

CLUSTERING WILAYAH DI INDONESIA BERDASARKAN KUALITAS PENDIDIKAN MENGGUNAKAN ALGORITMA FUZZY C-MEANS

Gema Fitria Anwar ¹⁾, Ulfa Khaira ²⁾, Pradita Eko Prasetyo Utomo ³⁾

¹⁾²⁾³⁾ Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Jambi

email: fitriaanwar43@gmail.com ¹⁾, ulfakhaira@unja.ac.id ²⁾, pradita.eko@unja.ac.id ³⁾

INFO ARTIKEL

Riwayat Artikel:

Diterima November, 2025

Revisi November, 2025

Terbit November, 2025

ABSTRAK

Pendidikan merupakan kebutuhan mendasar yang berperan penting dalam membentuk kualitas sumber daya manusia serta menentukan masa depan bangsa. Hasil PISA (*Program for International Student Assessment*) 2018 menunjukkan kualitas pendidikan Indonesia masih memprihatinkan dengan peringkat 74 dan skor 1.146. Meskipun pada PISA 2022 posisi Indonesia meningkat lima peringkat, skor turun menjadi 1.108 dan tetap berada di bawah rata-rata negara peserta. Upaya peningkatan kualitas pendidikan salah satunya dapat dilakukan melalui pemerataan kualitas di setiap provinsi. Penelitian ini bertujuan untuk mengidentifikasi pengelompokan provinsi di Indonesia berdasarkan kualitas pendidikan menggunakan algoritma *Fuzzy C-Means*. Data yang digunakan mencakup enam indikator Pendidikan APK, APM, APS, AMH, RLS, dan HLS pada 34 provinsi selama periode 2022–2024 yang diperoleh dari Badan Pusat Statistik (BPS). Hasil analisis menunjukkan bahwa jumlah *Cluster* optimal setiap tahun adalah dua *Cluster* berdasarkan nilai *Partition Coefficient Index*, yaitu 0.8905128 (2022), 0.8822183 (2023), dan 0.8602231 (2024). Tahun 2022 dan 2023 menunjukkan pola pengelompokan yang konsisten dengan 15 provinsi pada *Cluster* 1 dan 19 provinsi pada *Cluster* 2. Pada tahun 2024 terjadi perubahan komposisi, di mana *Cluster* 1 berkurang menjadi 14 provinsi dan *Cluster* 2 meningkat menjadi 20 provinsi.

Kata Kunci :

Kualitas Pendidikan; *Clustering*; *Fuzzy C-Means*; *Partition Coefficient*;

ABSTRACT

Education is a basic need that plays an important role in shaping the quality of human resources and determining the future of a nation. The results of PISA (*Programme for International Student Assessment*) 2018 show that the quality of education in Indonesia is still concerning, with a ranking of 74 and a score of 1,146. Although in PISA 2022 Indonesia's position improved by five ranks, the score dropped to 1,108 and remained below the average of participating countries. Efforts to improve the quality of education can be done, among other things, through the equalization of quality in each province. This study aims to identify the grouping of provinces in Indonesia based on education quality using the *Fuzzy C-Means* algorithm. The data used includes six education indicators APK, APM, APS, AMH, RLS, and HLS in 34 provinces during the period 2022–2024 obtained from the Central Statistics Agency (BPS). The results of the analysis show that the optimal number of *Clusters* each year is two *Clusters* based on the *Partition Coefficient Index* values, namely 0.8905128 (2022), 0.8822183 (2023), and 0.8602231 (2024). The years 2022 and 2023 show a consistent grouping pattern with 15 provinces in *Cluster* 1 and 19.

Penulis Korespondensi:

Gema Fitria Anwar
Program Studi Sistem Informasi,
Fakultas Sains dan Teknologi,
Universitas Jambi

Email:

fitriaanwar43@gmail.com

Keywords:

Educational Quality; *Clustering*; *Fuzzy C-Means*; *Partition Coefficient*

1. PENDAHULUAN

Pendidikan merupakan kebutuhan mendasar bagi setiap individu untuk menjalani kehidupan sekaligus mempersiapkan kualitas sumber daya manusia di suatu negara dan memiliki peran penting dalam menentukan masa depan sebuah bangsa [1]. Tanpa adanya kualitas pendidikan yang baik, maka akan kecil harapan meningkatnya indeks pembangunan yang bermutu. Selain itu, kualitas pendidikan juga menjadi tolak ukur sebuah negara untuk dijadikan kategori sebagai negara maju [2]. Terkait tentang kualitas Pendidikan di Indonesia, Kementerian Pendidikan dan Kebudayaan menyatakan bahwa hasil PISA 2018 menunjukkan kualitas pendidikan Indonesia berada pada level yang mengkhawatirkan, menempati peringkat 74 dengan total skor 1.146. Pada PISA 2022, posisi Indonesia naik lima peringkat, namun pencapaiannya tetap rendah karena skor turun menjadi 1.108 dan masih berada di bawah rata-rata negara peserta lainnya [3].

Sejalan dengan Rencana Strategis (Renstra) Kemendikbud 2020–2024 yang menitikberatkan pada pengembangan sumber daya manusia melalui peningkatan mutu pendidikan, penelitian ini diarahkan untuk mendukung pemerataan kualitas pendidikan di Indonesia. Penelitian dilakukan dengan menerapkan teknik *clustering* guna mengidentifikasi kondisi pendidikan di tiap provinsi berdasarkan sejumlah indikator pendidikan menggunakan pendekatan data mining. Data mining sendiri merupakan proses eksplorasi dan analisis data berukuran besar untuk menemukan pola atau informasi penting, sehingga menghasilkan wawasan yang lebih mendalam. Proses ini memanfaatkan metode statistik, matematis, dan kecerdasan buatan dalam tahapan yang dikenal sebagai *Knowledge Discovery Process* (KDD) [4].

