

COMPARISON OF GAUSSIAN MIXTURE MODEL AND FUZZY C-MEANS IN EARTHQUAKE CLUSTERING IN INDONESIA

PERBANDINGAN *GAUSSIAN MIXTURE MODEL* DAN *FUZZY C-MEANS* DALAM KLASTERISASI GEMPA BUMI DI INDONESIA

Indah Dzulkharifah¹, Helmi Falah^{2†}, M Taqy Abiyu Dzakwan³, Cameliya Ulya Hidayah⁴, Nyoman Kiranasassi Anara⁵ and Cici Suhaeni⁶

^{1,2,3,4,5}Statistics and Data Science Study Program, IPB University, Indonesia

[†]corresponding author: helmi.ipbfalah@apps.ipb.ac.id

Copyright © 2026 Indah Dzulkharifah, Helmi Falah, M Taqy Abiyu Dzakwan, Cameliya Ulya Hidayah, Nyoman Kiranasassi Anara and Cici Suhaeni. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Indonesia is a region with high seismic activity, making the identification of earthquake patterns essential for disaster mitigation. Although various clustering methods have been applied to earthquake data, studies comparing probabilistic and fuzzy soft clustering approaches remain limited. This study aims to compare the performance of Fuzzy C-Means (FCM) and Gaussian Mixture Model (GMM) in clustering 17,181 BMKG earthquake records from January 2025 to February 2026. The variables analyzed include latitude, longitude, depth, magnitude, phase count, and azimuth gap. The optimal number of clusters was determined using the internal evaluation metrics of each algorithm, resulting in two clusters for FCM and four clusters for GMM. Subsequently, both models were evaluated using the Silhouette Score, Calinski-Harabasz Index (CHI), and Davies-Bouldin Index (DBI), while cluster similarity was assessed using the Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI). The results show that FCM achieved slightly higher Silhouette Score and CHI values, whereas GMM obtained a lower DBI, indicating better cluster separation in overlapping data. Furthermore, GMM produced a more detailed representation of spatial and seismic variations, including distinctions in earthquake depth characteristics, whereas FCM tended to form broader macro-scale clusters. Therefore, GMM provides a more representative framework for supporting earthquake risk assessment and disaster mitigation strategies in Indonesia.

Keywords: earthquake, Fuzzy C-Means, Gaussian Mixture Model

Abstrak

Indonesia merupakan wilayah dengan aktivitas seismik tinggi sehingga diperlukan pemahaman pola kejadian gempa bumi untuk mendukung mitigasi bencana. Berbagai penelitian telah menerapkan metode klusterisasi pada data gempa, namun kajian yang secara khusus membandingkan pendekatan soft clustering berbasis probabilistik dan fuzzy pada data gempa bumi Indonesia masih terbatas. Penelitian ini bertujuan membandingkan metode Fuzzy C-Means (FCM) dan Gaussian Mixture Model (GMM) dalam klusterisasi 17181 data gempa bumi BMKG periode Januari 2025-Februari 2026. Variabel yang dianalisis meliputi lintang, bujur, kedalaman, magnitudo, phase count, dan azimuth gap. Penentuan jumlah kluster optimal dilakukan berdasarkan metrik evaluasi internal masing-masing algoritma yang menghasilkan dua kluster untuk FCM dan empat kluster untuk GMM. Selanjutnya, kedua model dievaluasi menggunakan Silhouette Score, Calinski-Harabasz Index (CHI), dan Davies-Bouldin Index (DBI), sedangkan tingkat kesamaan kluster diukur menggunakan Adjusted Rand Index (ARI) dan Normalized Mutual Information (NMI). Hasil menunjukkan FCM memiliki Silhouette Score dan CHI sedikit lebih tinggi, namun GMM unggul pada DBI yang mengindikasikan separasi antarkluster lebih baik pada data overlap. Selain itu, GMM menghasilkan struktur kluster yang lebih rinci dalam merepresentasikan variasi spasial dan seismik, termasuk perbedaan karakteristik kedalaman gempa. Dengan demikian, GMM dinilai lebih representatif untuk mendukung analisis risiko dan mitigasi bencana berbasis karakteristik kegempaan di Indonesia.

Kata Kunci: *Fuzzy C-Means, Gaussian Mixture Model, gempa bumi*

1. Pendahuluan

Indonesia merupakan salah satu wilayah dengan aktivitas seismik tertinggi di dunia karena posisinya pada *Ring of Fire*. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) mencatat lebih dari 12.000 aktivitas seismik sepanjang tahun 2024. Mahanani *et al.*, (2024) melaporkan bahwa gempa bermagnitudo 5,0 di Kabupaten Bandung dan 4,6 di Daruba pada 18 September 2024 menyebabkan korban jiwa serta kerusakan pada rumah, fasilitas umum, dan sarana usaha. Tingginya frekuensi dan dampak kejadian tersebut menuntut upaya mitigasi berbasis pemahaman karakteristik seismik yang lebih komprehensif guna mendukung pengambilan keputusan yang efektif.

Salah satu pendekatan yang dapat digunakan adalah analisis kluster, yaitu metode *data mining* yang melakukan klusterisasi objek berdasarkan

kemiripan karakteristik sehingga mampu mengungkap pola yang sulit diamati secara langsung (Prasetio *et al.*, 2023). Pada data gempa bumi, proses klasterisasi memerlukan analisis multidimensi yang mengintegrasikan variabel spasial (lintang, bujur, dan kedalaman), variabel seismik (magnitudo), serta metrik kualitas jaringan observasi seperti *phase count* dan *azimuth gap*. Willyana *et al.*, (2025) menunjukkan bahwa integrasi informasi spasial dan seismik mampu menghasilkan identifikasi wilayah rawan gempa yang lebih representatif.

