

IMPLEMENTASI FUZZY C-MEANS DALAM PENGELOMPOKAN TINGKAT KEMISKINAN PADA KABUPATEN/KOTA DI PROVINSI JAWA TENGAH

Rinaldo Dwi Faturahman*¹⁾, Nurtriana Hidayati²⁾

1. S1-Teknik Informatika, Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang
2. S1-Teknik Informatika, Fakultas Teknologi Informasi dan Komunikasi, Universitas Semarang

Article Info

Kata Kunci: *Clustering; Data Mining; Fuzzy C Means; Kemiskinan; Provinsi Jawa Tengah*

Keywords: *Central Java Province; Clustering; Fuzzy C Means; Data Mining; Poverty;*

Article history:

Received 02 Oktober 2024
Revised 10 November 2024
Accepted 1 Maret 2025
Available online 1 Maret 2025

DOI :

<https://doi.org/10.29100/jipi.v10i1.5747>

* Corresponding author.

Corresponding Author

E-mail address:

rinaldodwifaturahman@gmail.com

ABSTRAK

Kemiskinan merupakan salah satu permasalahan bagi negara berkembang khususnya di Indonesia. Setiap tahunnya kemiskinan di Provinsi Jawa Tengah terdapat kenaikan ataupun penurunan. Kemiskinan di Provinsi Jawa Tengah cukup tinggi. Dibuktikan pada data Badan Pusat Statistik Provinsi Jawa Tengah kemiskinan pada tahun 2022 sebesar 10,93% dan pada tahun 2021 sebesar 10,77%. Tujuan dari penelitian ini untuk mengklasifikasikan tingkat kemiskinan di Provinsi Jawa Tengah dengan metode *clustering* dengan menggunakan *fuzz c means*. *clustering* sendiri adalah salah satu teknik data mining. Dimana data ini merupakan sebuah metode dari data mining untuk mengelompokkan data menjadi beberapa kelompok berbeda berdasarkan karakteristik yang sama. Data penelitian yang digunakan diambil dari Badan Pusat Statistik Provinsi Jawa Tengah dari tahun 2021 – 2023 dengan total 106 data dan 5 atribut indikator kemiskinan yaitu garis kemiskinan (Rp/kapita/bln), jumlah penduduk miskin (ribu jiwa), pengeluaran, rata-rata pendidikan dan jumlah pengangguran. Hasil penelitian ini menghasilkan 5 cluster dengan beberapa data di dalamnya. Pada *cluster 0* dengan tingkat kemiskinan rendah terdapat 32 kabupaten/kota. Kemudian pada *cluster 1* terdapat 16 kabupaten/ kota dengan tingkat kemiskinan tinggi. *Cluster 2* terdapat 13 kabupaten/kota dengan tingkat kemiskinan sangat tinggi. *cluster 3* dengan 22 kabupaten/kota dengan kemiskinan yang sangat rendah. Dan *cluster 4* dengan 22 kabupaten/kota dengan kemiskinan yang sedang.

ABSTRACT

Poverty is one of the problems for developing countries, especially in Indonesia. Every year poverty in Central Java Province increases or decreases. Poverty in Central Java Province is quite high. As evidenced by the data from the Central Java Provincial Statistics Agency, poverty in 2022 was 10.93% and in 2021 it was 10.77%. The purpose of this study is to classify the level of poverty in Central Java Province using the clustering method using *fuzz c means*. *clustering* itself is one of the data mining techniques. Where this data is a method of data mining to group data into several different groups based on the same characteristics. The research data used was taken from the Central Java Provincial Statistics Agency from 2021 - 2023 with a total of 106 data and 5 poverty indicator attributes, namely the poverty line (IDR / capita / month), the number of poor people (thousand people), expenditure, average education and the number of unemployed. The results of this study produced 5 clusters with some data in them. In cluster 0 with a low poverty rate there are 32 districts/cities. Then in cluster 1 there are 16 districts/municipalities with a high poverty rate. Cluster 2 has 13 districts/municipalities with a very high poverty rate. cluster 3 has 22 districts/municipalities with a very low poverty rate. And cluster 4 with 22 districts/municipalities with moderate poverty.

I. PENDAHULUAN

SALAH satu permasalahan sosial yang kompleks dan beragam yang sering dihadapi banyak negara, termasuk Indonesia adalah kemiskinan. Kemiskinan mengacu pada ketidakmampuan individu atau komunitas untuk memenuhi kebutuhan penting, yang mencakup bidang-bidang seperti makanan, layanan kesehatan,

pendidikan, tempat tinggal, sumber daya alam, lingkungan, dan keselamatan. [1]. Pendapatan bukan satu-satunya faktor yang berkontribusi terhadap kemiskinan, faktor ini juga mencakup kendala pada pendidikan, kesehatan, pangan, air bersih, dan aspek lain yang berkaitan dengan kualitas hidup manusia. Sebagai salah satu provinsi terbesar di Indonesia, kemiskinan di Jawa Tengah bersifat kompleks dan berbeda-beda antar daerah/kota.

Menurut data BPS Provinsi Jawa Tengah (Badan Pusat Statistik) sendiri tingkat kemiskinan pada Provinsi Jawa Tengah Tahun 2022 sebanyak 10,93% dan pada tahun 2023 sebanyak 10,77%. Meskipun mengalami penurunan dibandingkan tahun sebelumnya, namun masih terdapat jarak yang signifikan dalam tingkat kemiskinan antar daerah. Sebagai contoh pada tahun 2022 presentase populasi yang paling terdapat di Kabupaten Wonosobo dengan presentase 16,17%, sementara di Kota Semarang hanya 4,25 %. Perbedaan tingkat kemiskinan antar daerah ini terpengaruhi oleh macam-macam faktor, seperti kondisi geografis, sumber daya alam, instruktur, akses terhadap pendidikan, kesehatan, serta karakteristik sosial-ekonomi masyarakat setempat. Oleh karena itu, upaya penanggulangan kemiskinan tiap daerah memerlukan strategi dan kebijakan yang disesuaikan dengan kondisi spesifik di setiap daerah.

Salah satu pendekatan yang digunakan untuk mengidentifikasi pola dan faktor-faktor yang mempengaruhi kemiskinan di setiap daerah adalah dengan memanfaatkan teknik data mining. Data mining sebuah proses mencari serta menganalisis data untuk mengidentifikasi pola yang menarik yang bertujuan menadapatkan informasi serta pengetahuan yang akurat dan potensial, agar mudah dipahami serta bermanfaat dalam membuat keputusan[2]. Proses pengumpulan dan ekstraksi informasi ini dapat dilakukan menggunakan software dengan bantuan perhitungan statistika, matematika ataupun menggunakan Artificial intelligence (AI). Data mining sendiri memiliki banyak sekali jenis algoritma yang dapat digunakan salah satunya adalah clustering. Clustering merupakan teknik untuk membagi data dalam beberapa grup yang memiliki karakteristik objek yang sama didalamnya[3]. Tujuan utama dari metode clustering ialah mengelompokkan beberapa data atau objek ke dalam cluster sehingga pada setiap cluster dapat berisikan data yang mirip atau sama. Terdapat banyak algoritma dalam metode clustering yang dapat digunakan pada studi sebelumnya seperti Fuzzy C-Means, K-Means, K-Medoids, Hierrarchical Agglomerative, dan sebagainya. Fuzzy C Means (FCM) merupakan teknik klustering data di mana tiap titik-titik data dalam suatu cluster ditentukan oleh derajat keanggotaannya[4]. Fuzzy C Means sendiri memiliki kelebihan, yaitu tingkat akurasi yang tinggi. Sehingga metode ini memungkinkan identifikasi pola-pola yang mungkin tersembunyi dalam data dan memberikan pemahaman yang lebih baik tentang distribusi kemiskinan di berbagai wilayah di Provinsi Jawa Tengah.

