

## Implementasi *Fuzzy C-Means* (FCM) dan *Generalized Fuzzy C-Means* (GFCM) dalam *Clustering* Produk Domestik Regional Bruto (PDRB) Menurut Pengeluaran

Anggun Yuliarum Qur'ani<sup>1</sup>, Ratna Sari Widiastuti<sup>2</sup>

<sup>1,2</sup>FMIPA, Prodi Matematika, Universitas Udayana

Email: [anggunyuliarum@unud.ac.id](mailto:anggunyuliarum@unud.ac.id)<sup>1</sup>, [ratnasariwidiastuti@unud.ac.id](mailto:ratnasariwidiastuti@unud.ac.id)<sup>2</sup>

### Article History:

Received: 30 Juli 2023

Revised: 04 Agustus 2023

Accepted: 05 Agustus 2023

**Keywords:** *FCM*, *GFCM*,  
*Clustering*, *PDRB*

**Abstract:** Pembagian objek ke dalam suatu kelompok yang mirip atau saling terkait satu sama lain dan berbeda dengan objek-objek dalam kelompok lain disebut dengan pengelompokan atau clustering. *Fuzzy Clustering* (FC) merupakan jenis clustering objek menjadi anggota dari setiap cluster dengan bobot keanggotaan antara 0 (tidak termasuk anggota) dan 1 (mutlak termasuk anggota). FC yang keberadaan atau letak suatu data dalam klaster tergantung pada derajat keanggotaan data tersebut disebut dengan *Fuzzy C-Means* (FCM) clustering. Modifikasi *Fuzzy C-Means* (FCM) dengan menambahkan sebuah parameter guna meningkatkan kecepatan konvergensi dan mendapatkan hasil yang lebih sedikit kesamaran dengan menyesuaikan matriks keanggotaan pada setiap iterasi disebut dengan *Generalized Fuzzy C-Means* (GFCM). Data Produk Domestik Regional Bruto (PDRB) merupakan statistik pendapatan regional yang dapat dimanfaatkan untuk mengevaluasi hasil pembangunan sekaligus sebagai dasar perencanaan pembangunan ekonomi di suatu wilayah tertentu. Pada penelitian ini ingin mengimplementasikan *Fuzzy C-Means* (FCM) dan *Generalized Fuzzy C-Means* (GFCM) dalam clustering Produk Domestik Regional Bruto (PDRB) menurut pengeluaran. Penelitian memberikan informasi bahwa jumlah klaster terbentuk adalah sebanyak 3 klaster dan hasil clustering memberikan kesimpulan bahwa FCM lebih tidak terlalu kabur (fuzzy). Informasi ini diperoleh dari nilai *Partition entropy* dan *XieBeni index* pada FCM lebih kecil daripada GFCM, dan nilai *Partition coeff* pada FCM lebih besar daripada GFCM.

### PENDAHULUAN

Pembagian objek ke dalam suatu kelompok yang mirip atau saling terkait satu sama lain dan berbeda dengan objek-objek dalam kelompok lain disebut dengan pengelompokan atau

*clustering*. Tujuan utama dari *clustering* adalah untuk membagi dataset ke dalam *cluster* yang mewakili objek yang mirip dengan *cluster* yang sama dan objek yang berbeda ke *cluster* yang berbeda (Rajkumar et al., 2019). Semakin besar kesamaan objek dalam suatu kelompok dan semakin besar perbedaan objek antar kelompok, maka semakin baik pengelompokannya. Pengelompokan dapat dianggap sebagai bentuk klasifikasi karena menciptakan pelabelan objek dengan label kelas (*cluster*).

Tan et al., (2019) membedakan jenis metode *clustering* menjadi 3 jenis, salah satunya *Fuzzy Clustering (FC)*. *FC* merupakan pengelompokan objek menjadi anggota dari setiap *cluster* dengan bobot keanggotaan antara 0 (tidak termasuk anggota) dan 1 (mutlak termasuk anggota). Dengan kata lain, *cluster* diperlakukan sebagai himpunan *fuzzy* yaitu himpunan sebuah objek yang termasuk ke dalam anggota dengan bobot antara 0 dan 1. *FC* yang keberadaan atau letak suatu data dalam kluster tergantung pada derajat keanggotaan data tersebut disebut dengan *Fuzzy C-Means (FCM) clustering*. Nilai derajat keanggotaan berada pada interval 0 dan 1. (Siringoringo & Jamaluddin, 2019). Modifikasi *Fuzzy C-Means (FCM)* dengan menambahkan sebuah parameter guna meningkatkan kecepatan konvergensi dan mendapatkan hasil yang lebih sedikit kesamaran dengan menyesuaikan matriks keanggotaan pada setiap iterasi disebut dengan *generalized Fuzzy C-Means (gFCM)*.