Karakteristik data gempa yang heterogen serta berpotensi memiliki *overlap* antarklaster menyebabkan batas klaster tidak selalu terbentuk secara tegas. Kondisi ini menjadikan pendekatan *soft clustering* lebih relevan karena mampu mengakomodasi ketidakpastian keanggotaan suatu observasi pada lebih dari satu klaster. Senkaya *et al.*, (2024) menunjukkan bahwa pendekatan *fuzzy clustering* efektif digunakan pada data seismik yang kompleks karena memungkinkan representasi keanggotaan yang lebih fleksibel dibandingkan pendekatan *hard clustering*.

Berbagai penelitian telah menerapkan klasterisasi data gempa bumi menggunakan pendekatan yang beragam. Ahmad dan Irsalinda (2020) menerapkan *Fuzzy K-Means Clustering* untuk melakukan klasterisasi gempa di Indonesia, sedangkan Rais *et al.*, (2025) menggunakan Bisecting K-Means untuk mengidentifikasi klaster kejadian gempa berdasarkan atribut spasial. Dwitianti *et al.*, (2024) membandingkan beberapa metode klasterisasi, yaitu K-Means, K-Medoids, DBSCAN, FCM, dan *K-Affinity Propagation* pada data gempa Indonesia. Di sisi lain, Huang *et al.*, (2025) menunjukkan bahwa *Gaussian Mixture Model* (GMM) mampu menghasilkan akurasi klasterisasi sinyal seismik yang tinggi, sementara Zhao *et al.*, (2024) menunjukkan efektivitas FCM dalam menangani data mikroseismik yang kompleks.

Penelitian terdahulu umumnya berfokus pada penerapan satu metode analisis klaster atau perbandingan metode secara umum. Kajian yang secara khusus membandingkan dua paradigma *soft clustering*, yaitu pendekatan probabilistik melalui *Gaussian Mixture Model* (GMM) dan pendekatan *fuzzy membership* melalui *Fuzzy C-Means* (FCM), pada data gempa bumi Indonesia masih terbatas. Selain itu, integrasi variabel spasial, seismik, dan metrik kualitas jaringan observasi dalam evaluasi kedua metode tersebut juga belum banyak dibahas. Oleh karena itu, penelitian ini bertujuan menerapkan serta membandingkan performa GMM dan FCM dalam melakukan klasterisasi data gempa bumi di Indonesia berdasarkan integrasi variabel spasial, seismik, dan kualitas jaringan observasi. Penelitian ini diharapkan dapat memberikan gambaran mengenai karakteristik klaster gempa bumi di Indonesia serta mengevaluasi efektivitas kedua metode dalam merepresentasikan pola aktivitas seismik.

2. Metodologi

2.1 Bahan dan Data

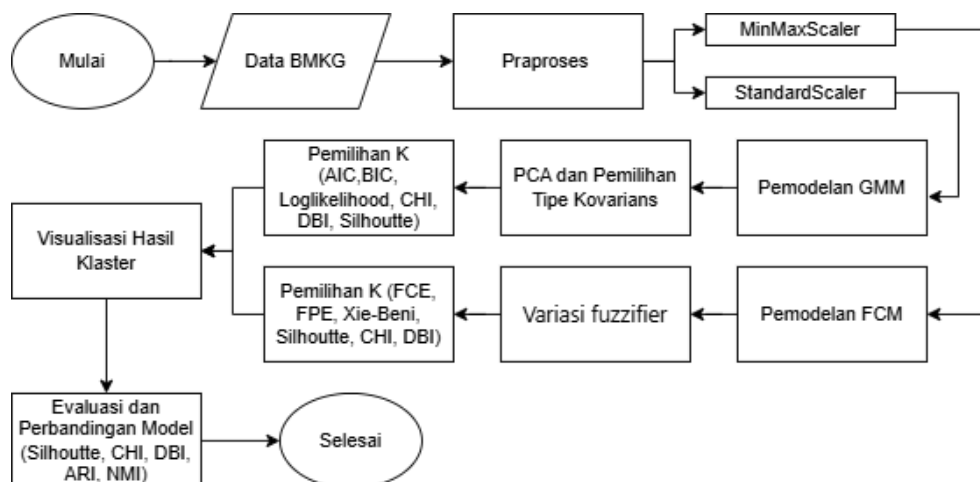
Penelitian ini menggunakan data sekunder yang diperoleh dari laman resmi Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) melalui Repositori Data Gempa (<https://repogempa.bmkg.go.id/>). Dataset mencakup seluruh kejadian gempa bumi di Indonesia selama periode Januari 2025–Februari 2026 dengan total data setelah *preprocessing* adalah 17181 data. Variabel yang digunakan dalam analisis kluster disajikan pada Tabel 1.

