

CLUSTERING KUALITAS PENDIDIKAN DI JAWA TIMUR DENGAN METODE FUZZY C-MEANS DAN INDEKS VALIDITAS PARTITION COEFFICIENT

Elizabeth Olivia Hermawan

Program Studi Matematika, FMIPA, Universitas Negeri Surabaya, Surabaya, Indonesia

e-mail : elizabetholivia.22036@mhs.unesa.ac.id

Raden Sulaiman

Program Studi Matematika, FMIPA, Universitas Negeri Surabaya, Surabaya, Indonesia

e-mail : radensulaiman@unesa.ac.id*

Abstrak

Penelitian ini bertujuan untuk mengelompokkan kualitas pendidikan kabupaten/kota di Jawa Timur menggunakan metode *Fuzzy C-Means* (FCM) serta menentukan jumlah *cluster* optimal dengan indeks validitas *Partition Coefficient* (PC). Data yang digunakan bersumber dari Badan Pusat Statistik Provinsi Jawa Timur dan mencakup indikator rata-rata lama sekolah, tingkat pendidikan tertinggi yang ditamatkan, angka partisipasi pendidikan, angka melek huruf, serta rasio murid terhadap guru. Pengujian dilakukan dengan variasi 2 hingga 5 *cluster*, dan hasil evaluasi menunjukkan bahwa dua *cluster* memberikan nilai PC tertinggi sehingga ditetapkan sebagai jumlah *cluster* optimal.

Hasil pengelompokan membagi kabupaten/kota menjadi dua kelompok, yaitu wilayah dengan dominasi kualitas pendidikan dasar dan wilayah dengan dominasi kualitas pendidikan menengah, dengan masing-masing terdiri atas 21 dan 17 kabupaten/kota. Perbedaan antar*cluster* terutama dipengaruhi oleh rata-rata lama sekolah dan capaian pendidikan menengah. Secara keseluruhan, metode FCM dengan evaluasi PC mampu menghasilkan pengelompokan yang jelas dan relevan sebagai dasar perumusan kebijakan pendidikan di Jawa Timur.

Kata Kunci: kualitas pendidikan, *Fuzzy C-Means*, *Partition Coefficient*, clustering, Jawa Timur.

Abstract

This research aims to classify the quality of education in districts/cities in East Java using the Fuzzy C-Means (FCM) method and determine the optimal number of clusters with the Partition Coefficient (PC) validity index. The data used is sourced from the East Java Provincial Statistics Agency and includes indicators such as average length of schooling, highest level of education completed, education participation rate, literacy rate, and student-teacher ratio. Testing was conducted with variations from 2 to 5 clusters, and the evaluation results showed that two clusters provided the highest PC value, thus determining the optimal number of clusters.

The clustering results divided the districts/cities into two groups, namely regions dominated by basic education quality and regions dominated by secondary education quality, with each consisting of 21 and 17 districts/cities, respectively. The differences between clusters were mainly influenced by the average length of schooling and secondary education attainment. Overall, the FCM method with PC evaluation was able to produce clear and relevant groupings as a basis for formulating more targeted education policies in East Java Province.

Keywords: education quality, *Fuzzy C-Means*, *Partition Coefficient*, clustering, East Java.

PENDAHULUAN

Pendidikan merupakan fondasi esensial dalam pembangunan bangsa, karena berperan penting dalam upaya peningkatan kualitas sumber daya manusia. Kualitas suatu negara dapat tercermin dari seberapa baik kualitas pendidikan yang diselenggarakan (Alodia dkk., 2021). Namun, hingga kini kualitas pendidikan di Indonesia masih tergolong rendah bila dibandingkan dengan negara lain. Berdasarkan hasil *Programme for International*

Student Assessment (PISA) tahun 2022 yang dirilis oleh *Organisation for Economic Co-operation and Development* (OECD), Indonesia menempati peringkat ke-69 dari 81 negara peserta. Posisi ini menunjukkan bahwa kualitas pembelajaran siswa Indonesia masih berada di bawah rata-rata negara anggota OECD (OECD, 2023). Kondisi ini mengindikasikan bahwa pemerataan akses serta kualitas pendidikan masih menjadi tantangan besar bagi Indonesia.

Permasalahan pemerataan pendidikan tidak hanya terjadi pada skala nasional, tetapi juga muncul secara nyata di berbagai daerah Indonesia. Dalam salah satu publikasinya, Kementerian Sekretariat Negara menekankan bahwa banyak wilayah terpencil di Indonesia masih belum memperoleh kualitas pendidikan yang setara dengan daerah perkotaan. Beragam tantangan seperti kondisi geografis yang sulit diakses, kurangnya tenaga pendidik yang memadai, keterbatasan fasilitas belajar, serta akses teknologi yang belum merata menjadi faktor yang memperlebar kesenjangan tersebut. Situasi ini menunjukkan bahwa pemerataan kualitas pendidikan masih menghadapi tantangan besar, sehingga diperlukan langkah kebijakan yang lebih menyeluruh agar seluruh daerah dapat menikmati pendidikan yang layak dan berkualitas.

Salah satu provinsi yang merepresentasikan tantangan pemerataan pendidikan di Indonesia adalah Provinsi Jawa Timur. Berdasarkan data Badan Pusat Statistik (BPS) tahun 2024, Provinsi Jawa Timur memiliki jumlah penduduk sebanyak 41,81 juta jiwa, menempati posisi kedua dengan jumlah penduduk terbanyak di Indonesia setelah Jawa Barat. Jumlah penduduk yang besar dan tingkat heterogenitas sosial ekonomi yang tinggi menyebabkan perbedaan dalam akses, sarana, dan mutu pendidikan antarwilayah, sehingga muncul kesenjangan capaian pendidikan di berbagai kabupaten/kota di provinsi ini. Selain itu, berdasarkan data Statistik Pendidikan Indonesia Tahun 2024 yang dipublikasikan oleh BPS, indikator Rata-rata Lama Sekolah (RLS) Provinsi Jawa Timur memiliki capaian sebesar 8,65 tahun, dimana nilai ini lebih rendah dibandingkan rata-rata nasional yang mencapai 9,22 tahun. Kondisi ini menunjukkan bahwa kualitas pendidikan di Jawa Timur secara umum masih berada di bawah standar nasional, sehingga analisis mengenai pola kualitas pendidikan antar daerah di provinsi ini menjadi relevan untuk mengidentifikasi faktor ketimpangan sekaligus peluang peningkatan kualitas pendidikan di Jawa Timur.

Data Badan Pusat Statistik Provinsi Jawa Timur tahun 2024 menunjukkan adanya ketimpangan (disparitas) kualitas pendidikan yang signifikan antarwilayah. Disparitas tersebut tercermin dari berbagai indikator kualitas pendidikan. Rata-rata Lama Sekolah (RLS) di Kota Madiun tahun 2024 mencapai 12,11 tahun, sedangkan di Kabupaten

Sampang hanya 5,08 tahun. Perbedaan ini berarti rata-rata masyarakat Sampang hanya menamatkan setara SD, sementara masyarakat Surabaya sudah hampir menyelesaikan pendidikan menengah atas. Pada indikator tingkat pendidikan tertinggi yang ditamatkan, Kabupaten Sidoarjo mencatat 42,38% penduduk menamatkan pendidikan hingga jenjang SMA dan hanya 13,83% berhenti di jenjang SD, sedangkan di Kabupaten Bangkalan, hanya 12,4% yang menamatkan SMA dan 39,65% penduduk berhenti di jenjang SD. Perbedaan lainnya tampak pada rasio murid terhadap guru, di Kabupaten Sumenep satu guru rata-rata mengajar sembilan murid, sedangkan di Kota Probolinggo satu guru mengajar hingga 17 murid. Disparitas juga tercermin dalam Indeks Pembangunan Manusia (IPM), dimana Kota Surabaya berada pada angka 84,69 (sangat tinggi), sedangkan Sampang hanya 66,72 (rendah). Data-data tersebut merepresentasikan adanya perbedaan pada capaian dan kualitas pendidikan antar beberapa wilayah di Jawa Timur.