Kanzawa et al., (2021) melakukan penelitian mengenai *clustering* menggunakan *generalized Fuzzy c-Means (gFCM)* yang mencakup algoritma *standard clustering* dan *exponential fuzzy c-means clustering*. Hasil kami penelitiannya menunjukkan bahwa klasifikasi *fuzzy* untuk *gFCM* menunjukkan perilaku yang mirip dengan dari *standard fuzzy c-means clustering* dan *exponential fuzzy c-means clustering*. (Suciati et al., 2021) menggunakan Metode *Fuzzy C-Means* pada Data COVID-19 di Provinsi Lampung dengan salah satu hasilnya adalah 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. Selain itu, (Rouza & Luth, 2020) mengimplementasikan *Fuzzy C-Means* dalam pengelompokan UKM di Kabupaten Rokan Hulu dengan salah satu hasilnya menyatakan bahwa *Fuzzy C-Means Clustering* memiliki tingkat akurasi yang tinggi sebesar 80-90 %.

Data Produk Domestik Regional Bruto (PDRB) merupakan statistik pendapatan regional yang dapat dimanfaatkan untuk mengevaluasi hasil pembangunan sekaligus sebagai dasar perencanaan pembangunan ekonomi di suatu wilayah tertentu (BPS, 2023). Penelitian dilakukan oleh Satria & Habibi, (2023) tentang pengaruh sektor Produk Domestik Regional Bruto (Basis dan Non Basis) terhadap pertumbuhan ekonomi dalam perspektif ekonomi Islam. Dari hasil penelitiannya, salah satu hasilnya adalah variabel sektor basis dan sektor non basis secara simultan berpengaruh positif dan signifikan terhadap pertumbuhan ekonomi di Kabupaten Pesisir Barat. Analisis mengenai PDRB dilakukan juga oleh Ahrizal, (2022) pada 34 provinsi di Indonesia pada masa pemerintahan Presiden Joko Widodo. Metode analisis yang digunakan adalah *Ordinary Least Square (OLS)*. Hasil penelitiannya menunjukkan bahwa pendidikan dan tenaga kerja berpengaruh signifikan terhadap PDRB 34 provinsi di Indonesia. Selain itu, Jaelanie, (2016) melakukan analisis kluster kabupaten/kota berdasarkan indikator Produk Domestik Regional Bruto (PDRB) menggunakan metode *Unsupervised Learning*. Penelitiannya menghasilkan 7 kluster sebagai yang terbaik menggunakan metode *Self-Organizing Maps (SOM)* dengan nilai validasi sebesar 0.9704.

Dari latar belakang tersebut, peneliti ingin mengimplementasikan *Fuzzy C-Means (FCM)* dan *Generalized Fuzzy C-Means (GFCM)* dalam *clustering* Produk Domestik Regional Bruto (PDRB) menurut pengeluaran.

## LANDASAN TEORI

Theodoridis & Koutroumbas, (2009) mendefinisikan *clustering* dengan definisi berikut. Diberikan suatu himpunan data  $X = \{x_1, x_2, \dots, x_N\}$ , dan didefinisikan sebanyak  $m$  *-clustering* dari  $X$ , dan  $\mathfrak{R}$  merupakan partisi  $X$  menjadi  $m$  himpunan (*cluster*), sehingga  $C_1, C_2, \dots, C_m$  memenuhi 3 kondisi berikut.

1.  $C_i \neq \emptyset, i = 1, 2, \dots, m$
2.  $\bigcup_{i=1}^m C_i = X$
3.  $C_i \cap C_j \neq \emptyset, i \neq j, i, j = 1, 2, \dots, m$

Istilah *fuzzy* digunakan untuk menyatakan *cluster* himpunan yang dapat dibedakan dengan *cluster* lain berdasarkan derajat keanggotaan secara samar atau tidak tegas. Dalam *fuzzy clustering*, setiap objek dapat dimiliki oleh beberapa klaster dengan peluang tertentu (Safitri et al., 2017). Sebuah *fuzzy clustering* dari himpunan data  $X = \{x_1, x_2, \dots, x_N\}$  ke  $m$  *cluster* disifatkan sebagai fungsi  $\mu_j$  sebanyak  $m$  dengan

$$\mu_j: X \rightarrow [0,1], j = 1, 2, \dots, m \quad (2.1)$$

dan

$$\sum_{j=1}^m \mu_j(x_i) = 1, i = 1, 2, \dots, N \quad 0 < \sum_{j=1}^m \mu_j(x_i) < N, j = 1, 2, \dots, m \quad (2.2)$$

Hal ini  $\mu_j$  disebut fungsi keanggotaan (Theodoridis & Koutroumbas, 2009).

