JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI e-ISSN: 2614-8811
Published by Departement of Mathematics, Hasanuddin University, Indonesia
https://journal.unhas.ac.id/index.php/jmsk/index
Vol. 22, No. 1, September 2025, pp. 78-89
DOI: 10.20956/j.v22i1.44314

Comparison Of FCM And FKNN Methods For Clustering Provinces In Indonesia Based On People's Welfare Indicators

Perbandingan Metode FCM dan FKNN Untuk Pengelompokkan Provinsi di Indonesia Berdasarkan Indikator Kesejahteraan Rakyat

Maghfiroh[1]*, Tony Yulianto[2]*, Faisol[3]*
[*] Study Program of Mathematics, Faculty Of Mathematics and Natural Sciences, Madura Islamic University
Email: *[1]iamviroh@gmail.com, [2]toniyulianto65@gmail.com, [3]faisol.munif@gmail.com*
Corresponding author

Abstract

Improving people's welfare is one of the main indicators of a country's successful development. Public welfare covers various aspects, such as population, social, economic, and labor, which continue to evolve along with the growth and changes in human life needs. In Indonesia, various challenges such as rapid population growth, uneven population distribution, unemployment, poverty, and low Human Development Index (HDI) require strategic solutions to ensure equitable welfare. This study compares the Fuzzy C-Means (FCM) and Fuzzy K-Nearest Neighbor (FKNN) methods to cluster provinces in Indonesia based on public welfare indicators using variables of population density, population growth rate, percentage of poverty rate, Human Development Index (HDI), and Open Unemployment Rate (TPT). The clustering results are expected to support the government in formulating development policies that are more targeted and effective. The results showed that the best method was the Fuzzy C-Means method as many as 2 clusters with the highest Silhouette coefficient value of 0,97656 which states that the cluster structure formed in this clustering is very good. Cluster 1 is categorized as prosperous and cluster 2 is categorized as less prosperous.

Keywords: cluster, Fuzzy C-Means, Fuzzy K-Nearest Neighbor, People's Welfare

Abstrak

Peningkatan kesejahteraan rakyat merupakan salah satu indikator utama keberhasilan pembangunan suatu negara. Kesejahteraan masyarakat mencakup berbagai aspek, seperti kependudukan, sosial, ekonomi, dan ketenagakerjaan, yang terus berkembang seiring dengan pertumbuhan dan perubahan kebutuhan hidup manusia. Di Indonesia, berbagai tantangan seperti pertumbuhan penduduk yang pesat, distribusi populasi yang tidak merata, tingkat pengangguran, kemiskinan, dan rendahnya Indeks Pembangunan Manusia (IPM) memerlukan solusi strategis untuk memastikan pemerataan kesejahteraan. Penelitian ini membandingkan - metode *Fuzzy C-*

Means (FCM) dan *Fuzzy K-Nearest Neighbor* (FKNN) untuk mengelompokkan provinsi di Indonesia berdasarkan indikator kesejahteraan rakyat dengan menggunakan variabel kepadatan penduduk, laju pertumbuhan penduduk, persentase tingkat kemiskinan, Indeks Pembangunan Manusia (IPM), dan Tingkat Pengangguran Terbuka (TPT). Hasil pengelompokan diharapkan dapat mendukung pemerintah dalam merumuskan kebijakan pembangunan yang lebih tepat sasaran dan efektif. Hasil penelitian menunjukkan bahwa metode terbaik terdapat pada metode *Fuzzy C-Means* sebanyak 2 *cluster* dengan nilai *Silhouette coefficient* tertinggi sebesar 0,97656 yang menyatakan bahwa struktur *cluster* yang terbentuk pada pengelompokan ini adalah sangat baik. *Cluster* 1 dikategorikan sebagai wilayah sejahtera dan *cluster* 2 dikategorikan sebagai wilayah yang kurang Sejahtera.

