

PENERAPAN ALGORITMA *FUZZY C-MEANS* UNTUK PENGELOMPOKKAN KEMISKINAN DI KABUPATEN/KOTA PROVINSI ACEH

Miftakhul Kharis¹, Indah Manfaati Nur²

^{1,2}Jurusan Statistika FSTP Universitas Muhammadiyah Semarang

Jalan Kedungmundu No.18, Kec. Tembalang, Kota Semarang, Jawa Tengah 50273,

Indonesia

¹ttakul54@gmail.com; ²indahnmur@unimus.ac.id

ABSTRACT

Poverty has become a familiar problem in Indonesia. However, poverty remains a persistent issue. This research aims to cluster the districts in Aceh province based on the poverty rate in 2021 using fuzzy c-means data mining. Aceh province is among the five poorest provinces in Indonesia. Data mining is known as a technique used to uncover hidden patterns in a dataset with the goal of generating new insights. One of the data mining techniques is clustering, which is used to group data based on their similarities into clusters. There are several algorithms used in clustering, one of which is the Fuzzy C-Means (FCM) clustering algorithm. The FCM algorithm can assist the government in poverty alleviation strategies by adapting their management strategies for specific groups. The results of the research show three clusters with different districts in each cluster. The validation of the FCM method demonstrates good quality with a Fuzzy Silhouette Index (FSI) value of 0.537. The Modified Partition Coefficient Index (MPCI) also indicates good cluster quality with a value of 0.441. The Davies Bouldin Index (DBI) shows a value of 1.074, indicating a reasonably good clustering result as the DBI validation value ranges from 0 to infinity.

Keywords : Poverty, Fuzzy C-Means, Cluster Validation.

ABSTRAK

Kemiskinan telah menjadi masalah yang tidak asing lagi bagi Indonesia. Meski begitu, masalah kemiskinan sampai saat ini masih menjadi masalah yang berlarut-larut. Penelitian ini bertujuan mengelompokkan kabupaten provinsi Aceh berdasarkan angka kemiskinan tahun 2021 menggunakan data mining fuzzy cmeans. Karena provinsi Aceh termasuk dalam 5 provinsi termiskin di Indonesia. Data mining dikenal sebagai teknik yang digunakan untuk mengungkap pola tersembunyi dalam kumpulan data dengan tujuan menghasilkan wawasan baru. Teknik data mining salah satunya yaitu clustering. Clustering adalah teknik data mining yang digunakan untuk mengelompokkan data berdasarkan kemiripannya ke dalam kelompok-kelompok yang dikenal dengan cluster. Terdapat beberapa algoritma yang digunakan dalam klasterisasi, salah satunya adalah algoritma *Fuzzy C-Means (FCM) Clustering*. Algoritma FCM dapat membantu pemerintah dalam strategi pengentasan kemiskinan dengan menyesuaikan strategi manajemen mereka untuk kelompok-kelompok tertentu. Hasil penelitian menunjukkan tiga kelompok dengan distrik-distrik yang berbeda di setiap kelompok. Validasi metode FCM menunjukkan kualitas yang baik dengan nilai Fuzzy Silhouette Index (FSI) yang mendekati nilai 1 yaitu sebesar 0,537. Modified Partition Coefficient Index (MPCI) juga menunjukkan kualitas kelompok yang baik dengan nilai

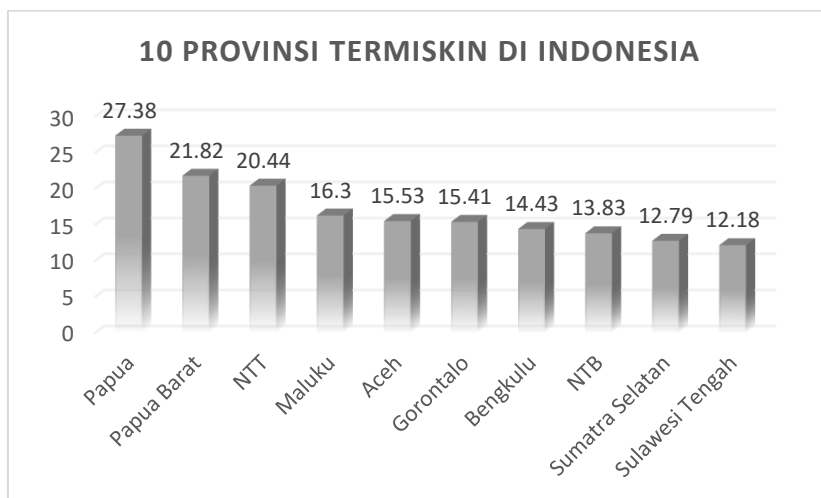
0,441. Davies Bouldin Index (DBI) menunjukkan nilai 1,074 yang menunjukkan hasil clustering cukup baik karena nilai validasi DBI berkisar antara 0 hingga tak terhingga.