Tabel 1: Variabel yang Digunakan dalam Analisis

No	Variabel	Satuan	Keterangan
1	Lintang	Derajat (°)	Posisi geografis utara-selatan hiposenter
2	Bujur	Derajat (°)	Posisi geografis timur-barat hiposenter
3	Kedalaman	km	Jarak hiposenter terhadap permukaan bumi
4	Magnitudo	MLv	Kekuatan gempa dalam skala magnitudo lokal komponen vertikal
5	<i>Phase Count</i>	-	Jumlah fase gelombang seismik yang digunakan dalam penentuan lokasi
6	<i>Azimuth Gap</i>	Derajat (°)	Sudut azimuth terbesar tanpa stasiun pencatat, indikator kualitas lokasi hiposenter

2.2 Metode Penelitian

Metode penelitian disusun secara sistematis untuk memperoleh hasil kluster yang optimal. Secara umum, tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1: Diagram alir tahapan analisis

2.2.1 Preprocessing data dan Deteksi Outlier

Tahap *preprocessing* diawali dengan memilih tipe magnitudo MLV untuk menjaga konsistensi pengukuran. Setelah proses seleksi dan penghapusan missing value, diperoleh 17181 data gempa. Kategori kedalaman dibentuk berdasarkan standar BMKG, yaitu dangkal (≤ 60 km), menengah (61-300 km), dan dalam (> 300 km) untuk interpretasi hasil. Variabel *phase count* dan *azimuth gap* tetap dipertahankan karena merepresentasikan kualitas observasi hiposenter yang relevan dalam evaluasi pola kejadian gempa. Seluruh variabel kemudian distandarisasi menggunakan *StandardScaler* pada GMM dan *MinMaxScaler* pada FCM agar perbedaan rentang nilai tidak mendominasi proses klasterisasi (Senkaya *et al.*, 2024), sedangkan koordinat gempa yang berada pada rentang geografis Indonesia (95°BT – 141°BT) tidak melintasi antimeridian sehingga tetap konsisten untuk analisis berbasis jarak Euclidean. Tahap selanjutnya adalah identifikasi outlier menggunakan metode nilai baku (Z-score). Pada pengujian ini, nilai syarat Z-score yang ditetapkan sebagai kriteria adalah ± 3 (Tileng, 2015 dalam Aiman, 2023).

2.2.2 Analisis Korelasi Spearman

Korelasi Spearman digunakan untuk mendeteksi redundansi antar variabel karena bersifat nonparametrik. Penelitian ini menggunakan *threshold* $|r| \geq 0.85$ (Gu *et al.*, 2019 dalam Senkaya *et al.*, 2024). Variabel dengan korelasi tinggi dianggap membawa informasi serupa sehingga salah satu variabel dapat dieliminasi (Yu & Liu, 2004 dalam Senkaya *et al.*, 2024).

2.2.3 Fuzzy C-Means

Fuzzy C-Means (FCM) merupakan metode *soft clustering* yang memungkinkan satu data memiliki derajat keanggotaan pada lebih dari satu klaster (Senkaya *et al.*, 2024). Algoritma bekerja secara iteratif dengan meminimalkan fungsi objektif dan memperbarui pusat kluster dan matriks keanggotaan berikut:

$$J(U, V) = \sum_{i=1}^N \sum_{j=1}^c (u_{ij})^m (d_{ij})^2 \quad (1)$$

$$v_{ij} = \frac{\sum_{i=1}^N (u_{ij})^m x_i}{\sum_{i=1}^N (u_{ij})^m}; \quad u_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{2/(m-1)}} \quad (2)$$

Dengan d_{ij} adalah jarak Euclidean antara data ke- i dan pusat kluster ke- j , x_i data ke- i , v_j pusat kluster ke- j , serta m adalah indeks *fuzziness* dengan rentang $m \in [1, \infty]$. Jumlah klaster optimal pada FCM

ditentukan menggunakan *Fuzzy Partition Coefficient* (FPC), *Fuzzy Partition Entropy* (FPE), *Xie-Beni Index* (XB), *Silhouette Score*, *Calinski-Harabasz Index* (CHI), dan *Davies-Bouldin Index* (DBI).

$$\begin{aligned}
 FPC &= \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^c (u_{ij})^2; FPE = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^c u_{ij} \ln(u_{ij}); \\
 XB &= \frac{\sum_{i=1}^N \sum_{j=1}^c (u_{ij})^m \|x_i - v_j\|^2}{N \cdot \min_{j \neq k} \|v_j - v_k\|^2} \\
 S(i) &= \frac{b(i) - a(i)}{\max(a(i), b(i))}; CHI = \frac{SS_B}{SS_W} \times \frac{N - k}{k - 1}; \\
 DBI &= \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij})
 \end{aligned} \tag{3}$$

Dengan u_{ij} merepresentasikan derajat keanggotaan data ke- i pada kluster ke- j . SS_B adalah total varians antarkluster, SS_W adalah total varians dalam kluster, R_{ij} merupakan rasio kemiripan jarak antarkluster i dan j . Adapun $a(i)$ dan $b(i)$ adalah rata-rata jarak suatu data terhadap sesama anggota klasternya dan terhadap kluster tetangga terdekatnya.