Algoritma *Fuzzy C-Means* (sering disingkat FCM) adalah algoritma iteratif yang menemukan kelompok dalam data dan menggunakan konsep keanggotaan *fuzzy*. *Fuzzy C-Means* mencoba menemukan *cluster* dalam data  $X = \{x_1, x_2, \dots, x_N\}$  dengan meminimalkan  $(\mu_{ij}, v_j)$  dari fungsi objektif berikut (Mahdi & Mahmood, 2014; Memon & Lee, 2017; Yang, 1993).

$$J_c(G_N, \mu, v) = \frac{1}{N} \sum_{j=1}^m \sum_{i=1}^N \mu_{ij}^c \|x_i - v_j\|^2 \quad (2.3)$$

dengan  $m$  merupakan banyaknya *cluster*;  $c$  dinotasikan sebagai parameter *fuzzy* yang bernilai pada  $1 < c < \infty$ ;  $N$  merupakan banyaknya pengamatan;  $\|x_i - v_j\|^2$  merupakan jarak *Euclidean* antara  $x_i$  dan  $v_j$ ;  $x_i$  merupakan nilai pengamatan ke- $i$ ;  $v_j$  merupakan pusat dari *cluster* ke- $j$ ;  $\mu_{ij}$  merupakan nilai derajat keanggotaan *fuzzy* dari titik  $x_i$  pada *cluster* ke- $j$ . Sehingga diperoleh dua persamaan berikut.

$$\mu_{ik} = \left( \frac{\left( \|x_i - v_k\|^{\frac{2}{(m-1)}} \right)}{\sum_{j=1}^m \left( \|x_i - v_j\|^{\frac{2}{(m-1)}} \right)} \right)^{-1}, k = 1, 2, \dots, m \quad (2.4)$$

dan

$$v_j = \frac{\sum_{i=1}^N u_{ij}^c \cdot x_i}{\sum_{i=1}^N u_{ij}^c} \quad (2.5)$$

*Generalized Fuzzy C-Means* (gFCM) merupakan modifikasi dari *Fuzzy C-Means* (FCM) dengan menambahkan parameter dalam mempercepat konvergensi dan mendapatkan hasil yang tidak terlalu kabur dengan menyesuaikan matriks keanggotaan pada setiap iterasi. Parameter  $\beta$  tambahan dibutuhkan untuk mengontrol efektivitas modifikasi. (Wu, 2022) mengatakan bahwa modifikasi hanya mempengaruhi persamaan yang memperbarui keanggotaan matriks.

$$\mu_{ik} = \frac{(\|x_i - v_k\|^2 - \beta_i)^{\left(\frac{-1}{m-1}\right)}}{\sum_{j=1}^m (\|x_i - v_j\|^2 - \beta_i)^{\left(\frac{-1}{m-1}\right)}}, k = 1, 2, \dots, m \quad (2.6)$$

dengan  $\beta_i = \min(\|x_i - v_k\|^2)$  dan  $0 \leq \beta_i \leq 1$ .

Setiap hasil *clustering* perlu dilakukan validasi. (Mahdi & Mahmood, 2014) menyatakan perlu melakukan 2 validasi, yakni yang pertama validasi hasil *clustering* berdasarkan partisi dan entropi partisi, yang kedua adalah validasi hasil *clustering* berdasarkan struktur sampel geometris. Partisi *fuzzy* digunakan untuk mengevaluasi kinerja *clustering* secara kuantitatif. Fungsi validitas berdasarkan partisi *fuzzy* merupakan koefisien partisi  $V_{pc}$  dan koefisien entropi partisi  $V_{pe}$  yang didefinisikan pada persamaan berikut:

$$V_{pc} = \frac{\sum_{j=1}^m \sum_{i=1}^N \mu_{ij}^2}{N} \quad (2.7)$$

dan

$$V_{pe} = \frac{-\sum_{j=1}^m \sum_{i=1}^N (\mu_{ij} \times \log \mu_{ij})}{N} \quad (2.8)$$

Sebagai hasil validasi, *clustering* dikatakan baik jika  $\frac{1}{m} \leq V_{pc} \leq 1$  dan  $\frac{1}{m} \leq V_{pe} \leq 2$ . Dan *clustering* dikatakan menjadi *clustering* terbaik jika nilai  $V_{pc}$  maksimum, atau  $V_{pe}$  minimum (Mahdi & Mahmood, 2014).

Validasi yang kedua adalah validasi hasil *clustering* berdasarkan struktur sampel geometris. Fungsi validitas yang didasarkan pada pengukuran struktur data geometris adalah sampel dalam satu partisi harus kompak dan sampel antara *cluster* yang berbeda harus terpisah. Untuk mengukur rasio variasi total dalam *cluster* dan pemisahan *cluster*, Xie dan Beni dalam (Mahdi & Mahmood, 2014) mengusulkan fungsi validitas Xie-Benie (XB) dan didefinisikan sebagai berikut.

$$V_{xb} = \frac{-\sum_{j=1}^m \sum_{i=1}^N \mu_{ij}^2 \|x_i - v_j\|^2}{N \cdot \left( \min_{j \neq k} (\|v_k - v_j\|^2) \right)} \quad (2.9)$$

