluarkan variabel  $X_5$  dan  $X_{10}$**

Variabel	VIF
$X_1$	3,0556
$X_2$	9,5038
$X_3$	8,5167
$X_4$	5,7620
$X_6$	2,0984
$X_7$	3,8144
$X_8$	5,3743
$X_9$	5,3293

### 3.5 Metode Pengelompokan Fuzzy Possibilistic C-Means (FPCM)

Tahapan-tahapan proses pengelompokan dengan menggunakan metode FPCM adalah sebagai berikut :

#### 1. Penentuan Nilai-Nilai Parameter Metode FPCM

Pada penelitian ini menggunakan banyak cluster ( $c$ ) = 2, 3, dan 4,  $w = 2$ ,  $\eta = 2$ ,  $MaxIter = 1000$ ,  $\varepsilon = 10^{-5}$ , dan  $P_0 = 0$ . Sebagai contoh perhitungan pada penelitian ini, maka akan ditampilkan pengelompokan FPCM dengan menggunakan  $c = 2$ .

#### 2. Memanggil Matriks Kekhasan Relatif (U) dan Pusat Cluster ( $V_{kj}$ ) Pada Hasil Akhir Algoritma FCM Untuk Menghitung Matriks Kekhasan Absolut (T)

**Tabel 6. Nilai Keanggotaan Kekhasan Relatif Akhir dengan  $c = 2$  pada Metode FCM**

No.	Provinsi	Nilai Keanggotaan	
		Cluster 1 ( $\mu_{i1}$ )	Cluster 2 ( $\mu_{i2}$ )
1.	Aceh	0,8888	0,1112
2.	Sumatera Utara	0,7728	0,2272
⋮	⋮	⋮	⋮
34.	Papua	0,4327	0,5673

**Tabel 7. Elemen Pusat Cluster Optimal dengan  $c = 2$  pada Metode FCM**

Variabel	Pusat Cluster	
	1	2
$X_1$	-0,23004	0,709911
$X_2$	-0,31713	1,263778
$X_3$	-0,20538	0,744523
$X_4$	-0,26312	1,155112
$X_6$	-0,19172	0,793262
$X_7$	-0,22541	0,990967
$X_8$	-0,34653	1,477075
$X_9$	-0,30142	1,408514

Langkah awal sebelum menghitung matriks kekhasan absolut (T) adalah

memanggil matriks kekhasan relatif (U) dan pusat cluster ( $V_{kj}$ ) pada hasil akhir algoritma FCM dengan bantuan software R. Adapun hasil akhir matriks kekhasan relatif (U) disajikan pada Tabel 6. Kemudian untuk hasil akhir pusat cluster ( $V_{kj}$ ) pada algoritma FCM disajikan pada Tabel 7.

#### 3. Menghitung Matriks Kekhasan Absolut (T)

Perhitungan nilai keanggotaan kekhasan absolut dengan  $c = 2$  menggunakan persamaan (6) dilakukan pada data ke-1 (Provinsi Aceh) hingga data nilai keanggotaan kekhasan absolut ke-34 (Provinsi Papua) untuk setiap cluster. Hasil perhitungan nilai keanggotaan kekhasan absolut disajikan pada Tabel 8.

**Tabel 8. Nilai Keanggotaan Kekhasan Absolut dengan  $c = 2$**

No.	Provinsi	Nilai Keanggotaan	
		Cluster 1 ( $t_{i1}$ )	Cluster 2 ( $t_{i2}$ )
1.	Aceh	0,0076	0,0296
2.	Sumatera Utara	0,0058	0,053
⋮	⋮	⋮	⋮
34.	Papua	0,0003	0,0119

#### 4. Menghitung Pusat Cluster ( $V_{kj}$ )

Setelah menghitung matriks kekhasan absolut (T), tahap selanjutnya adalah menghitung pusat cluster menggunakan persamaan (7). Nilai pusat cluster seluruh variabel disajikan pada Tabel 9.

**Tabel 9. Elemen Pusat Cluster Awal dengan  $c = 2$  pada Metode FPCM**

Variabel	Pusat Cluster	
	1	2
$X_1$	-0,2304	0,704
$X_2$	-0,3175	1,2508
$X_3$	-0,2055	0,7355
$X_4$	-0,2633	1,1417
$X_6$	-0,1919	0,7841
$X_7$	-0,2256	0,9809
$X_8$	-0,3467	1,4649
$X_9$	-0,3015	1,3964

#### 5. Menghitung Fungsi Objektif ( $P_t$ )

Perhitungan fungsi objektif pada iterasi pertama dengan jumlah  $c = 2$  dihitung menggunakan persamaan (8). Diperoleh nilai fungsi objektif iterasi 1 sebesar

145,0638 di mana nilai fungsi objektif awal ( $P_0$ ) adalah 0 sehingga  $|P_1 - P_0| = 145,0638 > \varepsilon = 10^{-5}$ . Selisih nilai fungsi objektif yang lebih besar dari nilai  $\varepsilon$  menyebabkan proses iterasi akan dilanjutkan.