2.2.4 Gaussian Mixture Model

Gaussian Mixture Model (GMM) merupakan metode klusterisasi probabilistik yang memodelkan data sebagai campuran beberapa distribusi Gaussian multivariat. GMM memungkinkan pembentukan kluster dengan ukuran dan orientasi berbeda melalui matriks kovarians serta menghasilkan probabilitas keanggotaan pada setiap observasi (Narulita & Kudus, 2022). Fungsi densitas GMM multivariat dirumuskan sebagai berikut.

$$p(x_i | \pi, \mu, \Sigma) = \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k), i = 1, 2, \dots, n \tag{5}$$

Dengan x_i merupakan data ke- i , μ_k dan Σ_k masing-masing menyatakan vektor rata-rata dan matriks kovarians komponen ke- k , sedangkan π_k merupakan proporsi campuran pada kluster ke- k . Struktur geometri data dievaluasi menggunakan *Principal Component Analysis* (PCA) pada ruang dua dimensi untuk mengidentifikasi pola isotropik maupun anisotropik sebagai dasar pemilihan tipe kovarians GMM. Pemilihan tipe kovarians dilakukan melalui perbandingan nilai *Akaike Information Criterion* (AIC) dan *Bayesian Information Criterion* (BIC).

$$AIC = 2p - 2 \ln(L); BIC = q \ln(n) - \ln L_c(\Psi) \tag{6}$$

Estimasi parameter GMM dilakukan menggunakan metode *Maximum Likelihood Estimation* (MLE). Namun, fungsi *log-likelihood* tersebut tidak memiliki solusi *closed-form*, sehingga estimasi parameter dilakukan menggunakan algoritma *Expectation-Maximization* (EM) (Narulita & Kudus, 2022). Fungsi likelihood dirumuskan sebagai berikut:

$$L(\Psi) = \prod_{i=1}^n \sum_{k=1}^K \pi_k \mathcal{N}(x_i | \mu_k, \Sigma_k) \quad (7)$$

Algoritma *Expectation-Maximization* (EM) terdiri atas tahap *Expectation* (*E-Step*) dan *Maximization* (*M-Step*) yang dilakukan secara iteratif hingga model konvergen (Ng *et al.*, 2019 dalam Narulita & Kudus, 2022). Pada iterasi ke- t , *E-Step* menghitung probabilitas posterior keanggotaan tiap observasi terhadap komponen Gaussian sebagai berikut:

$$\tau_k(x_i | \Psi^{(t)}) = \frac{\pi_k^{(t)} \mathcal{N}(x_i | \mu_k^{(t)}, \Sigma_k^{(t)})}{\sum_{j=1}^K \pi_j^{(t)} \mathcal{N}(x_i | \mu_j^{(t)}, \Sigma_j^{(t)})} \quad (8)$$

Dengan t menyatakan iterasi ke- t dan $\tau_k(x_i | \Psi^{(t)})$ merupakan probabilitas posterior bahwa data ke- i termasuk komponen ke- k . Untuk mempermudah pembaruan parameter pada distribusi Gaussian, perhitungan dirangkum dalam bentuk statistik yang memadai (Ng *et al.*, 2019 dalam Narulita & Kudus, 2022).

$$T_{k1}^{(t)} = \sum_{i=1}^n \tau_k(x_i | \Psi^{(t)}); \quad T_{k2}^{(t)} = \sum_{i=1}^n \tau_k(x_i | \Psi^{(t)}) x_i; \quad T_{k3}^{(t)} = \sum_{i=1}^n \tau_k(x_i | \Psi^{(t)}) x_i x_i^T \quad (9)$$

Pada *M-Step*, parameter Ψ diperbarui menjadi $\Psi^{(t+1)}$ menggunakan statistik tersebut sehingga diperoleh estimasi parameter campuran komponen normal sebagai berikut:

$$\pi_k^{(t+1)} = \frac{T_{k1}^{(t)}}{n}; \quad \mu_k^{(t+1)} = \frac{T_{k2}^{(t)}}{T_{k1}^{(t)}}; \quad \Sigma_k^{(t+1)} = \frac{T_{k3}^{(t)} - \left(T_{k1}^{(t)}\right)^{-1} T_{k2}^{(t)} T_{k2}^{(t)T}}{T_{k1}^{(t)}} \quad (10)$$

Iterasi dihentikan ketika perubahan nilai *log-likelihood* berada di bawah batas toleransi atau jumlah iterasi maksimum tercapai. Penentuan jumlah kluster optimal dilakukan menggunakan AIC, BIC, Silhouette Score, CHI, dan DBI.

2.2.5 Evaluasi dan Perbandingan model

Kualitas pembagian kluster diukur menggunakan Silhouette Score, CHI, dan DBI untuk menilai tingkat kekompakan dan jarak antarkluster. Selain itu, metrik *Adjusted Rand Index* (ARI) dan *Normalized Mutual Information* (NMI) digunakan untuk membandingkan kesamaan hasil antarmodel. Evaluasi ini bertujuan membandingkan secara kuantitatif fleksibilitas dan performa GMM dan FCM dalam analisis kluster gempa (Dwitiyanti *et al.*, 2024).