Kata kunci: *Cluster, Fuzzy C-Means, Fuzzy K-Nearest Neighbor,* Kesejahteraan Rakyat

1. Pendahuluan

Pembangunan suatu negara dapat dinilai dari semakin naik atau tidaknya kesejahteraan rakyatnya. Kesejahteraan dapat diartikan sebagai suatu keadaan dimana setiap warga negara selalu berada dalam kondisi serba cukup dalam segala kebutuhannya, baik material maupun spiritual. Semakin meningkat pembangunan di suatu negara maka kesejahteraan rakyatnya juga semakin meningkat. Indonesia sebagai salah satu negara yang memiliki ciri-ciri luhur dalam pembangunan kesejahteraan rakyat sebagaimana cita-cita tersebut disebutkan di dalam Undang-Undang Dasar 1945, maka perlu membuat suatu kebijakan dan program yang tepat untuk mewujudkan cita-cita luhur tesebut [6]. Kesejahteraan rakyat merujuk pada tingkat kesejahteraan ekonomi dan sosial masyarakat yang dikaji dari delapan bidang meliputi kemiskinan, kependudukan, ketenagakerjaan, perumahan dan lingkungan, kesehatan dan gizi, pendidikan, pola konsumsi, serta bidang sosial lainnya yang menjadi acuan dalam upaya peningkatan kualitas hidup [16]. Indikator-indikator ini bersifat dinamis karena dipengaruhi oleh perubahan kebutuhan hidup manusia dan kondisi sosial ekonomi yang terus berkembang [1].

Pada tahun 2023, kepadatan penduduk di Indonesia diperkirakan mencapai 147,27 orang per kilometer persegi, yang mengalami kenaikan sekitar 1,08% dibandingkan tahun sebelumnya. Pertumbuhan ini berdampak pada ketidakseimbangan distribusi penduduk dan menimbulkan tekanan terhadap fasilitas dan layanan dasar [4]. Menurut Badan Pusat Statistik (BPS), kualitas pertumbuhan ekonomi berpengaruh terhadap kesejahteraan masyarakat. Pertumbuhan ekonomi biasanya diikuti oleh pengurangan kemiskinan. Di samping itu pertumbuhan ekonomi juga mempengaruhi tingkat pengangguran [11]. IPM juga menjadi salah satu indicator penting dalam menilai Tingkat kesejahteraan Masyarakat di berbagai daerah [15]. Untuk mendapatkan pemetaan yang akurat terhadap kondisi kesejahteraan di seluruh provinsi di Indonesia, diperlukan metode pengelompokkn *(clustering)* yang mampu merepresentasikan karakteristik masing-masing wilayah berdasarkan variable-variabel seperti kepadatan penduduk, laju pertumbuhan penduduk, persentase penduduk miskin, IPM dan Tingkat pengangguran terbuka [6].

Metode *clustering* adalah suatu metode pengelompokkan data ke dalam kelas atau *cluster* berdasarkan suatu kemiripan atribut di antara kelompok data. Hasil *cluster* yang optimal dapat dipengaruhi oleh metode *clustering* yang digunakan [5]. Ada beberapa metode yang dapat digunakan untuk mengelompokkan data menjadi beberapa data, di antaranya adalah dengan menggunakan salah satu cabang dari ilmu matematika, yaitu logika *fuzzy*. Dalam teori *fuzzy*, keanggotaan sebuah data tidak diberi nilai secara tegas dangan nilai 1 (menjadi anggota) dan 0 (tidak menjadi anggota), melainkan dengan suatu nilai derajat keanggotaan yang jangkauan nilainya 0 sampai 1 [2,13]. Pada penelitian ini digunakan metode *Fuzzy C-Means* dan *Fuzzy K-Nearest Neighbor*.