Kata kunci : Kemiskinan, Fuzzy C-Means, Validasi Klaster.

I. PENDAHULUAN

Kemiskinan bagi Indonesia tidak lagi menjadi persoalan krisis yang baru terjadi. Sejak awal kemerdekaannya, bangsa Indonesia telah memberikan perhatian yang besar untuk mewujudkan masyarakat yang adil dan makmur sebagaimana termaktub dalam alinea keempat pembukaan UUD 1945. [1]. Program pembangunan yang dilakukan selama ini selalu memperhatikan upaya penanggulangan kemiskinan karena pada dasarnya pembangunan dilakukan untuk tujuan meningkatkan kesejahteraan rakyat. Meski begitu, masalah kemiskinan sampai saat ini masih menjadi masalah yang berlarut-larut. Sebenarnya sudah banyak program penanggulangan kemiskinan yang dilaksanakan oleh pemerintah, namun belum membawa perubahan yang signifikan.

Selain itu, tahun 2019 - 2020 merupakan tahun yang sulit bagi semua negara, termasuk Indonesia, akibat kerusakan parah akibat pandemi *covid 19* pada tahun tersebut. Dan tahun 2021 merupakan tahun transisi bagi negara Indonesia untuk pulih dari keterpurukan ekonomi. Dalam penelitian ini, peneliti ingin mengelompokkan kabupaten-kabupaten provinsi Aceh berdasarkan angka kemiskinan tahun 2021. Aceh bisa dibilang merupakan provinsi termiskin di Indonesia. Karena provinsi Aceh termasuk dalam 5 provinsi termiskin di Indonesia [2].



Gambar 1 : Grafik 10 Besar Provinsi Termiskin di Indonesia

Dalam penelitian ini, peneliti menggunakan variabel persentase kemiskinan, PDRB, harapan hidup, dan rata-rata lama sekolah untuk menentukan cluster. Karena menurut artikel Fitriana, variabel-variabel tersebut saling mempengaruhi secara signifikan [3]. Untuk melihat seberapa banyak kabupaten di provinsi Aceh yang miskin, diperlukan metode seperti *data mining* untuk mencari informasi tertutup dari data tersebut [4].

Data mining dikenal sebagai teknik yang digunakan untuk mengungkap pola tersembunyi dalam kumpulan data dengan tujuan menghasilkan wawasan baru. Metode yang digunakan dalam data mining bervariasi tergantung pada tujuan penggunaan kumpulan data. Ini termasuk estimasi, prediksi, klasifikasi, pengelompokan, dan asosiasi. Dalam pemetaan data, salah satu teknik yang digunakan adalah pengelompokan atau clustering. *Clustering* adalah teknik data mining yang digunakan untuk mengelompokkan data berdasarkan kemiripannya ke dalam kelompok-kelompok yang dikenal dengan cluster [5]. Terdapat beberapa algoritma yang digunakan dalam klasterisasi, salah satunya adalah algoritma *Fuzzy C-Means* (FCM) *Clustering*[6]. Proses pengelompokan *Fuzzy C-Means* (FCM) mencakup bobot tingkat keanggotaan dalam himpunan *fuzzy*. Metode *Fuzzy C-Means* (FCM) memungkinkan suatu data memiliki dua atau lebih kelompok. Dalam *Fuzzy C-Means* (FCM), objek cenderung ditempatkan ke dalam kelompok berdasarkan tingkat keanggotaannya.

Adapun penelitian terdahulu yang dilakukan oleh Nidyashofa pada tahun 2017 dengan judul Penerapan Algoritma *Fuzzy C-Means* untuk Pengelompokan Kabupaten/Kota di Jawa Tengah Berdasarkan Status Kesejahteraan Tahun 2015. Penelitian ini menghasilkan 3 kluster kabupaten/kota di Jawa Tengah berdasarkan indikator kesejahteraan rakyat.