Beberapa penelitian sebelumnya telah dilakukan mengenai kemiskinan itu sendiri. Pada penelitian bertajuk “Implementasi Data Mining untuk Menentukan Tingkat Kemiskinan Menggunakan Fuzzy C-Means” yang dilakukan oleh Nurrahman Fitriani Kahar, Lillyan Hadjaratie, Sitti Suhada dan Indhitiya R. Padiku, hasilnya dibagi menjadi 3 cluster yaitu 50% sangat miskin, 34% miskin, 16% hampir miskin[5].

Dalam analisis berjudul “Penerapan Algoritma Clustering Pengelompokan Tingkat Kemiskinan di Provinsi Banten” yang diterbitkan oleh Tb Ai Munandar. Dalam penelitian ini, tiga kelompok berbeda dihasilkan. Tingkat kemiskinan rendah terdapat di Kota Serang, Kota Cilegon, dan Kota Tangerang selatan, sedangkan kelompok miskin sedang terdapat di Kabupaten Pandenglang, Kabupaten Lebak, dan Kabupaten Serang. Kabupaten Tangerang, dan Kota Tangerang Selatan mempunyai kelompok yang memiliki tingkat kemiskinan yang tinggi[6].

Kajian yang dilakukan Nurhafizah Sepriyanti, Rahma Sani Nahampun, dan rekannya berfokus pada pemanfaatan K-means clustering untuk mengkategorikan tingkat kemiskinan di Provinsi Riau. Studi mereka menghasilkan tiga kelompok berbeda. Cluster 1 termasuk dalam kategori tingkat kemiskinan rendah karena jumlah penduduk miskin yang sedikit, tingkat pengangguran yang rendah, dan durasi pendidikan rata-rata yang moderat. Kelompok 2 terdiri dari individu-individu yang tergolong cukup miskin, ditandai dengan tingkat pengangguran yang lebih tinggi dan rata-rata lama pendidikan yang lebih lama. Di sisi lain, klaster 3 tergolong memiliki tingkat kemiskinan yang tinggi, karena mencakup jumlah penduduk miskin terbesar, tingkat pengangguran sedang, dan rata-rata lama pendidikan yang lebih rendah.[7].

Penelitian yang dilakukan oleh Edy Widodo, Putri Ermayani, dkk dengan judul “Pengelompokan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative clustering”. Hasil penelitian ini terdapat 3 cluster dengan cluster 1 sebanyak 25 daerah yang memiliki tingkat kemiskinan rendah, cluster 2 sebanyak 7 daerah yang memiliki menghadapi kemiskinan sedang, dan cluster 3 sebanyak 2 daerah yang memiliki tingkat kemiskinan tinggi[8].

Studi yang diselesaikan oleh F. Fajriani dengan judul “Persebaran Tingkat Kemiskinan di Sulawesi Selatan Menggunakan K-Means Clustering Analysis”. Hasil penelitian tersebut adalah 24 Kabupate/Kota di Sulawesi Selatan, yang memiliki kategori kemiskinan tinggi berjumlah 7 wilayah, sedangkan terdapat 18 daerah dengan kategori kemiskinan yang rendah[9].

Sehubungan pada studi sebelumnya, dalam studi ini relevan dengan studi sebelumnya yang telah menggunakan berbagai metode clustering untuk mengelompokkan tingkat kemiskinan di berbagai daerah, seperti Provinsi

Banten, Riau, dan Sulawesi Selatan, serta tingkat nasional di Indonesia. Penelitian sebelumnya menggunakan berbagai metode seperti K-Means, k-medoids, dan Hierarchical Agglomerative Clustering, menunjukkan efektivitas berbagai teknik dalam analisis kemiskinan. Namun, penelitian ini mengisi celah penting dengan fokus khusus pada Provinsi Jawa Tengah, menggunakan metode Fuzzy C-Means yang memiliki keunggulan dalam menangani data yang tidak terstruktur dan memberikan hasil clustering yang lebih halus. Selain itu, penelitian ini memperdalam analisis hingga ke tingkat kabupaten/kota, memberikan detail yang belum banyak dibahas dalam penelitian sebelumnya. Dengan demikian, penelitian ini tidak hanya memperluas penerapan metode Fuzzy C-Means tetapi juga memberikan kontribusi yang signifikan dalam pemahaman distribusi kemiskinan di wilayah yang berbeda, khususnya di Jawa Tengah.

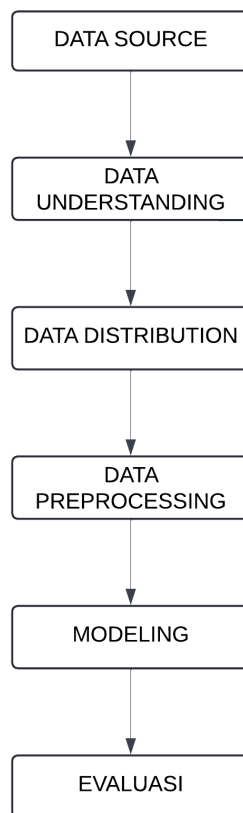
Tujuan pada studi ini ialah untuk menerapkan teknik *fuzzy c-means* dalam mengelompokkan strata kemiskinan di kabupaten atau kota di Provinsi Jawa Tengah berdasarkan indikator determinan. Klasifikasi seperti ini diharapkan dapat memberikan informasi berharga bagi instansi pemerintah yang memungkinkan mereka memiliki wawasan yang lebih baik mengenai pola dan sifat kemiskinan di setiap wilayah sehingga membantu mereka dalam merancang program dan kebijakan pengentasan kemiskinan yang lebih disesuaikan dan efisien. Selain itu, informasi tersebut akan membantu masyarakat melihat apa yang menyebabkan kemiskinan, nantinya mereka dapat menggunakan informasi tersebut untuk mengidentifikasi jalan keluar dari kemiskinan.

II. METODE PENELITIAN

Studi ini merupakan dari sebuah proses data mining dengan menggunakan sebuah metode clustering dan memanfaatkan algoritma Fuzzy C-Means.

A. Metode Fuzzy C-Means Clustering

Fuzzy C-Means (FCM) merupakan metode yang tergolong dalam kategori *unsupervised learning*[10]. Metode ini memungkinkan data untuk dimasukkan ke dalam cluster yang berbeda serta memiliki tingkat keanggotaan yang berbeda. Keuntungan dari fuzzy c-means ialah mereka menemukan pusat cluster lebih cepat dibandingkan metode clustering lainnya [11]. Pada gambar 1 menunjukan tahapan fuzzy c-means.



Gambar. 1. PROSES FUZZY C-MEANS

Pada gambar 1 tahapan fuzzy c-means clustering ini, dimana pada gambar ini terdapat beberapa tahapan mulai dari data source, dimana data source ini digunakan untuk pengambilan data untuk penelitian ini. Kemudian terdapat data understanding, yang digunakan untuk mengidentifikasi jenis data. Data distribution digunakan untuk melihat

persebaran data pada data set. Data preprocessing untuk mengubah data untuk di analisis guna melakukan modeling. Yang terakhir terdapat modeling yang menerapkan fuzzy c-means serta menentukan jumlah cluster dari data yang sudah siap.

B. Data Source

Studi ini memanfaatkan data indikator kemiskinan Provinsi Jawa Tengah dari tahun 2021-2023 yang diperoleh dari Badan Pusat Statistik (BPS)[12]. Populasinya yang digunakan dalam studi ini adalah wilayah yang berada di Provinsi Jawa Tengah. Dalam dataset ini terdiri dari 105 data dan 7 atribut. Ketujuh atribut tersebut antara lain kabupaen/kota, garis kemiskinan(Rp/kapita/bln), jumlah penduduk miskin (ribu jiwa), pengeluaran, rata-rata pendidikan, jumlah pengangguran, dan tahun.