Penelitian dengan teknik *clustering* yang dilakukan oleh [5] menunjukkan bahwa beberapa indikator pendidikan tingkat SMA, seperti angka partisipasi kasar (APK), angka partisipasi murni (APM), dan angka partisipasi sekolah (APS), memiliki keterkaitan dengan kualitas pendidikan masyarakat di suatu wilayah. Selain itu, BPS Indonesia menetapkan tiga indikator utama dalam aspek pendidikan pada Indeks Pembangunan Manusia (IPM), yaitu rata-rata lama sekolah (RLS), harapan lama sekolah (HLS), dan angka melek huruf (AMH).

Clustering merupakan teknik data mining yang digunakan untuk mengelompokkan data ke dalam beberapa *cluster*, di mana data yang memiliki karakteristik serupa akan tergabung dalam satu *cluster*, sementara data yang memiliki perbedaan ciri akan masuk ke *cluster* yang lain [6]. *Fuzzy C-Means* (FCM) merupakan metode *Non-hierarchical Clustering* yang pertama kali diperkenalkan oleh Dunn (1973) dan kemudian disempurnakan oleh Jim Bezdek pada 1981. Metode ini digunakan untuk mengelompokkan data ke dalam satu atau lebih *cluster* dengan menentukan derajat keanggotaan setiap data pada masing-masing *cluster* [7].

Berdasarkan hasil penelitian [8] metode *Fuzzy C-Means* terbukti lebih akurat dibandingkan *K-Means* dalam membentuk *cluster* provinsi di Indonesia berdasarkan data intensitas bencana alam. *Fuzzy C-Means* (FCM) terbukti lebih akurat dibandingkan *K-Means* dalam penelitian tersebut karena FCM termasuk dalam metode *soft clustering*, yang memungkinkan sebuah objek memiliki derajat keanggotaan di lebih dari satu *cluster*. Berbeda dengan *K-Means*, yang bersifat *hard clustering* dan hanya mengelompokkan objek ke dalam satu *cluster* tertentu secara tegas, FCM mempertimbangkan nilai keanggotaan setiap objek, yang berada dalam rentang 0 hingga 1. Keunggulan ini membuat FCM lebih fleksibel dalam menangkap struktur data yang kompleks, terutama pada data yang memiliki batasan antar *cluster* yang kurang jelas. Metode FCM memiliki hasil *Silhouette coefficient* yang lebih tinggi dengan nilai (0,8403194) dibandingkan dengan *K-Means* (0,8031255) yang menunjukkan kualitas pengelompokan FCM lebih baik.

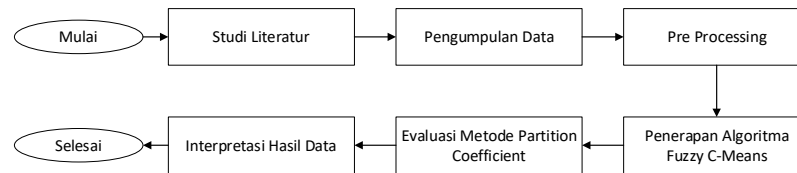
Penelitian serupa mengenai *clustering* menggunakan algoritma *Fuzzy C-Means* dilakukan oleh [2] yang bertujuan mengelompokkan kualitas pendidikan pada tingkat kota/kabupaten di Jawa Timur berdasarkan beberapa parameter, yaitu angka partisipasi kasar, angka partisipasi murni, persentase penduduk yang tamat SMA, dan rata-rata lama sekolah. Hasil penelitian tersebut menunjukkan bahwa dua *cluster* merupakan pengelompokan terbaik dengan nilai *Silhouette coefficient* sebesar 0,65. Selain itu, penelitian oleh [9] juga menerapkan metode *Fuzzy C-Means* untuk mengelompokkan jenis UKM di Rokan Hulu, dan berhasil menghasilkan tiga *cluster* dengan tingkat akurasi mencapai 80–90%.

Berdasarkan temuan-temuan tersebut, peneliti tertarik untuk melakukan penelitian menggunakan algoritma *Fuzzy C-Means* untuk mengelompokkan kualitas pendidikan antarprovinsi di Indonesia berdasarkan indikator pendidikan APK, APM, APS, AMH, RLS dan HLS sebagai dasar dalam proses *clustering*. Seluruh data indikator pendidikan diperoleh dari situs resmi BPS Indonesia.

2. METODOLOGI PENELITIAN

Penelitian ini menerapkan metode *Fuzzy C-Means Clustering* untuk melakukan pengelompokan kualitas pendidikan di Indonesia berdasarkan wilayah provinsi dengan menggunakan sejumlah indikator, yaitu Angka Partisipasi Kasar (SMA), Angka Partisipasi Murni (SMA), Angka Partisipasi Sekolah usia 16–18 tahun, Angka Melek Huruf, Rata-Rata Lama Sekolah, serta Harapan Lama Sekolah.

2.1 Tahapan Penelitian



Gambar 1. Flowchart tahapan penelitian.

2.2 Studi Literatur

Studi literatur dilakukan dengan tujuan untuk mengumpulkan teori dan referensi relevan yang dapat digunakan untuk menyelesaikan permasalahan dalam penelitian ini. Serta berfungsi sebagai bahan perbandingan terhadap hasil penelitian sebelumnya. Tujuan utama dari studi literatur adalah untuk memahami konsep-konsep dasar yang menjadi dasar dari penelitian ini.