Penelitian lain yang dilakukan oleh Kahar pada tahun 2015 dengan judul Implementasi Data Mining Dalam Penentuan Tingkat Kemiskinan Menggunakan *Fuzzy C-Means*. Penelitian ini berhasil mengklaster 100 (seratus) data sampel rumah tangga miskin di Kabupaten Bone Bolango, Provinsi Gorontalo dalam tiga kategori kemiskinan, dengan persentase setiap kategori adalah 50% sangat miskin, 34% hampir miskin, dan 16% sangat miskin.

Dan penelitian lain yang dilakukan oleh Ulfah pada tahun 2015 dengan judul Analisis Kinerja Algoritma *Fuzzy C-Means* dan *K-Means* pada Data Kemiskinan. Hasil penelitian ini memiliki hasil *clustering* menggunakan algoritma FCM memiliki tingkat akurasi hanya 50%, sedangkan untuk algoritma *K-means* memiliki tingkat akurasi lebih baik yaitu 83.33%.

II. METODE PENELITIAN

2.1. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah variabel input dan output. Dimana variabel inputnya adalah data indikator kemiskinan yaitu persentase kemiskinan, PDRB, harapan hidup dan rata-rata lama sekolah. Sedangkan variabel outputnya adalah jumlah kabupaten miskin dengan 3 fuzzy set (rendah, sedang, tinggi)[6].

2.2. Data Penelitian

Data yang digunakan dalam penelitian ini adalah berupa data yang sudah ada sebelumnya atau biasa disebut dengan data sekunder yang diperoleh dari BPS provinsi Aceh yaitu: data presentase kemiskinan, pdrb, angka harapan hidup dan rata-rata lama sekolah [7].

Dimana data tersebut adalah data data yang saling mempengaruhi satu sama lain sehingga relevan digunakan dalam menentukan *cluster-cluster* tiap wilayah kota/kabupaten Aceh.

2.3. Tahapan Analisis Data

Pada tahap ini, data akan dianalisis menggunakan KDD (*Knowledge Discovery in Databases*). Seluruh proses KDD dimulai dengan tiga langkah yang melibatkan pemrosesan data sebelum menghasilkan wawasan , yaitu[4]:

a) *Preprocessing*

Preprocessing data dilakukan dengan menyiapkan struktur database untuk memudahkan proses eksplorasi. Tahap preprocessing terdiri dari tiga hal utama, yaitu:

1. Seleksi Data

Seleksi data adalah proses pemilihan subset dari semua data yang relevan dan penting untuk analisis atau pemodelan saat ini.

2. *Cleaning*

Pada tahap ini yang harus dilakukan adalah membersihkan data yang tidak lengkap atau tidak dapat digunakan karena *noise*, *missing value*, dan *outlier*. Secara umum pada tahap ini merupakan langkah pemilihan, verifikasi dan pengecekan keakuratan data, karena data yang diperoleh dari sumber mungkin masih mengandung atribut yang tidak sesuai untuk penelitian ini [8].

3. Integrasi Data

Setelah menyelesaikan *cleaning* data secara keseluruhan, langkah selanjutnya adalah melakukan integrasi data. Integrasi data adalah proses mengumpulkan dan menggabungkan data dari berbagai sumber menjadi satu sumber data yang konsisten.

b) *Data Mining*

Data mining adalah konsep yang berhubungan dengan penemuan pengetahuan dalam basis data. Ini melibatkan penerapan statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengenali informasi berharga dan pengetahuan terkait dari database besar yang beragam [9]. Pada penelitian ini digunakan prosedur data mining yang meliputi penerapan teknik *clustering* menggunakan algoritma *Fuzzy C-Means* (FCM) dengan tujuan untuk mengidentifikasi dan memperoleh karakteristik dari data yang dikumpulkan.

c) *Fuzzy C-Means*

Penelitian ini menggunakan metode *Fuzzy C-Means* (FCM). *Fuzzy C-Means* (FCM) adalah algoritma yang termasuk dalam kategori pembelajaran tanpa pengawasan. Algoritma ini digunakan untuk mengolah kumpulan data yang mengandung outlier, dimana terdapat objek yang letaknya jauh dari cluster yang ada, sehingga tidak memungkinkan untuk mengklasifikasikannya secara tegas ke dalam cluster tertentu [10]. *Fuzzy clustering* adalah metode clustering yang menggunakan fungsi keanggotaan *fuzzy* untuk mengidentifikasi grup berdasarkan jarak. Sebenarnya teknik ini merupakan hasil pengembangan dari metode partisi dengan menambahkan elemen *fuzzy*, yang memungkinkan untuk mengelompokkan data

meskipun kelompok data tersebut memiliki distribusi yang tidak jelas atau tidak dapat dilihat secara jelas [8].