Hasil *clustering* yang optimal menghasilkan sampel yang berada dalam satu *cluster* dan sampel yang terpisah di antara *cluster* yang berbeda. Peminimalan nilai  $V_{pe}$  diharapkan dapat menghasilkan partisi yang baik. Selain itu, validasi kedua dapat menggunakan validasi *Silhouette Coefficient* (SC). SC digunakan untuk mengevaluasi apakah titik-titik dikelompokkan dengan baik dan dipisahkan dengan baik yang dapat dituliskan dalam persamaan berikut.

$$V_{sc} = \frac{b - \bar{v}}{\max(b, \bar{v})} \quad (2.10)$$

Di sini,  $b$  mewakili rata-rata jarak kluster terdekat untuk setiap sampel dan  $\bar{v}$  merupakan rata-rata jarak pusat kluster. Jika dua *cluster* saling berdekatan, nilai  $V_{sc}$  akan mendekati 0. Jika tumpang tindih, nilainya akan mendekati  $-1$  (Sinnott et al., 2016). Validasi kedua juga dapat menggunakan validasi *Explained Inertia* (EI). EI juga disebut dengan *the total within-cluster sum of square* (WCSS) (Wang et al., 2022). WCSS digunakan untuk melihat kebaikan dalam *clustering* yang dengan meminimalkan nilainya. (Cooper, 2018) menuliskan WCSS dalam persamaan berikut.

$$\text{WCSS/EI} = \min \sum_{j=1}^m \sum_{i=1}^N \|x_i - v_j\|^2 \quad (2.11)$$

## METODE PENELITIAN

Pada metode penelitian ini akan dijelaskan secara rinci jenis data, sumber data, dan prosedur analisis data.

### Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder yang terdiri dari 8 jenis pengeluaran dari persentase komponen pengeluaran PDRB Provinsi terhadap total 34 Provinsi atas dasar harga konstan 2010 dari tahun 2018 hingga tahun 2022. Data diperoleh dari (BPS, 2023) yang secara detail keterangan komponen pengeluaran PDRB pada Tabel 1.

**Tabel 1. Komponen Pengeluaran PDRB**

Pengeluaran ( $P_i$ )	Keterangan
$P_1$	Pengeluaran konsumsi rumah tangga
$P_2$	Pengeluaran konsumsi lembaga swasta nirlaba yang melayani rumah tangga
$P_3$	Pengeluaran konsumsi pemerintah
$P_4$	Pembentukan modal tetap bruto
$P_5$	Perubahan inventori
$P_6$	Ekspor barang dan jasa
$P_7$	Impor barang dan jasa
$P_8$	Total

Sumber: BPS 2023

### Prosedur Analisis Data

Berikut prosedur yang perlu dilakukan dalam *clustering* menggunakan FCM dan GFCM.

1. Menentukan jumlah *cluster* menggunakan 5 validasi pada FCM sehingga memenuhi 3 kondisi dalam *clustering*.

2. Menghitung faktor keanggotaan dalam FCM menggunakan Persamaan (2.4) dan GFCM menggunakan persamaan (2.6).
3. Menentukan pusat *cluster* menggunakan Persamaan (2.5)
4. Melakukan validasi pertama menggunakan menggunakan Persamaan (2.7) dan (2.8).
5. Melakukan validasi kedua menggunakan menggunakan Persamaan (2.9) sampai dengan Persamaan (2.11).
6. Pada penelitian ini, analisis menggunakan software R, *package* “geocmeans”, fungsi “CMeans” dan “GCMeans” pada Gelb, (2023) untuk analisis FCM dan GFCM; *package* “fclust”, fungsi “calcqualityIndexes” pada Giordani et al., (2022) untuk menghitung validasi hasil *clustering*.
7. Interpretasi hasil dan pembahasan.

## HASIL DAN PEMBAHASAN

Untuk mengeksplorasi dataset dan memilih jumlah kluster/kelompok ( $m$ ) yang tepat pada penelitian ini digunakan pemilihan jumlah kluster terbaik dari *clustering* menggunakan FCM. Setelah terpilih, jumlah kluster tersebut akan digunakan untuk analisis selanjutnya, yaitu *clustering* menggunakan GFCM. Pemilihan jumlah kluster yang akan dibentuk pada penelitian ini mengimplementasikan 5 jenis validasi seperti yang dituliskan pada Tabel 2.

**Tabel 2. Pemilihan Jumlah Kluster Terbaik yang Dapat Dibentuk dengan FCM**

Validasi	Cluster_2	Cluster_3	Cluster_4	Cluster_5	Cluster_6	Cluster_7
Partition entropy	0,03577582	0,01745691	0,1261155	0,1261155	0,2865278	0,3885229
Partition coeff	0,9790735	0,9931562	0,9322343	0,9322343	0,8463765	0,8071264
Silhouette index	0,8830504	0,9175638	0,7291939	0,7291939	0,5368794	0,3570123
XieBeni index	0,09123162	0,02969096	0,4147333	0,4147333	2,004608	2,650658
Explained inertia	0,4652929	0,8175191	0,9130635	0,9130635	0,9148182	0,8468207

Sumber: Pemilihan Jumlah Kluster Terbaik dengan FCM menggunakan Software R

: nilai terbaik menurut masing-masing jenis validasi.