6. Memperbaiki Matriks Kekhasan Relatif (U)

Langkah selanjutnya yaitu memperbaiki nilai keanggotaan kekhasan relatif dengan  $c = 2$  dengan menggunakan persamaan (9). Hasil nilai keanggotaan kekhasan relatif yang telah diperbarui disajikan pada Tabel 10.

**Tabel 10. Nilai Keanggotaan Kekhasan Relatif yang Telah Diperbarui dengan  $c = 2$**

No.	Provinsi	Nilai Keanggotaan	
		Cluster 1 ( $\mu_{i1}$ )	Cluster 2 ( $\mu_{i2}$ )
1.	Aceh	0,8873	0,1127
2.	Sumatera Utara	0,3432	0,6568
⋮	⋮	⋮	⋮
34.	Papua	0,4753	0,5247

7. Memperbaiki Matriks Kekhasan Absolut (T)

Setelah menghitung perubahan nilai keanggotaan pada matriks kekhasan relatif, dihitung pula perubahan nilai keanggotaan pada matriks kekhasan absolut dengan  $c = 2$  menggunakan persamaan (10). Hasil nilai keanggotaan kekhasan absolut yang telah diperbarui disajikan pada Tabel 11.

**Tabel 11. Nilai Keanggotaan Kekhasan Absolut yang Telah Diperbarui dengan  $c = 2$**

No.	Provinsi	Nilai Keanggotaan	
		Cluster 1 ( $t_{i1}$ )	Cluster 2 ( $t_{i2}$ )
1.	Aceh	0,0076	0,0032
2.	Sumatera Utara	0,0058	0,0369
⋮	⋮	⋮	⋮
34.	Papua	0,0003	0,0011

Setelah menghitung perubahan nilai keanggotaan kekhasan relatif (U) dan nilai keanggotaan kekhasan absolut (T), kemudian akan dihitung kembali pusat *cluster*, fungsi objektif, nilai keanggotaan kekhasan relatif dan nilai keanggotaan kekhasan absolut dengan  $c = 2$ . Iterasi berhenti ketika  $|P_t - P_{t-1}| < 10^{-5}$  atau  $t > 1000$ . Pada penelitian yang telah

dilakukan, proses perhitungan dihentikan pada iterasi ke-13. Adapun hasil akhir nilai keanggotaan kekhasan relatif disajikan pada Tabel 12.

**Tabel 12. Nilai Keanggotaan Kekhasan Relatif dengan  $c = 2$**

No.	Provinsi	Nilai Keanggotaan	
		Cluster 1 ( $\mu_{i1}$ )	Cluster 2 ( $\mu_{i2}$ )
1.	Aceh	0,8857	0,1143
2.	Sumatera Utara	0,766	0,234
⋮	⋮	⋮	⋮
34.	Papua	0,4333	0,5667

Kemudian untuk hasil akhir nilai keanggotaan kekhasan absolut dengan  $c = 2$  disajikan pada Tabel 13.

**Tabel 13. Nilai Keanggotaan Kekhasan Absolut dengan  $c = 2$**

No.	Provinsi	Nilai Keanggotaan		Cluster yang Diikuti
		Cluster 1 ( $t_{i1}$ )	Cluster 2 ( $t_{i2}$ )	
1.	Aceh	0,0075	0,0298	2
2.	Sumatera Utara	0,0057	0,0536	2
⋮	⋮	⋮	⋮	⋮
34.	Papua	0,0003	0,0116	2

**Tabel 14. Hasil Pengelompokan FPCM dengan  $c = 2$**

Cluster	Jumlah Anggota	Anggota Cluster
		Provinsi
1	19	Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat.
2	15	Aceh, Sumatera Utara, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Papua.

Berdasarkan nilai keanggotaan kekhasan absolut pada Tabel 13, data pengamatan ke-1 (Provinsi Aceh) mempunyai nilai keanggotaan sebesar 0,0075 pada *cluster*

pertama dan 0,0298 pada *cluster* kedua. Diperoleh nilai keanggotaan terbesar yaitu 0,0298 sehingga data pengamatan ke-1 lebih tepat menjadi anggota *cluster* kedua. Penentuan *cluster* yang diikuti dilakukan dengan cara yang sama pada data pengamatan ke-4 (Provinsi Riau) hingga data pengamatan ke-34 (Provinsi Papua), sehingga dapat diketahui hasil pengelompokan pada metode FPCM dengan  $c = 2$  seperti pada Tabel 14.