Penelitian terkait Metode FCM oleh oleh Isah dkk (2024) dengan judul "Perbandingan Metode *Fuzzy C-Means* Dan *Fuzzy Probabilistic C-Means* Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Indeks Pembangunan Manusia Tahun 2022", hasil yang diperoleh bahwa dari perbandingan nilai indeks validitas klaster *Partition Entropy* (PE) , *Partition Coefisien* (PC), dan *Modified Partition Coefisien* (MPC) pada metode *Fuzzy C-Means* dan *Fuzzy Probabilistic C-Means* diperoleh bahwa metode *Fuzzy C-Means* merupakan metode yang terbaik dalam kasus ini. Nilai evaluasi yang diperoleh, yaitu PE sebesar 0,3505417, PC sebesar 0,8170827, dan MPC sebesar 0,75de1103. *Partition Entropy* (PE) adalah evaluasi yang mengukur Tingkat kekaburan partisi klaster. *Partition Coefisien* (PC) adalah validitas dengan menghitung koefisien partisi sebagai evaluasi derajat keanggotaan data pada setiap klaster. *Modified Partition Coefisien* (MPC) adalah indeks validitas hasil perbaikan indeks PC yang memperhitungkan jumlah klasternya. Ketiga indeks validitas tersebut digunakan untuk mengukur tingkat akurasi pada metode FCM dan FPCM [7]. Sedangkan penelitian menggunakan metode FKNN dilakukan oleh Mas`udia dkk (2018) dengan judul "*Analysis of Comparison of Fuzzy KNN, C4.5 Algorithm, and Naïve Bayes Classification Method for Diabetes Mellitus Diagnosis*" yang memberikan hasil metode terbaik pada penelitian tersebut yaitu metode *Fuzzy K-Nearest Neighbor* dengan nilai rata-rata akurasi 96%, sedangkan klasifikasi menggunakan algoritma C4.5 nilai rata-rata akurasi 79,5%, dan klasifikasi menggunakan metode *Naïve Bayes* nilai rata-rata akurasi 87,5% [10].

Berdasarkan latar belakang tersebut, peneliti tertarik untuk melakukan perbandingan metode *Fuzzy C-Means* dan *Fuzzy K-Nearest Neighbor* dengan harapan diperoleh hasil pengelompokkan yang optimal sehingga pemerintah dapat mengambil atau memutuskan kebijakan dan strategi yang baik atau tepat sasaran dalam pembangunan,

2. METODE PENELITIAN

Studi Literatur

Pada langkah ini studi literatur yang dilakukan dengan cara mengkaji teori-teori yang akan mendukung dalam pemecahan masalah tentang indikator kesejahteraan rakyat dan metode *fuzzy* yaitu dengan cara mencari referensi yang menunjang penelitian berupa jurnal, buku-buku, skripsi, serta penelitian-penelitian sebelumnya yang berkaitan dengan indikator kesejahteraan rakyat ataupun tentang metode *fuzzy*.

Pengambilan Data

Pada tahap ini dilakukan proses pengambilan data sekunder mengenai indikator kesejahteraan rakyat yang terdiri dari kepadatan penduduk, laju pertumbuhan penduduk, persentase penduduk miskin, Indeks Pembangunan Manusia (IPM) dan Tingkat Pengangguran Terbuka (TPT) pada tahun 2024. Data ini diambil dari *website* resmi Badan Pusat Statistik Indonesia yaitu https://www.bps.go.id/id

Penerapan Metode *Fuzzy C-Means* (FCM)

Clustering dengan metode *Fuzzy C-Means* (FCM) didasarkan pada teori logika *fuzzy*. Teori ini pertama kali diperkenalkan oleh Lothfi Zadeh pada tahun 1965 dengan nama himpunan *fuzzy (fuzzy set*). Dalam teori *fuzzy*, keanggotaan sebuah data tidak diberikan nilai secara tegas dengan nilai 1 (menjadi anggota) dan 0 (tidak menjadi anggota), melainkan dengan suatu nilai derajat keanggotaan yang jangkauan nilainya 0 sampai 1 [2,13]. Kelebihan metode *Fuzzy C-Means* yaitu sederhana, mudah di implementasikan, memiliki kemampuan untuk mengelompokkan data yang besar. Sedangkan kelemahan metode *Fuzzy C-Means* membutuhkan banyaknya kelompok dan matriks

keanggotaan kelompok yang ditetapkan sebelumnya [19]. Algoritma *Fuzzy C-Means* sebagai berikut [6].

