

# ANALISIS *CLUSTER* KABUPATEN/KOTA BERDASARKAN TINGKAT PENGANGGURAN DI PROVINSI KALIMANTAN TIMUR DENGAN MENGGUNAKAN PENDEKATAN *FUZZY C- MEANS* (FCM)

Wa Dina<sup>1\*</sup>, Fidia Deny Tisna Amijaya<sup>2</sup>, Idriasri Raming<sup>3</sup>

<sup>1, 2, 3</sup>Program Studi Matematika, Universitas Mulawarman  
Jalan Barong Tongkok, Kampus MIPA, Samarinda, Indonesia

e-mail: <sup>1</sup>wdinaa12@gmail.com

Submitted: February 23, 2025

Revised: Oktober 12, 2025  
corresponding author\*

Accepted: January 7, 2026

## Abstrak

Metode *Fuzzy C-Means* (FCM) adalah salah satu metode *cluster* non hirarki yang memiliki tujuan untuk mengelompokkan data yang memiliki kesamaan atribut dalam satu kelompok dan atribut yang berbeda dikelompokkan dalam kelompok yang lain. Metode *Fuzzy C-Means* dapat diterapkan untuk mengelompokkan tingkat pengangguran. Penelitian ini membahas mengenai analisis *cluster* kabupaten/kota Provinsi Kalimantan Timur berdasarkan tingkat pengangguran dengan menggunakan pendekatan *Fuzzy C-Means* (FCM). Penelitian ini bertujuan untuk mengetahui hasil dari penerapan metode *Fuzzy C-Means* (FCM) untuk mengelompokkan Kabupaten/Kota di Provinsi Kalimantan Timur dengan tingkat pengangguran tinggi, sedang, dan rendah. Adapun variabel yang digunakan yaitu Produk domestik Regional Bruto ( $x_1$ ), Tingkat Partisipasi Angkatan Kerja ( $x_2$ ), Indeks Pembangunan Manusia ( $x_3$ ) dan jumlah penduduk usia 15 tahun ke atas ( $x_4$ ). Berdasarkan hasil dan pembahasan diperoleh tiga *cluster* tingkat pengangguran. *Cluster* 1 dikategorikan untuk wilayah dengan tingkat pengangguran sedang yang terdiri atas 5 kabupaten/kota, *cluster* 2 dikategorikan untuk wilayah dengan tingkat pengangguran tinggi yang terdiri atas 1 kabupaten/kota, dan *cluster* 3 dikategorikan untuk wilayah dengan tingkat pengangguran rendah yang terdiri atas 4 kabupaten/kota. Nilai validitas hasil *cluster* dengan menggunakan metode *Silhouette Coefficients* diperoleh sebesar 0,419568 yang berarti struktur hasil *cluster* tersebut lemah.

**Kata Kunci:** *Fuzzy C-Means*, *Silhouette Coefficients*, Pengangguran

## CLUSTER ANALYSIS OF REGENCIES/CITIES BASED ON UNEMPLOYMENT RATES IN EAST KALIMANTAN PROVINCE USING THE FUZZY C-MEANS (FCM) APPROACH

## Abstract

Fuzzy C-Means (FCM) method is one of the non-hierarchical clustering methods aimed at grouping data with similar attributes into one cluster, while attributes with differences are grouped into separate clusters. The Fuzzy C-Means method can be applied to classify unemployment rates. Unemployment is a group of people who are not working or a group of people who are looking for a job with the aim of getting employment. This study discusses the cluster analysis of Districts/Cities in East Kalimantan Province based on the unemployment rate using the Fuzzy C-Means (FCM) approach. The aim of this research is to determine the outcomes of applying the Fuzzy C-Means (FCM) method to cluster Districts/Cities in East Kalimantan Province according to high, moderate, and low levels of unemployment. The variables used include PDRB ( $x_1$ ), TPAK ( $x_2$ ), IPM ( $x_3$ ), and the total population aged 15 and above ( $x_4$ ). Based on the results and discussions, three clusters of unemployment levels were identified. Cluster 1 is categorized for areas with moderate unemployment rates, consisting of 5 districts/cities, cluster 2 is for areas with high unemployment rates, comprising 1 district/city, and cluster 3 is for areas with low unemployment rates, consisting of 4 districts/cities. The validity value of the cluster results using the *Silhouette Coefficients* method is obtained as 0,419568, indicating a weak structure of the cluster results.

**Keywords:** Fuzzy C-Means, *Silhouette Coefficients*, Unemployment

## 1. Pendahuluan

Pengangguran merupakan permasalahan yang selalu dihadapi negara-negara berkembang salah satunya Indonesia dalam pembangunan ekonomi (Aprizkiyandari et al., 2023). Menurut Adyaksa (2020) pengangguran adalah sekumpulan orang yang tidak bekerja atau seseorang yang sedang mencari suatu pekerjaan dengan tujuan untuk mendapatkan pekerjaan. Tingginya pengangguran sangat berdampak pada kehidupan perekonomian dan kehidupan sosial masyarakat. Pertumbuhan ekonomi yang menurun dan tingkat kesejahteraan masyarakat yang menurun merupakan salah satu dampak pengangguran. Dampak pengangguran terhadap perekonomian dan kehidupan sosial yaitu menurunkan aktivitas perekonomian, menurunkan pertumbuhan ekonomi dan pertumbuhan perkapita, biaya sosial menurun, menurunkan tingkat keterampilan, penerimaan negara menurun, bertambahnya tindakan kriminal, dan meningkatnya pengamen dan pengemis (Sejati, 2020).