Untuk memahami disparitas tersebut, beberapa penelitian terdahulu telah menerapkan pendekatan pengelompokan (*clustering*) kualitas pendidikan antar daerah di Jawa Timur. Penelitian tersebut adalah penelitian oleh Jayuning, Kusumaningtyas, dan Basofi (2019) dimana dalam penelitiannya mengenai pemetaan pendidikan SMA sederajat menggunakan metode *Fuzzy Sugeno* dan membentuk tiga kelompok daerah berdasarkan tingkat kualitas pendidikan. Sementara itu, Maulana, Ulinnuha, dan Novitasari (2023) dengan metode *Fuzzy C-Means* membagi kualitas pendidikan Jawa Timur menjadi dua *cluster* besar, yakni 14 daerah dengan kualitas tinggi dan 24 daerah dengan kualitas rendah. Hasil dari penelitian-penelitian tersebut menunjukkan bahwa penerapan metode *clustering* efektif dalam merepresentasikan perbedaan dan ketimpangan kualitas pendidikan antar wilayah. Namun, fokus penelitian sebelumnya umumnya terbatas pada jenjang pendidikan menengah atas (SMA/sederajat) serta perbedaan jumlah *cluster* pada penelitian-penelitian tersebut menunjukkan belum adanya konsistensi terkait pola pengelompokan kualitas pendidikan di Jawa Timur. Ketidakkonsistenan jumlah *cluster* ini mengindikasikan bahwa jumlah *cluster* yang paling tepat dan representatif terhadap kondisi kualitas pendidikan di Jawa Timur masih belum teridentifikasi.

Melihat keterbatasan penelitian terdahulu, diperlukan kajian yang lebih komprehensif dengan cakupan indikator dan jenjang pendidikan yang lebih luas. Dalam penelitian ini, analisis diperluas hingga mencakup jenjang SD, SMP, dan SMA, karena kualitas pendidikan yang ditinjau tidak dibatasi pada satu jenjang tertentu, melainkan dipahami sebagai kondisi pendidikan secara menyeluruh yang terbentuk dari keterkaitan dan kesinambungan antarjenjang pendidikan dasar hingga menengah. Untuk menentukan jumlah *cluster* kualitas pendidikan yang paling optimal, penelitian ini menerapkan indeks validitas *clustering* yang memberikan ukuran matematis terhadap kejelasan pemisahan antar*cluster*, sehingga jumlah *cluster* yang dipilih dapat merepresentasikan pola data secara lebih akurat.

Dalam penentuan jumlah *cluster* optimal, berbagai indeks validitas telah banyak digunakan, di antaranya *Partition Entropy* (PE), *Xie-Beni Index* (XB), dan metode *Elbow*. *Partition Entropy* mengukur tingkat ketidakpastian dalam pembagian *cluster*, dimana nilai yang lebih rendah menunjukkan struktur *cluster* yang lebih baik. Namun, indeks ini cenderung sensitif terhadap jumlah *cluster* dan sering merekomendasikan jumlah *cluster* yang terlalu besar, sehingga hasilnya kurang stabil pada data dengan variasi derajat keanggotaan yang tinggi. *Xie-Beni Index* mengevaluasi kualitas *clustering* berdasarkan tingkat kekompakan dalam *cluster* dan tingkat pemisahan antar pusat *cluster*, dimana nilai yang lebih kecil menunjukkan hasil *clustering* yang lebih baik. Meskipun demikian, indeks ini sangat sensitif terhadap keberadaan *outlier* serta asumsi bentuk *cluster* yang cenderung bulat (*spherical*), sehingga kurang optimal ketika diterapkan pada data dengan distribusi yang kompleks. Metode *Elbow* merupakan pendekatan visual yang menentukan jumlah *cluster* optimal berdasarkan titik kelengkungan pada grafik *within-cluster sum of squares* (WCSS). Akan tetapi, metode ini bersifat subjektif karena penentuan titik “siku” sangat bergantung pada interpretasi peneliti, sehingga kurang sesuai untuk analisis yang memerlukan kriteria evaluasi yang konsisten dan berbasis matematis. Berdasarkan keterbatasan tersebut, penelitian ini menggunakan indeks validitas *Partition Coefficient* (PC). *Partition Coefficient* merupakan indeks validitas pada *fuzzy clustering* yang mengukur tingkat kejelasan pembagian *cluster*

berdasarkan derajat keanggotaan data terhadap masing-masing *cluster*. Nilai PC yang semakin tinggi menunjukkan tingkat *overlapping* antar *cluster* yang semakin rendah, sehingga kualitas hasil *clustering* yang diperoleh menjadi semakin baik dan jumlah *cluster* yang dipilih lebih merepresentasikan pola data yang terbentuk.

Selain itu, penelitian ini menggunakan metode *Fuzzy C-Means* (FCM) sebagai metode *clustering* karena FCM memiliki keunggulan yaitu dapat memberikan fleksibilitas derajat keanggotaan, sehingga satu daerah dapat memiliki kedekatan dengan lebih dari satu *cluster* (Ulinnuha dan Novitasari, 2023). Pendekatan ini memberikan representasi yang lebih realistis terhadap kondisi pendidikan yang kompleks dan tidak sepenuhnya terpisah antarwilayah. Dengan demikian, *Fuzzy C-Means* menjadi metode yang tepat untuk menganalisis data pada penelitian ini karena mampu menangani ketidakpastian dan memberikan hasil pengelompokan yang fleksibel dan akurat (Jipkate dan Gohokar, 2012).

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengelompokkan kualitas pendidikan di Jawa Timur dengan metode *Fuzzy C-Means* menggunakan indikator kualitas pendidikan dari jenjang SD, SMP, hingga SMA serta indeks validitas PC untuk menentukan jumlah *cluster* yang paling optimal. Hasil *clustering* diharapkan dapat memberikan gambaran yang lebih akurat mengenai disparitas kualitas pendidikan antarwilayah di Jawa Timur, sekaligus menjadi rekomendasi bagi pemerintah dalam merumuskan kebijakan pemerataan pendidikan di Jawa Timur.

KAJIAN TEORI

PENDIDIKAN

Pendidikan merupakan suatu proses yang diselenggarakan secara sadar dan terencana dengan tujuan menciptakan kondisi belajar yang memungkinkan peserta didik untuk mengembangkan potensi dirinya secara optimal (Munandar dkk., 2022). Secara etimologis, kata pendidikan berasal dari bahasa Latin *educatum*, yang terdiri atas kata *E* berarti “keluar” dan *Duco* berarti “memimpin” atau “mengembangkan”, sehingga pendidikan dapat diartikan sebagai proses pengembangan potensi dalam diri individu. Dalam bahasa Yunani, istilah pendidikan dikenal dengan

kata *pedagogi*, yang berasal dari *paid* berarti “anak” dan *agogos* berarti “menuntun”, yang menunjukkan bahwa pendidikan merupakan proses menuntun peserta didik menuju kedewasaan dan perkembangan pribadi yang menyeluruh. Menurut Kamus Besar Bahasa Indonesia (KBBI), pendidikan merupakan proses pengubahan sikap dan perilaku seseorang atau sekelompok orang dalam usaha mendewasakan manusia melalui kegiatan pengajaran dan pelatihan. Dengan demikian, pendidikan merupakan proses membimbing manusia melalui pengajaran moral dan pengembangan kecerdasan untuk mengoptimalkan potensinya.

INDIKATOR KUALITAS PENDIDIKAN

Indikator kualitas pendidikan merupakan statistik yang mencerminkan aspek-aspek kunci dari sistem pendidikan, berfungsi untuk mengevaluasi dan memantau kinerja sekolah, guru, program, maupun peserta didik, serta menilai pengaruhnya terhadap hasil belajar. Indikator ini mencakup dimensi input, proses, dan output, sehingga memungkinkan analisis efektivitas dan kualitas pendidikan pada berbagai tingkat, mulai dari individu, kelas, sekolah, hingga sistem pendidikan secara keseluruhan (Creemers dan Kyriakides, 2009). Definisi masing-masing indikator kualitas pendidikan didasarkan pada publikasi Statistik Pendidikan Indonesia Tahun 2024 yang diterbitkan oleh Badan Pusat Statistik (BPS).

1. Rata-rata Lama Sekolah (RLS)

Rata-rata Lama Sekolah (RLS) merupakan indikator yang menunjukkan rata-rata jumlah tahun pendidikan formal yang telah ditempuh penduduk di suatu wilayah. Semakin tinggi nilai RLS, semakin tinggi tingkat pendidikan yang ditempuh penduduk.

Rumus yang digunakan untuk menghitung indikator ini adalah sebagai berikut:

$$RLS = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

di mana,

x_i = Lama sekolah penduduk ke- i yang berusia 15 tahun ke atas

n = Jumlah penduduk usia 15 tahun ke atas

2. Pendidikan Tertinggi yang Ditamatkan (PTT)

Pendidikan Tertinggi yang Ditamatkan (PTT) merupakan indikator yang menunjukkan

jenjang pendidikan formal tertinggi yang telah diselesaikan oleh penduduk usia 10 tahun ke atas dan dibuktikan dengan ijazah atau sertifikat. Indikator ini menggambarkan distribusi tingkat pendidikan penduduk sebagai cerminan kualitas sumber daya manusia di suatu wilayah. PTT umumnya diklasifikasikan ke dalam jenjang SD, SMP, dan SMA.