1. Masukkan data yang akan di-*cluster*, berupa matriks partisi awal U berukuran $n \times m$, dimana $n =$ banyaknya sampel data dan $m =$ banyaknya atribut setiap data, x_{ij} adalah elemen dalam matriks data sampel ke-i ($i = 1,2,3,\dots,n$), varieabel ke-j ($j = 1,2,3,\dots,m$).
2. Tentukan :
   - Jumlah *cluster* yang akan dibentuk = $c(\geq 2)$
   - Pangkat pembobot = $w(> 1)$
   - Maksimum iterasi = *(MaxIter)*
   - Kriteria penghentian = ξ (nilai positif yang sangat kecil)
   - Iterasi awal $t = 1$ dan $\Delta = 1$
3. Menentukan bilangan random ($\mu_{ik}, i = 1,2,3,\dots,c; k = 1,2,3,\dots,n$), sebagai elemen matriks partisi awal U.
$$U_{n\times c} = \begin{bmatrix} \mu_{11}(x_1) & \mu_{12}(x_2) & \cdots & \mu_{1c}(x_c) \\ \mu_{21}(x_1) & \mu_{22}(x_2) & & \mu_{2c}(x_c) \\ \vdots & & \ddots & \vdots \\ \mu_{n1}(x_1) & \mu_{n2}(x_2) & \cdots & \mu_{nc}(x_c) \end{bmatrix} \quad (2.1)$$
   Entry-entry dalam kondisi matriks awal $U_{n\times c}$ pada *fuzzy clustering* harus memenuhi kondisi sebagai berikut:
$$\mu_{ik} = [0,1]$$
$$\sum_{i=1}^{n} \mu_{ik} = 1$$
4. Menghitung pusat *cluster* ke-k (v_{kj}), dengan $k = 1,2,3,\dots,c$; dan $j = 1,2,3..,m$. v_{kj} adalah titik-titik dari pusat setiap *cluster*, dengan
$$v_{kj} = \frac{\sum_{i=1}^{n}((\mu_{ik})^w \times x_{ij})}{\sum_{i=1}^{n}(\mu_{ik})^w} \quad (2.2)$$
   dimana v_{kj} ialah pusat cluster ke-i, μ_{ik} ialah nilai bobot ke-k dengan pusat cluster ke-i, x_{ij} ialah data sampel ke-i dan variabel ke-j, w ialah pembobot, dan n ialah jumlah objek penelitian.
5. Menghitung fungsi objektif p_t pada iterasi ke-t dengan persamaan sebagai berikut:
$$p_t = \sum_{k=1}^{n}\sum_{i=1}^{c}\left(\left[\sum_{j=1}^{m}\left(x_{kj} - v_{ij}\right)^2\right](\mu_{ik})^w\right) \quad (2.3)$$
   Untuk p_t ialah fungsi objektif pada iterasi ke-t, x_{ij} ialah data sampel ke- i dan variabel ke- j, μ_{ik} ialah nilai bobot ke- k dengan pusat *cluster* ke- i, c ialah banyaknya cluster dan m ialah banyaknya data pengamat.
6. Memperbarui nilai μ_{ik}:
$$\mu_{ik} = \frac{\left[\sum_{j=1}^{m}(x_{ij} - v_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{i=1}^{c}\left[\sum_{j=1}^{m}(x_{ij} - v_{kj})^2\right]^{\frac{-1}{w-1}}} \quad (2.4)$$
7. Cek kondisi berhenti:
   a. Jika nilai mutlak dari ($|p_t - p_{t-1}| < \xi$) atau $t >$ banyaknya iterasi maksimal, maka iterasi berhenti
   b. Jika tidak; maka $t = t + 1$ kemudian ulangi langkah ke-4

Penerapan Metode *Fuzzy K-Nearest Neighbor* (FKNN)