2.3 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data yang diperoleh secara tidak langsung oleh peneliti, melainkan sudah ada sebelumnya dari pihak lain, yang dikenal sebagai data sekunder. Dalam penelitian ini data diperoleh dari situs resmi Badan Pusat Statistik (BPS) Indonesia, mencakup berbagai variabel indikator pendidikan dari tahun 2022 hingga 2024. Variabel penelitian adalah suatu konsep yang dapat diukur dan berfungsi untuk merepresentasikan suatu fenomena dalam sebuah penelitian [10]. Penelitian ini akan menggunakan variabel indikator-indikator pendidikan yang berperan dalam mengukur kualitas pendidikan di suatu wilayah meliputi:

1. Angka Partisipasi Kasar adalah perbandingan antara siswa pada jenjang pendidikan SMA/Sederajat dengan penduduk usia sekolah jenjang tersebut [2].
2. Angka Partisipasi Murni adalah perbandingan antara siswa usia sekolah SMA/Sederajat (16-18 tahun) dibagi dengan jumlah penduduk usia 16-18 tahun [2].
3. Angka Partisipasi Sekolah adalah proporsi dari semua anak yang masih sekolah pada umur 16-18 tahun dibagi jumlah penduduk dengan kelompok umur sesuai [2].
4. Angka Melek Huruf merupakan persentase penduduk berusia 15 tahun ke atas yang mampu membaca dan menulis [11].
5. Rata-Rata Lama Sekolah merupakan ukuran terhadap jumlah tahun pendidikan formal yang telah diselesaikan oleh penduduk berusia 25 tahun ke atas [11].
6. Harapan Lama Sekolah merupakan harapan lamanya tahun pendidikan yang akan dijalani oleh seorang anak berusia 7 tahun di masa yang akan datang [11].

Tabel 1. Variabel penelitian.

No	Variabel	Satuan Data	Kode Variabel
1	Angka Partisipasi Kasar (SMA/Sederajat)	Persentase	APK
2	Angka Partisipasi Murni (SMA/Sederajat)	Persentase	APM
3	Angka Partisipasi Sekolah (Umur 16-18 Tahun)	Persentase	APS
4	Angka Melek Huruf	Persentase	AMH
5	Rata-Rata Lama Sekolah	Tahun	RLS
6	Harapan Lama Sekolah	Tahun	HLS

Variabel APK, APM, dan APS dipilih karena berdasarkan penelitian sebelumnya [5], ketiga variabel tersebut memiliki pengaruh terhadap tingkat kualitas pendidikan di suatu daerah. Sementara itu, variabel RLS, HLS, dan AMH digunakan sebagai indikator untuk menilai dimensi pendidikan berdasarkan komponen Indeks Pembangunan Manusia (IPM) menurut BPS [11].

2.4 Pre-processing

Pre-processing data merupakan tahap awal analisis yang mencakup proses pembersihan, pengubahan, dan penyiapan data agar sesuai dengan kebutuhan analisis [12]. Pada penelitian ini, tahapan pra-pemrosesan dilakukan sebagai berikut:

1. Data Cleaning

Tahap ini bertujuan membersihkan data dengan memeriksa keberadaan *missing value* serta data duplikat. Pada penelitian ini dilakukan pengecekan untuk memastikan tidak terdapat nilai hilang maupun data ganda dalam *dataset*.

2. Data Selection

Data Selection dilakukan untuk memilih variabel yang relevan dengan kebutuhan analisis. Penelitian ini menggunakan enam indikator pendidikan, yaitu APK, APM, APS, AMH, RLS, dan HLS.

3. Data Transformation

Transformasi data dilakukan untuk menyesuaikan data dengan kebutuhan analisis, salah satunya melalui normalisasi. Penelitian ini menggunakan metode *min-max normalization* untuk menyamakan skala data agar berada dalam rentang seragam, sehingga memudahkan proses pengolahan dan analisis lebih lanjut, Persamaan (1) digunakan untuk menghitung *Min-Max Normalization* [13].

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Keterangan:

- x = nilai tertentu yang akan dinormalisasi
- x' = nilai hasil normalisasi
- x_{min} = nilai minimal dari sebuah atribut
- x_{max} = nilai maksimal dari sebuah atribut

2.5 Algoritma Fuzzy C-Means

Fuzzy C-Means (FCM) pertama kali dikemukakan oleh Dunn (1973) yang kemudian dikembangkan oleh Jim Bezdek pada tahun 1981, merupakan salah satu metode *Non-hierarchical Clustering* yang digunakan untuk mengelompokkan data ke dalam satu *cluster* atau lebih [7]. Berbeda dengan metode *clustering* biasa, FCM menentukan keanggotaan setiap titik data dalam *cluster* berdasarkan derajat keanggotaan, sehingga satu data dapat termasuk dalam beberapa *cluster* sekaligus dengan tingkat keanggotaan yang berbeda-beda. Adapun algoritma dari metode *Fuzzy C-Means* adalah sebagai berikut [14]:

1. *Input* data yang akan di-*cluster* berupa matriks X berukuran $n \times m$ dimana (n = jumlah sampel data, dan m = atribut setiap data), sedangkan X_{ij} = data sampel ke- i ($i = 1, 2, \dots, n$), dan atribut ke- j ($j = 1, 2, \dots, m$).
2. Menentukan input yang diperlukan dalam menghitung FCM:
 - a. Jumlah *cluster* ($c > 2$)
 - b. Pangkat pembobot ($w > 1$). Menurut Pal dan Bezdek, nilai pangkat pembobot FCM direkomendasikan berada pada rentang m (1.5, 2.5) [15].
 - c. Maksimum iterasi (*MaxIter*)
 - d. *Error* terkecil (ϵ)
 - e. Fungsi objektif ($P_0 = 0$)
 - f. Iterasi awal ($t = 1$)
3. Bangkitkan bilangan random μ_{ik} sebagai elemen matriks partisi awal U .

$$\sum_{j=1}^c \mu_{ik} = 1 \quad (2)$$

4. Hitung pusat *Cluster* ke- k : V_{kj} Dengan $k = 1, 2, \dots, c$ dan $j = 1, 2, \dots, m$.

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n ((\mu_{ik})^w)} \quad (3)$$

dengan:

- V_{kj} = pusat *cluster*
 i = Iterasi
 w = Pangkat pembobot
 μ_{ik} = Elemen matriks partisi awal U ke- i *cluster* ke- k
 X_{ij} = Data sampel ke- i variabel ke- j

5. Hitung fungsi objektif pada iterasi ke- t .

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

6. Hitung perubahan derajat keanggotaan matriks partisi setiap data pada setiap *cluster*.

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{\frac{-1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{\frac{-1}{w-1}}} \quad (5)$$

7. Cek kondisi berhenti,
a. Jika: $(|P_t - P_{t-1}| < \varepsilon)$ maka berhenti;
b. Jika tidak: $t = t + 1$, ulangi langkah ke-4.