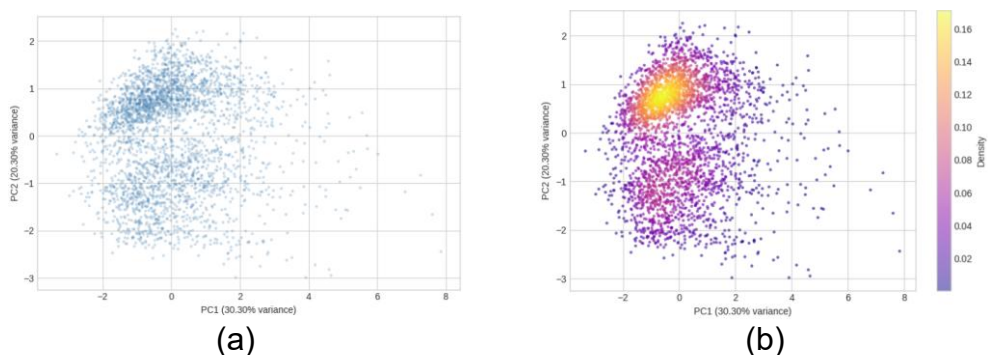
3. Hasil dan Pembahasan

3.1 Eksplorasi Data

Pemeriksaan outlier dilakukan pada variabel yang berkaitan dengan variabel fisik serta operasional yaitu pada variabel kedalaman, magnitudo, *phase count*, dan *azimuth gap*. Variabel koordinat merepresentasikan lokasi spasial sehingga seluruh nilai valid dalam wilayah studi dan tidak perlu dilakukan pemeriksaan outlier. Hasil menunjukkan 3.11% data teridentifikasi sebagai *outlier* dari keseluruhan data. Outlier tetap dipertahankan karena masih merepresentasikan variasi alami aktivitas seismik dan bukan kesalahan data, melainkan dapat mengandung informasi penting terkait karakteristik kejadian gempa bumi (Hirano *et al.*, 2022).

Hasil korelasi Spearman menunjukkan tidak ada hubungan antarvariabel yang terlalu kuat karena seluruh nilai korelasi berada pada $|r| \leq 0.85$. Korelasi terbesar terdapat pada magnitudo dan *phase count* sebesar 0.54 yang menandakan gempa bermagnitudo lebih besar akan cenderung memiliki lebih banyak fase. Oleh karena itu, seluruh variabel tetap digunakan karena masih merepresentasikan informasi yang berbeda dalam klasterisasi gempa bumi.

Hasil tahap PCA pada Gambar 3 menunjukkan kontribusi variansi PC1 dan PC2 masing-masing sebesar 30.30% dan 20.30% dengan rasio variansi sebesar 1.493, yang mengindikasikan struktur data cenderung mendekati isotropik. Berdasarkan visualisasi PCA yang menunjukkan sebaran data kontinu serta plot densitas yang mengidentifikasi subpopulasi dengan variasi kepadatan, *Gaussian Mixture Model* (GMM) dinilai sesuai untuk memodelkan karakteristik data ini.



Gambar 2: Visualisasi hasil PCA (a) Scatter plot 2D (b) Estimasi Kernel Density

3.2 Pemilihan Tipe Kovarians dan K-Optimal pada Gaussian Mixture Model

Penentuan tipe kovarians metode GMM dilakukan sebelum proses klasterisasi dengan menggunakan $K = 3$ sebagai nilai referensi. Hasil

analisis menunjukkan bahwa tipe kovarians *full* memberikan nilai terbaik dengan AIC sebesar 223055.67 dan BIC sebesar 223699.05. Hal ini menunjukkan bahwa tipe kovarians *full* sesuai karena mampu mengakomodasi perbedaan bentuk, ukuran, dan orientasi sebaran pada setiap cluster gempa yang terbentuk.

Selanjutnya, penentuan jumlah klaster optimal dalam rentang $K = 2$ hingga $K = 10$ pada Tabel 2 menunjukkan bahwa nilai AIC dan BIC terus menurun, sedangkan *log-likelihood* meningkat seiring bertambahnya jumlah klaster. Meskipun demikian, peningkatan kompleksitas model tidak selalu merepresentasikan struktur data yang lebih baik (Coraggio & Coretto, 2023).

Tabel 2: Evaluasi Jumlah Klaster Optimal pada *Gaussian Mixture Model* (GMM).

K	BIC	AIC	Log-Likelihood	Silhouette	CHI	DBI
2	240162.12	239735.79	-119812.90	0.17	2453.55	2.41
3	223699.05	223055.67	-111444.84	0.16	2682.56	2.00
4	209092.28	208231.85	-104004.93	0.20	2952.72	1.72
5	191091.34	190013.87	-94867.94	0.14	2634.56	1.65
6	185110.88	183816.37	-91741.19	0.13	2267.26	1.99
7	175651.27	174139.72	-86874.86	0.14	2278.52	1.89
8	173382.43	171653.84	-85603.92	0.14	2096.32	1.84
9	169835.89	167890.24	-83694.12	0.14	1978.80	1.97
10	166839.79	164677.11	-82059.55	0.13	1865.69	2.03

Silhouette Score dan CHI mencapai nilai optimum pada $K=4$, sedangkan DBI minimum pada $K=5$. Karena mayoritas metrik evaluasi mendukung $K=4$, jumlah klaster tersebut dipilih sebagai jumlah klaster optimal pada GMM. Pemilihan ini mempertimbangkan keseimbangan antara kualitas klaster dan interpretabilitas hasil, sehingga struktur klaster yang terbentuk tetap mampu merepresentasikan variasi karakteristik seismik secara rinci (Coraggio & Coretto, 2023).

3.3 Pemilihan K-Optimal pada *Fuzzy C-Means* (FCM)

Pemilihan jumlah klaster optimal pada metode FCM pada Tabel 3 menunjukkan bahwa FPC, FPE, *Silhouette score*, dan CHI secara konsisten menghasilkan nilai optimum pada $K = 2$. *Xie-Beni index* dan *Davies-Bouldin index* menunjukkan hasil optimum masing-masing pada $K = 6$ dan $K = 10$. Namun, berdasarkan mayoritas metrik evaluasi, $K = 2$ dipilih sebagai jumlah klaster optimal pada FCM. Hasil ini menunjukkan bahwa FCM cenderung membentuk klasterisasi yang lebih umum dengan dua klaster utama berdasarkan karakteristik kegempaan.