Pada Tabel 2, menurut nilai terbaik pada masing-masing validasi menggunakan FCM, terdapat 3 jenis validasi yang menunjukkan bahwa 3 kluster cukup mewakili keragaman data. Selanjutnya, dengan menentukan maksimal jumlah iterasi dan pada penelitian ini ditentukan maksimal sebanyak 500 iterasi. Dari hasil *clustering* menggunakan FCM diperoleh pusat kluster ( $v_j$ ) yang ditunjukkan pada Tabel 3.

**Tabel 3. Pusat Kluster ( $v_j$ )**

Pengeluaran( $P_i$ )	Klaster 1	Klaster 2	Klaster 3
$P_1$	11,70273	17,97526	1,181587
$P_2$	7,951743	29,71059	1,345626

Pengeluaran( $P_i$ )	Klaster 1	Klaster 2	Klaster 3
$P_3$	6,928041	26,13451	1,702224
$P_4$	9,111981	20,80792	1,444678
$P_5$	29,32162	-19,9749	-0,02172
$P_6$	9,083085	16,98454	1,62211
$P_7$	8,63263	23,56751	1,464987
$P_8$	10,58599	16,43461	1,401573

Sumber: Pusat Klaster menggunakan Software R


Pusat klaster ( $v_j$ ) pada FCM tidak mengalami perbedaan pada GFCM. Titik perbedaan FCM dan GFCM hanya memberikan pengaruh pada nilai faktor keanggotaan ( $\mu_{ik}$ ). Setelah diperoleh pusat klaster, tahap selanjutnya adalah menentukan  $\beta$  terbaik menggunakan 3 kriteria validasi yakni dari *Silhouette index*, *XieBeni index*, dan *Explained inertia*. Secara detail nilai  $\beta$  dari ketiga validasi dapat dilihat pada Tabel 4.

**Tabel 4. Pemilihan  $\beta$  terbaik**

Silhouette Index	XieBeni Index	Explained Inertia	beta
0,911052	0,031268	0,833381	0
0,910742	0,031039	0,834637	0,05
0,910417	0,030813	0,835868	0,1
0,910076	0,030589	0,837075	0,15
0,909718	0,030368	0,838257	0,2
0,909344	0,03015	0,839414	0,25
0,908952	0,029935	0,840546	0,3
0,908545	0,029725	0,841654	0,35
0,908121	0,02952	0,842735	0,4
0,907684	0,029321	0,843789	0,45
0,907234	0,02913	0,844814	0,5
0,906776	0,028947	0,845807	0,55
0,906312	0,028776	0,846764	0,6
0,905848	0,028618	0,847679	0,65
0,905391	0,028476	0,848544	0,7
0,904949	0,028352	0,849348	0,75
0,729214	0,276968	0,733176	0,8
0,723786	0,287415	0,735822	0,85
0,720372	0,294009	0,7378	0,9
0,718264	0,298239	0,739085	0,95

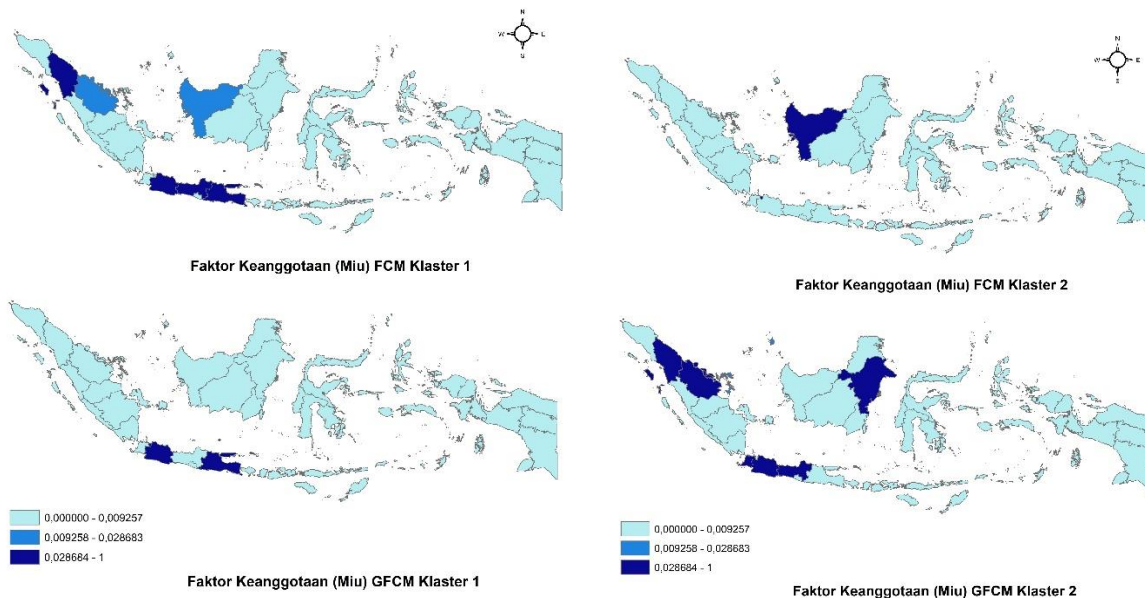
Silhouette Index	XieBeni Index	Explained Inertia	beta
0,6614	0,690887	0,715154	1