2.6 Validitas *Partition Coefficient Index*

Partition Coefficient Index (PCI) merupakan salah satu indeks validitas *cluster* yang digunakan untuk menilai hasil pengelompokan berdasarkan derajat keanggotaan setiap data, terhadap *cluster* yang terbentuk [16]. Nilai PCI berada pada rentang 0 hingga 1, di mana semakin mendekati 1, mengindikasikan bahwa kualitas yang dihasilkan semakin baik. Metode *Partition Coefficient Index* (PCI) didefinisikan dengan persamaan (6) [16].

$$PCI = \frac{1}{n} \sum_{k=1}^c \sum_{i=1}^n \mu_{ik}^2 \quad (6)$$

Keterangan:

- n = Banyak objek penelitian
 c = Banyak *cluster*
 μ_{ik} = Derajat keanggotaan data ke- i pada *cluster* ke- k

2.7 Interpretasi Hasil *Cluster*

Tahapan selanjutnya adalah interpretasi dari hasil *clustering* yang sudah diperoleh. Interpretasi data bertujuan untuk memberikan pemahaman yang lebih menyeluruh terhadap hasil analisis yang telah dilakukan, dengan mengubah data numerik atau statistik menjadi informasi yang lebih informatif dan bermakna. Hasil interpretasi data ini akan ditampilkan dalam bentuk peta visualisasi untuk mempermudah peneliti mengenali pola antar *cluster* serta memahami hasil *clustering* secara lebih jelas mengenai kualitas pendidikan di berbagai provinsi di Indonesia, sehingga dapat ditarik kesimpulan yang logis dan sesuai dengan konteks penelitian yang dilakukan.

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini mencakup indikator APK, APM, APS, AMH, RLS, dan HLS pada 34 provinsi di Indonesia, dengan total 612 data. Seluruh data tersebut digunakan untuk proses pengelompokan provinsi periode 2022–2024 dan diperoleh dari situs resmi Badan Pusat Statistik (BPS).

3.1 Pre-processing

Berikut langkah-langkah persiapan data yang dilakukan dalam penelitian ini sebelum data diolah dan dilakukan analisis lebih lanjut.

1. Data Cleaning

Untuk memeriksa kebersihan data, digunakan fungsi “*is.na()*” untuk mendeteksi adanya nilai yang hilang (*missing values*) serta fungsi “*duplicated()*” untuk mengidentifikasi data yang terduplikasi pada proses pengolahan di RStudio.

2. Data Selection

Setelah proses tersebut, dilakukan seleksi data yaitu proses pemilihan variabel yang relevan untuk digunakan dalam proses analisis. Adapun variabel yang dipilih adalah data yang bertipe numerik, yaitu seluruh data dalam kode variabel indikator pendidikan kecuali variabel Provinsi yang bertipe kategorikal.

3. Data Transformation

Dalam normalisasi *Min-Max*, fungsi “*scale()*” digunakan dengan penyesuaian parameter center dan *scale* sesuai rumus normalisasi *Min-Max*, sementara fungsi “*min()*” dan “*max()*” digunakan untuk mendapatkan nilai minimum dan maksimum dari setiap variabel.

Tabel 2. Data sebelum normalisasi.

NO	APK	APM	APS	AMH	RLS	HLS
1	92.53	71.16	83.1	99.94	9.44	14.37
2	97.23	68.27	78.66	99.92	9.71	13.31
...
33	97.71	63.66	80.56	99.66	7.84	13.21
34	77.06	47.63	65.93	92.04	7.02	11.14

Tabel 3. Data setelah normalisasi

NO	APK	APM	APS	AMH	RLS	HLS
1	0.766877	0.868266	0.714821	0.996217	0.564103	0.716186
2	0.978398	0.761624	0.529975	0.993695	0.62704	0.481153
...
33	1.000000	0.591513	0.609076	0.960908	0.191142	0.458980
34	0.070657	0.000000	0.000000	0.000000	0.000000	0.000000

Tabel 2., menampilkan data indikator pendidikan tahun 2022 sebelum normalisasi, sedangkan Tabel 3., menunjukkan hasil setelah dinormalisasi dengan metode *Min-Max*, yang mengubah skala data ke rentang 0–1. Proses normalisasi yang sama juga diterapkan pada data tahun 2023 dan 2024.

3.2. Algoritma Fuzzy C-Means

Penerapan formula FCM dilakukan pada data tahun 2022 sebagai contoh:

- Langkah awal yang harus dilakukan adalah menginput data hasil normalisasi berjumlah 34 dengan 6 indikator yang digunakan tahun 2022.
- Menentukan input yang diperlukan dalam menghitung FCM:
 - Jumlah *cluster* (c) = 2 (digunakan berdasarkan hasil validasi menggunakan PCI pada Tabel 5).
 - Pangkat (w) = 1.5 (digunakan berdasarkan hasil validasi menggunakan PCI pada Tabel 5).
 - Maksimum iterasi = 1000, yang didasarkan dari hasil penelitian [14].
 - Error* terkecil (ϵ) = 0.001, sebagai kriteria penghentian yang didasarkan dari hasil penelitian [14] dan [17].
 - Fungsi objektif ($P_0 = 0$)
 - Iterasi awal ($t = 1$)
- Bangkitkan bilangan random μ_{ik} dengan ($i = 1, 2, \dots, c$; $k = 1, 2, \dots, n$) yang digunakan sebagai komponen matriks partisi awal U.
- Menghitung pusat *cluster* ke- k : dengan persamaan V_{kj} diperoleh pusat *cluster* yang terbentuk pada iterasi pertama:

Tabel 4. Hasil pusat *cluster* iterasi pertama.