C. Data Understanding

Data understanding merupakan tahapan awal, untuk melakukan pengecekan pada data yang digunakan, mengumpulkan data awal, dan mengidentifikasi kualitas data. Dalam tahapan data understanding sebagai berikut:

1) Missing Value Identification

Untuk mengidentifikasi adanya missing value (data hilang) pada data yang ada. Biasanya *missing value* ini terdapat pada sebuah kolom. Dimana *missing value* ini dapat di atasi dengan beberapa metode yaitu dengan metode imputasi dengan mengisikan rata-rata atau dengan median dari kolom tersebut atau juga dapat di eliminasi data.

2) Duplicated Data Identification.

Untuk mengidentifikasi adanya Duplicate data (data yang sama) pada data yang ada. Biasanya *duplicated data* ini ditemukan data yang sama pada sebuah kolom. *Duplicated data* ini dapat di atasi dengan menghapus data pada kolom yang terdapat data yang sama

D. Data Distribution

Data distribusi (data distribution) adalah cara data tersebar atau terdistribusi di dalam suatu rentangan nilai. Ini menggambarkan frekuensi atau probabilitas kemunculan nilai-nilai tertentu dalam suatu set data. Distribusi data penting untuk memahami karakteristik dasar dari data dan untuk melakukan analisis statistik yang lebih lanjut.

E. Data Preprocessing

Preprocessing data merupakan penyiapan data mentah atau data yang sudah diperoleh, berbagai bentuk data yang berbeda menjadi data biasa atau menghilangkan data yang tidak diperoleh atau tidak sesuai dengan bentuk lain agar data tersebut dapat diolah menjadi data yang mudah untuk dikenali atau diterapkan.[13]. Dalam proses data preprocessing ada beberapa tahapan seperti :

1) Correlation Data

Suatu ukuran statistik yang menggambarkan sejauh mana dua variable berhubungan satu sama lain.

2) Scalling Data

Proses transformasi data sehingga data tersebut berada dalam skala atau rentang tertentu. Studi ini menggunakan z-score dalam melakukan proses scalling data karena data yang digunakan memiliki distribusi yang normal dan skala yang berbeda.

F. Modeling

Pemodelan ini dilakukan untuk mengidentifikasi dan menerapkan teknik dan algoritma data mining dengan menentukan parameter dengan nilai optimal untuk digunakan. [14]. Modeling ini menggunakan fuzzy c-means yang merupakan pengembangan dari jenis clustering k-means, dimana setiap data menjadi anggota dari semua cluster dengan derajat keanggotaan tertentu. Dasar dari menghitung teknik fuzzy c-means adalah sebagai berikut[15]:

1) *Input data yang akan di-cluster X, berupa matriks berukuran $n \times p$ (n =jumlah sampel data, p =atribut setiap data). X_{kj} = data sampel ke- k ($k = 1, 2, \dots, n$), atribut ke- j ($j = 1, 2, 3, \dots, m$).*

2) Menentukan:

Fuzzy C-Means (FCM) adalah metode clustering yang memerlukan pemilihan parameter penting seperti c (jumlah cluster) dan m (pangkat pembobot). Nilai c ditentukan berdasarkan pengetahuan domain atau melalui analisis eksploratif seperti metode siluet atau elbow untuk menemukan jumlah cluster yang optimal. Parameter m , yang biasanya dipilih dalam rentang 1.5 hingga 2.5, mengontrol tingkat fuzzy keanggotaan data dalam cluster. Pemilihan nilai m umumnya dilakukan melalui eksperimen, dengan nilai yang lebih rendah mendekati 1 membuat FCM lebih deterministik seperti K-Means.

- a) Jumlah Cluster = c ;
- b) Pangkat pembobot = m ;
- c) Maksimum iterasi = $MaxIter$;
- d) Error terkecil yang diharapkan = ξ ;
- e) Fungsi object awal = $P_0 = 0$;
- f) Iterasi awal = $t = 1$

3) Bangkitakan bilangan acak ($\mu_{ik}, I = 1, 2, \dots, c; k = 1, 2, \dots, n$) sebagai elemen matriks partisi awal U

$$U_0 = \begin{bmatrix} \mu_{11}(X_1) & \mu_{12}(X_2) & \dots & \mu_{1c}(X_c) \\ \vdots & \vdots & & \vdots \\ \mu_{11}(X_1) & \mu_{12}(X_2) & \dots & \mu_{1c}(X_c) \end{bmatrix}$$

Matriks partisi pada *fuzzy clustering* harus memenuhi kondisi sebagai berikut:

$$\mu_{ik} = [0,1]; (1 \leq i \leq c; 1 \leq k \leq n)$$

$$\sum_{i=1}^n \mu_{ik} = 1; 1 \leq i \leq c$$

$$0 < \sum_{i=1}^c \mu_{ik} = c; 1 \leq k \leq n$$

Hitung jumlah setiap kolom (atribut):

$$Q_j = \sum_{i=1}^c \mu_{ik}$$

Dengan $j = 1, 2, 3, \dots, m$

Kemudian hitung :

$$\mu_{ik} = \frac{\mu_{ik}}{Q_j}$$

Keterangan:

μ_{ik} : elemen-elemen dari matriks partisi U atau fungsi keanggotaan data ke- k ($k = 1, 2, 3, \dots, n$) pada cluster ke- i ($i = 1, 2, 3, \dots, c$).

n : jumlah data

c : jumlah cluster

k : fungsi keanggotaan

4) Hitung pusat cluster ke- k : V_{ij} , dimana $i=1, 2, 3, \dots, c$ dan $j=1, 2, 3, \dots, m$

$$V_{ij} = \frac{\sum_{k=1}^n ((\mu_{ik})^m * X_{kj})}{\sum_{k=1}^n (\mu_{ik})^m}$$

$$V = \begin{bmatrix} V_{11} & \dots & V_{1m} \\ \vdots & \ddots & \vdots \\ V_{c1} & \dots & V_{cm} \end{bmatrix}$$

Keterangan:

v : matriks pusat cluster

n : jumlah data

μ_{ik} : elemen-elemen dari matriks partisi U atau fungsi keanggotaan data ke- k ($k = 1, 2, 3, \dots, n$) pada cluster ke- i ($i = 1, 2, 3, \dots, c$).

k : fungsi keanggotaan

X_{kj} : data sampel ke- k ($k = 1, 2, \dots, n$), atribut ke- j ($j = 1, 2, 3, \dots, m$).

5) Hitung fungsi objektif pada iterasi ke- t , P_t dengan menggunakan persamaan sebagai berikut:

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

Keterangan:

- P : atribut setiap data
- k : fungsi keanggotaan
- i : cluster ke-i
- j : atribut ke-j
- n : jumlah data
- c : jumlah cluster
- m : pangkat pembobot
- X_{kj} : data sampel ke-k ($k = 1, 2, \dots, n$), atribut ke-j ($j = 1, 2, 3, \dots, m$).

6) Hitung perubahan matriks partisi:

$$\mu_{ik} = \frac{[\sum_{j=1}^p (X_{kj} - V_{ij})^2]^{-\frac{1}{p-1}}}{\sum_{i=1}^c [\sum_{j=1}^p (X_{kj} - V_{ij})^2]^{-\frac{1}{p-1}}}$$

Keterangan:

- μ_{ik} : elemen-elemen dari matriks partisi U atau fungsi keanggotaan data ke-k ($k = 1, 2, 3, \dots, n$) pada cluster ke-i $i = 1, 2, 3, \dots, c$
- P : atribut setiap data
- j : atribut ke-j
- X_{kj} : data sampel ke-k ($k = 1, 2, \dots, n$), atribut ke-j ($j = 1, 2, 3, \dots, m$).
- i : cluster ke-i

7) Cek kondisi berhenti:

- a) Jika $(|P_t - P_{t-1}| < \xi$ atau $(t < \text{iterasi maksimal})$ maka berhenti;
- b) Jika tidak: maka $t=t+1$ kemudian ulangi langkah ke-4;

G. Evaluasi

Tahapan evaluasi dalam penelitian ini untuk mengevaluasi hasil, meninjau proses serta menentukan langkah selanjutnya[16]. Metode pengelompokan fuzzy berbeda dari metode lain karena pendekatannya yang unik. Dalam metode ini, setiap titik data diberi nilai derajat keanggotaan, sehingga memungkinkannya menjadi anggota beberapa cluster secara bersamaan. Ini berarti bahwa suatu titik data mungkin tidak secara eksklusif dimiliki oleh satu kluster, namun memiliki derajat keanggotaan dalam beberapa kluster. Akibatnya, indeks validitas yang dirancang khusus untuk fuzzy c-means menjadi diperlukan[17]. Berikut metode evaluasi yang digunakan:

1) Partition Coefficient (PC)

Clustering fuzzy dapat divalidasi dengan menghitung koefisien partisi (PC) untuk mengevaluasi keanggotaan data di setiap cluster. Kualitas cluster lebih baik dengan nilai yang lebih tinggi, yaitu berkisar antara 0 hingga 1 untuk PC.[18].