Rumus yang digunakan untuk menghitung indikator ini adalah sebagai berikut:

$$PTT_A = \frac{\text{Jumlah penduduk dengan pendidikan tertinggi } A}{\text{Jumlah penduduk usia 10 tahun ke atas}} \times 100 \quad (2)$$

di mana,

PTT_A = Persentase penduduk dengan pendidikan tertinggi pada jenjang A
 $A \in \{SD, SMP, SMA\}$

3. Angka Partisipasi Murni (APM)

Angka Partisipasi Murni (APM) merupakan indikator yang menunjukkan persentase penduduk usia sekolah pada suatu kelompok usia tertentu yang bersekolah pada jenjang pendidikan yang sesuai dengan usianya. APM dapat ditinjau berdasarkan tiga kelompok usia resmi yang merepresentasikan tiga jenjang pendidikan, yaitu usia 7–12 tahun (SD), 13–15 tahun (SMP), dan 16–18 tahun (SMA).

Rumus yang digunakan untuk menghitung indikator ini adalah sebagai berikut:

$$APM_A = \frac{JM_A}{JP_A} \times 100 \quad (3)$$

di mana,

APM_A = Angka Partisipasi Murni pada jenjang pendidikan ke- A

JM_A = Jumlah murid pada jenjang A yang berada pada kelompok usia resmi jenjang tersebut

JP_A = Jumlah penduduk pada kelompok usia resmi jenjang A

$A \in \{SD, SMP, SMA\}$

4. Angka Melek Huruf (AMH)

Angka Melek Huruf (AMH) merupakan indikator yang menunjukkan persentase penduduk usia 10 tahun ke atas yang mampu membaca dan menulis huruf Latin dan/atau huruf lainnya. Indikator ini digunakan untuk menggambarkan tingkat kemampuan dasar literasi di suatu wilayah, karena kemampuan membaca dan menulis menjadi fondasi utama dalam pengembangan wawasan dan pengetahuan masyarakat.

Rumus yang digunakan untuk menghitung indikator ini adalah sebagai berikut:

$$AMH = \frac{MH_{10+}}{P_{10+}} \times 100\% \quad (4)$$

di mana,

MH_{10+} = Penduduk berusia 10 tahun keatas yang melek huruf

P_{10+} = Jumlah penduduk berusia 10 tahun keatas

5. Rasio Murid terhadap Guru (RMG)

Rasio Murid terhadap Guru (RMG) merupakan indikator yang menunjukkan perbandingan antara jumlah peserta didik dan jumlah guru pada setiap jenjang pendidikan. Indikator ini digunakan untuk menilai tingkat pemerataan serta efektivitas layanan pendidikan, karena semakin kecil nilai rasio, semakin optimal perhatian dan bimbingan yang dapat diberikan guru kepada setiap murid. Berdasarkan Peraturan Pemerintah Nomor 74 Tahun 2008 tentang Guru Pasal 17, pada jenjang SD/MI, SMP/MTs, dan SMA/MA idealnya satu guru bertanggung jawab terhadap 15-20 murid. Dalam penelitian ini, RMG dihitung untuk jenjang SD, SMP, dan SMA.

Rumus yang digunakan untuk menghitung indikator ini adalah sebagai berikut:

$$RMG_A = \frac{JM_A}{JG_A} \quad (5)$$

di mana,

RMG_A = Rasio murid terhadap guru pada jenjang pendidikan A

JM_A = Jumlah murid pada jenjang pendidikan A

JG_A = Jumlah guru pada jenjang pendidikan A

$A \in \{SD, SMP, SMA\}$

CLUSTERING

Clustering adalah suatu teknik untuk mengklasifikasikan data kedalam kelompok-kelompok tertentu berdasarkan kesamaan sifat atau karakter (Valente de Oliveira dan Pedrycz, 2007). Tujuan utama dari *clustering* adalah mengelompokkan data ke dalam beberapa *cluster* yang terpisah, di mana setiap data dalam satu *cluster* memiliki kesamaan karakteristik, sedangkan data dengan karakteristik yang berbeda akan dikelompokkan ke dalam *cluster* lain. Metode *clustering* terbagi menjadi dua kategori, yaitu *fuzzy clustering* dan *hard clustering* (Prabowo dan

Kurniawan, 2019). Pada penelitian ini, metode *clustering* yang digunakan adalah *fuzzy clustering*, tepatnya dengan menerapkan algoritma *Fuzzy C-Means*.

HIMPUNAN FUZZY

Himpunan fuzzy merupakan perluasan dari himpunan tegas (*crisp set*). Pada himpunan tegas, fungsi keanggotaan suatu elemen hanya memiliki dua kemungkinan nilai, yaitu 0 dan 1. Sementara itu, fungsi keanggotaan suatu elemen pada himpunan fuzzy, nilainya berada pada interval $[0,1]$ (Febriyanti dkk., 2023). Konsep ini dikembangkan untuk mengatasi keterbatasan himpunan tegas yang bersifat diskrit (biner) serta memodelkan ketidakpastian dan ketidakjelasan. Konsep himpunan fuzzy dapat didefinisikan sebagai berikut:

Misalkan X adalah himpunan semesta tak kosong dan anggota dari X dilambangkan dengan x . Himpunan fuzzy A pada X didefinisikan sebagai berikut:

$$A = \{(x, \mu_A(x)): x \in X\}$$

dengan μ_A adalah fungsi keanggotaan himpunan fuzzy A , dimana $\mu_A: X \rightarrow [0,1]$ dan $\mu_A(x)$ adalah derajat keanggotaan x pada himpunan fuzzy A , dengan ketentuan nilainya $0 \leq \mu_A(x) \leq 1$ untuk setiap $x \in X$.

FUZZY C-MEANS

Metode *Fuzzy C-Means* (FCM) pertama kali dikemukakan oleh Dunn (1973) kemudian dikembangkan oleh Jim Bezdek pada tahun 1981 (Buana, 2023). FCM merupakan salah satu teknik pengelompokan data dimana keberadaan tiap titik data dalam suatu *cluster* ditentukan oleh derajat keanggotaan (Suwirmayanti, 2018). Konsep dasar FCM adalah menentukan pusat *cluster* yang berfungsi untuk mengelompokkan setiap data ke dalam kelompok tertentu. Pada tahap awal, posisi pusat *cluster* dan derajat keanggotaan tiap data terhadap *cluster* tersebut masih belum akurat. Oleh karena itu, proses pembaruan dilakukan secara berulang (iteratif) hingga posisi pusat *cluster* berpindah ke lokasi yang paling sesuai dan nilai keanggotaannya menjadi stabil (Setiawan dkk., 2018).

Berikut ini merupakan tahapan dalam algoritma *Fuzzy C-Means*:

1. Memasukkan data yang akan dikelompokkan.

Data direpresentasikan dalam bentuk matriks X berukuran $n \times m$, dengan n menunjukkan banyak data, dan m menunjukkan banyak atribut atau variabel yang digunakan. Elemen matriks x_{ij} menyatakan nilai dari data ke- i dan atribut ke- j , dengan $i = 1, 2, \dots, n$ dan $j = 1, 2, \dots, m$.

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

2. Menentukan parameter/nilai awal :

Jumlah cluster = c

Pangkat = w

Maksimum iterasi = $MaxIter$

Error terkecil yang diharapkan = ε

Fungsi Objektif awal = $P_0 = 0$

Iterasi awal = $t = 1$

3. Membentuk matriks partisi awal $U = [\mu_{ik}]$. Matriks U dibentuk dengan membangkitkan bilangan random μ_{ik} untuk $i = 1, 2, \dots, n$ dan $k = 1, 2, \dots, c$ dengan μ_{ik} berada pada interval $[0, 1]$. μ_{ik} diperoleh dengan:

- a. Menghitung jumlah setiap baris pada matriks U .

$$Q_i = \sum_{k=1}^c \mu'_{ik} \quad (6)$$

di mana,

μ'_{ik} = derajat keanggotaan awal data ke- i terhadap cluster ke- k yang dibangkitkan secara acak

Q_i = jumlah derajat keanggotaan data ke- i terhadap semua cluster

- b. Menghitung derajat keanggotaan data ke- i pada cluster ke- k (μ_{ik}).

$$\mu_{ik} = \frac{\mu'_{ik}}{Q_i} \quad (7)$$

4. Menghitung pusat cluster ke- k untuk atribut ke- j (v_{kj}), di mana $k = 1, 2, \dots, c$ dan $j = 1, 2, \dots, m$ dengan menggunakan persamaan sebagai berikut:

$$v_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w x_{ij})}{\sum_{i=1}^n (\mu_{ik})^w} \quad (8)$$

di mana,

μ_{ik} = Nilai keanggotaan data ke- i dan cluster ke- k .

x_{ij} = Data ke- i dan atribut ke- j .

w = Pangkat

5. Menghitung fungsi objektif pada iterasi ke- t , yaitu P_t dengan $t = 1, 2, \dots, d$, d merupakan banyak iterasi.

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (x_{ij} - v_{kj})^2 \right] (\mu_{ik})^w \right) \quad (9)$$

di mana,

x_{ij} = Data ke- i dan atribut ke- j .

v_{kj} = Pusat cluster ke- k untuk atribut ke- j

μ_{ik} = Nilai keanggotaan data ke- i dan cluster ke- k .

w = Pangkat

6. Menghitung perubahan matriks partisi

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (x_{ij} - v_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (x_{ij} - v_{kj})^2 \right]^{\frac{-1}{w-1}}} \quad (10)$$

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

7. Cek kondisi berhenti :

- a. Jika $(|P_t - P_{t-1}| < \varepsilon)$ atau $(t > MaxIter)$ terpenuhi maka perhitungan selesai.
- b. Jika tidak, maka $t = t + 1$ dan ulangi langkah d.