Berikut adalah deskripsi algoritma *Fuzzy C-Means* (FCM) menurut Kusumadewi dan Purnomo [4]:

1. Tentukan data yang akan berada di *cluster* X, berupa matriks berukuran $n \times m$ ($n = \text{jumlah sampel data}, m = \text{atribut masing - masing data}$). $X_{ij} = \text{data sampel ke } i$ ($i = 1, 2, \dots, n$), $j = \text{atribut}$ ($j = 1, 2, \dots, m$). dengan batasan :
 - a. Jumlah *cluster* (c)
 - b. Pangkat (w)
 - c. Maksimum iterasi (Maxit)
 - d. Error terkecil yang diharapkan (ξ)
 - e. Fungsi obyektif awal ($P_0 = 0$)
 - f. Iterasi awal ($t = 1$)
2. Membangkitkan bilangan random secara acak untuk $\mu_{ik}, i = 1, 2, \dots, n; k = 1, 2, \dots, c$; sebagai elemen-elemen matrik partisi awal U, setiap baris memiliki jumlah nilai elemen kolom yang sama yaitu 1 (satu).

$$\sum_{i=1}^c \mu_{ci} = 1 \quad (1)$$

3. Menghitung pusat *cluster* ke- k : V_{kj} dengan $k = 1, 2, \dots, c$; dan $j = 1, 2, \dots, m$.

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \times X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (2)$$

4. Menghitung fungsi obyektif pada iterasi ke- t , P_t :
Fungsi tujuan digunakan sebagai syarat perulangan untuk mendapatkan pusat *cluster* yang tepat. Untuk mendapatkan kecenderungan data masuk ke *cluster* mana pada langkah terakhir. Untuk iterasi awal, nilai $t = 1$.

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (3)$$

5. Menghitung perubahan matrik partisi:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}} \quad (4)$$

dengan $i = 1, 2, \dots, n$; dan $k = 1, 2, \dots, c$

6. Memeriksa kondisi berhenti:
 - a. Jika : $(|P_t - P_{t-1}| < \xi)$ atau $(t > MakIter)$ maka berhenti;
 - b. Jika tidak : $t = t + 1$, ulangi langkah ke- d (menghitung V_{kj}).

d) Postprocessing

Post-processing bertujuan untuk membantu pengguna lebih memahami informasi yang dihasilkan. Karena masih sulit untuk menginterpretasikan pusat kluster secara langsung, maka diperlukan suatu metode bernama *Simple Additive Weighting* (SAW) yang dapat mengklasifikasikan nilai akhir dari pusat kluster. Metode *simple additive weighting* (SAW) juga dikenal sebagai metode penjumlahan tertimbang. Konsep dasar metode SAW adalah menghitung penjumlahan terbobot dari skor kinerja setiap alternatif untuk semua atribut. Berikut langkah-langkah metode SAW [4]:

1. Membuat matriks keputusan Z berukuran $m \times n$, dimana m = alternatif yang akan dipilih dan n = kriteria.
2. Memberikan nilai x setiap alternatif (i) pada setiap kriteria (j) yang sudah ditentukan, dimana, $i = 1, 2, \dots, m$ dan $j = 1, 2, \dots, n$ pada matriks keputusan Z,

$$\begin{bmatrix} x_{11} & x_{12} & x_{1j} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_{i1} & x_{i2} & x_{ij} \end{bmatrix}$$

3. Memberikan nilai bobot preferensi (W) oleh pengambil keputusan untuk masing-masing kriteria yang sudah ditentukan.

$$W = W_1 W_2 W_3 \dots W_j \tag{5}$$

4. Melakukan normalisasi matriks keputusan Z dengan cara menghitung nilai rating kinerja ternormalisasi (r_{ij}) dari alternatif A_i pada atribut C_j ;

$$r_{ij} = \begin{cases} \frac{x_{ij}}{\text{MAX}_i(x_{ij})} & \text{jika } j \text{ adalah atribut keuntungan} \\ \frac{\text{MIN}_i(x_{ij})}{x_{ij}} & \text{jika } j \text{ adalah atribut biaya} \end{cases} \tag{6}$$

dengan ketentuan:

- a. Atribut keuntungan dapat diidentifikasi ketika atribut tersebut memberikan manfaat yang signifikan bagi pengambil keputusan. Atribut biaya, di sisi lain, adalah atribut yang berkontribusi pada peningkatan biaya atau biaya dengan nilai yang lebih besar bagi pembuat keputusan.