2) Partition Entropy (PE)

Teknik pertama yang berhubungan dengan algoritma fuzzy c-means adalah tentang menentukan jumlah cluster yang optimal. Hal ini dapat dipahami dengan melihat nilai suatu ukuran tertentu yang terjadi jika mendekati angka 0, dan nyatanya semakin rendah nilai yang diperoleh maka semakin baik kualitas cluster tersebut.[19]

3) Modified Partition Coefficient (MPC)

Salah satu metrik evaluasi untuk mengukur kualitas partisi hasil pengelompokan data mining[20]. MPC menggambarkan seberapa baik objek-objek yang serupa dikelompokkan bersama dalam sebuah cluster dan seberapa berbeda cluster-cluster yang berbeda di antara mereka. Nilai MPC berkisar dari -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan partisi yang lebih baik atau kualitas clustering yang lebih baik.

III. HASIL DAN PEMBAHASAN PENELITIAN

Banyaknya data yang digunakan untuk studi ini adalah sebanyak 105 data dan atribut data yang di dapatkan dari website BPS (Badan Pusat Statistik) Provinsi Jawa Tengah mulai tahun 2021-2023. Atribut data yang dijadikan indikator kemiskinan ini adalah Kabupaten/Kota, garis kemiskinan (Rp/Kapita/Bln) atau pendapatan, jumlah penduduk miskin(ribu jiwa), pengeluaran, rata-rata pendidikan, jumlah pengangguran, serta tahun. Data dapat dilihat dalam Tabel 1.

TABEL I

DATA KEMISKINAN

Kabupaten/Kota	Garis Kemiskinan (Rp/Kapita/Bln)	Jumlah Penduduk Miskin (Ribu Jiwa)	Pengeluaran	Rata-rata pendidikan	Jumlah Pengangguran	Tahun
Kabupaten Cilacap	363367.00	201.71	830026.00	7.09	82714.00	2021
Kabupaten Banyumas	417086.00	232.91	1023973.00	7.63	52390.00	2021
Kabupaten Purbalingga	384183.00	153.08	900102.00	7.25	30450.00	2021
Kabupaten Banjarnegara	328679.00	150.19	975274.00	6.75	29281.00	2021
Kabupaten Kebumen	390599.0	212.92	920590.0	7.55	37408.00	2021
Kota Semarang	642456.00	80.53	1787507.00	10.81	55656.00	2023
Kota Pekalongan	565998.00	21.36	1348573.00	9.29	9065.00	2023
Kota Tegal	623617.00	19.22	1750336.00	9.24	8805.00	2023

Pemilihan atribut di atas seperti garis kemiskinan, jumlah penduduk miskin, pengeluaran, rata-rata pendidikan dan jumlah pengangguran sangat penting dalam mengukur tingkat kemiskinan karena setiap atribut mencerminkan aspek penting dari kondisi ekonomi dan sosial yang menyebabkan kemiskinan. Garis kemiskinan menentukan ambang batas minimum yang diperlukan untuk memenuhi kebutuhan dasar, sedangkan jumlah penduduk miskin secara langsung mengukur besarnya kemiskinan. Pengeluaran mencerminkan kemampuan finansial individu atau keluarga, sementara pendidikan merupakan faktor kunci dalam mobilitas sosial dan dapat mengurangi kemiskinan antargenerasi. Banyaknya pengangguran menunjukkan kemampuan pasar tenaga kerja untuk menyerap angkatan kerja yang ada, dan tingginya angka pengangguran seringkali berhubungan langsung dengan kemiskinan.

A. Data Understanding

Data yang di kumpulkan harus melewati tahapan data understanding terlebih dahulu dimana pada proses data understanding melibatkan pengecekan pada data untuk mengetahui apakah terdapat data yang missing value (data yang hilang) dan duplicated data (data yang sama).

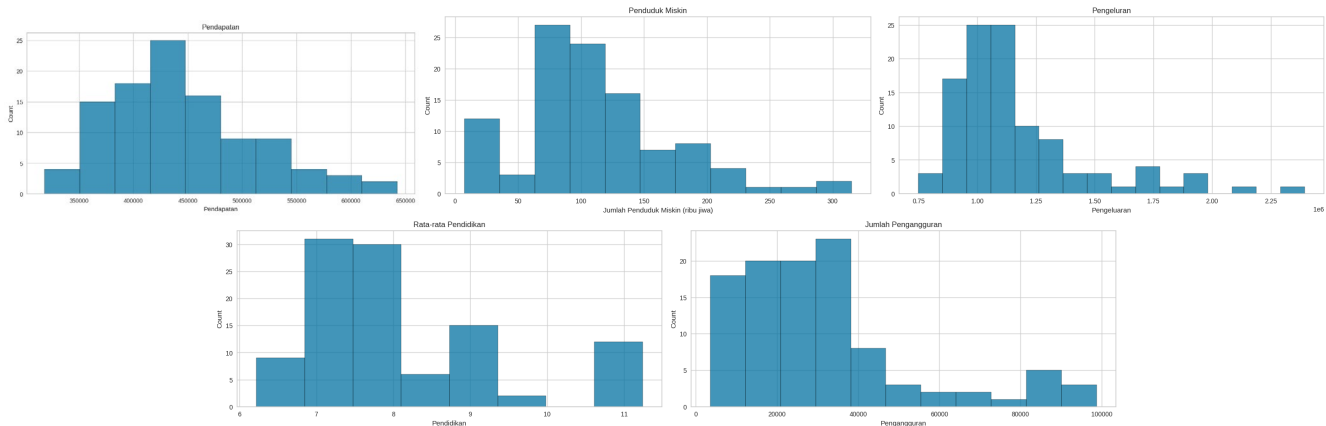
TABEL II Informasi Dataset

Column	Non-Null Count	Dtype
Kabupaten / Kota	105 non-null	object
Garis Kemiskinan (Rp/Kapita/Bln)	105 non-null	float64
Jumlah Penduduk Miskin (ribu jiwa)	105 non-null	float64
Pengeluaran	105 non-null	float64
Rata-rata pendidikan	105 non-null	float64
Jumlah Pengangguran	105 non-null	float64
Tahun	105 non-null	Int64

Pada table 2 menunjukkan atribut dan 107 baris data. Dari semua atribut sudah sesuai dengan tipe data tersebut. Terdapat 1 atribut dengan kategorikal dan atribut dengan numerik. Atribut kategorikal adalah kabupaten/kota, sedangkan atribut dengan numerik adalah garis kemiskinan(Rp.kapita/bln), jumlah penduduk miskin (ribu jiwa), pengeluaran, rata-rata pendidikan, jumlah pengangguran, serta tahun. pada data understanding ini tidak ditemukan missing value dan duplicated data.

B. Data Distribution

Tahapan ini digunakan untuk melihat tersebarnya data-data dari indikator kemiskinan. Disini penulis menggunakan hisplot untuk melihat penyebarannya. Hisplot sendiri merupakan alat yang kuat dalam eksplorasi data dan analisis deskriptif, serta memberikan gambaran visual tentang bagaimana data itu tersebar. Berikut adalah gambaran dari data distribution menggunakan hisplot pada indikator kemiskinan.