8. Mengulang langkah b sampai dengan langkah g untuk cluster yang berbeda.

NORMALISASI DATA

Normalisasi data merupakan proses penskalaan nilai pada setiap variabel agar berada dalam rentang atau skala yang sebanding, sehingga tidak terdapat variabel yang mendominasi analisis akibat perbedaan satuan atau besaran nilai. Normalisasi sangat diperlukan dalam proses pengelompokan, terutama ketika metode yang digunakan berbasis pengukuran jarak antar objek, seperti *Euclidean Distance*, karena metrik tersebut sensitif terhadap perbedaan skala antar variabel (Richia Putri dkk., 2023). Normalisasi data pada penelitian ini menggunakan normalisasi Z-Score. Normalisasi ini dilakukan pada masing-masing variabel berdasarkan rata-rata dan standar deviasi dari variabel tersebut. Rumus dari normalisasi z-score adalah sebagai berikut:

$$z_{x_{ij}} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \quad (11)$$

dengan $\bar{x}_j$ adalah rata-rata nilai dari variabel ke- j dan σ_j adalah standar deviasi dari variabel ke- j .

INDEKS VALIDITAS

Indeks validitas merupakan suatu ukuran yang digunakan untuk menentukan jumlah *cluster* yang paling optimal, sehingga dapat secara akurat merepresentasikan struktur dan pola yang terdapat dalam data (Wang dan Zhang, 2007). Pada penelitian ini, indeks validitas yang digunakan adalah indeks validitas *Partition Coefficient* (PC). PC merupakan salah satu indeks validitas pada *fuzzy clustering* yang digunakan untuk mengukur jumlah *overlapping* antar *cluster*. Semakin tinggi nilai PC, maka semakin baik kualitas *cluster* yang dihasilkan. Indeks ini direkomendasikan oleh Bezdek dan Dunn (1975) sebagai salah satu metode evaluasi kualitas *cluster*.

Rumus yang digunakan untuk menghitung *Partition Coefficient* adalah sebagai berikut:

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

di mana,

c = Jumlah *cluster*

n = Jumlah data

μ_{ik} = Nilai keanggotaan pada data sampel ke- i dan *cluster* ke- k

$PC(c)$ = Nilai indeks PC untuk c *cluster*

Semakin dekat nilai PC dengan satu, mengindikasikan bahwa kualitas *cluster* yang dihasilkan semakin baik.

METODE

DATA PENELITIAN

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) Provinsi Jawa Timur tahun 2024. Unit analisis dalam penelitian ini adalah seluruh kabupaten/kota di Provinsi Jawa Timur. Penelitian ini menggunakan 11 indikator kualitas pendidikan yang bersumber dari publikasi resmi Badan Pusat Statistik (BPS) Provinsi Jawa Timur tahun 2024. Daftar variabel penelitian disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Indikator
X_1	Rata-rata Lama Sekolah (RLS)
X_2	Pendidikan Tertinggi yang Ditamatkan SD (PTT SD)
X_3	Pendidikan Tertinggi yang Ditamatkan SMP (PTT SMP)
X_4	Pendidikan Tertinggi yang Ditamatkan SMA (PTT SMA)

X_5	Angka Partisipasi Murni SD (APM SD)
X_6	Angka Partisipasi Murni SMP (APM SMP)
X_7	Angka Partisipasi Murni SMA (APM SMA)
X_8	Angka Melek Huruf (AMH)
X_9	Rasio Murid terhadap Guru SD (RMG SD)
X_{10}	Rasio Murid terhadap Guru SMP (RMG SMP)
X_{11}	Rasio Murid terhadap Guru SMA (RMG SMA)

TAHAPAN ANALISIS

1. Preprocessing Data.

Data dinormalisasi menggunakan metode *z-score* untuk menghilangkan perbedaan skala antarvariabel. Normalisasi ini bertujuan untuk menghindari dominasi variabel yang memiliki rentang nilai lebih besar dibandingkan variabel lainnya, sehingga seluruh variabel berada pada skala yang sebanding.

2. Klasterisasi Menggunakan Metode Fuzzy C-Means.

Proses klasterisasi dilakukan menggunakan metode *Fuzzy C-Means* (FCM) dengan bantuan perangkat lunak RStudio untuk beberapa variasi jumlah *cluster*. Hal ini bertujuan untuk memperoleh jumlah *cluster* yang optimal berdasarkan indeks validitas yang digunakan.

3. Evaluasi Hasil Klasterisasi.

Untuk menentukan jumlah *cluster* yang optimal, digunakan indeks validitas *Partition Coefficient* (PC). Nilai PC dihitung untuk beberapa alternatif jumlah *cluster*, kemudian dipilih nilai *cluster* dengan indeks PC terbaik sebagai jumlah *cluster* optimal.

4. Visualisasi Hasil Cluster Optimal.

Hasil *cluster* optimal divisualisasikan dalam bentuk peta tematik (*choropleth map*) untuk menggambarkan sebaran kualitas pendidikan kabupaten/kota di Provinsi Jawa Timur.

5. Interpretasi Hasil Cluster yang Optimal.

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini terdiri atas 38 kabupaten/kota di Provinsi Jawa Timur dengan 11 variabel indikator kualitas pendidikan yang bersumber dari Badan Pusat Statistik (BPS). Data tersebut disajikan pada Tabel 2 dan selanjutnya melalui tahap *preprocessing* sebelum dilakukan proses klasterisasi.

Tabel 2. Data Penelitian

No.	Kab/Kota	X_1	X_2	X_3	...	X_{11}
1	Kab. Pacitan	7,90	38,09	23,84	...	13
2	Kab. Ponorogo	7,80	24,48	21,73	...	14
3	Kab. Trenggalek	7,92	34,93	22,62	...	16
4	Kab. Tulungagung	8,68	29,14	22,91	...	18

5	Kab. Blitar	7,87	28,09	24,64	...	17
6	Kab. Kediri	8,26	28,09	20,63	...	18
⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	Kota Batu	9,87	24,94	20,97	...	16

1. Preprocessing Data.

Data penelitian pada Tabel 2 selanjutnya dinormalisasi menggunakan Persamaan (11). Hasil normalisasi disajikan pada Tabel 3.

Tabel 3. Hasil Normalisasi Data

No.	Kab/Kota	X_1	X_2	X_3	...	X_{11}
1	Kab. Pacitan	-0,34	1,72	1,40	...	-1,42
2	Kab. Ponorogo	-0,40	-0,21	0,60	...	-1,00
3	Kab. Trenggalek	-0,32	1,27	0,93	...	-0,13
4	Kab. Tulungagung	0,13	0,45	1,04	...	0,80
5	Kab. Blitar	-0,35	0,30	1,70	...	0,33
6	Kab. Kediri	-0,12	0,30	0,18	...	1,04
⋮	⋮	⋮	⋮	⋮	⋮	⋮
38	Kota Batu	0,84	-0,15	0,31	...	0,00

2. Klasterisasi Menggunakan Metode Fuzzy C-Means.

- a. Memasukan data yang akan di-cluster, dari Tabel 3, ke dalam matriks X berukuran 38×11 .

$$X = \begin{bmatrix} -0,34 & 1,72 & 1,40 & \dots & -1,42 \\ -0,40 & -0,21 & 0,60 & \dots & -1,00 \\ -0,32 & 1,27 & 0,93 & \dots & -0,13 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0,84 & -0,15 & 0,31 & \dots & -0,00 \end{bmatrix}$$

- b. Menentukan parameter/nilai awal

Pada penelitian ini menggunakan jumlah cluster (c) 2 sampai dengan 5, $w = 2$, $MaxIter = 100$, $\varepsilon = 0,00001$ atau 10^{-5} , $P_0 = 0$, dan $t = 1$. Pada penelitian ini contoh perhitungan dilakukan dengan menggunakan $c = 2$.

- c. Membentuk matriks partisi awal U

Setelah menentukan nilai-nilai parameter yang digunakan, langkah selanjutnya adalah melakukan pembangkitan bilangan acak μ_{ik} sebagai elemen matriks partisi awal U . Nilai keanggotaan awal data ke- i pada cluster ke- k (μ_{ik}) pada masing-masing cluster ditunjukkan pada Tabel 4.