- b. Jika atribut tersebut merupakan atribut keuntungan, maka nilai (x_{ij}) dari setiap kolom atribut akan dibagi dengan nilai maksimum $(MAX x_{ij})$ dari setiap kolom. Namun, jika atribut tersebut adalah atribut biaya, maka nilai minimum $(MIN x_{ij})$ dari setiap kolom atribut akan dibagi dengan nilai (x_{ij}) dari setiap kolom.
5. Hasil dari nilai rating kinerja ternormalisasi (r_{ij}) membentuk matriks ternormalisasi (N) ;

$$N = \begin{bmatrix} r_{11} & r_{12} & r_{1j} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ r_{i1} & r_{i2} & r_{ij} \end{bmatrix}$$

6. Proses perankingan dilakukan dengan mengalikan matriks yang telah dinormalisasi (N) dengan nilai bobot preferensi (W) .
7. Untuk menentukan nilai preferensi dari setiap alternatif (V_i) , dilakukan dengan menjumlahkan hasil perkalian antara matriks yang telah dinormalisasi (N) dengan nilai bobot preferensi (W) .

$$V_i = \sum_{j=1}^n w_j r_{ij} \quad (7)$$

Apabila nilai V_i lebih tinggi, hal tersebut menunjukkan bahwa alternatif A_i adalah alternatif terbaik.

e) Evaluasi hasil pengelompokan

Setelah menerapkan *Fuzzy C-Means* (FCM) pada data, beberapa hasil akan diamati seperti pengelompokan dan jumlah maksimal iterasi yang dilakukan. Faktor *silhouette* adalah metode yang digunakan untuk mengevaluasi kualitas dan kekuatan *cluster*. Metode ini menggabungkan dua pendekatan, yaitu metode kohesi mengukur jarak antar objek dalam suatu cluster dan metode pemisahan mengukur sejauh mana cluster terpisah dari cluster lainnya. Nilai indeks *Silhouette* berkisar dari 0 hingga 1, dimana nilai yang lebih tinggi menunjukkan kualitas kluster yang lebih baik. Nilai positif menunjukkan bahwa objek tersebut terkcluster dengan baik, sedangkan nilai negatif menunjukkan bahwa objek tersebut mungkin lebih cocok untuk cluster lain [11][12].

$$S_i = \left(\frac{b_i - a_i}{\max(a_i, b_i)} \right) \quad (8)$$

Dimana:

S_i = silhouette coefficient

a_i = rerata jarak dari objek ke $- i$ dengan seluruh objek di cluster

b_i = rerata jarak dari objek ke $- i$ dengan cluster apa pun yang tidak mengandung objek

Terdapat beberapa langkah untuk menunjukkan kluster yang telah dilakukan. Tahapan pertama melibatkan penghitungan beberapa indeks validasi. Untuk indeks validasi yang digunakan, yaitu *Fuzzy Silhouette Index* (FSI), *Modified Partition Coefficient Index* (MPCI) dan *Davies Bouldin Index* (DBI) [13][14]. Kriteria indeks ini untuk kluster disajikan dalam Tabel 2.

Tabel 2 : Nilai Kriteria Subjektif Pengukuran Kluster

Nilai	Interpretasi
0,71 - 1,00	Struktur Kluster Sangat Baik
0,51 - 0,70	Struktur Kluster Baik
0,26 - 0,50	Struktur Kluster Lemah
<0,25	Struktur Kluster Buruk

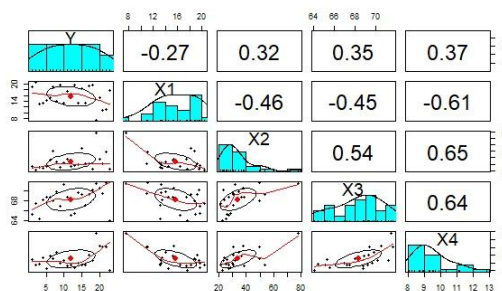
Nilai kriteria subjektif pengukuran kluster digunakan untuk mengukur nilai validasi dari FSI, MPCI dan DBI. Dapat dilihat untuk nilai validasi pada FSI yang sangat baik adalah yang mendekati nilai 1. Sedangkan untuk nilai validasi MPCI dan DBI nilai kriteria pengukuran yang memiliki struktur kluster sangat baik adalah yang mendekati nilai 0 [15].