Sumber: Pemilihan  $\beta$  terbaik menggunakan Software R

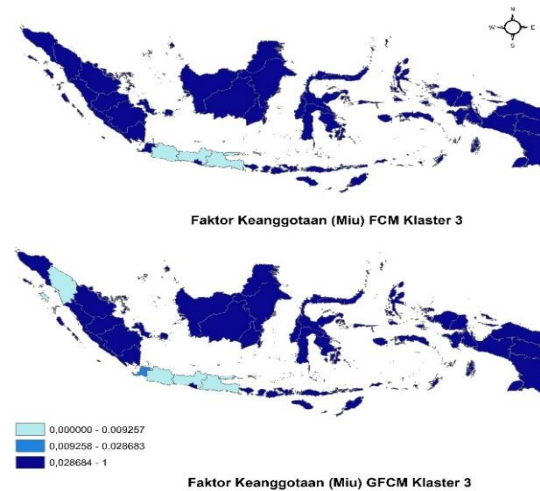
 : nilai terbaik menurut rata-rata semua jenis validasi.

Dari Tabel 4, kita dapat melihat berdasarkan rata-rata terkecil dari *Silhouette index*, *XieBeni index*, dan *Explained inertia*. Dan mengambil kesimpulan bahwa nilai  $\beta$  terbaik sebesar 0,8. Nilai  $\beta$  ini yang digunakan sebagai parameter tambahan dalam *clustering* menggunakan GFCM.

Setelah kita mendapatkan nilai  $\beta$ , bisa digunakan untuk *clustering* menggunakan GFCM. Dari kedua *clustering* kita mendapatkan nilai faktor keanggotaan ( $\mu_{ik}$ ) untuk setiap provinsi. Jika dipetakan, nilai faktor keanggotaan ( $\mu_{ik}$ ) ketiga klaster untuk *clustering* menggunakan FCM dan GFCM ditunjukkan pada Gambar 1.







**Gambar 1. Probabilitas faktor keanggotaan ( $\mu_{ik}$ ) setiap klaster pada setiap provinsi menggunakan FCM dan GFCM**

Berdasarkan Gambar 1, kita dapat memperoleh informasi bahwa untuk *clustering* menggunakan FCM, probabilitas provinsi di Pulau Jawa paling besar berada pada klaster 1, **kecuali** Provinsi DKI Jakarta, Banten, dan DI Yogyakarta. Provinsi Sumatera Utara mendominasi keanggotaannya pada klaster 1 dan 3. Untuk klaster 2 pada FCM didominasi oleh Provinsi Kalimantan Barat dan DKI Jakarta. Selain provinsi yang telah disebutkan, memiliki probabilitas terbesar masuk keanggotaan ke dalam klaster 3. Untuk *clustering* menggunakan GFCM, klaster 1 didominasi keanggotaannya oleh Provinsi DKI Jakarta, Jawa Barat, dan Jawa Timur. Klaster 2 didominasi keanggotaannya oleh Provinsi Kalimantan Timur, Sumatera Utara, Riau, Banten, Jawa Barat, dan Jawa Tengah. Klaster 3 didominasi seluruh provinsi di Indonesia **kecuali** Provinsi Sumatera Utara, Banten, DKI Jakarta, Jawa Barat, Jawa Tengah, dan Jawa Timur.

Selanjutnya, kita dapat mengetahui keanggotaan akhir setiap provinsi di Indonesia pada ketiga klaster menggunakan FCM dan GFCM yang dipetakan secara detail pada Gambar 2 dan Gambar 3.