Tabel 4. Nilai Keanggotaan Awal dengan $c = 2$

Kabupaten/Kota	Nilai Keanggotaan Awal	
	Cluster 1	Cluster 2
Kab. Pacitan	0,4747395	0,5252605
Kab. Ponorogo	0,7729005	0,2270995
Kab. Trenggalek	0,7411997	0,2588003
Kab. Tulungagung	0,6805195	0,3194805
Kab. Blitar	0,6944861	0,3055139
⋮	⋮	⋮
Kota Batu	0,4957495	0,5042505

- d. Menghitung pusat cluster ke- k untuk atribut ke- j (v_{kj})

Setelah diperoleh μ_{ik} , langkah selanjutnya adalah menghitung pusat cluster awal dengan menggunakan Persamaan (8). Berikut diberikan contoh perhitungan pusat cluster ke-1 untuk atribut atau variabel ke-1 (v_{11}) dan pusat cluster ke-2 untuk atribut atau variabel ke-1 (v_{21}).

$$\begin{aligned} v_{11} &= \frac{\sum_{i=1}^{38} ((\mu_{i1})^2 x_{i1})}{\sum_{i=1}^{38} (\mu_{i1})^2} \\ &= \frac{-0,823546158}{14,24204524} \\ &= -0,057824992 \end{aligned}$$

$$\begin{aligned} v_{21} &= \frac{\sum_{i=1}^{38} ((\mu_{i2})^2 x_{i1})}{\sum_{i=1}^{38} (\mu_{i2})^2} \\ &= \frac{2,12215614}{9,484958736} \\ &= 0,223739101 \end{aligned}$$

Dengan proses perhitungan yang sama juga dilakukan untuk atribut atau variabel lainnya pada cluster ke-1 dan cluster ke-2. Pusat cluster awal dari masing-masing variabel disajikan pada Tabel 5.

Tabel 5. Pusat Cluster Awal dengan $c = 2$

Variabel	$k = 1$	$k = 2$
X_1	-0,057824992	0,223739101
X_2	0,119344528	-0,320896143
X_3	0,136156739	-0,201553332
X_4	-0,129879122	0,331288644
X_5	-0,075264903	0,187441504
X_6	0,002353539	0,028090002
X_7	-0,237773160	0,370069254
X_8	-0,092252589	0,193444737
X_9	-0,024328727	0,076918479
X_{10}	-0,077978954	0,150024759
X_{11}	-0,113038314	0,184125387

- e. Menghitung fungsi objektif

Setelah menghitung pusat cluster awal, selanjutnya menghitung nilai fungsi objektif awal ($t = 1$) dengan menggunakan data pada Tabel 3, μ_{ik} pada Tabel 4, v_{kj} pada Tabel 5, dan Persamaan (9). Hasil perhitungan fungsi objektif pada iterasi ke-1 (P_1) adalah sebagai berikut:

$$\begin{aligned} P_1 &= \sum_{i=1}^{38} \sum_{k=1}^2 \left(\left[\sum_{j=1}^{11} (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^2 \right) \\ &= 7,3703476 + 3,2743811 + \dots + 2,4584566 \\ &= 256,7700355 \end{aligned}$$

- f. Menghitung perubahan matriks partisi

Setelah menghitung nilai fungsi objektif pada iterasi ke-1, selanjutnya menghitung perubahan nilai keanggotaan dengan menggunakan Persamaan (10). Contoh perhitungan untuk $i = 1$ dan $k = 1$.

$$\begin{aligned}\mu_{11} &= \frac{[\sum_{j=1}^{11}(x_{1j}-v_{1j})^2]^{-\frac{1}{2-1}}}{\sum_{k=1}^2[\sum_{j=1}^{11}(x_{1j}-v_{kj})^2]^{-\frac{1}{2-1}}} = \frac{[\sum_{j=1}^{11}(x_{1j}-v_{1j})^2]^{-1}}{\sum_{k=1}^2[\sum_{j=1}^{11}(x_{1j}-v_{kj})^2]^{-1}} \\ &= \frac{(11,678441874)^{-1}}{[(11,678441874)^{-1} + (17,174022697)^{-1}]} \\ &= \frac{0,0856278612}{0,0856278612 + 0,0582274763} \\ &= \frac{0,0856278612}{0,1438553375} \\ &= 0,59523589933905\end{aligned}$$

Contoh perhitungan untuk $i = 1$ dan $k = 2$

$$\begin{aligned}\mu_{i2} &= \frac{[\sum_{j=1}^{11}(x_{1j}-v_{2j})^2]^{-\frac{1}{2-1}}}{\sum_{k=1}^2[\sum_{j=1}^{11}(x_{1j}-v_{kj})^2]^{-\frac{1}{2-1}}} = \frac{[\sum_{j=1}^{11}(x_{1j}-v_{2j})^2]^{-1}}{\sum_{k=1}^2[\sum_{j=1}^{11}(x_{1j}-v_{kj})^2]^{-1}} \\ &= \frac{(17,174022697)^{-1}}{[(11,678441874)^{-1} + (17,174022697)^{-1}]} \\ &= \frac{0,0582274763}{0,0856278612 + 0,0582274763} \\ &= \frac{0,0582274763}{0,1438553375} \\ &= 0,40476410066095\end{aligned}$$

Hasil nilai keanggotaan yang telah diperbarui disajikan pada Tabel 6.

Tabel 6. Nilai Keanggotaan yang Telah Diperbarui

Kabupaten/Kota	Nilai Keanggotaan	
	Cluster 1	Cluster 2
Kab. Pacitan	0,5952359	0,4047641
Kab. Ponorogo	0,6132306	0,3867694
Kab. Trenggalek	0,5814214	0,4185786
Kab. Tulungagung	0,5547058	0,4452942
⋮	⋮	⋮
Kota Batu	0,3594798	0,6405202

g. Cek kondisi berhenti

Karena $|P_1 - P_0| > \varepsilon$ yaitu $256,7700355 > 0,00001$ dan $t < MaxIter$ yaitu $1 < 100$ maka proses perhitungan dilanjutkan ke iterasi berikutnya.

Pada penelitian ini, proses iterasi dua *cluster* berhenti pada iterasi ke-33. Hasil akhir berupa pusat *cluster* dan nilai derajat keanggotaan disajikan pada Tabel 7 dan Tabel 8.

Tabel 7. Pusat Cluster Akhir dengan $c = 2$

Variabel	$k = 1$	$k = 2$
X_1	-0,6386198	0,7330259
X_2	0,5666836	-0,6608650
X_3	0,02363018	-0,12532404
X_4	-0,6481129	0,7402856
X_5	-0,2899200	0,3065286
X_6	-0,4857975	0,4423197
X_7	-0,4817426	0,5384548
X_8	-0,6187781	0,6786680
X_9	-0,3868508	0,4556234
X_{10}	-0,4240201	0,3838354
X_{11}	-0,4235979	0,3923676

Tabel 8. Nilai Keanggotaan Akhir dengan $c = 2$

Kabupaten/Kota	Nilai Keanggotaan	
	Cluster 1	Cluster 2
Kab. Pacitan	0,7666844	0,2333156
Kab. Ponorogo	0,7630547	0,2369453
Kab. Trenggalek	0,6767284	0,3232716
Kab. Tulungagung	0,4777911	0,5222089
⋮	⋮	⋮
Kota Batu	0,1343349	0,8656651

Berdasarkan nilai keanggotaan akhir pada Tabel 8, Kabupaten Pacitan memiliki nilai keanggotaan sebesar 0,7666844 pada *cluster* pertama dan 0,2333156 pada *cluster* kedua. Karena nilai terbesar terdapat pada *cluster* pertama, maka Kabupaten Pacitan ditetapkan sebagai anggota *cluster* pertama. Penentuan keanggotaan untuk kabupaten/kota lainnya dilakukan dengan prinsip yang sama, yaitu berdasarkan nilai derajat keanggotaan maksimum. Prosedur FCM yang diterapkan pada pembentukan dua *cluster* juga dilakukan dengan tahapan yang sama pada pembentukan tiga, empat, dan lima *cluster*.

3. Evaluasi Hasil Klasterisasi.

Evaluasi hasil *clustering* dilakukan menggunakan indeks validitas *Partition Coefficient* (PC) yang telah dihitung sebelumnya. Nilai PC dihitung menggunakan Persamaan (12). Contoh perhitungan nilai PC untuk $c = 2$ adalah sebagai berikut:

$$\begin{aligned}PC(2) &= \frac{1}{38} \sum_{i=1}^{38} \sum_{k=1}^2 (\mu_{ik})^2 \\ &= \frac{11,4171343 + 12,3026679}{38} \\ &= \frac{23,7198023}{38} \\ &= 0,624205323\end{aligned}$$

Hasil perhitungan nilai indeks validitas PC untuk jumlah *cluster* (c) 2 sampai dengan 5 disajikan pada Tabel 9.