III. HASIL DAN PEMBAHASAN

Menganalisis proses *data mining* dengan variabel data angka kemiskinan, PDRB, angka harapan hidup dan rata-rata lama sekolah provinsi Aceh tahun 2021, 92 data diolah untuk menghitung jarak ke setiap titik pusat cluster. Kemudian dilakukan clustering untuk masing-masing Kabupaten/Kota serta validasi untuk metode *Fuzzy C-Means* (FCM). Perhitungan dilakukan dengan menggunakan sintaks Rstudio.

3.1. Korelasi Antar Atribut

Korelasi atribut mengacu pada hubungan statistik antara dua atau lebih variabel atau atribut dalam kumpulan data. Korelasi ini mengukur derajat perubahan hubungan antara perubahan satu variabel dengan perubahan variabel lainnya. Korelasi dapat membantu kita memahami pola dan hubungan antar atribut dalam sebuah dataset. Korelasi ditunjukkan pada Gambar 2.



Gambar 2 : Korelasi Antar Atribut

Gambar di atas menunjukkan hasil korelasi antar properti beserta diagram dan grafiknya. Jika hasil korelasi mendekati 1, berarti korelasi antar variabel menguat. Kemudian untuk hasil negatif menunjukkan korelasi terbalik, yaitu jika salah satu nilainya naik, maka korelasinya turun. Sementara hasil positif menunjukkan bahwa korelasi berjalan ke arah yang sama, jika salah satu nilainya naik, maka korelasinya naik.

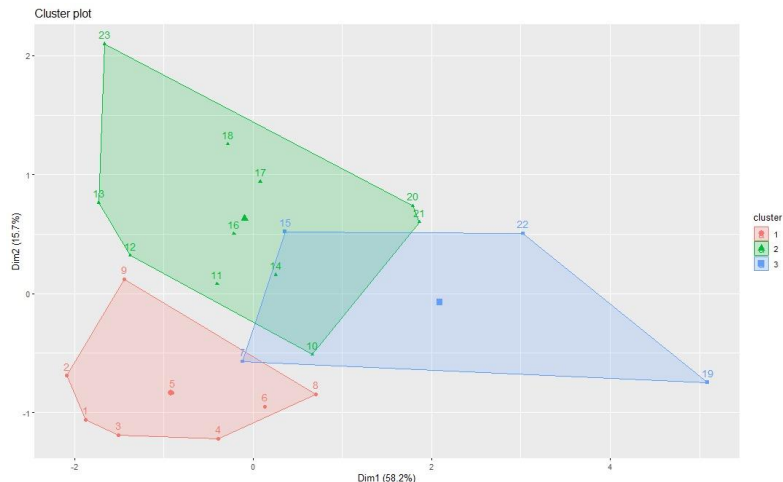
3.2. Melakukan *Clustering Fuzzy C-Means (FCM)*

Sebelum melanjutkan dengan clustering, terlebih dahulu kita harus menentukan jumlah cluster atau klaster. Dalam penelitian ini, 3 *cluster* ($k = 3$) digunakan. Berikut adalah Tabel 3 yang merupakan hasil matriks keanggotaan yang menunjukkan keanggotaan pada masing-masing kelompok.

Tabel 3 : Matriks Derajat Keanggotaan

Kabupaten	<i>Membership Degrees Matrix</i>		
	<i>Cluster 1</i>	<i>Cluster 2</i>	<i>Cluster 3</i>
Simeulue	0.86386059	0.11311466	0.023034754
Aceh Singkil	0.79895248	0.16856881	0.032478713
Aceh Selatan	0.87166003	0.10969763	0.018642339
Aceh Tenggara	0.89079890	0.09538853	0.013812575
Aceh Timur	0.94139217	0.05223571	0.006372112
Aceh Tengah	0.50639285	0.39919624	0.094410903
Aceh Barat	0.18389972	0.21978883	0.596311449
Aceh Besar	0.18389972	0.46816061	0.058468133
Pidie	0.77102103	0.21219948	0.061779484
Bireuen	0.37854266	0.58503021	0.039127132
Aceh Utara	0.29490456	0.58856910	0.116526346
Aceh Barat Daya	0.45009001	0.52869198	0.021218015
Gayo Lues	0.34462229	0.62839678	0.026980930
Aceh Tamiang	0.09380941	0.89472491	0.011465678
Nagan Raya	0.03645417	0.05524763	0.908298206
Aceh Jaya	0.05949898	0.93025512	0.010245904
Bener Meriah	0.09263413	0.88860360	0.018762262
Pidie Jaya	0.29715841	0.65679172	0.046049875
Banda Aceh	0.12310672	0.15300080	0.723892479
Sabang	0.16050479	0.69601957	0.143475640
Langsa	0.15094650	0.79160691	0.057446592
Lhokseumawe	0.09023643	0.17476395	0.734999616
Subulussalam	0.28238882	0.64526925	0.072341932