Dampak tingkat pengangguran juga dirasakan oleh provinsi di Indonesia salah satunya Provinsi Kalimantan Timur yang tingkat penganggurannya mengalami fase naik dan turun setiap tahunnya. Hal ini, disebabkan oleh sempitnya lapangan pekerjaan yang membuat jumlah pengangguran semakin meningkat, ditambah pertumbuhan penduduk yang tinggi membuat kebutuhan lapangan pekerjaan juga semakin meningkat. Meningkatnya tingkat pengangguran di Kalimantan Timur menjadi masalah serius yang harus segera diatasi (Junaidi dan Junaidi, 2023). Menurut Badan Pusat Statistik (BPS) tingkat pengangguran di Kalimantan Timur tahun 2023 sebesar 5,31%. Berdasarkan informasi tersebut, perlu dilakukan identifikasi terhadap kabupaten dan kota yang menjadi penyumbang terbesar terhadap tingkat pengangguran di Provinsi Kalimantan Timur. Informasi tingkat pengangguran ini dapat dicari dengan menggunakan analisis *cluster*.

*Cluster* adalah salah satu metode dalam data mining (Sari et al., 2020). Menurut Afira dan Wijiyanto (2021) analisis *cluster* adalah teknik multivariat yang mempunyai tujuan untuk mencari kesamaan dalam data dan menempatkan data yang sama kedalam kelompok-kelompok. Analisis *cluster* dapat diselesaikan dengan menggunakan Metode *Fuzzy C-Means* (FCM). Menurut Maulana et al. (2023) Algoritma *Fuzzy C-Means* adalah salah satu metode pengklasteran data dan

menentukan data mana yang merupakan bagian terbesar dari keanggotaannya. Kelebihan Metode *Fuzzy C-Means* adalah dalam penempatan pusat cluster lebih tepat, tingkat akurasi yang tinggi dan waktu komputasi yang tepat dibandingkan dengan metode pengklasteran yang lain (Rouza dan Fimawahib, 2020).

Pada penelitian terdahulu yang membahas analisis cluster dengan menggunakan Metode *Fuzzy C-Means* yaitu Azzahra et al (2023) meneliti tentang pengelompokan tingkat pengangguran, kemiskinan, dan pendapatan di kabupaten atau kota Provinsi Jawa Barat menggunakan metode K-Means dan metode *Fuzzy C-Means*. Hasil dari penelitiannya adalah bahwa setiap wilayah memiliki kondisi yang berbeda-beda, tetapi kemiripannya dapat dilihat berdasarkan kluster-kluster yang dihasilkan. Novianti et al. (2022) juga meneliti tentang penerapan Algoritma *Fuzzy C-Means* dalam pengelompokan provinsi di Indonesia berdasarkan indikator penyakit menular manusia. Adapun hasil penelitiannya adalah sebanyak 3 *cluster* tinggi, sedang dan rendah dengan menggunakan pengujian validasi dengan menggunakan *Shillouette Coefficient*. Selanjutnya Ramadhan et al. (2019) meneliti tentang implementasi algoritma *Fuzzy C-Means* untuk pengelompokan wilayah bencana banjir. Hasil penelitiannya diperoleh 3 *cluster* tinggi, sedang dan rendah dengan menggunakan 2 pengujian validasi *Partition Coefficient Index* dan *Partition Entrophy Index*. Selanjutnya penelitian yang dilakukan oleh Rouza dan Fimawahib (2020) mengenai implementasi *Fuzzy C-Means clustering* dalam pengelompokan UKM di Kabupaten Rokan Hulu. Adapun hasil penelitiannya adalah penerapan metode *Fuzzy C-Means Clustering* mampu mengelompokkan jenis UKM dengan menggunakan variabel berupa omset, asset dan jumlah tenaga kerja dan hasil uji coba dari sejumlah data yang diinputkan menunjukkan bahwa metode *Fuzzy C-Means Clustering* mempunyai nilai validasinya rata-rata hampir mendekati angka 1, dengan ini menunjukkan bahwa *Fuzzy C-Means Clustering* memiliki tingkat akurasi yang tinggi sebesar 90 %. Wardoyo dan Triuspita (2020) meneliti penentuan *cluster* optimum pada tingkat pengangguran dan tingkat kemiskinan di Jawa Timur menggunakan algoritma *Fuzzy C-Means*. Hasil penelitiannya diperoleh *cluster* optimum dengan hasil 4 *cluster* berdasarkan nilai *davies bouldin index* 0.8737.

Tujuan yang ingin dicapai pada penelitian ini adalah Mengetahui hasil analisis *cluster*

kabupaten/kota Provinsi Kalimantan Timur berdasarkan tingkat pengangguran dengan menggunakan pendekatan *Fuzzy C-Means* (FCM). Mengetahui validitas hasil analisis cluster kabupaten/kota Provinsi Kalimantan Timur berdasarkan tingkat pengangguran dengan menggunakan *Silhouette Coefficient*. Mengetahui interpretasi hasil analisis *cluster* kabupaten/kota Provinsi Kalimantan Timur dengan menggunakan pendekatan *Fuzzy C-Means* (FCM).