**Gambar 2. Hasil *Clustering* Menggunakan FCM**

Secara deskriptif, dari Gambar 2 dan Lampiran A1 hingga A3 diperoleh kesimpulan dari hasil *clustering* menggunakan FCM bahwa pada

- Klaster 1 jika dilihat nilai mean, maka dapat disimpulkan bahwa memiliki keanggotaan paling tinggi **kecuali** untuk jenis pengeluaran PDRB ( $P_5$ ) yaitu perubahan inventori. Selain itu, dari klaster 1 mencakup Provinsi DKI Jakarta.
- Klaster 2, memiliki keanggotaan paling tinggi untuk jenis pengeluaran PDRB ( $P_5$ ) yaitu perubahan inventori dan mencakup Provinsi Jawa Barat, Jawa Tengah, dan Jawa Timur.
- Klaster 3, memiliki keanggotaan hanya sebagian kecil saja untuk setiap jenis pengeluaran PDRB dan mencakup selain seluruh provinsi yang telah disebutkan.

Kemudian kita juga mendapatkan informasi secara deskriptif mengenai *clustering* menggunakan GFCM.



**Gambar 3. Hasil Clustering Menggunakan GFCM**

Secara deskriptif, dari Gambar 3 dan Lampiran B1 hingga B3 diperoleh kesimpulan dari hasil *clustering* menggunakan GFCM bahwa pada

- Klaster 1 jika dilihat nilai mean, maka dapat disimpulkan seperti halnya pada *clustering* menggunakan FCM yakni memiliki keanggotaan paling tinggi **kecuali** untuk jenis pengeluaran PDRB ( $P_5$ ) yaitu perubahan inventori. Selain itu, dari klaster 1 mencakup Provinsi DKI Jakarta, Jawa Barat, dan Jawa Timur.
- Klaster 2, memiliki keanggotaan paling tinggi untuk jenis pengeluaran PDRB ( $P_5$ ) yaitu perubahan inventori dan mencakup Provinsi Sumatera Utara, Riau, Kalimantan Timur, Banten, dan Jawa Tengah.
- Klaster 3, memiliki keanggotaan hanya sebagian kecil saja untuk setiap jenis pengeluaran PDRB dan mencakup selain seluruh provinsi yang telah disebutkan.

Setelah mengetahui hasil *clustering* untuk FCM maupun GFCM, kita perlu melihat validasi dari masing-masing metode analisis *clustering* menggunakan 5 jenis validasi yang dapat dilihat secara detail pada Tabel 5.

**Tabel 5. Validasi Hasil Clustering 3 Klaster**

Validasi	GFCM	FCM
<i>Partition.entropy</i>	0,021	0,017
<i>Partition.coeff</i>	0,992	0,993
<i>Silhouette.index</i>	0,317	0,918
<i>XieBeni.index</i>	3,074	0,03
<i>Explained.inertia</i>	0,539	0,818

Sumber: Validasi hasil *clustering* menggunakan Software R

: nilai terbaik menurut masing-masing jenis validasi.

Berdasarkan Tabel 5, kita dapat melihat hasil validasi *clustering* menggunakan FCM dan GFCM yang memberikan kesimpulan bahwa FCM lebih tidak terlalu kabur(*fuzzy*) karena nilai *Partition.entropy* dan *XieBeni.index* pada FCM lebih kecil daripada GFCM, dan nilai *Partition.coeff* pada FCM lebih besar daripada GFCM.

## KESIMPULAN DAN SARAN

Berdasarkan hasil *clustering*, jumlah klaster yang sudah mencakup keragaman data adalah sebanyak 3 klaster. Hasil *clustering* menggunakan FCM dan GFCM dapat memberikan kesimpulan bahwa FCM lebih tidak terlalu kabur(*fuzzy*). Informasi ini diperoleh dari nilai *Partition.entropy* dan *XieBeni.index* pada FCM lebih kecil daripada GFCM, dan nilai *Partition.coeff* pada FCM lebih besar daripada GFCM.

Untuk hasil pembentukan *clustering* menggunakan FCM maupun GFCM berdasarkan sudut pandang jenis pengeluaran PDRB memberikan informasi bahwa klaster 1 didominasi 7 dari 8 jenis pengeluaran PDRB, klaster 2 didominasi oleh perubahan inventori, dan klaster 3 tidak ada satupun jenis pengeluaran PDRB yang mendominasi.

Dari sudut pandang provinsi, *clustering* menggunakan FCM menghasilkan klaster 1 yang beranggotakan Provinsi DKI Jakarta; klaster 2 beranggotakan Provinsi Jawa Barat, Jawa Tengah, dan Jawa Timur; dan selainnya masuk keanggotaan pada klaster 3. Dan untuk *clustering* menggunakan GFCM menghasilkan klaster 1 yang beranggotakan Provinsi DKI Jakarta, Jawa Barat, dan Jawa Timur; klaster 2 beranggotakan Provinsi Sumatera Utara, Riau, Kalimantan Timur, Banten, dan Jawa Tengah; dan selainnya masuk keanggotaan pada klaster 3.