Gambar. 2 DATA DISTRIBUTION PADA INDIKATOR KEMISKINAN

C. Data Preprocessing

Pada tahapan ini, proses yang dilakukan adalah untuk persiapan data dilakukan dengan menyesuaikan *dataset* untuk memenuhi syarat yang akan digunakan saat tahapan modeling. pada tahapan ini terdiri dari:

1) Correlation Matrix

Correlation matrix ini digunakan untuk merepresentasikan visual dari korelasi antar pasangan dalam sebuah dataset. Pada penelitian ini korelasi matriks dapat digunakan untuk memahami hubungan antar variable-variable dalam indikator tingkat kemiskinan. Korelasi adalah ukuran statistik yang diwakili oleh koefisien korelasi Pearson, yang bervariasi antara -1 dan 1. Nilai 1 menunjukkan korelasi positif mutlak; sebaliknya, -1 menunjukkan korelasi negatif mutlak, sedangkan 0 menunjukkan tidak ada hubungan antar variabel.



Gambar. 3 CORRELATION MATRIX

Berdasarkan gambar 3 menunjukkan bahwa corelasi antara pendapatan dan pengeluaran dengan nilai 0.8 menunjukkan hubungan positif yang kuat yang artinya bahwa pendapatan lebih tinggi akan lebih banyak dalam pengeluaran. Corelasi antara pendapatan dan pendidikan dengan nilai 0.7 menunjukkan hubungan positif yang kuat artinya bahwa dengan pendapatan lebih tinggi cenderung memiliki tingkat pendidikan yang lebih tinggi secara umum. Kemudian corelasi jumlah penduduk miskin dan pengangguran memiliki hubungan yang cukup kuat karena memiliki nilai 0.7 yang menunjukkan bahwa kenaikan jumlah penduduk miskin berbanding lurus dengan kenaikan tingkat pengangguran. Sedangkan corelasi pengeluaran dan pendidikan memiliki hubungan yang kuat dengan nilai 0.8 yang artinya bahwa pengeluaran yang lebih tinggi cenderung dikaitkan dengan tingkat pendidikan yang lebih tinggi.

2) Scaling data

Scaling data ini digunakan untuk proses transformasi fitur dalam dataset sehingga berada dalam skala atau rentang yang sama. Dimana *scaling ini* menggunakan *StandartScaler*. Penggunaan *StandartScaler* karena data yang dimiliki mempunyai distribusi data yang normal. Hasil dari ini adalah untuk menghitung mean dan standar deviasi dari dataset tersebut. Berikut hasil dari *scaling data*:

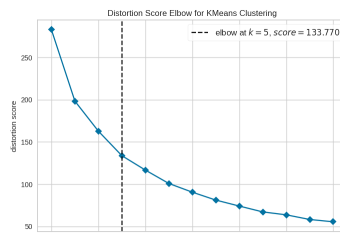
TABEL III HASIL DATA SETELAH MELAKUKAN SCALING DATA

Kabupaten/Kota	Garis Kemiskinan (Rp/Kapita/Bln)	Jumlah Penduduk Miskin (Ribu Jiwa)	Pengeluaran	Rata-rata pendidikan	Jumlah Pengangguran	Tahun
Kabupaten Cilacap	-1.208027	1.454606	-1.163019	-0.820296	82714.00	2021
Kabupaten Banyumas	-0.415618	1.959039	-0.497379	-0.392335	52390.00	2021
Kabupaten Purbalingga	-0.900970	0.668369	-0.922513	-0.693493	30450.00	2021
Kabupaten Banjarnegara	-1.719709	0.621645	-0.664517	-1.089753	29281.00	2021
Kabupaten Kebumen	-0.806327	1.635846	-0.852197	-0.455737	37408.00	2021

Kota Semarang	2.908815	-0.504599	2.123125	2.127879	1.122017	2023
Kota Pekalongan	1.780983	-1.461244	0.616672	0.923248	-1.029742	2023
Kota Tegal	2.630921	-1.495843	1.995551	0.883622	-1.041750	2023

D. Modeling

Pada tahap ini, pola data mining yang dihasilkan melalui pemodelan *clustering* adalah algoritma *fuzzy c means*. Langkah awal adalah menetapkan berapa banyak *cluster* yang ada. Untuk menghitung jumlah *cluster* penulis menggunakan *elbow* seperti yang di tunjukkan oleh gambar 4 di bawah ini.



Gambar. 4 MENANTUKAN JUMLAH CLUSTER DENGAN ELBOW

Pada grafik gambar 4 dapat dilihat hasil dari grafik tersebut. *Cluster* yang diperoleh sebanyak 5 *cluster* dengan skor 133,770. Ini dapat diidentifikasi dengan adanya perubahan yang signifikan dalam penurunan *distortion score* sebelum titik tersebut dan perlambatan penurunan setelah titik tersebut. Skor dan *cluster* didapatkan melalui *elbow*, jadi penelitian ini menggunakan 5 *cluster* sebagai nilai *k* untuk pemodelan *fuzzy c means*. selah menentukan *cluster*, tahap selanjutnya adalah pemodelan dengan menggunakan algoritma *fuzzy c means* pada dataset ditunjukkan pada gambar 5.

```
fcm = FCM(n_clusters=5, m=2.0, max_iter=300, random_state=42)
fcm.fit(data2)
```

Gambar. 5 SOURCODE PEMODELAN FUZZY C MEANS

Pada gambar 5 merupakan sourcode untuk mendapatkan hasil yang optimal. Parameter *fuzziner (m)* memiliki nilai 2.0, nilai ini sering digunakan karena memberikan keseimbangan yang baik antar kejelasan klaster. *Stopping criteria(max_iter)* dipilih untuk memastikan bahwa algoritma memiliki cukup iterasi untuk mencapai konvergensi. Dan *random_state* untuk memastikan hasil yang dapat dibuat.

```
fcmModel = FCM(n_clusters=5, m=2.0, max_iter=300, random_state=42)
fcmModel.fit(data2)
center = fcmModel.centers
center
```

Gambar. 6 Sourcode Cluster center

Pada gambar 6 ini digunakan untuk mengelompokkan dataset pada data2 menjadi 5 *cluster*. Algoritma ini diinisialisasi dengan beberapa parameter, termasuk jumlah klaster (*n_clusters=5*), parameter fuzziness (*m=2.0*), jumlah iterasi maksimum (*max_iter=300*), dan *random_state=42* untuk memastikan hasil yang konsisten. Setelah model FCM diinisialisasi, model tersebut di-fit atau dilatih menggunakan dataset data2 dengan perintah *fcmModel.fit(data2)*. Setelah proses pelatihan selesai, pusat-pusat klaster (*cluster centers*) diekstraksi menggunakan *fcmModel.centers* dan disimpan dalam variabel *center*. Pusat-pusat klaster ini merupakan titik-titik yang mewakili posisi sentral dari masing-masing klaster dalam ruang data, yang dapat digunakan untuk analisis lebih lanjut atau visualisasi hasil pengelompokan.