Tabel 9. Nilai Indeks Validitas Masing-masing Cluster

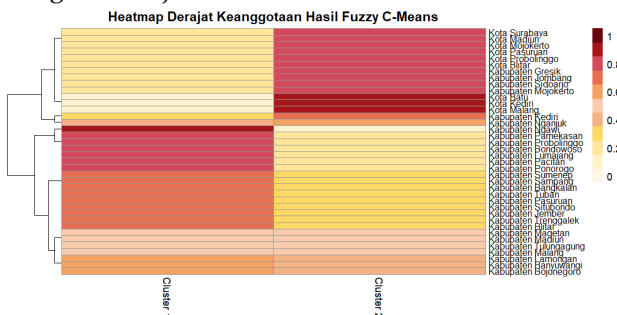
Jumlah Cluster	Nilai Indeks Validitas PC
2	0,624205323
3	0,518109643
4	0,423234941
5	0,378885145

Jumlah *cluster* yang optimal ditentukan berdasarkan nilai indeks *Partition Coefficient* (PC) tertinggi. Hasil evaluasi menunjukkan bahwa *clustering* dengan 2 *cluster* memiliki nilai PC paling besar dibandingkan jumlah *cluster* lainnya.

Selain nilai indeks *Partition Coefficient*, pola pemisahan *cluster* dianalisis melalui visualisasi derajat keanggotaan hasil *Fuzzy C-Means*. Visualisasi

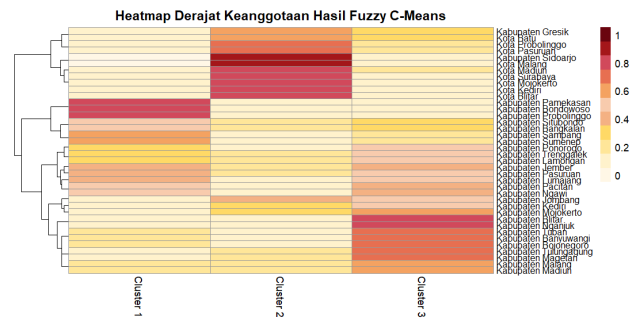
ini digunakan untuk mengamati tingkat kejelasan pemisahan antar *cluster* serta mengidentifikasi wilayah yang menunjukkan indikasi ketidakpastian keanggotaan pada berbagai variasi jumlah *cluster*.

Pada penelitian ini, ketidakpastian tersebut dibedakan menjadi dua kondisi. Pertama, *overlapping*, yaitu ketika suatu wilayah memiliki lebih dari satu *cluster* dengan nilai derajat keanggotaan maksimum yang sama, sehingga kecenderungan pengelompokannya benar-benar seimbang. Kedua, wilayah tanpa dominasi *cluster*, yaitu ketika nilai derajat keanggotaan tertinggi belum melampaui ambang tertentu, yang menunjukkan bahwa meskipun terdapat *cluster* dengan nilai terbesar, tingkat kecenderungannya belum cukup kuat. Kedua pendekatan ini membantu memberikan gambaran yang lebih komprehensif mengenai ketegasan hasil pengelompokan. Visualisasi yang ditampilkan merupakan representasi dari nilai derajat keanggotaan yang telah dipaparkan pada bagian sebelumnya. Dalam penyajiannya, nilai-nilai tersebut dibulatkan sampai satu angka di belakang koma agar pola kedekatan dan dominasi *cluster* dapat diamati dengan lebih jelas.



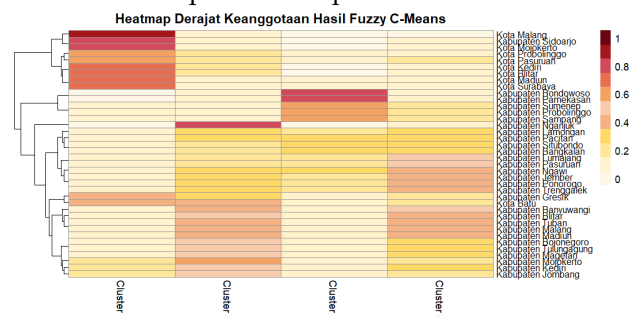
Gambar 1. Visualisasi Pemisahan 2 Cluster

Berdasarkan visualisasi dua *cluster*, sebagian besar wilayah telah menunjukkan kecenderungan keanggotaan yang jelas pada salah satu *cluster*. Namun, beberapa wilayah masih belum terkelompok secara tegas. Berdasarkan nilai derajat keanggotaan, wilayah yang mengalami *overlapping* adalah Kabupaten Tulungagung, Kabupaten Malang, Kabupaten Madiun, dan Kabupaten Magetan, karena nilai keanggotaannya relatif seimbang antar *cluster* sehingga berada pada area transisi.



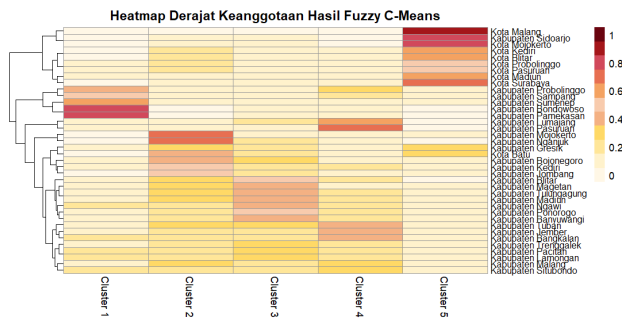
Gambar 2. Visualisasi Pemisahan 3 Cluster

Berdasarkan visualisasi tiga *cluster*, sebagian wilayah telah menunjukkan kecenderungan keanggotaan yang lebih jelas, meskipun beberapa wilayah masih belum memiliki dominasi yang kuat. Dengan ambang batas derajat keanggotaan 0,6, wilayah yang teridentifikasi mengalami *overlapping* meliputi Kab. Pacitan, Kab. Ponorogo, Kab. Trenggalek, Kab. Kediri, Kab. Lumajang, Kab. Jember, Kab. Situbondo, Kab. Pasuruan, Kab. Jombang, Kab. Ngawi, Kab. Lamongan, dan Kab. Bangkalan, karena nilai keanggotaannya masih relatif tersebar pada beberapa *cluster*.



Gambar 3. Visualisasi Pemisahan 4 Cluster

Berdasarkan visualisasi empat *cluster*, sebagian wilayah belum menunjukkan kecenderungan kuat pada satu *cluster* karena nilai keanggotaannya masih tersebar. Terdapat wilayah dengan nilai maksimum yang sama pada lebih dari satu *cluster*, yaitu Kab. Pacitan, Kab. Malang, Kab. Situbondo, Kab. Madiun, Kab. Tuban, Kab. Lamongan, Kab. Bangkalan, serta Kota Batu. Selain itu, dengan ambang dominasi 0,6, sebanyak 22 kabupaten/kota belum memiliki keanggotaan yang dominan, sehingga pemisahan empat *cluster* dinilai belum tegas.



Gambar 4. Visualisasi Pemisahan 5 Cluster

Berdasarkan visualisasi lima *cluster*, sebagian wilayah menunjukkan kecenderungan ke satu kelompok, namun pemisahannya belum merata. Kab. Malang dan Kab. Gresik teridentifikasi mengalami *overlapping* karena memiliki nilai maksimum yang sama pada lebih dari satu *cluster*. Selain itu, dengan ambang dominasi 0,6, sebanyak 24 kabupaten/kota belum memiliki keanggotaan yang kuat, sehingga pengelompokan lima *cluster* dinilai belum tegas.

Secara umum, semakin banyak jumlah *cluster*, semakin meningkat pula wilayah yang mengalami *overlapping*. Fenomena ini terlihat pada Kabupaten Pacitan, yang pada dua *cluster* masih menunjukkan keanggotaan tegas, namun pada tiga hingga lima *cluster* nilai keanggotaannya semakin terbagi sehingga tidak lagi dominan. Hal ini menunjukkan bahwa penambahan *cluster* membuat batas antar kelompok semakin kabur.

Di sisi lain, sejumlah kabupaten/kota menunjukkan konsistensi berada pada *cluster* yang sama meskipun jumlah *cluster* berubah, seperti Kota Malang, Kab. Sidoarjo, Kota Mojokerto, Kota Surabaya, Kota Kediri, Kota Blitar, Kota Madiun, Kota Probolinggo, dan Kota Pasuruan selalu berada di *cluster* yang sama. Konsistensi serupa juga ditunjukkan oleh Kab. Bondowoso, Kab. Pamekasan, Kab. Probolinggo, dan Kab. Sampang juga selalu berada pada *cluster* yang sama. Hal ini menunjukkan bahwa wilayah-wilayah tersebut memiliki karakteristik yang relatif serupa sehingga tetap tergabung dalam kelompok yang sama.