Cluster	Hasil Pusat Cluster V					
	APK	APM	APS	AMH	RLS	HLS
1	0.06390639	0.4889299	0.2568693	0.9949559	1.0000000	0.4301552
2	0.53600360	0.5287823	0.2169026	0.9911728	0.2703963	0.3547672

5. Menghitung fungsi objektif pada iterasi pertama:

$$P_t = \sum_{i=1}^{34} \sum_{k=1}^2 ([\sum_{j=1}^{1.5} X_{ij} - V_{kj}]^2) (\mu_{ik})^{1.5} = 6.406367$$

6. Menghitung perubahan derajat keanggotaan matriks partisi setiap data pada setiap *Cluster* melalui persamaan (7)

$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}} \quad (7)$$

7. Cek kondisi berhenti:

Karena $|P_1 - P_0| = |6.406367 - 0| = 6.406367$ artinya selisih antara P_1 dan P_0 masih lebih besar dari batas error ($P_1 - P_0 > \text{nilai error } 0.001$) maka proses dilanjutkan pada iterasi ke-2, 3, ..., n hingga nilai *error* stabil atau berada di bawah 0.001. Hasil iterasi maksimal yang diperoleh dengan bantuan *tools* RStudio diperlukan 10 iterasi untuk memenuhi kondisi berhenti, dengan nilai fungsi objektif akhir sebesar 4.635865.

3.3. Validitas *Partition Coefficient Index*

Perhitungan untuk evaluasi *cluster* terbaik dilakukan dengan jumlah *cluster* mulai dari 2 hingga 5, dengan variasi pangkat pembobot 1.5, 2, dan 2.5 menggunakan RStudio. Selanjutnya, dari setiap hasil pengelompokan, dipilih *cluster* yang paling optimal berdasarkan nilai *Partition Coefficient Index* (PCI). Nilai PCI berada pada rentang 0 hingga 1, dimana semakin mendekati 1, mengindikasikan bahwa kualitas yang dihasilkan semakin baik. Tabel 5., menampilkan hasil optimal perhitungan PCI di RStudio berdasarkan data masing-masing tahun.

Tabel 5. *Partition Coefficient Index* data tahun 2022.

<i>Cluster</i>	Pangkat Pembobot (m)	<i>Partition Entropy Index</i>
2	1.5	0,8905128
	2	0,7563037
	2.5	0,6591485
3	1.5	0,8009625
	2	0,6027081
	2.5	0,4810454
4	1.5	0,7722352
	2	0,5327229
	2.5	0,4065894
5	1.5	0,7752177
	2	0,5210971
	2.5	0,3756385

Tabel 6. *Partition Coefficient Index* data tahun 2023.

<i>Cluster</i>	Pangkat Pembobot (m)	<i>Partition Entropy Index</i>
2	1.5	0,8822183
	2	0,8822183
	2.5	0,7517556
3	1.5	0,6563197
	2	0,8822183
	2.5	0,7517556
4	1.5	0,6563197
	2	0,8026299
	2.5	0,6167642
5	1.5	0,4935232
	2	0,7740387
	2.5	0,5376639

Tabel 7. *Partition Coefficient Index* data tahun 2024.

<i>Cluster</i>	<i>Pangkat Pembobot (m)</i>	<i>Partition Entropy Index</i>
2	1.5	0,8602231
	2	0,7153300
	2.5	0,6231952
3	1.5	0,8018588
	2	0,5873155
	2.5	0,4651989
4	1.5	0,7655700
	2	0,5171910
	2.5	0,3844866
5	1.5	0,7702118
	2	0,4368202
	2.5	0,3200901

Berdasarkan hasil perhitungan *Partition Coefficient Index* (PCI), hasil terbaik diperoleh pada jumlah *cluster* 2 dengan pangkat pembobot 1.5. Hal ini ditunjukkan oleh nilai PC tertinggi sebesar 0.8822183 untuk tahun 2022, 0.8822183 untuk tahun 2023 dan 0,8602231 untuk tahun 2024. Nilai PC yang tinggi mencerminkan kekompakan yang baik dalam masing-masing *cluster*. Dengan demikian, *cluster* 2 memberikan kualitas pengelompokan yang paling optimal untuk data tahun 2022 hingga 2024.

Tabel 8. *Clustering Fuzzy C-Means* 2022.

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

Tabel 9. *Clustering Fuzzy C-Means* 2023.

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

Tabel 10. *Clustering Fuzzy C-Means* 2024.

<i>Cluster</i>	<i>Jumlah Anggota Cluster</i>	<i>Provinsi Anggota Cluster</i>
1	14	Aceh, Sumatera Utara, Sumatera Barat, Riau, Bengkulu, Kepulauan Riau, Daerah Istimewa Yogyakarta, Bali, Nusa Tenggara Barat, Kalimantan Timur, Kalimantan Utara, Sulawesi Tenggara, Maluku, Maluku Utara.
2	20	Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Gorontalo, Sulawesi Barat, Papua Barat, Papua.

3.4. Interpretasi Hasil *Fuzzy C-Means Clustering*

Untuk memperjelas distribusi hasil pengelompokan, visualisasi *cluster* disajikan dalam bentuk peta. Pada peta, warna hijau merepresentasikan *cluster* 1 yang menunjukkan kelompok provinsi dengan kualitas pendidikan tinggi, sedangkan warna merah menunjukkan *cluster* 2 dengan kualitas pendidikan rendah. Penerapan pola warna ini dilakukan secara konsisten pada peta tahun 2022, 2023, dan 2024 sehingga memudahkan dalam menganalisis perubahan kualitas pendidikan antarwilayah dari waktu ke waktu. Berikut merupakan interpretasi hasil pengelompokan kualitas pendidikan di Indonesia menggunakan algoritma *Fuzzy C-Means*.