Tabel 3: Evaluasi Jumlah Kluster Optimal pada *Fuzzy C-Means* (FCM).

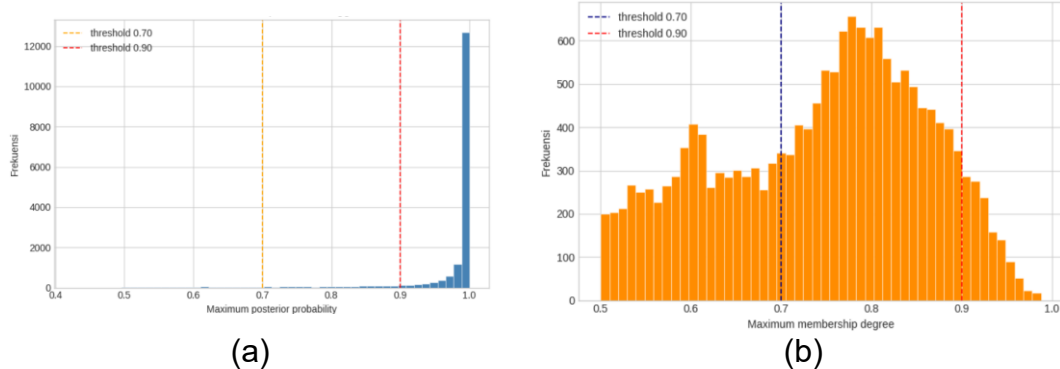
K	FPC	FPE	Xie-Beni	Silhouette	CHI	DBI
2	0.65	0.53	0.52	0.31	7574.91	1.33
3	0.48	0.89	1.47	0.26	6055.48	1.58
4	0.44	1.04	0.83	0.28	7104.67	1.29
5	0.35	1.30	1.34	0.23	4771.70	1.64
6	0.37	1.32	0.50	0.29	6758.97	1.27
7	0.33	1.47	0.72	0.28	6389.08	1.23
8	0.32	1.55	0.71	0.27	6067.83	1.26
9	0.29	1.66	0.64	0.25	5685.86	1.24
10	0.28	1.74	0.52	0.27	5634.17	1.19

3.4 Evaluasi dan Perbandingan Model

Model *Gaussian Mixture Model* (GMM) dengan tipe kovariansi penuh menghasilkan empat kluster optimal dengan proporsi dominan pada Kluster 3 sebesar 40%, sedangkan algoritma *Fuzzy C-Means* (FCM) menghasilkan partisi yang lebih sederhana menjadi dua kluster dengan distribusi tidak seimbang, yaitu Kluster 0 (61.6%) dan Kluster 1 (38.4%). Perbedaan jumlah kluster ini menunjukkan kapabilitas GMM dalam menangkap struktur subpopulasi data yang lebih kompleks, sementara FCM menghasilkan klusterisasi yang lebih sederhana.

Berdasarkan evaluasi kualitas kluster, FCM memperoleh *Silhouette Score* sebesar 0.203 dan CHI sebesar 3292.38, sedikit lebih tinggi dibandingkan GMM dengan *Silhouette Score* sebesar 0.200 dan CHI sebesar 2952.72. Sebaliknya, GMM unggul pada DBI dengan nilai 1.724 dibandingkan FCM sebesar 2.006, yang mengindikasikan separasi antarkluster lebih baik pada data yang berpotensi *overlap*. Meskipun FCM memperoleh nilai *Silhouette Score* dan CHI yang sedikit lebih tinggi, perbedaannya relatif kecil, sedangkan GMM mampu menghasilkan struktur kluster yang lebih rinci. Nilai *Silhouette* kedua model tergolong rendah, sejalan dengan karakteristik data seismik berdimensi tinggi yang bersifat kontinu.

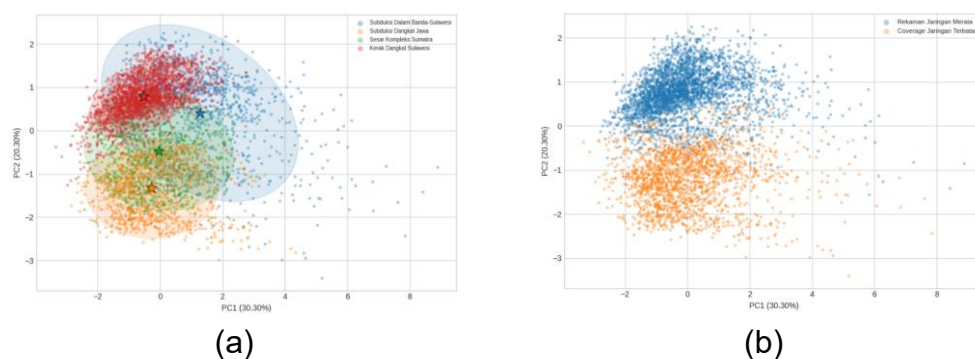
Tingkat kesepakatan antarmodel yang diukur menggunakan ARI = 0.38 dan NMI = 0.39 tergolong moderat, menunjukkan bahwa kedua algoritma menghasilkan struktur kluster yang berbeda. Perbedaan tersebut juga tercermin pada karakteristik keanggotaannya (Gambar 4), di mana lebih dari 90% data pada GMM memiliki probabilitas keanggotaan di atas 0.90 sehingga menghasilkan kluster yang relatif tegas. Sebaliknya, FCM mempertahankan derajat keanggotaan yang lebih tersebar pada sebagian observasi, mencerminkan karakteristik *fuzzy clustering* yang mengakomodasi ketidakpastian pada wilayah transisi antarkluster.