Fuzzy K-Nearest Neighbor merupakan sebuah metode yang akan mencari nilai derajat keanggotaan data uji pada setiap kelasnya, yang kemudian akan mengambil nilai derajat keanggotaan terbesar yang dihasilkan. Nilai derajat keanggotaan terbesar akan digunakan sebagai kelas hasil dari klasifikasi [14]. FKNN mempunyai kelebihan yaitu setiap objek akan memiliki

derajat keanggotaan pada setiap kelas sehingga akan memberikan kekuatan atau keyakinan yang lebih terhadap suatu objek dalam suatu kelas [8]. Sedangkan kelemahan metode FKNN yaitu antara sejumlah k tetangga dianggap sama pentingnya padahal belum tentu, hal tersebut diatasi dengan perhitungan derajat keanggotaan data pada tiap kelas. Selain itu pada data latih belum diketahui kekuatan keanggotaannya pada sebuah kelas dan untuk mengatasi hal tersebut menggunakan inisalisasi *fuzzy* [17]. Adapun tahap dari perhitungan *Fuzzy K-Nearest Neighbor* adalah sebagai berikut [18].

1. Melakukan proses menghitung jarak dari kedekatan tiap data dengan menggunakan jarak *euclidean* 2.5

$$d(x_1, x_2) = \sqrt{\sum_r^n (a_r(x_1) - a_r(x_2))^2} \quad (2.5)$$

   Untuk x_1, x_2 ialah dua *record* dengan n atribut, n ialah banyaknya data, dan a_r ialah nilai atribut ke-r pada *record.*
2. Melakukan Pengurutan Jarak terkecil hingga terbesar
3. Melakukan proses inisialisasi *fuzzy* pada data latih berdasarkan persamaan berikut:

$$u_{(x_j,c_i)} \begin{cases} 0{,}51 + \left(\frac{n_j}{K}\right) * 0{,}49 \,, jika\ j = i \\ \left(\frac{n_j}{K}\right) * 0{,}49 \qquad\quad , jika\ j \neq i \end{cases} \quad (2.6)$$

   dengan $u_{(x_j,c_i)}$ ialah nilai keanggotaan data, n_j ialah jumlah anggota kelas-j pada himpunan K tetangga terdekat data, K ialah jumlah tetangga terdekat, i ialah kelas data, dan j ialah kelas target.
4. Melakukan proses menghitung nilai keanggotaan data menggunakan persamaan berikut:

$$\mu_i(x) = \frac{\sum_{j=1}^{k} \mu_{ij} \left(1/\|x - x_j\|^{\frac{2}{m_1 - 1}}\right)}{\sum_{j=1}^{k} \left(1/\|x - x_j\|^{\frac{2}{m_1 - 1}}\right)} \quad (2.7)$$

   Untuk $\mu_i(x)$ ialah nilai keanggotaan data ke x ke kelas c_i, $x - x_j$ ialah selisih jarak data x ke x_j dalam K tetangga terdekat, dan m_1 ialah bobot pangkat yang besarnya $m > 1$
5. Menentukan hasil klasifikasi dilihat dari nilai terbesar pada nilai keanggotaan yang didapatkan

Simulasi

Pada langkah ini, metode *Fuzzy C-Means* dan metode *Fuzzy K-Nearest Neighbor* akan diterapkan pada data dengan bantuan software *Matlab* R2019a.

Validasi Hasil Simulasi

Berdasarkan hasil simulasi menggunakan *Matlab* R2019a dengan penerapan metode *Fuzzy C-Means* dan *Fuzzy K-Nearest Neighbor* pada pengelomopokkan kesejahteraan rakyat di Indonesia, maka cek validasi hasil simulasi dilakukan dengan menggunakan nilai *Silhouette coefficient.* Adapun *Silhouette coefficient* adalah metode untuk menentukan jumlah *cluster* optimal dengan menggunakan pendekatan nilai rata-rata *silhouette*. Metode ini mengevaluasi kualitas hasil pengelompokkan dalam analisis *cluster* menggunakan perbandingan jarak rata-rata suatu titik yang termasuk dalam *cluster* yang sama dengan *cluster* yang lain. Semakin tinggi nilai *silhouette coefficient* untuk suatu jumlah K, semakin optimal jumlah *cluster* yang dibentuk dan hasil pengelompokkan akan semakin baik [3]. Adapun nilai *silhouette coefficient* terhadap suatu objek amatan ke-i ditentukan melalui langkah-langkah sebagai berikut:

1. Menghitung jarak rata-rata data ke-i dengan seluruh data dalam satu *cluster* yang sama melalui persamaan berikut:

(2.8)

$$a_i(k) = \frac{1}{A_k - 1} \sum_{j-1, j \neq 1}^{A_k} \left| d_{i,k} - d_{j,k} \right|, j \neq i$$

dengan $a_i(k)$ adalah nilai jarak rata-rata antara data ke-i dengan seluruh data dalam *cluster* yang sama, A_k ialah banyaknya anggota pada satu cluster yang sama, dan $\left| d_{i,k} - d_{j,k} \right|$ ialah jarak data ke-i terhadap data ke-j pada suatu cluster yang sama.

2. Menghitung jarak rata-rata pada data ke-i terhadap *cluster* yang berbeda melalui persamaan berikut:

$$d_i(k, C) = \frac{1}{C} \sum_{j-1, j \neq 1}^{A_k} \left| d_{i,k} - d_{j,C} \right|, k \neq C \qquad (2.9)$$

kemudian mengambil nilai minimum menggunakan persamaan berikut:

$$b_i(k) = \min(d_i(k, C)) \qquad (2.10)$$

dengan $d_i(k, C)$ ialah jarak rata-rata pada data ke-i terhadap semua data yang berada pada *cluster* yang berbeda, C ialah banyaknya anggota *cluster* yang berbeda, dan $b_i(k)$ ialah nilai minimum dari $d_i(k, C)$.

3. Menghitung nilai dari *silhouette index* untuk setiap data ke-i dengan persamaan berikut:

$$SI_{i,k} = \frac{b_i(k) - a_i(k)}{\max(a_i(k), b_i(k))} \qquad (2.11)$$

dengan $SI_{i,k}$ ialah nilai *index sillhouette* data ke-i yang berada pada satu *cluster*

4. Menghitung nilai *index sillhouette* pada seluruh data dalam satu *cluster* dengan persamaan berikut:

$$SI_k = \frac{1}{A_k} \sum_{i=1}^{A_k} SI_{i,k} \qquad (2.12)$$

dengan SI_k ialah rata-rata *index silhouette* pada *cluster k*

5. Menghitung nilai *sillhouette coefficient* global pada persamaan berikut;

$$SC = \frac{\sum_{k=1}^{K} (A_k \times SI_k)}{\sum_{k=1}^{K} A_k} \qquad (2.13)$$

dengan SC ialah rata-rata *index silhouette* dari seluruh dataset

Menurut Rousseeuw tahun 1987, kriteria pengukuran *sillhouette coefficient* sebagai berikut [12].

sillhouette coefficient	Interpretasi yang disesuaikan
$0{,}7 < SC \leq 1{,}0$	: Susunan sangat baik
$0{,}5 < SC \leq 0{,}7$	: Susunan baik
$0{,}25 < SC \leq 0{,}5$	: Susunan lemah
$SC \leq 0{,}25$	: Susunan buruk

Penarikan Kesimpulan

Langkah ini merupakan tahapan terakhir dalam menyelesaikan penelitian ini. Setelah peneliti mendapatkan hasil akhir dari *Clustering* kesejahteraan rakyat tiap provinsi menggunakan metode *Fuzzy C-Means* dan metode *Fuzzy K-Nearest Neighbor* maka bisa ditarik kesimpulan dari penelitian ini menggunakan interpretasi *cluster*. Tahap interpretasi meliputi pengujian pada masing-masing *cluster* yang terbentuk untuk memberikan nama atau keterangan secara tepat sebagai gambaran sifat dari *cluster* tersebut, menjelaskan bagaimana bisa berbeda secara relavan pada tiap dimensi. Ketika memulai proses interpretasi digunakan rata-rata *(centroid)* setiap *cluster* pada setiap variable [9]. Adapun cara menghitung *centroid* disajikan pada persamaan (2.14)

(2.14)

$$C_{i,j} = \frac{\sum_{k=1}^{n \in C_i} X_{k,j}}{N_i}$$

dengan $C_{i,j}$ ialah nilai rata-rata (centroid) cluster, $X_{k,j}$ ialah data wilayah provinsi indikator kesejahteraah rakyat, dan N_i ialah banyak cluster ke-i.