TABEL IV Hasil cluster center

Cluster	Jumlah Kabupaten/Kota	Garis Kemiskinan (Rp/Kapita/Bln)	Jumlah Penduduk Miskin (Ribu Jiwa)	Pengeluaran	Rata-rata pendidikan	Jumlah Pengangguran
0	32	-0.77221302	-0.1397652	-0.51620771	-0.47730891	-0.39041275
1	16	1.65900819	-1.34982953	1.87926848	1.97488421	-0.80404669
2	13	-0.13481072	1.66455591	-0.41968252	-0.79649645	1.87486378
3	22	0.39854432	-0.44446166	0.18760854	0.36487557	-0.36739458
4	22	-0.34089817	0.54652151	-0.48691599	-0.44582189	0.18534651

Dari hasil tabel 4 memperlihatkan bahwa algoritma FCM membentuk 5 cluster dan menghasilkan *cluster 0* mempunyai data sebanyak 32, *cluster 1* memiliki data sebanyak 16, *cluster 2* memiliki data sebanyak 13, *cluster 3* dengan data sebanyak 22, dan *cluster 4* dengan data sebanyak 22

TABEL V Cluster berdasarkan Kabupaten/Kota

Kabupaten/Kota	Garis Kemiskinan (Rp/Kapita/Bln)	Jumlah Penduduk Miskin (Ribu Jiwa)	Pengeluaran	Rata-rata pendidikan	Jumlah Pengangguran	Tahun	Cluster
Kabupaten Cilacap	-1.208027	1.454606	-1.163019	-0.820296	82714.00	2021	2
Kabupaten Banyumas	-0.415618	1.959039	-0.497379	-0.392335	52390.00	2021	2
Kabupaten Purbalingga	-0.900970	0.668369	-0.922513	-0.693493	30450.00	2021	4
Kabupaten Banjarnegara	-1.719709	0.621645	-0.664517	-1.089753	29281.00	2021	0
Kabupaten Kebumen	-0.806327	1.635846	-0.852197	-0.455737	37408.00	2021	4
Kota Semarang	2.908815	-0.504599	2.123125	2.127879	1.122017	2023	1
Kota Pekalongan	1.780983	-1.461244	0.616672	0.923248	-1.029742	2023	1
Kota Tegal	2.630921	-1.495843	1.995551	0.883622	-1.041750	2023	1

Dari hasil nilai *cluster center* dapat dilihat bahwa *cluster 0* memiliki kemiskinan yang rendah dengan ciri-ciri garis kemiskinan(Rp/Kapita/Bln) sangat rendah, Jumlah penduduk miskin rendah, pengeluaran rendah, rata-rata pendidikan rendah, jumlah pengangguran rendah. *Cluster 1* memiliki kemiskinan yang tinggi dengan ciri-ciri garis kemiskinan(Rp/Kapita/Bln) sangat tinggi, Jumlah penduduk miskin sangat rendah, pengeluaran sangat tinggi, rata-rata pendidikan sangat tinggi, jumlah pengangguran sangat rendah. *Cluster 2* memiliki kemiskinan yang sangat tinggi dengan ciri-ciri garis kemiskinan(Rp/Kapita/Bln) rendah, Jumlah penduduk miskin sangat tinggi, pengeluaran rendah, rata-rata pendidikan sangat rendah, jumlah pengangguran sangat tinggi. *Cluster 3* memiliki kemiskinan yang sangat rendah dengan ciri-ciri garis kemiskinan(Rp/Kapita/Bln) tinggi, Jumlah penduduk miskin rendah, pengeluaran tinggi, rata-rata pendidikan sedang, jumlah pengangguran rendah. *Cluster 4* memiliki kemiskinan yang sedang dengan ciri-ciri garis kemiskinan(Rp/Kapita/Bln) rendah, Jumlah penduduk miskin tinggi, pengeluaran rendah, rata-rata pendidikan rendah, jumlah pengangguran sedang.

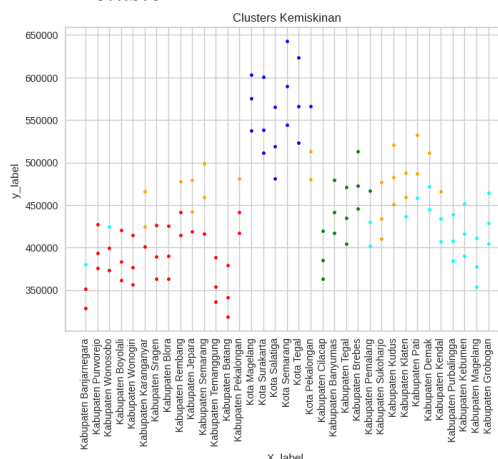
```
labels = np.argmax(fcm.u, axis=1)
labels

array([2, 2, 4, 0, 4, 0, 0, 4, 0, 4, 3, 0, 0, 0, 4, 0, 0, 4, 3, 0, 4, 0,
       0, 4, 0, 0, 4, 2, 2, 1, 1, 1, 1, 3, 1, 2, 2, 4, 0, 4, 0, 0, 4, 0,
       3, 3, 0, 3, 0, 4, 0, 0, 3, 3, 3, 4, 3, 0, 4, 0, 0, 4, 2, 2, 1, 1,
       1, 1, 3, 1, 2, 2, 4, 4, 4, 0, 4, 4, 0, 3, 3, 0, 3, 0, 4, 0, 3, 3,
       3, 3, 3, 3, 0, 3, 0, 3, 2, 2, 2, 1, 1, 1, 1, 1, 1])
```

Gambar. 7 labels cluster

Gambar 7 ini untuk merupakan hasil dari proses klastering menggunakan algoritma Fuzzy C-Means (FCM) pada sebuah dataset. Setiap nilai dalam array tersebut mewakili label klaster untuk setiap data point setelah proses klastering selesai dilakukan. Misalnya, nilai 0 menunjukkan bahwa data point tersebut termasuk dalam klaster pertama, nilai 1 menunjukkan klaster kedua, dan seterusnya.

Kemudian setelah tahapan pemodelan *cluster* menggunakan *fuzzy c means*. terdapat 5 titik *cluster* yang terbentuk dari hasil pemodelan tersebut dengan tiap titik *cluster* dengan berbagai warna yang merupakan penyebaran datanya. Berikut gambar 8 penyebaran data *cluster*.



Gambar. 8 PENYEBARAN DATA CLUSTER

Setelah pengujian keseluruhan selesai, hasil informasi *cluster* yang menunjukkan area mana saja yang termasuk dalam *cluster* penyebaran serta tiap area dengan masing masing *cluster*. *Cluster 0* menunjukkan tingkat kemiskinan yang rendah, *cluster 1* menunjukkan tingkat kemiskinan yang tinggi, *cluster 2* menunjukkan tingkat kemiskinan yang sedang, dan *cluster 3* menunjukkan tingkat kemiskinan yang saat tinggi. Untuk penyebaran *cluster* berdasarkan kabupaten/kota dapat dilihat pada tabel 3