Meskipun terdapat beberapa wilayah yang stabil, secara keseluruhan pola pemisahan menunjukkan bahwa semakin banyak *cluster* yang digunakan, semakin besar pula tingkat ketidakjelasan keanggotaan yang muncul. Temuan visual ini selaras dengan hasil indeks *Partition Coefficient* yang menunjukkan nilai tertinggi pada dua *cluster*, kondisi

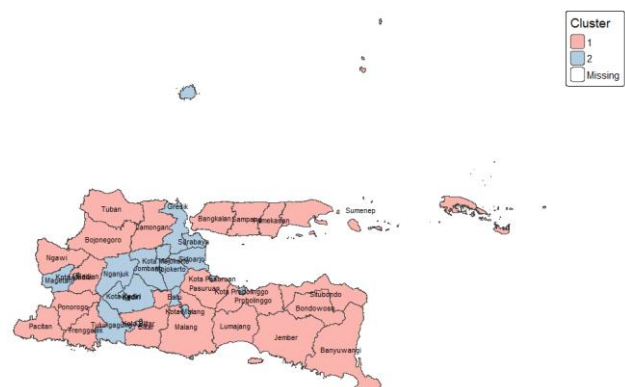
ini mengindikasikan bahwa dua *cluster* memberikan pemisahan yang paling stabil dan ditetapkan sebagai jumlah *cluster* yang optimal. Pada pembentukan dua *cluster* tersebut, *cluster* 1 terdiri atas 21 kabupaten/kota, sedangkan *cluster* 2 terdiri atas 17 kabupaten/kota. Adapun anggota dari masing-masing *cluster* optimal disajikan pada Tabel 10.

Tabel 10. Cluster Optimal Kualitas Pendidikan Jawa Timur

Cluster	Anggota Cluster
Cluster 1	Kab. Pacitan, Kab. Ponorogo, Kab. Trenggalek, Kab. Blitar, Kab. Malang, Kab. Lumajang, Kab. Jember, Kab. Banyuwangi, Kab. Bondowoso, Kab. Situbondo, Kab. Probolinggo, Kab. Pasuruan, Kab. Madiun, Kab. Ngawi, Kab. Bojonegoro, Kab. Tuban, Kab. Lamongan, Kab. Bangkalan, Kab. Sampang, Kab. Pamekasan, Kab. Sumenep.
Cluster 2	Kab. Tulungagung, Kab. Kediri, Kabupaten Sidoarjo, Kab. Mojokerto, Kab. Jombang, Kab. Nganjuk, Kab. Magetan, Kab. Gresik, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu.

4. Visualisasi Hasil Cluster Optimal.

Berdasarkan hasil perhitungan indeks PC, jumlah *cluster* optimal yang diperoleh adalah dua *cluster*. Untuk memperjelas hasil pengelompokan tersebut, berikut disajikan visualisasi hasil *clustering* kabupaten/kota di Provinsi Jawa Timur.



Gambar 5. Visualisasi Hasil Cluster Optimal

Visualisasi hasil clustering menunjukkan bahwa Cluster 1 (21 kabupaten/kota) tersebar lebih luas dan didominasi wilayah non-perkotaan, sedangkan Cluster 2 (17 kabupaten/kota) terkonsentrasi pada wilayah perkotaan dan pusat pertumbuhan. Perbedaan tersebut merepresentasikan pengelompokan wilayah berdasarkan kemiripan karakteristik indikator yang dianalisis.

5. Interpretasi Hasil *Cluster* yang Optimal.

Interpretasi hasil *clustering* dilakukan dengan menganalisis karakteristik masing-masing *cluster* berdasarkan nilai rata-rata setiap indikator yang membentuk *cluster* tersebut. Dengan data setiap indikator pada masing-masing *cluster* disajikan pada Tabel 11 dan Tabel 12.

Tabel 11. Data Indikator Kualitas Pendidikan *Cluster* 1

No.	Kab/Kota	X_1	X_2	X_3	...	X_{11}
1	Kab. Pacitan	7,90	38,09	23,84	...	13
2	Kab. Ponorogo	7,80	24,48	21,73	...	14
3	Kab. Trenggalek	7,92	34,93	22,62	...	16
⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	Kab. Sumenep	6,10	27,56	17,05	...	12

Tabel 12. Data Indikator Kualitas Pendidikan *Cluster* 2

No.	Kab/Kota	X_1	X_2	X_3	...	X_{11}
1	Kab. Tulungagung	8,68	29,14	22,91	...	18
2	Kab. Kediri	8,26	28,09	20,63	...	18
3	Kab. Sidoarjo	10,91	13,83	18,29	...	18
⋮	⋮	⋮	⋮	⋮	⋮	⋮
17	Kota Batu	9,87	24,94	20,97	...	16

Berdasarkan Tabel 11 dan Tabel 12 didapatkan rata-rata setiap indikator pada masing-masing *cluster* yang disajikan pada Tabel 13.

Tabel 13. Rata-rata Setiap Indikator Kualitas Pendidikan Pada Masing-masing *Cluster*

<i>Cluster</i>	Rata-rata										
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
1	7,2	30,4	20,4	20,0	98,5	82,0	61,0	92,6	12,5	14,4	15,3
2	10,0	20,5	19,9	33,4	99,1	87,6	72,9	97,3	15,0	16,5	17,1

Berdasarkan nilai rata-rata masing-masing indikator pada Tabel 13 dan pusat *cluster* pada Tabel 7, kedua *cluster* menunjukkan perbedaan karakteristik kualitas pendidikan yang cukup jelas. Berdasarkan Tabel 4.28, *cluster* 1 menunjukkan nilai rata-rata lama sekolah (X_1) serta pendidikan tertinggi yang ditamatkan pada jenjang SMA (X_4) yang relatif lebih rendah dibandingkan *cluster* 2, sedangkan untuk indikator pendidikan tertinggi yang ditamatkan pada jenjang SD (X_2) serta pendidikan tertinggi yang ditamatkan pada jenjang SMP (X_3) nilainya lebih tinggi dibandingkan *cluster* 2. Hal ini sejalan dengan Tabel 4.12 yang menunjukkan bahwa pada *cluster* 1 variabel yang dominan adalah X_2 dan X_3 . Selain itu, Angka Partisipasi Murni (APM) pada jenjang SMP dan SMA (X_6 dan X_7) pada *cluster* ini juga cenderung lebih rendah, yang mengindikasikan masih terbatasnya keberlanjutan pendidikan pada jenjang menengah. Meskipun demikian, APM SD (X_5) dan Angka Melek

Huruf (X_8) pada *cluster* ini telah berada pada tingkat yang relatif tinggi, menunjukkan bahwa akses terhadap pendidikan dasar pada wilayah-wilayah dalam *cluster* ini relatif telah terpenuhi. Berdasarkan karakteristik tersebut, *Cluster* 1 diberi label *Cluster* dengan Dominasi Kualitas Pendidikan Dasar.

Sebaliknya, *Cluster* 2 menunjukkan kualitas pendidikan yang relatif lebih baik, yang tercermin dari nilai rata-rata lama sekolah (X_1), pendidikan tertinggi yang ditamatkan SMA (X_4), serta APM SMP dan SMA (X_6 dan X_7) yang lebih tinggi dibandingkan *cluster* 1. Hal ini mengindikasikan tingkat partisipasi dan keberlanjutan pendidikan menengah yang lebih optimal. Namun demikian, *cluster* ini juga memiliki rasio murid terhadap guru (X_9 , X_{10} , dan X_{11}) yang relatif lebih tinggi, yang menunjukkan adanya beban pengajaran yang lebih besar seiring dengan tingginya partisipasi pendidikan. Oleh karena itu, *Cluster* 2 diberi label *Cluster* dengan Dominasi Kualitas Pendidikan Menengah.

PENUTUP

SIMPULAN

Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Jawa Timur berdasarkan kualitas pendidikan menggunakan metode *Fuzzy C-Means* dengan penentuan jumlah *cluster* optimal berdasarkan indeks validitas *Partition Coefficient*. Berdasarkan hasil pengelompokan kualitas pendidikan kabupaten/kota di Provinsi Jawa Timur menggunakan metode FCM, dilakukan pembentukan *cluster* dengan variasi jumlah *cluster* sebanyak 2, 3, 4, dan 5 *cluster*. Pada pembentukan dua *cluster*, kabupaten/kota di Provinsi Jawa Timur terbagi ke dalam dua kelompok dengan jumlah anggota masing-masing 21 dan 17 kabupaten/kota. Pada pembentukan tiga *cluster*, wilayah terbagi menjadi tiga kelompok dengan jumlah anggota 9, 11, dan 18 kabupaten/kota. Selanjutnya, pada pembentukan empat *cluster*, pembagian wilayah menghasilkan empat kelompok dengan jumlah anggota 11, 11, 6, dan 10 kabupaten/kota, sedangkan pada pembentukan lima *cluster* terbentuk lima kelompok dengan jumlah anggota 5, 6, 10, 7, dan 10 kabupaten/kota. Peningkatan jumlah *cluster* menyebabkan peningkatan tumpang tindih (*overlapping*) antar*cluster*, sehingga batas antar kelompok menjadi kurang tegas.