Tujuan yang ingin dicapai pada penelitian ini adalah Mengetahui hasil analisis *cluster* kabupaten/kota Provinsi Kalimantan Timur berdasarkan tingkat pengangguran dengan menggunakan pendekatan *Fuzzy C-Means* (FCM). Mengetahui validitas hasil analisis cluster kabupaten/kota Provinsi Kalimantan Timur berdasarkan tingkat pengangguran dengan menggunakan *Silhouette Coefficient*. Mengetahui interpretasi hasil analisis *cluster* kabupaten/kota Provinsi Kalimantan Timur dengan menggunakan pendekatan *Fuzzy C-Means* (FCM).

### 1.1. Data Mining

Data mining adalah suatu rangkaian proses yang dirancang untuk menggali nilai tambah dalam bentuk informasi yang tidak diambil secara manual dari database. Data mining telah ada sejak tahun 1990-an sebagai metode yang akurat dan tepat untuk mengambil pola dan informasi yang digunakan untuk mencari hubungan antar data dan mengelompokkan data ke dalam satu atau lebih *cluster* sehingga objek dalam *cluster* tersebut memiliki kemiripan yang tinggi satu sama lain (Harahap dan Sulindawaty, 2019). Istilah data mining dikenal juga dengan nama Knowledge Discovery in Database (KDD). KDD adalah suatu kegiatan yang menggunakan dan mengumpulkan data histori untuk menemukan kumpulan dataset tentang desain atau hubungan (Dewayanti, 2018)

### 1.2. Analisis Cluster

*Cluster* digunakan untuk mengelompokkan anggota setiap partisi dengan kesamaan berdasarkan matriks tertentu. Analisis klaster atau analisis kelompok adalah suatu teknik analisis data mining yang bertujuan untuk mengelompokkan individu atau objek ke dalam kelompok-kelompok yang mempunyai ciri-ciri yang berbeda antar kelompok, sehingga individu atau objek dalam satu kelompok akan mempunyai ciri-ciri yang relatif seragam (Anggraini dan Arum, 2022). Cluster adalah proses pengelompokan sekumpulan objek ke dalam kelas/pola objek yang sama. *Cluster* merupakan teknik yang sangat penting dalam

analisis data. *Cluster* melibatkan pembagian data menjadi beberapa kelompok atau cluster berdasarkan kesamaan atribut dalam sekumpulan data. Atribut serupa ini direpresentasikan sebagai titik dalam ruang multidimensi (Kurniawan et al, 2020).

$$P = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \cdots & \mu_{1n}(x_n) \\ \mu_{21}(x_1) & \mu_{22}(x_2) & \cdots & \mu_{2n}(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{c1}(x_1) & \mu_{c2}(x_2) & \cdots & \mu_{cn}(x_n) \end{bmatrix} \quad (3)$$

### 1.3. Normalisasi Data

Normalisasi data adalah menskalakan nilai data dalam rentang yang ditentukan untuk memudahkan dalam tahapan-tahapan perhitungan seperti perhitungan kesamaan atau operasi cluster. Salah satu metode yang digunakan dalam menormalisasikan data adalah Algoritma Min-Max (Saputra dan Chusyairi, 2020). Algoritma Min-Max merupakan metode sederhana yang secara khusus cocok untuk data dalam batasan yang ditetapkan sebelumnya. Metode ini akan menghasilkan rentang nilai dari 0 sampai 1. Adapun algoritma ini dirumuskan seperti berikut (Suyanto, 2018):

$$a_{kj} = \frac{x_{kj} - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Keterangan:

$a_{kj}$  = nilai baru hasil normalisasi untuk data ke- $k$  atribut ke- $j$

$x_{kj}$  = data ke- $k$  atribut ke- $j$

$x_{min}$  = data minimum

$x_{max}$  = data maximum

### 1.4. Fuzzy C-Means

Adapun menurut Rahakbauw et al. (2019) algoritma yang digunakan pada metode *Fuzzy C-Means* adalah sebagai berikut:

1. Memasukan data yang akan di-*cluster* ke dalam bentuk matriks  $A$  dengan ukuran  $n \times m$ , dengan  $n$  adalah jumlah sampel data dan  $m$  adalah atribut setiap data. Adapun bentuk matriksnya sebagai berikut:

$$A = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (2)$$

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

2. Menentukan parameter awal:
  - (a) Jumlah *cluster* (c)

- (b) Pangkat pembobot ( $w$ )
- (c) Maksimum iterasi (*Maksiter*)
- (d) Error terkecil yang diharapkan ( $\varepsilon$ )
- (e) Fungsi objektif awal ( $R^0$ ) = 0
- (f) Iterasi awal ( $t$ ) = 1

3. Membangkitkan bilangan *random*  $\mu_{kl}$ ,  $k = 1, 2, \dots, n$ ;  $l = 1, 2, \dots, c$  sebagai elemen-elemen matriks partisi awal  $P$  berukuran  $c \times n$ .  $l$  menyatakan jumlah *cluster*.  $\mu_{kl}$  adalah bilangan *random* dari 0 sampai 1 yang menyatakan derajat keanggotaan dari setiap data yang digunakan. Berikut matriks partisi awal  $P$

Hitung jumlah setiap kolom dengan menggunakan rumus berikut:

$$X_k = \sum_{l=1}^c \mu_{kl} \text{ dengan } l = 1, 2, \dots, c \quad (4)$$

Selanjutnya menghitung elemen matriks dari kolom dalam setiap baris harus bernilai mendekati atau sama dengan 1 dengan menggunakan rumus:

$$\mu_{kl_{baru}} = \frac{\mu_{kl}}{X_k} \quad (5)$$

4. Menghitung pusat *cluster* ke- $l$ ,  $V_{lj}$  dengan  $l = 1, 2, \dots, c$  dan  $j = 1, 2, \dots, m$

$$V_{lj} = \frac{\sum_{k=1}^n ((\mu_{kl})^w \cdot x_{kj})}{\sum_{k=1}^n (\mu_{kl})^w} \quad (6)$$