Gambar 4: Distribusi tingkat kepastian hasil kluster (a) probabilitas keanggotaan posterior pada GMM dan (b) derajat keanggotaan pada FCM.

3.5 Interpretasi Karakteristik Kluster

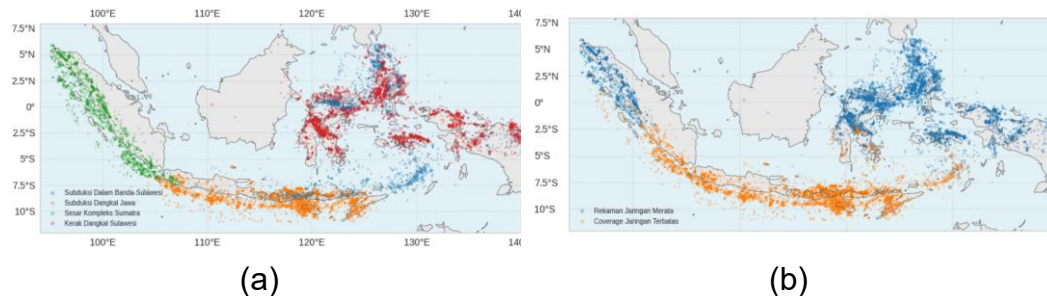
Karakteristik kluster dianalisis berdasarkan visualisasi pada ruang *Principal Component Analysis* (PCA), distribusi spasial, serta distribusi magnitudo dan kedalaman gempa untuk membandingkan pola klusterisasi yang dihasilkan oleh *Gaussian Mixture Model* (GMM) dan *Fuzzy C-Means* (FCM). Gambar 5 menunjukkan bahwa GMM membentuk empat kluster dengan pola distribusi yang relatif berbeda meskipun masih terdapat beberapa area *overlap* antar kluster. Hal ini mengindikasikan bahwa data gempa memiliki struktur yang kompleks dan tidak terpisah secara tegas. Sementara itu, FCM menghasilkan dua kluster utama dengan pola pemisahan yang lebih sederhana.



Gambar 5: Visualisasi hasil kluster (a) GMM dan (b) FCM pada ruang PCA.

Distribusi spasial pada Gambar 6 memperlihatkan bahwa GMM menghasilkan persebaran kluster yang lebih rinci pada beberapa wilayah di Indonesia karena mampu membedakan empat segmentasi kegempaan utama di Indonesia. Kluster 0 menggambarkan zona subduksi dalam Banda-Sulawesi dengan kedalaman rata-rata 110.1 km, *phase count* tertinggi 37.2, dan *azimuth gap* terendah 115.8°. Kombinasi ini mengindikasikan gempa berkedalaman menengah dengan kualitas

rekaman seismik paling baik. Klaster 1 merepresentasikan zona subduksi dangkal Jawa dengan kedalaman rata-rata 25 km dan *azimuth gap* tertinggi 170.6°. Nilai *azimuth gap* yang besar mencerminkan posisi episenter di selatan Jawa yang menyebabkan cakupan stasiun hanya dari arah daratan, sehingga kualitas lokasi hiposenter relatif terbatas. Klaster 2 mencerminkan zona sesar kompleks Sumatra dengan standar deviasi kedalaman terbesar 41.4 km, mengindikasikan heterogenitas sumber gempa akibat campuran dari zona subduksi dan sistem Sesar Besar Sumatra dalam satu wilayah. Klaster 3 menunjukkan gempa kerak dangkal Sulawesi dengan kedalaman rata-rata 11.9 km dan *phase count* terendah 19.0. Kombinasi kedalaman dangkal yang homogen dan jumlah fase rekaman terbatas mencerminkan karakteristik gempa kerak lokal yang menghasilkan gempa dengan rentang kedalaman terkontrol namun terekam oleh stasiun terbatas.



Gambar 6: Distribusi spasial hasil klaster (a) GMM dan (b) FCM.

Sementara itu, FCM menghasilkan pembagian yang lebih sederhana menjadi dua klaster. Klaster 0 mencakup gempa dengan kedalaman rata-rata 33.2 km, *azimuth gap* 123.5°, dan *phase count* relatif rendah 23.8. Klaster ini mengindikasikan gempa dengan distribusi stasiun yang melingkupi episenter dengan cukup baik namun jumlah fase yang terekam terbatas, didominasi gempa di wilayah Sulawesi. Klaster 1 merepresentasikan gempa dengan kedalaman rata-rata lebih besar 47.6 km, *azimuth gap* lebih tinggi 164.4°, dan *phase count* lebih tinggi 31.0. Kombinasi ini mencerminkan gempa di zona dengan kedalaman lebih besar yang terekam oleh lebih banyak stasiun namun dengan cakupan kurang simetris, seperti bagian selatan kepulauan Indonesia yang menghadap Samudera Hindia. Dengan demikian, GMM lebih sesuai untuk menjelaskan variasi spasial-seismik secara detail karena membedakan empat zona seismik dengan profil variabel dan makna geologis yang berbeda nyata, sedangkan FCM hanya menonjolkan pola makro berdasarkan kombinasi kedalaman, kualitas rekaman, dan distribusi jaringan stasiun.