Output dari tabel matriks keanggotaan menunjukkan derajat keanggotaan dari masing-masing variabel. Tingkat keanggotaan tertinggi, atau mendekati nilai satu, masuk ke dalam cluster. Misalnya pada data pertama, nilai keanggotaan terbesar yaitu 0,863 terdapat pada *cluster* 1, artinya data pertama masuk pada cluster ketiga dan seterusnya. Hasil *clustering* dengan *Fuzzy C-Means* (FCM) dapat dilihat pada Gambar 3.



Gambar 3 : Hasil Plot *Clustering*

Berdasarkan *output* hasil clustering dengan *Fuzzy C-Means* (FCM) berupa diagram cluster di atas, terdapat 3 cluster pada penelitian ini. Kluster pertama berwarna merah dengan anggota Simeulue, Aceh Singkil, Aceh Selatan, Aceh Tenggara, Aceh Timur, Aceh Tengah, Aceh Besar dan Pidie. Kluster kedua berwarna biru dengan anggota Bireun, Aceh Utara, Aceh Barat Daya, Gayo Lues, Aceh Tamiang, Aceh Jaya, Bener Meriah, Pidie Jaya, Sabang, Langsa dan Subulussalam. Kluster ketiga berwarna biru dengan anggota Aceh Barat, Nagan Raya, Banda Aceh dan Lhokseumawe.

Validasi selanjutnya akan dilakukan dengan menghitung *Fuzzy Silhouette Index* (FSI), kemudian Partition Entropy Index (PEI), Partition Factor Index (PCI), Partition Factor Index Modified (MPCI) dan Davies Bouldin Index (DBI). Nilai validasi cluster disajikan pada Tabel 4.

Tabel 4 : Nilai Validasi *Cluster*

Nilai Validasi <i>Cluster</i>	
<i>Fuzzy Silhouette Index</i> (FSI)	0.537
<i>Modified Partition Coefficient Index</i> (MPCI)	0.441
<i>Davies Bouldin Index</i> (DBI)	1.074

Berdasarkan data pada Tabel 4, hasil uji *Fuzzy C-Means clustering test* (FCM) memiliki kualitas yang baik dengan nilai FSI sebesar 0,537. Namun hasil validasi MPCI menunjukkan nilai 0,441 yang menunjukkan kualitas cluster sangat baik jika mendekati nol karena menunjukkan bahwa

elemen cluster memang bersebelahan dan terpisah dari *cluster* yang lain. Sedangkan untuk DBI diperoleh nilai 1,074 yang menunjukkan hasil clustering cukup baik karena nilai validasi DBI berkisar antara 0 hingga tak terhingga.

IV. KESIMPULAN

Hasil agregat penelitian ini dibagi menjadi tiga kelompok. Kluster pertama berwarna merah dengan anggota Simeulue, Aceh Singkil, Aceh Selatan, Aceh Tenggara, Aceh Timur, Aceh Tengah, Aceh Besar dan Pidie. Kluster kedua berwarna biru dengan anggota Bireun, Aceh Utara, Aceh Barat Daya, Gayo Lues, Aceh Tamiang, Aceh Jaya, Bener Meriah, Pidie Jaya, Sabang, Langsa dan Subulussalam. Kluster ketiga berwarna biru dengan anggota Aceh Barat, Nagan Raya, Banda Aceh dan Lhokseumawe.

Hasil validasi agregasi data dengan *metode Fuzzy C-Means* (FCM) memiliki kualitas yang baik dengan nilai FSI sebesar 0,537. Namun hasil validasi MPCl menunjukkan nilai 0,441 yang menunjukkan kualitas cluster sangat baik jika mendekati nol karena menunjukkan bahwa elemen *cluster* memang bersebelahan dan terpisah dari *cluster* lain. Sedangkan untuk DBI diperoleh nilai 1,074 yang menunjukkan hasil *clustering* cukup baik karena nilai validasi DBI berkisar antara 0 hingga tak terhingga.