### 3.6 Menghitung Indeks Validitas

Penentuan metode terbaik untuk metode FPCM dilakukan dengan membandingkan hasil pengelompokan oleh masing-masing *cluster* dengan menghitung nilai MPC menggunakan persamaan (12) di mana semakin besar nilai MPC maka semakin baik *cluster* yang dihasilkan. Adapun hasil perhitungan nilai indeks validitas untuk  $c = 2, 3, \text{ dan } 4$  disajikan pada Tabel 15.

**Tabel 15. Nilai Indeks Validitas untuk Keseluruhan Cluster Pada Metode FPCM**

Banyak Cluster ( $c$ )	2	3	4
MPC	0,6506	0,5975	0,4953

Berdasarkan Tabel 15 dapat dilihat bahwa nilai MPC dengan  $c = 2$  merupakan nilai MPC terbesar dibandingkan dengan  $c = 3$  dan 4 sehingga diperoleh *cluster* optimal adalah sebanyak 2 *cluster*.

### 3.7 Interpretasi Hasil Cluster

Setelah *cluster* terbentuk, langkah selanjutnya yaitu menghitung nilai rata-rata semua variabel untuk setiap *cluster*. Hasil perhitungan rata-rata disajikan pada Tabel 16.

**Tabel 16. Nilai Rata-rata Variabel Setiap Cluster**

Variabel	Cluster ke- $c$	
	1	2
$X_1$	-0,4244	0,5375
$X_2$	0,4485	0,5681
$X_3$	-0,2852	0,3613
$X_4$	-0,3229	0,4090
$X_6$	-0,3636	0,4605
$X_7$	-0,3556	0,4504
$X_8$	-0,4932	0,6247
$X_9$	-0,3997	0,5064

Berdasarkan Tabel 16, dapat diketahui rata-rata variabel dari masing-masing *cluster* di mana terdapat perbedaan rata-rata pada keseluruhan variabel dari kedua *cluster* yang

telah terbentuk. Dari segi jumlah kejadian bencana banjir, korban meninggal, korban hilang, korban terluka, korban mengungsi, kerusakan rumah, kerusakan fasilitas pendidikan, dan kerusakan fasilitas kesehatan pada *cluster* 2 memiliki nilai rata-rata yang lebih tinggi dibandingkan dengan *cluster* 1. Selain itu, menurut data [15] dan [16] pada tahun 2018 dan 2021 provinsi yang termasuk dalam anggota *cluster* 2 didominasi oleh provinsi dengan jumlah desa/kelurahan dengan keberadaan permukiman di bantaran sungai yang lebih tinggi dibandingkan *cluster* 1. Rata-rata jumlah desa/kelurahan dengan keberadaan permukiman di bantaran sungai tahun 2021 pada *cluster* 1 adalah sebanyak 256 desa/kelurahan, sedangkan pada *cluster* 2 sebanyak 590 desa/kelurahan, di mana faktor ini merupakan salah satu penyebab terjadinya bencana banjir karena dengan berkembangnya bantaran sungai sebagai kawasan permukiman dapat berdampak pada turunnya nilai fungsi sungai yang akan mengakibatkan daerah tangkapan air (DTA) menyempit sehingga tidak ada lagi daerah resapan air hujan [17]. Sementara pada *cluster* 1 didominasi oleh provinsi di Kepulauan Sumatera yang sebagian besar kawasannya terdiri dari dataran tinggi dan pegunungan, salah satunya adalah Pegunungan Bukit Barisan yang memiliki luas 63,8% dari total keseluruhan luas Pulau Sumatera [18]. Selain itu, *cluster* 1 juga didominasi oleh provinsi dari Kepulauan Papua dan Maluku dengan jumlah penduduk yang kecil dibandingkan provinsi lain.

## 4. Penutup

### 4.1 Kesimpulan

Kesimpulan yang dapat diambil berdasarkan hasil penelitian dan pembahasan yang telah dilakukan adalah sebagai berikut :

1. Berdasarkan hasil indeks validitas MPC jumlah *cluster* yang optimal untuk mengelompokkan Provinsi di Indonesia berdasarkan data jumlah kejadian dan dampak bencana banjir menggunakan metode FPCM adalah sebanyak 2 *cluster* dengan nilai MPC terbesar adalah 0,6506.
2. Hasil pengelompokan optimal menggunakan metode FPCM pada data jumlah kejadian dan dampak bencana banjir menurut Provinsi di Indonesia

dengan nilai MPC sebesar 0,6506 pada jumlah *cluster* sebanyak 2 yaitu terdapat 19 Provinsi yang masuk dalam anggota *cluster* 1 dan 15 Provinsi yang masuk dalam anggota *cluster* 2. Adapun provinsi yang termasuk dalam anggota *cluster* 1 yaitu Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Bangka Belitung, Kepulauan Riau, DI Yogyakarta, Banten, Bali, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat. Sedangkan provinsi yang termasuk dalam anggota *cluster* 2 yaitu Aceh, Sumatera Utara, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Papua.