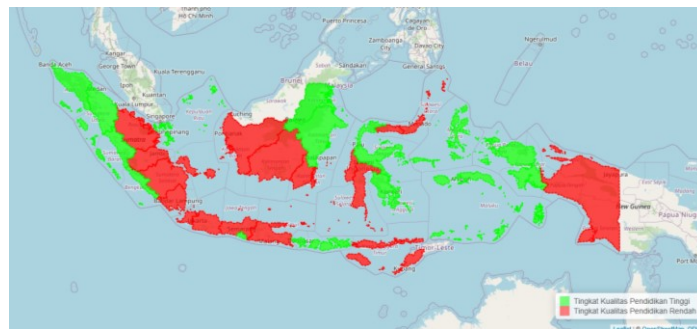
3.4.1. Interpretasi Data Tahun 2022

Proses *clusterisasi* selesai pada iterasi ke-11 dengan nilai fungsi objektif akhir 4.635865. *Cluster* 1 terdiri dari 15 provinsi, sementara *cluster* 2 memuat 19 provinsi. Tabel 11., menampilkan pusat akhir dari masing-masing *cluster* hasil *Fuzzy C-Means*.

Tabel 11. Hasil pusat *cluster* setelah konvergensi data tahun 2022.

<i>Cluster</i>	APK	APM	APS	AMH	RLS	HLS
<i>Cluster</i> 1	0.7741649	0.7513298	0.5989487	0.9817824	0.5329409	0.5741204
<i>Cluster</i> 2	0.3847312	0.4251668	0.2063975	0.9432107	0.3290319	0.3787236

Hasil perhitungan *centroid*, menunjukkan bahwa *cluster* 1 mempunyai nilai yang relatif lebih tinggi pada seluruh indikator pendidikan jika dibandingkan dengan nilai pada *cluster* 2. Kondisi ini menunjukkan bahwa *cluster* 1 menggambarkan wilayah dengan tingkat kualitas pendidikan yang tinggi. Sementara itu, *cluster* 2 menggambarkan wilayah dengan tingkat kualitas pendidikan yang rendah.



Gambar 2. Pemetaan *cluster* kualitas pendidikan Indonesia tahun 2022.

Tahun 2022, *cluster* 1 yang digambarkan dengan warna hijau merepresentasikan provinsi-provinsi dengan kualitas pendidikan tinggi mencakup 15 provinsi yaitu: Aceh, Sumatera Utara, Sumatera Barat, Bengkulu, Kepulauan Riau, Daerah Istimewa Yogyakarta, Bali, Nusa Tenggara Barat, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Tenggara, Maluku, Maluku Utara, Papua Barat. Sementara itu, *cluster* 2 yang ditandai warna merah merupakan wilayah dengan kualitas pendidikan yang lebih rendah mencakup 19 provinsi yaitu: Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Selatan, Gorontalo, Sulawesi Barat, Papua.

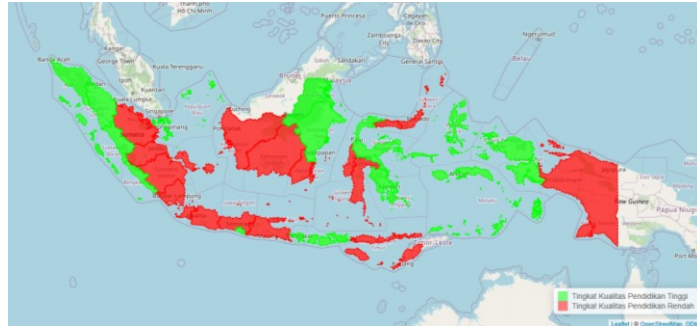
3.4.2. Interpretasi Data Tahun 2023

Proses *clusterisasi* selesai pada iterasi ke-15 dengan nilai fungsi objektif akhir 4.428688. *Cluster* 1 terdiri dari 15 provinsi, sementara *cluster* 2 memuat 19 provinsi. Tabel 12., menampilkan pusat akhir dari masing-masing *cluster* hasil *Fuzzy C-Means*.

Tabel 12. Hasil pusat *cluster* setelah konvergensi data tahun 2023.

<i>Cluster</i>	APK	APM	APS	AMH	RLS	HLS
<i>Cluster</i> 1	0.8112601	0.7328139	0.6191803	0.9786836	0.5246800	0.5871070
<i>Cluster</i> 2	0.4698151	0.4159186	0.2618702	0.9439638	0.3205602	0.3880046

Hasil perhitungan *centroid*, menunjukkan bahwa *cluster* 1 mempunyai nilai yang relatif lebih tinggi pada seluruh indikator pendidikan jika dibandingkan dengan nilai pada *cluster* 2. Kondisi ini menunjukkan bahwa *cluster* 1 menggambarkan wilayah dengan tingkat kualitas pendidikan yang tinggi. Sementara itu, *cluster* 2 menggambarkan wilayah dengan tingkat kualitas pendidikan yang rendah.



Gambar 3. Pemetaan *cluster* kualitas pendidikan Indonesia tahun 2023.

Tahun 2023, *cluster* 1 yang digambarkan dengan warna hijau merepresentasikan provinsi-provinsi dengan kualitas pendidikan tinggi mencakup 15 provinsi yaitu: Aceh, Sumatera Utara, Sumatera Barat, Bengkulu, Kepulauan Riau, Daerah Istimewa Yogyakarta, Bali, Nusa Tenggara Barat, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Sulawesi Tenggara, Maluku, Maluku Utara, Papua Barat. Sementara itu, *cluster* 2 yang ditandai warna merah merupakan wilayah dengan kualitas pendidikan yang lebih rendah mencakup 19 provinsi yaitu: Riau, Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Selatan, Gorontalo, Sulawesi Barat, Papua.