## DAFTAR REFERENSI

- Ahrizal, G. R. (2022). Analisis PDRB 34 Provinsi di Indonesia Pada Masa Pemerintahan Presiden Joko Widodo. *Jurnal Ilmu Ekonomi JIE*, 6(3), 396–408. <https://doi.org/10.22219/jie.v6i3.21653>
- BPS. (2023). *Produk Domestik Regional Bruto Provinsi-Provinsi Di Indonesia Menurut Pengeluaran 2018-2022*. Badan Pusat Statistik RI.
- Cooper, C. (2018). *Clustering Algorithms* [Lecturing Slide]. <https://nms.kcl.ac.uk/colin.cooper/teachingmaterial/CSMWAL/CSMWAL/Lectures/ClusterSlides.pdf>
- Gelb, J. (2023). Implementing Methods for Spatial Fuzzy Unsupervised Classification. *CRAN R*, 1–57.
- Giordani, P., Ferraro, M. B., & Serafini, A. (2022). Fuzzy Clustering. *CRAN R*, 1–81.
- Jaelanie, A. (2016). *Analisis Klaster Kabupaten/Kota Berdasarkan Indikator Produk Domestik Regional Bruto (PDRB) Menggunakan Metode Unsupervised Learning* [Tesis]. Institut Teknologi Sepuluh November.
- Kanzawa, Y., Miyamoto, S., Shibaura Institute of Technology 3-7-5 Toyosu, Koto, Tokyo 135-

- 8548, Japan, & University of Tsukuba 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan. (2021). Generalized Fuzzy c-Means Clustering and its Property of Fuzzy Classification Function. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 25(1), 73–82. <https://doi.org/10.20965/jaciii.2021.p0073>
- Mahdi, S. S., & Mahmood, R. S. (2014). MR Brain Image Segmentation Using Spatial Fuzzy C-Means Clustering Algorithm. *Journal of Engineering*, 20(09), 78–89. <https://doi.org/10.31026/j.eng.2014.09.06>
- Memon, K. H., & Lee, D. (2017). Generalised fuzzy c-means clustering algorithm with local information. *IET Image Processing*, 11(1), 1–12. <https://doi.org/10.1049/iet-ipr.2016.0282>
- Rajkumar, K. V., Yesubabu, A., & Subrahmanyam, K. (2019). Fuzzy clustering and fuzzy c-means partition cluster analysis and validation studies on a subset of citescore dataset. *International Journal of Electrical and Computer Engineering (IJECE)*, 9(4), 2760. <https://doi.org/10.11591/ijece.v9i4.pp2760-2770>
- Rouza, E., & Luth, F. (2020). Implementasi Fuzzy C-Means Clustering dalam Pengelompokan UKM Di Kabupaten Rokan Hulu. *Techno.COM*, 19(4), 481–495.
- Safitri, Q. U., Huda, A. F., & Awaludin, A. S. (2017). Segmentasi Citra Menggunakan Algoritma Fuzzy c-Means (FCM) Dan Spatial Fuzzy c-Means (sFCM). *Kubik: Jurnal Publikasi Ilmiah Matematika*, 2(1), 22–34. <https://doi.org/10.15575/kubik.v2i1.1471>
- Satria, A. D., & Habibi, A. (2023). Pengaruh Sektor Produk Domestik Regional Bruto (Basis dan Non Basis) Terhadap Pertumbuhan Ekonomi dalam Perspektif Ekonomi Islam. *Jurnal Ilmiah Ekonomi Islam*, 9(1), 1213–1226.
- Sinnott, R. O., Duan, H., & Sun, Y. (2016). Chapter 15—A Case Study in Big Data Analytics: Exploring Twitter Sentiment Analysis and the Weather. In *Big Data Principles and Paradigms*. Elsevier Inc.
- Siringoringo, R., & Jamaluddin. (2019). Peningkatan Performa Cluster Fuzzy c-Means Pada Pengklasteran Sentimen Menggunakan Particle Swarm Optimization. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 6(4), 349–354.
- Suciati, L., Herawati, N., Subian, S., & Widiarti. (2021). Analisis Klaster Menggunakan Metode Fuzzy C-Means pada Data COVID-19 di Provinsi Lampung. *Prosiding SN-SMIAP-VI FMIPA Unila*.
- Tan, P.-N., Steinbach, M., Karpatne, A., & Kumar, V. (2019). Cluster Analysis: Basic Concepts and Algorithms. In *Introduction to Data Mining* (2nd ed.). Pearson.
- Theodoridis, S., & Koutroumbas, K. (2009). *Pattern recognition* (4th ed). Academic Press.
- Wang, K., Yang, R., Liu, C., Samarasinghalage, T., & Zang, Y. (2022). Extracting Electricity Patterns from High-dimensional Data: A comparison of K-Means and DBSCAN algorithms. *IOP Conference Series: Earth and Environmental Science*, 1101(2), 022007. <https://doi.org/10.1088/1755-1315/1101/2/022007>
- Yang, M. S. (1993). Convergence Properties of The Generalized Fuzzy C-Means Clustering Algorithms. *Computers Math. Applic.*, 25(12), 3–10.