4. Simpulan

Metode Gaussian Mixture Model dengan tipe kovarians *full* menunjukkan performa yang lebih representatif dalam menggambarkan variasi

karakteristik seismik dibandingkan FCM. Walaupun FCM memiliki *Silhouette Score* dan Calinski-Harabasz *Index* sedikit lebih tinggi, selisihnya terbilang marginal. Sebaliknya, GMM memisahkan data *overlap* secara lebih optimal yang ditunjukkan oleh nilai Davies-Bouldin Index sebesar 1.724, lebih rendah dari FCM yang berada pada angka 2.006. Keunggulan ini didukung oleh sifat algoritmanya yang probabilistik, sehingga memberikan kepastian klasifikasi yang tinggi dengan 90.9% data memiliki probabilitas keanggotaan di atas 0.90. Hal ini kontras dengan algoritma FCM yang membentuk keanggotaan berdasarkan kedekatan terhadap pusat kluster sehingga masih menyisakan banyak ambiguitas klasifikasi pada area transisi data. Secara karakteristik, pembentukan empat kluster pada GMM sukses mendiskriminasi variasi profil seismik secara rinci, khususnya dalam membedakan batas gempa dangkal, menengah, dan dalam, dibandingkan FCM yang hanya mengerucut pada dua kluster makro yang terlalu umum. Oleh karena itu, pendekatan probabilistik GMM dinilai lebih representatif untuk mendukung kebijakan analisis risiko dan mitigasi bencana.

Daftar Pustaka

- Ahmad, A. & Irsalinda, N. (2020). Cluster analysis of earthquake's data clustering in Indonesia using fuzzy K-means clustering. *Proceedings of the International Conference on Science and Engineering*, 3: 3–7.
- Aiman, M. K. (2023). Pemetaan kerentanan gempa di Kecamatan Pacitan Kabupaten Pacitan menggunakan pengukuran mikrotremor dengan metode HVSr. Tugas Akhir. Yogyakarta(ID): Universitas Islam Indonesia.
- Coraggio, L. & Coretto, P. (2023). Selecting the number of clusters, clustering models, and algorithms: A unifying approach based on the quadratic discriminant score. *Journal of Multivariate Analysis*, 196, p.105181. DOI:10.1016/j.jmva.2023.105181.
- Dwitiyanti, N., Kumala, S. A. & Handayani, S. D. (2024). Comparative study of earthquake clustering in Indonesia using K-medoids, K-means, DBSCAN, Fuzzy C-means and K-AP algorithms. *Jurnal RESTI*, 8(6): 768–778. DOI: 10.29207/resti.v8i6.551.
- Hirano, S., Kawakata, H., & Doi, I. (2022). A matched-filter technique with an objective threshold. *Scientific Reports*, 12(1): 22090. <https://doi.org/10.1038/s41598-022-25839-2>.
- Huang, Z., Yang, Z., Pang, B., Wu, Z. & Feng, L. (2025). Characterizing and clustering debris flow and environmental noise seismic signals using unsupervised deep learning. *Geophysical Journal International*, 243(1):

1–18. DOI: 10.1093/gji/ggaf353.

Mahanani, E., Djuhardi, L. & Malays, E. (2024). Kajian tentang bencana dan dampak gempa bumi (suatu pembelajaran pasca gempa bumi Cianjur, Jawa Barat). *IKRAITH-ABDIMAS*, 8(3): 246–253. DOI: 10.37817/ikraithabdimas.v8i3.4115.

Narulita, R. & Kudus, A. (2022). Penerapan gaussian mixture model pada klasterisasi desa di Kalimantan Barat berdasarkan indeks desa membangun tahun 2021. *Bandung Conference Series: Statistics*, 2(2). DOI: 10.29313/bcss.v2i2.3877.

Prasetio, A., Effendi, M. M. & Dwi M, M. N. (2023). Analisis gempa bumi di Indonesia dengan metode clustering. *Bulletin of Information Technology (BIT)*, 4(3): 338–343. DOI: 10.47065/bit.v4i3.820.

Rais, Z., Hafid, H. & Risqi, S. (2025). Application of bisecting k-means method in grouping earthquake data (case study: earthquakes in Indonesia 2023). *Inferensi*, 8(3).

Senkaya, M., Karslı, H. & Doğru, A. (2024). Seismic activity clustering using fuzzy C-means: A case study for Turkey. *Natural Hazards*, 120(3): 2871–2894. DOI: 10.1007/s11069-023-06297-x.

Willyana, I. P., Hadiana, A. I. & Ilyas, R. (2025). Analisis klaster daerah rawan gempa di Indonesia menggunakan K-Means dan DBSCAN berbasis data historis BMKG. *Jurnal Teknologi Informasi Komunikasi (Tematik)*, 12(1): 59–71.

Zhao, X., Chen, H., Li, B., Yang, Z. & Li, H. (2024). Using fuzzy c-means clustering to determine first arrival of microseismic recordings. *Sensors*, 24(5). DOI: 10.3390/s24051682.