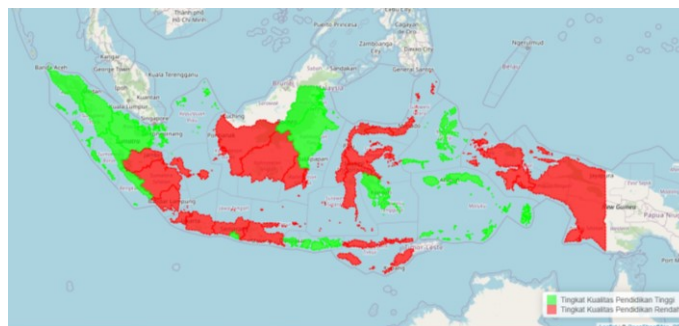
3.4.3. Interpretasi Data Tahun 2024

Proses *clusterisasi* selesai pada iterasi ke-11 dengan nilai fungsi objektif akhir 6.091311. *Cluster* 1 terdiri dari 14 provinsi, sementara *cluster* 2 memuat 20 provinsi. Tabel 13., menampilkan pusat akhir dari masing-masing *cluster* hasil Fuzzy C-Means.

Tabel 13. Hasil pusat *cluster* setelah konvergensi data tahun 2024.

Cluster	APK	APM	APS	AMH	RLS	HLS
Cluster 1	0.7741649	0.7513298	0.5989487	0.9817824	0.5329409	0.5741204
Cluster 2	0.3847312	0.4251668	0.2063975	0.9432107	0.3290319	0.3787236

Hasil perhitungan *centroid*, menunjukkan bahwa *cluster* 1 mempunyai nilai yang relatif lebih tinggi pada seluruh indikator pendidikan jika dibandingkan dengan nilai pada *cluster* 2. Kondisi ini menunjukkan bahwa *cluster* 1 menggambarkan wilayah dengan tingkat kualitas pendidikan yang tinggi. Sementara itu, *cluster* 2 menggambarkan wilayah dengan tingkat kualitas pendidikan yang rendah.



Gambar 4. Pemetaan *cluster* kualitas pendidikan Indonesia tahun 2024.

Tahun 2024, *cluster* 1 yang digambarkan dengan warna hijau merepresentasikan provinsi-provinsi dengan kualitas pendidikan tinggi mencakup 14 provinsi yaitu: Aceh, Sumatera Utara, Sumatera Barat, Riau, Bengkulu, Kepulauan Riau, Daerah Istimewa Yogyakarta, Bali, Nusa Tenggara Barat, Kalimantan Timur, Kalimantan Utara, Sulawesi Tenggara, Maluku, Maluku Utara. Sementara itu, *cluster* 2 yang ditandai warna merah merupakan wilayah dengan kualitas pendidikan yang lebih rendah mencakup 20 provinsi yaitu: Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Gorontalo, Sulawesi Barat, Papua Barat, Papua.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa jumlah *cluster* yang paling tepat untuk mengelompokkan provinsi berdasarkan tingkat kualitas pendidikan pada periode 2022 hingga 2024 adalah 2 *cluster*. Penetapan ini diperoleh melalui penerapan algoritma *Fuzzy C-Means*, yang dievaluasi menggunakan *Partition Coefficient Index* yang menunjukkan bahwa penggunaan 2 *cluster* dengan pangkat pembobot 1.5 memberikan performa yang paling optimal. Berdasarkan hasil *clustering* tahun 2022, 2023, dan 2024, *cluster* 1 secara konsisten merepresentasikan wilayah dengan kualitas pendidikan tinggi, sedangkan *cluster* 2 menunjukkan wilayah dengan kualitas pendidikan rendah.

Pada tahun 2022 dan 2023, jumlah anggota *cluster* tidak berubah: *cluster* 1 berisi 15 provinsi dan *cluster* 2 mencakup 19 provinsi. Namun, pada tahun 2024 terjadi pergeseran komposisi, di mana *cluster* 1 berkurang menjadi 14 provinsi dan *cluster* 2 meningkat menjadi 20 provinsi. Secara umum, sejumlah provinsi tetap stabil berada dalam zona hijau selama tiga tahun berturut-turut, yaitu Aceh, Sumatera Utara, Sumatera Barat, Bengkulu, Kepulauan Riau, DI Yogyakarta, Bali, Nusa Tenggara Barat, Kalimantan Timur, Kalimantan Utara, Sulawesi Tenggara, Maluku, dan Maluku Utara. Sebaliknya, provinsi yang konsisten berada pada zona merah antara lain Jambi, Sumatera Selatan, Lampung, Kepulauan Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Banten, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Sulawesi Utara, Sulawesi Selatan, Gorontalo, Sulawesi Barat, dan Papua. Beberapa wilayah mengalami perubahan kategori, seperti Sulawesi Tengah dan Papua Barat yang berpindah dari zona hijau pada 2023 menjadi zona merah pada 2024. Sebaliknya, Provinsi Riau menunjukkan peningkatan dengan bergeser dari zona merah pada 2023 menjadi zona hijau pada 2024.

Dari hasil analisis terlihat bahwa beberapa provinsi secara konsisten berada pada *cluster* 2 dengan kategori rendah dari tahun 2022 hingga 2024. Kondisi ini mengindikasikan bahwa upaya pemerataan pendidikan oleh pemerintah masih belum optimal karena pada tahun 2024 jumlah provinsi yang masuk kedalam *cluster* 2 bertambah. Oleh karena itu, pemerintah perlu memberikan perhatian khusus untuk meningkatkan kualitas pendidikan di provinsi-provinsi yang terus berada dalam kategori tersebut. Intervensi strategis seperti penyediaan program beasiswa, peningkatan akses transportasi menuju sekolah, serta penguatan kemampuan literasi dan numerasi perlu menjadi prioritas guna mendorong perbaikan pada indikator-indikator yang selama ini menunjukkan nilai terendah. Selain itu, kolaborasi antara pemerintah pusat dan daerah dalam penerapan kebijakan merdeka belajar juga perlu diperkuat agar peningkatan mutu pendidikan dapat dirasakan secara merata di seluruh wilayah Indonesia.

Sebagai upaya pengembangan penelitian di masa mendatang, penulis menyarankan beberapa hal, antara lain memperluas variabel indikator yang digunakan untuk mengukur kualitas pendidikan provinsi di Indonesia, menerapkan teknik klasterisasi lain seperti DBSCAN atau *K-Medoids* untuk membandingkan hasil analisis, serta menggunakan metode evaluasi tambahan guna menilai tingkat optimalitas *cluster* yang terbentuk.