5. Menghitung fungsi objektif pada iterasi ke- $t$ ,  $R^t$

$$R^t = \sum_{k=1}^n \sum_{l=1}^c \left( \left[ \sum_{j=1}^m (x_{kj} - V_{lj})^2 \right] (\mu_{kl})^w \right) \quad (7)$$

6. Menghitung perubahan matriks partisi

$$\mu_{kl}^t = \frac{\left[ \sum_{j=1}^m (x_{kj} - V_{lj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{l=1}^c \left[ \sum_{j=1}^m (x_{kj} - V_{lj})^2 \right]^{\frac{-1}{w-1}}} \quad (8)$$

dengan:  $k = 1, 2, \dots, n$ ;  $l = 1, 2, \dots, c$ .

7. Mengecek kondisi iterasi berhenti

Jika  $(|R^t - R^{t-1}| < \varepsilon)$  atau  $(t > \text{Maksiter})$  maka iterasi berhenti;

Jika tidak, maka iterasi dilanjutkan dengan mengulangi langkah ke-4 dengan  $t = t + 1$ .

### 1.5 Silhouette Coefficients

*Silhouette Coefficient* dalam perhitungannya mempertimbangkan jarak antara titik data dengan titik-titik dalam yang sama dan jarak antara titik data dengan titik-titik data dalam yang lain. *Silhouette Coefficients* dirumuskan seperti berikut (Huddin et al., 2023). Menurut Dewi et al. (2019) *Silhouette Coefficients* dirumuskan seperti berikut:

$$S_i = \frac{b_i - d_i}{\max(b_i; d_i)} \quad (9)$$

Keterangan:

$S_i$  = *Silhouette Coefficients*

$b_i$  = rata-rata jarak data ke- $i$  terhadap semua data yang tidak dalam satu *cluster* dengan data ke- $i$

$d_i$  = rata-rata jarak data ke- $i$  terhadap semua data dalam satu *cluster*

Berikut adalah rumus perhitungan untuk mendapatkan nilai  $d_i$  dan  $b_i$ .

$$b_i = \text{Min}_{q=1, \dots, c, q \neq j} \left\{ \frac{1}{n_q} \sum_{r=1}^{n_q} a(x_{i,j}, x_{r,q}) \right\} \quad (10)$$

$$d_i = \frac{1}{n_j - 1} \sum_{r=1}^{n_j} a(x_{i,j}, x_{r,j}) \quad (11)$$

Keterangan:

$j$  = *cluster*

$i$  = index data ( $i = 1, 2, \dots, n_j$ )

$b_i$  = rata-rata jarak data ke- $i$  terhadap semua data yang tidak dalam satu *cluster*

dengan data ke- $i$

$d_i$  = rata-rata jarak data ke- $i$  terhadap semua data dalam satu *cluster*

$n_q$  = jumlah data dalam *cluster* ke- $q$

$n_j$  = jumlah data dalam *cluster* ke- $j$

$a(x_{i,j}, x_{r,q})$  = jarak data ke- $i$  dengan data ke- $j$  dalam satu *cluster*  $q$

$a(x_{i,j}, x_{r,j})$  = jarak data ke- $i$  dengan data ke- $r$  dalam satu *cluster*  $j$

Menurut Fahriya dan Yustanti (2021) ada beberapa kriteria subjektif yang digunakan dalam pengukuran *cluster* berdasarkan Metode *Silhouette Coefficients* yaitu nilai *Silhouette Coefficients* 0,71 – 1,00 struktur hasil yang kuat. Nilai *Silhouette Coefficients* 0,51 – 0,70 struktur hasil yang baik. Nilai *Silhouette Coefficients* 0,26 – 0,50 struktur hasil yang lemah. *Silhouette Coefficients*  $\leq 0,25$  struktur hasil yang sangat buruk.

### 1.6 Pengangguran di Provinsi Kalimantan Timur

Pengangguran juga dirasakan oleh Provinsi Kalimantan Timur. Pengangguran di Provinsi Kalimantan Timur selain disebabkan oleh sempitnya lapangan kerja juga disebabkan oleh berdatangnya imigrasi para pekerja dari luar daerah untuk mencari suatu pekerjaan yang diinginkannya (Tumilaar et al., 2022). Provinsi Kalimantan Timur merupakan salah satu Provinsi terluas kedua setelah Papua, memiliki potensi sumber daya alam yang melimpah dimana sebagian besar potensi tersebut belum dimanfaatkan secara optimal. Provinsi ini



merupakan penghasil devisa utama bagi negara, khususnya dari sektor pertambangan, kehutanan dan hasil lainnya. (BPS Kalimantan Timur, 2024).

## 2. Metode Penelitian

Metode yang digunakan pada penelitian ini adalah metode penelitian kuantitatif dalam pengolahan datanya. Penelitian kuantitatif adalah jenis penelitian menggunakan angka dan statistika matematika atau komputasi dalam pengumpulan serta analisis data yang dapat diukur. Data yang digunakan pada penelitian ini adalah data sekunder dan merupakan penelitian non eksperimen karena datanya diambil di BPS Provinsi Kalimantan Timur.