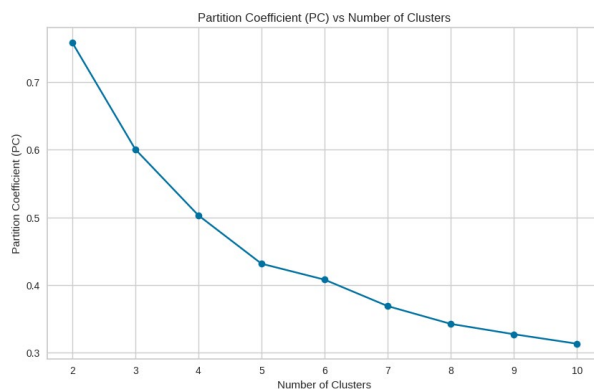
TABEL VI PENYEBARAN KELOMPOK DATA KABUPATEN/KOTA BERDASARKAN CLUSTERNYA

Cluster	Kabupaten/Kota	Kriteria
Cluster 0 (32 data)	Kabupaten Banjarnegara (2021), Kabupaten Banjarnegara (2022), Kabupaten Batang (2021), Kabupaten Batang (2022), Kabupaten Batang (2023), Kabupaten Blora (2021), Kabupaten Blora (2022), Kabupaten Blora (2023), Kabupaten Boyolali (2021), Kabupaten Boyolali (2022), Kabupaten Boyolali (2023), Kabupaten Jepara (2021), Kabupaten Karanganyar (2021), Kabupaten Pekalongan (2021), Kabupaten Pekalongan (2022), Kabupaten Purworejo (2021), Kabupaten Purworejo (2022), Kabupaten Purworejo (2023), Kabupaten Rembang (2021), Kabupaten Rembang (2022), Kabupaten Semarang (2021), Kabupaten Sragen (2021), Kabupaten Sragen (2022), Kabupaten Sragen (2023), Kabupaten Temanggung (2021), Kabupaten Temanggung (2022), Kabupaten Temanggung (2023), Kabupaten Wonogiri (2021), Kabupaten Wonogiri (2022), Kabupaten Wonogiri (2023), Kabupaten Wonosobo (2021), Kabupaten Wonosobo (2022),	Rendah
Cluster 1 (16 data)	Kota Magelang (2021), Kota Magelang (2022), Kota Magelang (2023), Kota Pekalongan (2023), Kota Salatiga (2021), Kota Salatiga (2022), Kota Salatiga (2023), Kota Semarang (2021), Kota Semarang (2022), Kota Semarang (2023), Kota Surakarta (2021), Kota Surakarta (2022), Kota Surakarta (2023), Kota Tegal (2021), Kota Tegal (2022), Kota Tegal (2022)	Tinggi
Cluster 2 (13 data)	Kabupaten Banyumas (2021), Kabupaten Banyumas (2022), Kabupaten Banyumas (2023), Kabupaten Brebes (2021), Kabupaten Brebes (2022), Kabupaten Brebes (2023), Kabupaten Cilacap (2021), Kabupaten Cilacap (2022), Kabupaten Cilacap (2023), Kabupaten Pemasang (2023), Kabupaten Tegal (2021), Kabupaten Tegal (2022), Kabupaten Tegal (2023),	Sangat Tinggi
Cluster 3 (22 Data)	Kabupaten Demak (2023), Kabupaten Jepara (2022), Kabupaten Jepara (2023), Kabupaten Karanganyar (2022), Kabupaten Karanganyar (2023), Kabupaten Kendal (2023), Kabupaten Klaten (2022), Kabupaten Klaten (2023), Kabupaten Kudus (2021), Kabupaten Kudus (2022), Kabupaten Kudus (2023), Kabupaten Pati (2022), Kabupaten Pati (2023), Kabupaten Pekalongan (2023), Kabupaten Rembang (2023), Kabupaten Semarang (2022), Kabupaten Semarang (2023), Kabupaten Sukoharjo (2021), Kabupaten Sukoharjo (2022), Kabupaten Sukoharjo (2023), Kota Pekalongan (2021), Kota Pekalongan (2022)	Sangat Rendah
Cluster 4 (22 Data)	Kabupaten Banjarnegara (2022), Kabupaten Demak (2021), Kabupaten Demak (2022), Kabupaten Grobogan (2021), Kabupaten Grobogan (2022), Kabupaten Grobogan (2023), Kabupaten Kebumen (2021), Kabupaten Kebumen (2022), Kabupaten Kebumen (2023), Kabupaten Kendal (2021), Kabupaten Kendal (2022), Kabupaten Klaten (2021), Kabupaten Magelang (2021), Kabupaten Magelang (2022), Kabupaten Magelang (2023), Kabupaten Pati (2021), Kabupaten Pemasang (2021), Kabupaten Pemasang (2022), Kabupaten Purbalingga (2021), Kabupaten Purbalingga (2022), Kabupaten Purbalingga (2023), Kabupaten Wonosobo (2023),	Sedang

E. Evaluasi

Pada bagian ini, mengevaluasi *cluster* yang optimal yang dihasilkan oleh *elbow method* sebagai berikut:

1) Partition Coefficient(PC)

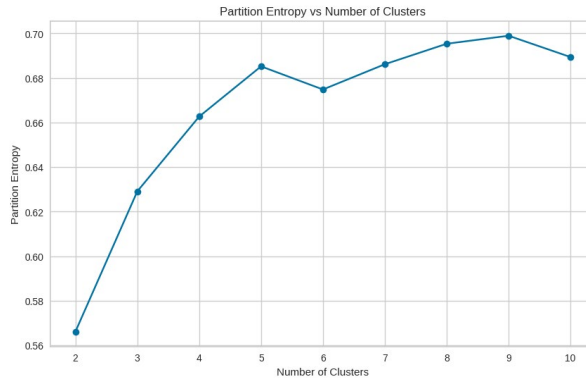


Gambar. 9 Grafik nilai Partition Coefficient(PC)

Partition Coefficient(PC) mengukur sejauh mana partisi data yang dihasilkan oleh FCM jelas atau tidak. Grafik ini memperlihatkan penurunan nilai PC dari sekitar 0.75 pada 2 cluster hingga sekitar 0.3 pada 10 cluster, yang

menunjukkan bahwa kejelasan cluster berkurang seiring bertambahnya jumlah cluster. Penurunan signifikan terlihat antara 2 hingga 5 cluster, setelah itu penurunan menjadi lebih lambat. Hal ini mengindikasikan bahwa ada titik optimal di mana menambah jumlah cluster lebih lanjut hanya memberikan sedikit peningkatan dalam kejelasan cluster. Dengan demikian, dalam memilih jumlah cluster yang optimal, perlu dipertimbangkan trade-off antara jumlah cluster dan kejelasan clustering, dengan pertimbangan optimal mungkin berada di sekitar 4 hingga 6 cluster

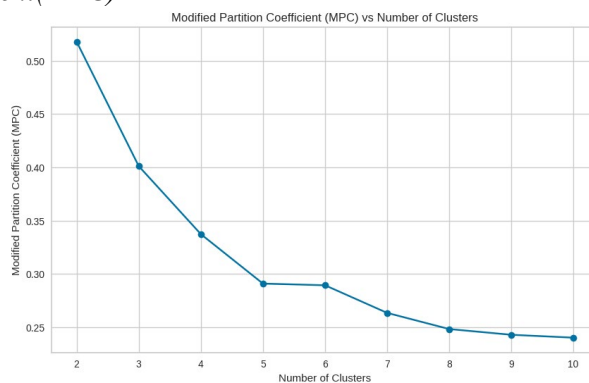
2) *Partition Entropy(PE)*



Gambar. 10 Grafik nilai *Partition Entropy(PE)*

Partition Entropy(PE) mengukur ketidakpastian dalam partisi yang dihasilkan oleh FCM. Grafik menunjukkan bahwa nilai PE meningkat seiring dengan bertambahnya jumlah cluster, mencapai puncak pada 5 cluster, sebelum sedikit menurun dan kemudian kembali meningkat. Idealnya, jumlah cluster optimal adalah pada nilai PE terendah, yang menunjukkan ketidakpastian minimum..

3) *Modified Partition Coefficient(MPC)*



Gambar. 11 Grafik nilai *Modified Partition Coefficient(MPC)*

Modified Partition Coefficient(MPC) mengukur seberapa baik data terbagi ke dalam *cluster* tertentu. Grafik menunjukkan bahwa nilai MPC cenderung menurun seiring bertambahnya jumlah cluster, yang merupakan fenomena umum karena data menjadi lebih tersebar dengan lebih banyak cluster. Dari grafik ini, terlihat bahwa nilai MPC mulai stabil dan mendatar pada sekitar 5-6 cluster, yang mengindikasikan bahwa jumlah cluster yang optimal mungkin berada di kisaran tersebut, karena setelah itu penurunan MPC tidak lagi signifikan.

F. *Comparison*

Hasil penelitian ini menunjukan kelebihan dari FCM dalam menandai data serta variable yang lebih banyak dibandingkan dengan penelitian menggunakan *K-menas* sebelumnya yang dilakukan oleh [7]. Penelitian dengan FCM ini menggunakan 5 indikator yang telah menapatkan nilai dari hasil terbaik dari niali *cluster center*. Kelima indikator tersebut Garis Kemiskinan (Rp/Kapita/Bln) sebesar 1.65900819, Jumlah Penduduk Miskin (Ribu Jiwa) sebesar 1.66455591, Pengeluaran Rata-rata sebesar 1.87926848, Pendidikan sebesar 1.97488421, dan Jumlah Pengangguran sebesar 1.87486378. Hasil dari metode FCM menunjukkan bahwa indikator Pengeluaran Rata-rata dan Pendidikan memiliki kontribusi positif yang signifikan dalam menentukan kelompok kemiskinan, sementara Jumlah Penduduk Miskin dan Jumlah Pengangguran memberikan kontribusi negatif. Untuk penelitian sebelumnya dengan menggunakan metode *K-means* hanya menggunakan 3 indikator didalamnya. Dengan tiga indikator hasil terbaik dari nilai *cluster center*: Jumlah Penduduk Miskin (0.712), Jumlah Pengangguran (1), dan Pendidikan (1). Penelitian sebelumnya menemukan bahwa Jumlah Pengangguran adalah faktor paling berpengaruh dalam mengelompokkan data kemiskinan. Namun, dengan tambahan dua indikator, hasil pengelompokan menjadi lebih akurat dan memberikan gambaran yang lebih lengkap tentang kondisi kemiskinan.