Untuk menentukan jumlah *cluster* yang paling optimal dalam merepresentasikan struktur data kualitas pendidikan di Jawa Timur, digunakan indeks validitas *Partition Coefficient* (PC). Hasil analisis menunjukkan bahwa pembentukan dua *cluster* memberikan nilai PC tertinggi dibandingkan jumlah *cluster* lainnya, sehingga dua *cluster* dianggap sebagai jumlah *cluster* yang paling optimal dalam merepresentasikan struktur data. *Clustering* optimal tersebut menghasilkan dua kelompok wilayah, yaitu *cluster* dengan kualitas pendidikan rendah dan *cluster* dengan kualitas pendidikan tinggi, dengan jumlah anggota masing-masing *cluster* sebanyak 21 dan 17 kabupaten/kota. Perbedaan karakteristik antar*cluster* terlihat jelas berdasarkan nilai rata-rata indikator pendidikan yang digunakan dalam penelitian ini.

Cluster dengan dominasi kualitas pendidikan dasar ditandai oleh nilai rata-rata lama sekolah, tingkat pendidikan menengah, serta angka partisipasi murni pada jenjang SMP dan SMA yang lebih rendah. Sebaliknya, *cluster* dengan dominasi kualitas pendidikan menengah menunjukkan nilai yang lebih tinggi pada indikator-indikator tersebut, yang mengindikasikan tingkat partisipasi dan keberlanjutan pendidikan yang lebih baik. Namun demikian, *cluster* dengan kualitas pendidikan tinggi juga memiliki rasio murid terhadap guru yang relatif lebih tinggi, yang menunjukkan adanya potensi ketidakseimbangan beban pengajaran seiring dengan meningkatnya partisipasi pendidikan. Secara umum, indikator rata-rata lama sekolah, capaian pendidikan menengah, dan partisipasi pendidikan berperan sebagai indikator pembeda utama dalam membedakan kualitas pendidikan antar*cluster* di Provinsi Jawa Timur.

SARAN

Hasil pengelompokan yang diperoleh dalam penelitian ini diharapkan dapat menjadi bahan pertimbangan bagi pemangku kebijakan pendidikan daerah dan instansi terkait dalam perumusan kebijakan pendidikan di Provinsi Jawa Timur. Perhatian khusus perlu diarahkan pada wilayah yang tergolong dalam *cluster* kualitas pendidikan rendah, terutama dalam upaya meningkatkan partisipasi dan keberlanjutan pendidikan pada jenjang menengah. Di sisi lain, wilayah yang termasuk dalam *cluster* kualitas pendidikan tinggi juga memerlukan perhatian, khususnya terkait

tingginya rasio murid terhadap guru, yang menunjukkan perlunya penyeimbangan distribusi tenaga pendidik untuk menjaga dan meningkatkan kualitas proses pembelajaran.

Untuk penelitian selanjutnya, disarankan untuk melakukan pengembangan dengan menambahkan variabel lain yang relevan seperti Angka Partisipasi Kasar (APK), serta menggunakan pendekatan evaluasi *clustering* yang lebih beragam, sehingga hasil pengelompokan kualitas pendidikan dapat memberikan gambaran yang lebih komprehensif dan akurat.

DAFTAR PUSTAKA

- Alodia, D. A., Fialine, A. P., Endriani, D., & Widodo, E. (2021). Implementasi Metode K-Medoids Clustering untuk Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan. *Sepren*, 2(2), 1-13. <https://doi.org/10.36655/sepren.v2i2.606>
- Badan Pusat Statistik Indonesia. (2024). *Statistik Pendidikan Indonesia* 2024. <https://www.bps.go.id/id/publication/2024/11/22/c20eb87371b77ee79ea1fa86/statistik-pendidikan-2024.html>
- Bezdek, J. C., & Dunn, J. C. (1975). Optimal Fuzzy Partitions: A Heuristic for Estimating the Parameters in a Mixture of Normal Distributions. *IEEE Transactions on Computers*, C-24(8), 835-838. <https://doi.org/10.1109/T-C.1975.224317>
- Buana, E. D. (2023). *Penerapan Fuzzy C-Means Dan Fuzzy Subtractive Clustering Pada Pariwisata Di Jawa Timur Skripsi Oleh: Egra Dwi Buana Nim. 19610077 Program Studi Matematika Fakultas Sains Dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang* 2023.
- Creemers, B. P. M., & Kyriakides, L. (2009). Validity of Educational Indicators. Dalam *International Encyclopedia of Education, Third Edition* (hlm. 193-199). Elsevier. <https://doi.org/10.1016/B978-0-08-044894-7.00274-8>
- Febriyanti, N. A., Goejantoro, R., & Prangga, S. (2023). Optimasi Fuzzy C-Means Menggunakan Particle Swarm Optimization Untuk Pengelompokan Kabupaten/Kota Di Pulau Kalimantan (Studi Kasus: Data Indikator Kesejahteraan Rakyat Tahun 2020). *EKSPONENSIAL*, 14(1), 31. <https://doi.org/10.30872/eksponensial.v14i1.1095>

- Jayuning, F., Kusumanigtyas, M., & Basofi, A. (2019). *Politeknik Elektronika Negeri Surabaya Dinas Pendidikan Provinsi Jawa Timur*. 12(2), 1–8.
- Jipkate, B. R., & Gohokar, V. V. (2012). “ A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms .” *International Journal of Computational Engineering Research*, 2(3), 737–739.
- Munandar, S. A., Fitriani, A., Karlina, Y., Yumriani, & BP, A. R. (2022). Pengertian pendidikan, ilmu pendidikan, dan unsur-unsur pendidikan. *Al Urwatul Wutsqa: Kajian Pendidikan Islam*, 2(1), 1–8.
https://www.academia.edu/download/91021639/7757_24249_1_PB_2_.pdf
- Nurmin, D., Hayati, M. N., & Goejantoro, R. (2023). Penerapan Metode Fuzzy C-Means Pada Pengelompokan Kabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Kesejahteraan Rakyat Tahun 2020. *Ekspansional*, 14(1), 41.
<https://doi.org/10.30872/ekspansional.v14i1.1002>
- OECD. (2023). *PISA 2022 Results (Volume I)*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Prabowo, E., & Kurniawan, R. (2019). Fuzzy Clustering Algorithm Optimization Using Forest Optimization Algorithm. *Information System Development*, 4(1).
- Richia Putri, M., Gibran Satya Nugraha, & Ramaditia Dwiyanaputra. (2023). Pengelompokan Provinsi di Indonesia Berdasarkan Indikator Pendidikan Menggunakan Metode K-Means Clustering. *Journal of Computer Science and Informatics Engineering (J-Cosine)*, 7(1).
<https://doi.org/10.29303/jcosine.v7i1.509>
- Setiawan, R., Ernawati, E., & Efendi, R. (2018). Klasifikasi Kawasan Permukiman Tingkat Kelurahan untuk Pembangunan Sistem Berbasis Data Kualitas Permukiman (Studi Kasus: 67 Kelurahan di Kota Bengkulu). *Pseudocode*, 5(1), 45–55.
<https://doi.org/10.33369/pseudocode.5.1.45-55>
- Sumantri, G., Novianto, M. D., & Prihastuti, P. P. (2023). *Implementasi Fuzzy C-Means dalam Pengelompokan Provinsi di Indonesia untuk Pemerataan Kualitas Pendidikan*. 8(2721).
- Suwirmayanti, N. L. G. P. (2018). Penerapan Metode Fuzzy C-Means untuk Pengelompokan Data Kredit. *STMIK Pontianak Online Proceedings*, 390–395.
- Ulinnuha, N., & Novitasari, D. C. R. (2023). Penerapan Fuzzy C-Means Untuk Pengelompokan Tingkat Kualitas Pendidikan Di Jawa Timur. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 14(2), 419–426.
<https://doi.org/10.24176/simet.v14i2.9442>
- Valente de Oliveira, J., & Pedrycz, W. (2007). Advances in Fuzzy Clustering and its Applications. Dalam *Advances in Fuzzy Clustering and its Applications*.
<https://doi.org/10.1002/9780470061190>
- Wang, W., & Zhang, Y. (2007). On fuzzy cluster validity indices. *Fuzzy Sets and Systems*, 158(19), 2095–2117.
<https://doi.org/10.1016/j.fss.2007.03.004>