Populasi dalam Penelitian ini adalah seluruh indikator sosial ekonomi. Adapun sampel yang digunakan adalah data jumlah Produk Domestik Regional Bruto (PDRB), Tingkat Partisipasi Angkatan Kerja (TPAK), laju pertumbuhan Indeks Pembangunan Manusia (IPM), dan jumlah penduduk usia 15 tahun ke atas kabupaten/kota Provinsi Kalimantan Timur tahun 2023. Teknik pengumpulan data yang digunakan dalam penelitian ini adalah teknik *scraping* yang diperoleh langsung dari *website* BPS Provinsi Kalimantan Timur diakses di <https://kaltim.bps.go.id/>

Adapun langkah-langkah dalam teknik analisis data yang digunakan pada penelitian ini adalah

normalisasi data dengan menggunakan Algoritma Min-Max, analisis *cluster* dengan menggunakan *Fuzzy C-Means*, Validitas hasil *cluster* dengan menggunakan *Silhouette Coefficients*, dan interpretasi data dengan menggunakan profilisasi *cluster*.

## 3. Hasil dan Pembahasan

Pada bagian ini akan disajikan hasil analisis yang telah dilakukan berdasarkan data yang tersedia, serta pembahasan yang bertujuan untuk menginterpretasikan temuan tersebut. Analisis dilakukan dengan pendekatan klasterisasi untuk mengelompokkan wilayah berdasarkan karakteristik dari variabel yang digunakan seperti PDRB, TPAK, IPM, dan jumlah penduduk usia 15 tahun ke atas di Povinsi Kalimantan Timur. Hasil klasterisasi ini selanjutnya dianalisis untuk mengetahui hubungan antara karakteristik wilayah dengan tingkat pengangguran yang terjadi.

### 3.1. Normalisasi Data

Sebelum melakukan perhitungan dengan menggunakan *Fuzzy C-Means*, dilakukan normalisasi data dengan menggunakan Algoritma Min-Max agar data yang digunakan memiliki rentang yang sama, sehingga data tidak ada yang besar maupun kecil. Dari hasil perhitungan dengan menggunakan Persamaan (1) diperoleh normalisasi data pada Tabel 1

Tabel 1. Data Normalisasi

No	Kabupaten/Kota Provinsi Kalimantan Timur	PDRB	TPAK	IPM	Pendudk Usia 15 Tahun Ke Atas
1	Paser	0.277009	0	0.357423	0.307997
2	Kuta Barat	0.78326	0.627386	0.32443	0.196666
3	Kutai Kartanegara	1	0.19668	0.448547	0.939107
4	Kutai Timur	0.818309	0.271369	0.423409	0.546341
5	Berau	0.237611	0.328631	.520031	0.277621
6	Penajam Pser Utara	0.068678	0.273859	0.291438	0.215032
7	Mahakam Ulu	0	1	0	0
8	Balikpapan	0.693712	0.043983	0.948154	0.780305
9	Samarinda	0.426275	0.210788	1	1
10	Bontang	0.321127	0.448133	0.940299	0.095083

### 3.2 Analisis Cluster Menggunakan Pendekatan Fuzzy C-Means (FCM)

Adapun analisis *cluster* kabupaten/kota dengan menggunakan Pendekatan FCM adalah sebagai berikut:

1. Memasukan data yang telah dinormalisasikan kedalam bentuk matriks  $A$  yang berukuran  $10 \times 4$  dikarenakan ada 10 kabupaten/kota dengan 4 variabel yang digunakan. Adapun bentuk matriksnya seperti berikut.

$$A = \begin{bmatrix} 0,277009 & 0 & 0,357423 & 0,307997 \\ 0,78326 & 0,627386 & 0,32443 & 0,19666 \\ 1 & 0,19668 & 0,448547 & 0,939107 \\ \vdots & \vdots & \vdots & \vdots \\ 0,321127 & 0,44813 & 0,940299 & 0,095083 \end{bmatrix}$$

2. Menentukan Parameter Awal

Sebelum melakukan analisis cluster dengan menggunakan Algoritma *Fuzzy C-Means* ada beberapa parameter yang harus ditentukan terlebih dahulu.

Parameter yang ditentukan yaitu jumlah *cluster* ( $c$ ) = 3, Pangkat pembobot ( $w$ ) = 2, Maksimum iterasi ( $Maksiter$ ) = 100, *Error* terkecil yang diharapkan ( $\epsilon$ ) =  $10^{-5}$ , Fungsi objektif awal ( $R^0$ ) = 0 dan Iterasi awal ( $t$ ) = 1

3. Membangkitkan Derajat Keanggotaan Pusat Cluster Secara Random

Adapun matriks partisi awal  $P$  yang berukuran  $3 \times 10$  sebagai berikut

$$P = \begin{bmatrix} 0,822 & 0,5002 & \dots & 0,6119 \\ 0,0766 & 0,4310 & \dots & 0,1709 \\ 0,1213 & 0,0689 & \dots & 0,2172 \end{bmatrix}$$

4. Menghitung Pusat Cluster

Adapun hasil perhitungan pusat *cluster* pada iterasi pertama dengan menggunakan Persamaan (6) dapat dilihat pada Tabel 2.

Tabel 2. Pusat Cluster

Pusat Cluster	$x_1$	$x_2$	$x_3$	$x_4$
1	0.236862	0.282785	0.470526	0.259688
2	0.048981	0.90459	0.086762	0.047452
3	0.712	0.164797	0.753745	0.821692

Pada Tabel 2 untuk  $x_1$  adalah PDRB,  $x_2$  adalah TPAK,  $x_3$  adalah IPM dan  $x_4$  adalah usia penduduk 15 tahun ke atas.