DAFTAR PUSTAKA

- [1] N. Prawoto, "Memahami Kemiskinan Dan Strategi Penanggulangannya", dalam Jurnal Ekonomi dan Studi Pembangunan," vol. 9, no. 1, pp. 56–68, 2009.
- [2] Kompas, "10 Daerah dengan Tingkat Kemiskinan Tertinggi di Indonesia, Mana Saja?" <https://www.kompas.com/tren/read/2022/03/03/203000365/10-daerah-dengan-tingkat-kemiskinan-tertinggi-di-indonesia-mana-saja-?page=all>
- [3] F. Ar, N. Salwa, and S. Wijayanti, "Analisis Faktor-Faktor yang Memengaruhi Tingkat Kemiskinan di Provinsi Aceh menggunakan Regresi Data Panel," vol. 5158, pp. 46–59, 2022, doi: 10.24815/jp.v11i1.32947.
- [4] Nurjanah, A. Farmadi, and F. Indriani, "Implementasi Metode Fzzy C-Means Pada Sistem Clustering Data Varietas Padi," *Kumpul. J. Ilmu Komput.*, vol. 01, no. 01, pp. 23–32, 2014.
- [5] P. R. Arum and I. M. Nur, "Clustering The Level Of Education Quality In The Central Java Province Using C-Means Method," pp. 508–513, 2020, doi: 10.4108/eai.28-9-2019.2291056.
- [6] D. L. Rahakbauw, F. J. Rianekuay, and Y. A. Lesnussa, "Penerapan Metode Fuzzy Mamdani Untuk Memprediksi Jumlah Produksi Karet (Studi Kasus: Data Persediaan Dan Permintaan Produksi Karet Pada Ptp Nusantara Xiv (Persero) Kebun Awaya, Teluk

Elpaputih, Maluku-Indonesia),” *J. Ilm. Mat. Dan Terap.*, vol. 16, no. 1, pp. 51–59, 2019, doi: 10.22487/2540766x.2019.v16.i1.12764.

- [7] B. P. Statistik, “BADAN PUSAT STATISTIK PROVINSI ACEH.”
- [8] S. Agung Priambodo and A. Zakki Falani, “Pemanfaatan Data Mining Untuk Klasterisasi Potensi Produksi Beras Di Kabupaten Blitar Dengan Menggunakan Metode Fuzzy C-Means,” *Nopember*, vol. 12, no. 2, pp. 30–36, 2020.
- [9] S. Mulyani, B. N. Sari, and A. A. Ridha, “Clustering Productivity of Rice in Karawang Regency Using the Fuzzy C-Means Method,” *Indones. J. Artif. Intell. Data Min.*, vol. 3, no. 2, pp. 103–112, 2020.
- [10] M. Orisa and A. Faisol, “Journal of Computer Networks , Architecture and High Performance Computing Researchers Productivity Level Clustering Based On H-Index and Citation Using The Fuzzy C-Means Algorithm Journal of Computer Networks , Architecture and High Performance Computing,” vol. 5, no. 1, pp. 18–26, 2023.
- [11] S. Solimun and A. A. R. Fernades, “Cluster Analysis on Various Cluster Validity Indexes with Average Linkage Method and Euclidean Distance (Study on Compliant Paying Behavior of Bank X Customers in Indonesia 2021),” *Math. Stat.*, vol. 10, no. 4, pp. 747–753, 2022, doi: 10.13189/ms.2022.100405.
- [12] D. A. I. C. Dewi and D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.
- [13] F. I. Dwiana, W. D. Utami, and A. Hamid, “Implementasi K-Means pada Klasterisasi Jenis Disabilitas,” vol. 6, no. 1, pp. 92–101, 2023.
- [14] M. F. Febrian, “PENERAPAN FUZZY C-MEANS DAN FUZZY TIME SERIES UNTUK ANALISIS TINGKAT NILAI SAHAM BEI TAHUN 2015-2020,” 2021.
- [15] S. Mashfuufah and D. Istiawan, “Penerapan Partition Entropy Index, Partition Coefficient Index dan Xie Beni Index untuk Penentuan Jumlah Klaster Optimal pada Algoritma Fuzzy C-Means dalam Pemetaan Tingkat Kesejahteraan Penduduk Jawa Tengah,” *Proceeding of The URECOL*, pp. 51–60, 2018.