IV. KESIMPULAN

Menurut hasil perhitungan algoritma *fuzzy c-means* dengan dataset sebanyak 105 data serta 4 atribut yang menjadi indikator kemiskinan pada Provinsi Jawa Tengah yang digunakan. Pengelompokan tingkat kemiskinan pada Provinsi Jawa Tengah memperoleh hasil 5 *cluster*. Dimana *cluster 0* terdapat 32 kabupatèn/kota yang artinya tingkat kemiskinannya rendah. Kemudian *cluster 1* terdapat 16 kabupaten/kota yang tingkat kemiskinannya tinggi, *Cluster 2* terdapat 13 kabupaten/kota yang tingkat kemiskinannya sangat tinggi. *Cluster 3* terdapat 22 kabupaten/kota dengan tingkat kemiskinan yang sangat rendah dan *Cluster 4* terdapat 22 kabupaten/kota dengan tingkat kemiskinan yang sedang.

DAFTAR PUSTAKA

- [1] G. Haumahu and M. Y. Matdoan, "Algoritma K-Medoids Clustering Untuk Mengelompokkan Tingkat Kemiskinan Pada Kabupaten Dan Kota Di Kepulauan Maluku Dan Papua," *Var. J. Stat. Its Appl.*, vol. 4, no. 2, pp. 81–88, 2023, doi: 10.30598/variancevol4iss2page81-87.
- [2] L. Setiyani, M. Wahidin, D. Awaludin, and S. Purwani, "Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naïve Bayes : Systematic Review," *Fakt. Exacta*, vol. 13, no. 1, p. 35, 2020, doi: 10.30998/faktorexacta.v13i1.5548.
- [3] A. Bahauddin, A. Fatmawati, and F. Permata Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *J. Manaj. Inform. dan Sist. Inf.*, vol. 4, no. 1, pp. 1–8, 2021, doi: 10.36595/misi.v4i1.216.
- [4] W. WIJAYANTI, I. RAHMI HG, and F. YANUAR, "Penggunaan Metode Fuzzy C-Means Untuk Pengelompokan Provinsi Di Indonesia Berdasarkan Indikator Kesehatan Lingkungan," *J. Mat. UNAND*, vol. 10, no. 1, p. 129, 2021, doi: 10.25077/jmu.10.1.129-136.2021.
- [5] I. R. P. Nurrahmah Fitirani Kahar, Lillyan Hadjaratie, Sitti Suhada, "Tingkat Kemiskinan Di Indonesia Tahun 2007," *J. Informatics*, vol. 14, no. 2, pp. 27–38, 2020.
- [6] T. A. Munandar, "Penerapan Algoritma Clustering Untuk Pengelompokan Tingkat Kemiskinan Provinsi Banten," *JSil (Jurnal Sist. Informasi)*, vol. 9, no. 2, pp. 109–114, 2022, doi: 10.30656/jsii.v9i2.5099.
- [7] S. Nurhafiza, H. Rahma, Sani Nahampun Muhammad, and I. A. A. Rahmadeyan, "Penerapan K-Means Clustering Untuk Mengelompokkan Tingkat Kemiskinan di Provinsi Riau," *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, vol. 1, no. 1, pp. 59–65, 2022.
- [8] E. Widodo, P. Ermayani, L. N. Laila, and A. T. Madani, "Pengelompokan Provinsi di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Analisis Hierarchical Agglomerative Clustering (Indonesian Province Grouping Based on Poverty Level Using Hierarchical Agglomerative Clustering Analysis)," *Semin. Nas. Off. Stat.*, vol. 2, pp. 557–566, 2021.
- [9] F. Fajriani, "Persebaran Tingkat Kemiskinan di Sulawesi Selatan Menggunakan K-Means Clustering Analysis," *Pros. Semant.*, vol. 2, no. 1, pp. 276–281, 2019, [Online]. Available: <https://journal.uncp.ac.id/index.php/semantik/article/view/1526/1337%0Ahttps://journal.uncp.ac.id/index.php/semantik/article/view/1526>.
- [10] M. K. Sukardi and Indah Manfaati Nur, "Penerapan Algoritma Fuzzy C-Means Untuk Pengelompokan Kemiskinan Di Kabupaten/kota Provinsi Aceh," *J. Ilm. Mat. Dan Terap.*, vol. 20, no. 2, pp. 115–126, 2023, doi: 10.22487/2540766x.2023.v20.i2.16494.
- [11] P. Algoritma, K. Fuzzy, P. Jawa, and C. Zonyfar, "411-Article Text-836-1-10-20220713," vol. III, pp. 1–8, 2022.
- [12] "BPS Provinsi Jawa Tengah." <https://jateng.bps.go.id/> (accessed May 28, 2024).
- [13] N. Nofiyani and W. Wulandari, "Implementasi Electronic Data Processing Untuk Meningkatkan Efektifitas dan Efisiensi Pada Text Mining," *J. Media Inform. Budidarma*, vol. 6, no. 3, p. 1621, 2022, doi: 10.30865/mib.v6i3.4332.
- [14] D. K. Sitingjak, B. N. Sari, and I. Maulana, "Clustering Daerah Penyumbang Sampah Berdasarkan Provinsi di Indonesia Menggunakan Algoritma K-Means," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 16, pp. 137–146, 2022, [Online]. Available: <https://doi.org/10.5281/zenodo.7059032>.
- [15] W. Sanusi, A. Zaky, and B. N. Afni, "Analisis Fuzzy C-Means dan Penerapannya Dalam Pengelompokan Kabupaten/Kota di Provinsi Sulawesi Selatan Berdasarkan Faktor-faktor Penyebab Gizi Buruk," *J. Math. Comput. Stat.*, vol. 2, no. 1, p. 47, 2020, doi: 10.35580/jmathcos.v2i1.12458.
- [16] F. Rahmadayanti, I. Anggraini, and T. Susanti, "Pengklasterasian Data Penyakit Hipertensi dengan Menggunakan Metode K-Means," *J. Inf. Syst. Res.*, vol. 4, no. 2, pp. 737–741, 2023, doi: 10.47065/josh.v4i2.2905.
- [17] G. Sumantri, M. D. Novianto, and P. P. Prihastuti, "Implementasi Fuzzy C-Means dalam Pengelompokan Provinsi di Indonesia untuk Pemerataan Kualitas Pendidikan," *Pros. Semin. Pendidik. Mat. dan Mat.*, vol. 8, no. 2721, 2023, doi: 10.21831/pspm.v8i2.310.
- [18] A. Siregar, A. Buono, and K. Priandana, "Perbandingan Algoritma K-Means dan Fuzzy C-Means untuk Clustering Citra Daun Melon," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1503–1510, 2022, doi: 10.47065/bits.v4i3.2534.
- [19] I. Nabilla Audy, T. Nur Padilah, and B. Nurina Sari, "Pengelompokan Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma Fuzzy C-Means," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 4, pp. 2799–2803, 2024, doi: 10.36040/jati.v7i4.7205.
- [20] G. F. Aghyari and A. Kudus, "Penerapan Algoritma Fuzzy Possibilistic C-Means (FPCM) pada Pengelompokan Kabupaten/Kota di Indonesia Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2022," *Bandung Conf. Ser. Stat.*, vol. 3, no. 2, pp. 113–120, 2023, doi: 10.29313/bcss.v3i2.7321.