5. Menghitung Fungsi Objektif

Hasil perhitungan fungsi objektif digunakan untuk mengecek kondisi pemberhentian iterasi. Adapun hasil perhitungan fungsi objektif dengan menggunakan Persamaan (7) pada iterasi pertama diperoleh  $R^1 = 0,86855$ .

6. Memperbaiki Derajat Keanggotaan

Langkah selanjutnya adalah memperbaiki derajat keanggotaan dengan menggunakan Persamaan (8) pada setiap data dalam setiap pusat *cluster* untuk mengetahui setiap data masuk ke dalam *cluster* 1, 2 dan 3. Adapun hasil perhitungannya dapat dilihat pada matriks  $P^1$  yang berukuran  $3 \times 10$  seperti berikut.

$$P^1 = \begin{bmatrix} 0,802033 & 0,501385 & \dots & 0,61212 \\ 0,076525 & 0,42966 & \dots & 0,170633 \\ 0,121443 & 0,068954 & \dots & 0,21744 \end{bmatrix}$$

7. Mengecek Kondisi Berhenti

Iterasi dikatakan berhenti jika memenuhi 1 atau 2 kondisi pemberhentian yaitu jika  $(|R^t - R^{(t-1)}| < \epsilon)$  atau  $(t > Maksiter)$ . Pada kondisi ini  $t = 1 < 100$  dan  $|R^t - R^{(t-1)}| = |R^1 - R^0| = |0,86855 - 0| = 0,86855 > 10^{-5}$  artinya tidak memenuhi kondisi pemberhentian iterasi, hal ini iterasi dilanjutkan dengan mengulangi langkah ke-3. Perhitungan iterasi 2 sampai 100 dilakukan dengan menggunakan program Python sehingga diperoleh hasil akhir *cluster* pada Tabel 3.

Tabel 3. Hasil Perhitungan Fuzzy C-Means

Cluster 1	Cluster 2	Cluster 3
Paser	Mahakam Ulu	Kutai
		Kartanegara
Kutai Barat		Kutai Timur
Berau		Balikpapan
Penajam		Samarinda
Paser Utara		
Bontang		

### 3.3 Validitas Hasil Cluster Menggunakan Silhouette Coeffients

Data yang sudah dikelompokkan pada Tabel 3, selanjutnya dilakukan pengujian *cluster* dengan menggunakan *Silhouette Coefficients* untuk mengetahui seberapa baik suatu data yang di*cluster* dengan menggunakan Persamaan (9). Dalam mencari nilai *Silhouette Coefficients*, diperlukan nilai  $d_i$  dan  $b_i$ . Nilai  $d_i$  diperoleh dari perhitungan jarak rata-rata

data ke data yang lain dalam satu *cluster* yang sama dan nilai  $b_i$  diperoleh dari perhitungan jarak rata-rata terkecil data ke data yang lain dalam *cluster* yang berbeda. Perhitungan nilai

$d_i$  dan  $b_i$  setiap *cluster* dengan menggunakan Persamaan (11) dan Persamaan (10) dapat dilihat pada Tabel 4.

**Tabel 4. Perhitungan Nilai  $S_i$**

No.	Kabupaten/Kota	$d_i$	$b_i$	$S_i$
1	Paser	0.536211	0.870403	0.383951
2	Kutai Barat	0.513386	0.560869	0.08466
3	Kutai Kartanegara	0.621388	1.121491	0.445928
4	Kutai Timur	0.63475	0.7634	0.168522
5	Berau	0.379483	0.854237	0.555764
6	Penajam Paser Utara	0.439596	0.814354	0.460191
7	Mahakam Ulu	0	0.916035	1
8	Balikpapan	0.548029	0.968635	0.434226
9	Samarinda	0.673318	1.018392	0.338841
10	Bontang		0.975809	0.3236
<i>Silhouette Coefficients</i> =				0.419568

Hasil akhir dari *Silhouette Coefficients* adalah perhitungan rata-rata nilai  $S_i$  dari 10 data yaitu sebesar 0,419568. Nilai *Silhouette Coefficients* sebesar ini mengindikasikan bahwa struktur hasil *cluster* lemah.

### 3.4 Interpretasi Hasil Analisis Cluster

Berdasarkan Tabel 3 merupakan hasil dari pengelompokan data tingkat pengangguran di Kabupaten/Kota Provinsi Kalimantan Timur

tahun 2023. Selanjutnya akan ditentukan karakteristik masing-masing *cluster*. Menurut Mayasari dan Nugraha (2023) bahwa dalam menentukan karakteristik dari masing-masing *cluster* dilakukan dengan cara profilisasi. Profilisasi dilakukan dengan cara menghitung nilai rata-rata variabel terhadap setiap *cluster* yang sudah terbentuk. Perhitungan nilai rata-rata variabel dapat dilihat pada Tabel 5.