#### 4.2 Saran

Berdasarkan permasalahan yang ditemui, saran yang dapat diberikan peneliti adalah sebagai berikut :

1. Peneliti selanjutnya dapat menggunakan metode *Principal Component Analysis* (PCA) untuk mengatasi multikolinearitas.
2. Peneliti selanjutnya dapat membandingkan dengan metode *fuzzy clustering* lainnya, seperti *Possibilistic Fuzzy C-Means* (PFCM).

#### Referensi

- [1] I. Werdiningsih, B. Nuqoba, and Muhammadun, "Data Mining Menggunakan Android, Weka, dan SPSS", Surabaya: Airlangga University Press, 2020.
- [2] S. Santoso, "Menguasai Statistik Multivariat", Jakarta : PT Elex Media Komputindo, 2015.
- [3] M. Musfiani, "Analisis Cluster Dengan Menggunakan Metode Partisi Pada Pengguna Alat Kontrasepsi Di Kalimantan Barat", *Bimaster : Buletin Ilmiah Matematika, Statistika Dan Terapannya*, 8(4), 893–902, 2019.
- [4] O. J. Oyelade, O. O. Oladipupo, and I. C. Obagbuwa, "Application of K Means Clustering Algorithm for Prediction of Students Academic Performance", 7, 292–295, 2010.
- [5] G. R. Apsari, M. S. Pradana, and N. E. Chandra, "Implementasi Fuzzy C-Means dan Possibilistik C-Means Pada Data Performance Mahasiswa", *Unisda Journal of Mathematics and Computer Science (UJMC)*, 6(2), 39–48, 2020.
- [6] G. N. S. Putri, D. Ispriyanti, and T. Widiharih, "Implementasi Algoritma Fuzzy C-Means Dan Fuzzy Possibilistics C-Means Untuk Klasterisasi Data Tweets Pada Akun Twitter Tokopedia", *Jurnal Gaussian*, 11(1), 86–98, 2022.
- [7] K. H. Nuansari and Kariyam. "Perbandingan Metode Fuzzy C-Means, Fuzzy Possibilistic C-Means, dan Possibilistic C-Means Pada Noisy Data (Studi Kasus : Fasilitas Tempat Tinggal Setiap Provinsi Tahun 2017)", Universitas Islam Indonesia, 2018.
- [8] S. Putriana, E. Ernawati, and D. Andreswari, "Clustering Data Titik Gempa Dengan Metode Fuzzy Possibilistic C-Means (Studi Kasus: Titik Gempa Pulau Sumatera Tahun 2013-2018)", *Rekursif: Jurnal Informatika*, 9(1), 2021.
- [9] Bakornas PB, "Pengenalan karakteristik bencana dan upaya mitigasinya di Indonesia : Vol. II", Jakarta : Direktorat Mitigasi, Lahar Bakornas PB, 2007.
- [10] A. Findayani, "Kesiap Siagaan Masyarakat Dalam Penanggulangan Banjir Di Kota Semarang", *Jurnal Geografi : Media Informasi Pengembangan Dan Profesi Kegeografian*, 12(1), 102–114, 2015.
- [11] U. Islamy, D. R. Nursidah, I. S. Narendra, M. L. Anshori, and W. Edi., "Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Dampak Bencana Banjir Tahun 2017-2020", *Bimaster*, 11(2), 381–388, 2022.
- [12] Suyanto, "Data Mining Untuk Klasifikasi dan Klasterisasi Data", Bandung: Informatika, 2017.
- [13] D. Gujarati, "Ekonometri Dasar. Terjemahan: Sumarno Zain", Jakarta: Erlangga, 2003.
- [14] E. Prasetyo, "Data Mining Mengelola Data Menjadi Informasi Menggunakan

- Matlab Edisi Baru”, Yogyakarta : Penerbit CV. ANDI, 2014.
- [15] Badan Pusat Statistik, “Banyaknya Desa/Kelurahan Menurut Keberadaan Permukiman di Bantaran Sungai (Desa), 2014-2018”, Jakarta : Badan Pusat Statistik, 2019.
- [16] Badan Pusat Statistik, “Statistik Potensi Desa 2021”, Jakarta : Badan Pusat Statistik, 2022.
- [17] Ramadhani, A. Siswanto, and L. Teddy, “Analisis Penyebab Terjadinya Banjir Pada Pemukiman Kumuh di Kecamatan Ilir Barat I Palembang”, *Seminar Nasional AVoER XII 2020*, 439-444, 2020.
- [18] E. Hermawan, “Analisis Perilaku Curah Hujan Di Atas Kototabang Saat Bulan Basah dan Bulan Kering”, *Prosiding Seminar Nasional Penelitian, Pendidikan, dan Penerapan MIPA 2009*, 415-424, 2009.