DAFTAR PUSTAKA

- [1] Z. I. Alfianti, "Pengelompokan Tingkat Pendidikan Berdasarkan Jumlah Sekolah Di Provinsi Jawa Barat Menggunakan Algoritma K-Means," *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, 2022, [Online]. Available: <https://referensi.data.kemdikbud.go.id>
- [2] N. Ulinnuha and D. C. R. Novitasari, "Penerapan Fuzzy C-Means Untuk Pengelompokan Tingkat Kualitas Pendidikan Di Jawa Timur," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 14, no. 2, pp. 419–426, 2023, doi: 10.24176/simet.v14i2.9442.
- [3] Kementerian Pendidikan Dan Kebudayaan, "Kemendikbudristek Harap Skor PISA Indonesia Segera Membaik," Pusat Asesmen Pendidikan, Aug. 03, 2023. [Online]. Available: <https://pisa2025.id/berita/read/pisa-di-indonesia/2/kemendikbudristek-harap-skor-pisa-indonesia-segera-membaik.html>
- [4] P. Rahayu et al., *Buku Ajar Data Mining*, vol. 1, no. January 2024. 2024.
- [5] I. G. Lenama, Nony; Kleden, Maria A.; Pasangka, "K-MEANS CLUSTERING ANALYSIS PADA PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI NUSA TENGGARA TIMUR BERDASARKAN INDIKATOR PENDIDIKAN," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 23, pp. 656–664, 2023, doi: <https://doi.org/10.53625/jcijurnalcakrawalailmiah.v2i9.5653>.

- [6] R. R. Prianggono, Analisis Konvergensi Intuitionistic Fuzzy C-Means Clustering Pada Machine Learning Dataset, vol. 75, no. 17. Universitas Brawijaya, 2021. [Online]. Available: https://repository.ub.ac.id/id/eprint/184880/1/-RIFQI_RIVANA_PRIANGGONO.pdf
- [7] F. T. Permadi, A. Nilogiri, and U. A. Rosyidah, "Algoritma Fuzzy C-Means dan Metode Davies Bouldin Index (DBI) untuk Mengelompokkan Kabupaten / Kota di Jawa Timur Berdasarkan Sanitasi Layak, Air Minum Layak, dan Rumah Layak Huni," Jurnal Smart Teknologi, vol. 1, no. 1, pp. 1–10, 2021, [Online]. Available: <https://jurnal.unmuhjember.ac.id/index.php/JST/article/view/16230>
- [8] H. Firdaus, "ANALISA CLUSTER MENGGUNAKAN K-MEANS DAN FUZZY C-MEANS DALAM PENGELOMPOKAN PROVINSI MENURUT DATA INTESITAS BENCANA ALAM DI INDONESIA TAHUN 2017-2021," MATHunesa Jurnal, vol. 10, no. 01, pp. 50–60, 2022, doi: <https://doi.org/10.26740/mathunesa.v10n1.p50-60>.
- [9] L. F. Erni Rouza, "Implementasi Fuzzy C-Means Clustering dalam Pengelompokan UKM Di Kabupaten Rokan Hulu," Techno.COM, vol. 19, no. 4, pp. 481–495, 2020, doi: <https://doi.org/10.33633/tc.v19i4.4101>.
- [10] O. Marlina Susianti, "Perumusan Variabel Dan Indikator Dalam Penelitian Kuantitatif Kependidikan," Jurnal Pendidikan Rokania, vol. 9, p. 18, 2024, doi: <https://doi.org/10.37728/jpr.v9i1.1066>.
- [11] BPS, "Indeks Pembangunan Manusia 2023," Badan Pusat Statistik, vol. 18, pp. 1–282, 2023.
- [12] F. Marisa, "Educational Data Mining (Konsep Dan Penerapan)," Jurnal Teknologi Informasi, vol. 4, no. 35, pp. 90–97, 2020, doi: 10.36382/jti-tki.v4i2.108.
- [13] A. Nurzida, I. T. Utami, and M. Y. Rochayani, "Perbandingan metode fuzzy c-means dan gustafson-kessel dalam penentuan cluster tingkat risiko penularan tuberculosis terhadap penyakit di jawa timur 1,2,3," vol. 13, pp. 373–382, 2024, doi: 10.14710/j.gauss.13.2.373-382.
- [14] I. H. Zahro, U. A. Rosyidah, and L. Handayani, "Implementasi Algoritma Fuzzy C-Means untuk Pengelompokan Provinsi di Indonesia Berdasarkan Kualitas Perguruan Tinggi," BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer, vol. 5, no. 1, pp. 80–86, 2024, doi: 10.37148/bios.v5i1.102.
- [15] C. H. Tambuwun, Y. A. R. Langi, and A. J. Rindengan, "Estimasi Bobot Parameter M Pada Fuzzy C-Means Menggunakan Analisis Robust Dengan Simulasi Data Spasial," d'CARTESIAN, vol. 9, no. 1, p. 50, 2020, doi: 10.35799/dc.9.1.2020.27600.
- [16] A. Ramadhan and Z. Efendi, "Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling," pp. 219–226, 2017, [Online]. Available: <https://ejournal.uin-suska.ac.id/index.php/SNTIKI/article/view/3268/2156>
- [17] P. F. Farradilla, "ALGORITMA FUZZY C-MEANS DAN CLASSIFICATION AND REGRESSION TREES UNTUK PENGELOMPOKAN DAN PENGKLASIFIKASIAN KOLEKTIBILITAS CALON DEBITUR," 2019. [Online]. Available: http://scioteca.caf.com/bitstream/handle/123456789/1091/RED2017-Eng-8ene.pdf?sequence=12&isAllowed=y%0Ahttp://dx.doi.org/10.1016/j.regsciurbeco.2008.06.005%0Ahttps://www.researchgate.net/publication/305320484_SISTEM_PEMBETUNGAN_TERPUSAT_STRATEGI_MELESTARI