**Tabel 5. Profilisasi Cluster**

	PDRB	TPAK	IPM	Penduduk Usia 15 Tahun Ke atas
Cluster 1	0.21655	0.335602	0.486724	0.21848
Cluster 2	0	1	0	0
Cluster 3	0.734574	0.180705	0.705027	0.816438

Hasil profilisasi *cluster* setiap variabel pada Tabel 5 menunjukkan bahwa *cluster* 1 memiliki nilai rata-rata yang sedang pada variabel PDRB, TPAK, IPM dan jumlah penduduk usia 15 tahun ke atas. *Cluster* 2 memiliki nilai rata-rata yang tinggi pada variabel TPAK dan nilai rata-rata yang rendah pada variabel PDRB, IPM dan jumlah penduduk usia 15 tahun ke atas. *Cluster* 3 memiliki nilai rata-rata yang tinggi pada variabel PDRB, IPM dan jumlah penduduk usia 15 tahun ke atas dan nilai rata-rata yang rendah pada variabel TPAK. Semakin tinggi nilai PDRB, TPAK, IPM, dan jumlah

penduduk usia 15 tahun ke atas maka tingkat pengangguran semakin berkurang.

*Cluster* 1 didominasi oleh nilai rata-rata variabel yang sedang maka *cluster* 1 termasuk karakteristik *cluster* yang dikategorikan untuk wilayah dengan tingkat pengangguran sedang. *Cluster* 2 didominasi oleh nilai rata-rata variabel yang rendah maka *cluster* 2 termasuk karakteristik *cluster* yang dikategorikan untuk wilayah dengan tingkat pengangguran tinggi. *Cluster* 3 didominasi oleh nilai rata-rata variabel yang tinggi maka *cluster* 3 termasuk karakteristik *cluster* yang dikategorikan untuk wilayah dengan tingkat pengangguran rendah.

Kategorisasi wilayah terhadap tingkat pengangguran juga tergantung pada makna dan arah hubungan setiap variabel terhadap pengangguran. jika kondisi rata-rata variabel PDRB Tinggi, rata-rata variabel IPM Tinggi, rata-rata variabel TPAK Sedang dan rata-rata variabel Penduduk usia 15 ke atas Sedang *cluster* ini cenderung mencerminkan wilayah dengan tingkat pengangguran rendah. Rata-rata variabel PDRB dan IPM merupakan indikasi wilayah berkembang dan kualitas suber daya manusia baik sedangkan TPAK dan penduduk usia kerja tidak membebani secara ekstrem pada suatu wilayah. Namun secara umum dengan asumsi bahwa semakin tinggi nilai dari ke-4 variabel tersebut semakin rendah tingkat pengangguran.

#### 4. Kesimpulan

Berdasarkan Hasil dan pembahasan pada penelitian ini, dapat disimpulkan sebagai berikut.

1. *Cluster* tingkat pengangguran berdasarkan 4 variabel yang digunakan yaitu PDRB, TPAK, IPM, dan jumlah penduduk usia 15 tahun ke atas di Kabupaten/Kota Provinsi Kalimantan Timur tahun 2023 di peroleh cluster pertama yaitu Kabupaten Paser, Kabupaten Kutai Barat, Kabupaten Berau, Kabupaten Penajam Paser Utara, dan Kota Bontang, *cluster* kedua yaitu Kabupaten Mahakam Ulu, *cluster* ketiga yaitu Kabupaten Kutai Kartanegara, Kabupaten Kutai Timur, Kota Balikpapan dan Kota Samarinda.
2. Validitas *cluster* tingkat pengangguran di kabupaten/kota Provinsi Kalimantan Timur diperoleh nilai rata-rata *Silhouette Coefficients* sebesar 0,419568, ini menunjukkan bahwa struktur hasil *cluster* lemah.
3. *Cluster* pertama dikategorikan untuk wilayah dengan tingkat pengangguran sedang, *cluster* kedua dikategorikan untuk wilayah dengan tingkat pengangguran tinggi, dan *cluster* ketiga dikategorikan untuk wilayah dengan tingkat pengangguran rendah.

#### Daftar Pustaka

- Adyaksa F. F. 2020. Analisis Faktor-Faktor Yang Mempengaruhi Pengangguran Terdidik di Indonesia Tahun 2018. *Jurnal Ilmiah Mahasiswa FEB*, 8(2).
- Afira, N., dan Wijayanto, A. W. (2021). Analisis Cluster dengan Metode Partitioning dan Hierarki pada Data Informasi Kemiskinan Provinsi di Indonesia Tahun 2019. *Komputika: Jurnal Sistem Komputer*, 10(2), 101-109.
- Anggraini, L., dan Arum, P. R. (2022). Analisis Cluster Menggunakan Algoritma K-Means Pada Provinsi Sumatera Barat Berdasarkan Indeks Pembangunan Manusia Tahun 2021. *Prosiding Seminar Nasional Publikasi Hasil-Hasil Penelitian Dan Pengabdian Masyarakat*, 5, 636–646.
- Aprizkiyandari, S., Satyahadewi, N., Pratama, A. N., Rivaldo, R., Nurdiansyah, S. I., dan Helena, S. (2023). Implementasi K-Means Cluster untuk Menentukan Persebaran Tingkat Pengangguran. *Empiricism Journal*, 4(2), 400–406.
- Azzahra, Q. N., Pratama, M. N., Priongo, E. A., Nurillatiffah, T., Pravitasari, A. A., dan Indrayatna, F. (2023). Pengelompokan Tingkat Pengangguran, Kemiskinan, Dan Pendapatan Di Kabupatenkota Provinsi Jawa Barat. *BIAStatistics: Jurnal Statistika Teori Dan Aplikasi: Biomedics, Industry dan Business And Social Statistics*, 2, 67–80.
- BPS Kalimantan Timur. (2024). *Indikator Strategi*. kaltim.bps.go.id. (Online), (<https://kaltim.bps.go.id/quickMap.html>) diakses 20-01-2024
- Dewayanti, A. A. (2018). Penerapan Data Mining Pada Data Nilai Siswa Dengan Menggunakan Algoritma Asosiasi Rule Metode *Apriori* (Studi Kasus di SMP N 36 Semarang). *Prosiding Seminar Nasional Matematika Dan Pendidikan Matematika*, 1–10.
- Dewi, D. A. I. C., dan Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Jurnal Matrix*, 9(3), 102-109.



- Harahap, P. N., Sulindawaty. (2019). Implementasi Data Mining Dalam Memprediksi Transaksi Penjualan Menggunakan Algoritma Apriori (Studi Kasus PT.Arma Anugerah Abadi Cabang Sei Rampah). *Matics:Jurnal Ilmu Komputer Dan Teknologi Informasi*, 11(2), 46–50.
- Huddin, S., Haerani, E., dan Oktavia, L. (2023). Penerapan Fuzzy C-Means Pada Klasterisasi Penerima Bantuan Pangan Non Tunai. *KLIK: Kajian Ilmiah Informatika Komputer*, 4(1), 453–461.
- Junaidi, dan Junaidi, A. (2023). Pengaruh Produk Domestik Regional Bruto Dan Pendidikan Serta Upah Terhadap Tingkat Pengangguran. *KINERJA: Jurnal Ekonomi Dan Manajemen*, 20(3), 455–466.
- Kurniawan, H., Defit, S., dan Sumijan. (2020). Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal. *Journal of Applied Computer Science and Technology*, 1(2) 80–89.
- Maulana, A. R., Ulinnuha, N., dan Novitasari, D. C. R. (2023). Penerapan Fuzzy CMeans Untuk Pengelompokan Tingkat Kualitas Pendidikan Di Jawa Timur. *Simetris: Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 14(2), 419–426.
- Mayasari, S. N., dan Nugraha, J. (2023). Implementasi K-Means Cluster Analysis Untuk Mengelompokkan Kabupaten/Kota Berdasarkan Data Kemiskinan Di Provinsi Jawa Tengah Tahun 2022. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, 3(2), 317–329.
- Novianti, F., Aisyah Yasmin, Y. R., dan Novitasari, D. C. R. (2022). Penerapan Algoritma Fuzzy C-Means (FCM) dalam Pengelompokan Provinsi di Indonesia berdasarkan Indikator Penyakit Menular Manusia. *UMANJI (Jurnal Masyarakat Informatika Unjani)*, 6(1), 23–33.
- Rahakbauw, D. L., Ilwaru, V. Y. I., dan Hahury, M. H. (2017). Implementasi Fuzzy C-Means Clustering Dalam Penentuan Beasiswa. *Jurnal Ilmu Matematika Dan Terapan*, 11(1), 1–11.
- Suyanto. (2018). *Machine Learning Tingkat Dasar dan Lanjut*. Bandung: informatika Bandung.
- Ramadhan, A., Mustakim, M., dan Handinata, R. (2019). Implementasi Algoritma Fuzzy C-Means untuk Pengelompokan Wilayah Bencana Banjir. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI)*, 11, 171–177.
- Rouza, E., dan Fimawahib, L. (2020). Implementasi Fuzzy C-Means Clustering dalam Pengelompokan UKM Di Kabupaten Rokan Hulu. *Techno.COM*, 19(4), 481–495.
- Saputra, P. R. N., dan Chusyairi, A. (2020). Perbandingan Metode Clustering Dalam Pengelompokan Data Puskesmas Pada Cakupan Imunisasi Dasar Lengkap. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 4(6), 1077–1084.
- Sarbaini, Saputri, W., Nazaruddin, dan Muttakin, F. (2022). Cluster Analysis Menggunakan Algoritma Fuzzy K-Means Untuk Tingkat Pengangguran Di Provinsi Riau. *Jurnal Teknologi Dan Manajemen Industri Terapan (JTMIT)*, 1(2), 78–84.
- Sari, D. N., Oktaviana, H., dan Saifudin, I. (2022). Penerapan Data Mining Untuk Klasifikasi Gaya Belajar Siswa Menggunakan Algoritma C4.5. *Jurnal Smart Teknologi*, 3(4), 413–422.
- Sejati, D. P. (2020). Pengangguran Serta Dampaknya Terhadap Pertumbuhan Dan Pembangunan Ekonomi. *AKSELERASI: Jurnal Ilmiah Nasional*, 2(3), 98–105
- Syadidan, A. (2023). *Profil Provinsi Kalimantan Timur: Letak Geografis dan Sejarah*, (Online), ([Jurnalkaltim.Com.https://jurnalkaltim.com/kalimantan-timur/profil-provinsi-kalimantan-timur/](https://jurnalkaltim.com/kalimantan-timur/profil-provinsi-kalimantan-timur/)) diakses 18-01-2024.
- Tumilaar, T. v., Maramis, M. B., dan F. Dj. Siwu, H. F. D. (2022). Pengaruh Jumlah Penduduk, Pendidikan, Dan Upah Minimum Terhadap Tingkat Pengangguran Terbuka Di Kabupaten/Kota Provinsi Kalimantan Timur. *urnal Berkala Ilmiah Efisiensi*, 22(5), 61–72.
- Wardoyo, A. E., dan Triuspita, N. (2020). Penentuan Cluster Optimum pada Tingkat Pengangguran dan Tingkat Kemiskinan di Jawa Timur Menggunakan Algoritma Fuzzy C-Means. *IOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 1(2), 40–47.