3. HASIL DAN PEMBAHASAN

Statistika Deskriptif

Statistika deskriptif pada penelitian ini berisi tentang rata-rata *(mean)*, standar deviasi, data terendah (minimum) dan data tertinggi (maksimum) variabel kepadatan penduduk x_1, laju pertumbuhan penduduk x_2, persentase penduduk miskin semester 1 (Maret) x_3, persentase pendduk miskin semester 2 (September) x_4, Indeks Pembangunan Manusia (IPM) x_5, dan Tingkat Pengangguran Terbuka (TPT) x_6. Terlihat pada Tabel 1.

Tabel 1. Deskripsi Variabel Penelitian

Variabel	Minimum	Maksimum	$mean$	Standar deviasi
Kepadatan Penduduk	11	16165	755,9	2711,6
Laju Pertumbuhan	0,31	1,93	1,27	0,32
Persentase Penduduk Miskin (Maret)	4	21,7	9,57	4,5
Persentase Penduduk Miskin (September)	3,8	21,09	9,17	4,41
Indeks Pembangun Manusia	67,69	84,15	74,7	3,21
Tingkat Pengangguran Terbuka	3,66	13,7	8,92	2,55

Untuk mendukung proses pengelompokkan provinsi di Indonesia berdasarkan indicator kesejahteraan rakyat, digunakan enam variable utama. Kepadatan penduduk pada wilayah yang diamati bervariasi dari 11 hingga 16.165 jiwa per km². Rata-rata kepadatan adalah 755,9 jiwa/km², dengan standar deviasi sebesar 2.711,6 jiwa/km. Laju pertumbuhan penduduk menunjukkan nilai minimum 0,31% dan maksimum 1,93% dengan rata-rata 1,27% serta standar deviasi 0,32. Persentase penduduk miskin pada bulan maret berkisar 4% hingga 21,7% dengan rata-rata 9,57% dan standar deviasi 4,5. Sedangkan pada bulan September berkisar 3,8%-21,09%, dengan rata-rata 9,17% dan standar deviasi 4,41. IPM berada pada rentang 67,69 sampai dengan 84,15 dengan nilai rata-rata 74,7 dan satndar deviasi 3,21. Sementara itu, TPT tercatat antara 3,66% hingga 13,7% dengan rata-rata 8,92% dan standar deviasi 2,55.

Penerapan Metode *Fuzzy C-Means*

Proses perhitungan pengelompokkan atau *clustering* akan dimulai dengan menginisialiasi jumlah *cluster* terlebih dahulu dengan nilai *cluster* atau $k = 2$, $w = 4$, *MaxIter* = 100, dan $\xi = 10^{-5}$. Awal perhitungan pada FCM akan dihitung dengan matriks random U dengan ukuran 36×4. Pusat *cluster* juga menggunakan matriks random awal perhitungan yang akan mempengaruhi pusat *cluster*. Dengan bantuan perhitungan *software Matlab* didapatkan pusat *cluster* pada iterasi ke-14 (terakhir) sebagai berikut:

Tabel 2. Pusat *Cluster* FCM

	X_1	X_2	X_3	…	X_{33}	X_{34}
Cluster 1	0,1524	0,5383	0,0782	⋮	0,6477	0,547
Cluster 2	0,8258	0,9961	0,4427	⋮	0,4509	0,2963

Nilai-nilai pada Tabel 2 menunjukkan bahwa suatu objek bisa memiliki keanggotaan lebih dari satu *cluster*, namun derajatnya berbeda. Contohnya, x_1 memiliki keanggotaan paling tinggi sebesar

0,8258 pada *cluster* C2, yang berarti bahwa objek x_1 sangat kuat berada di *cluster* C2. Matriks partisi ini diperbarui secara bertahap selama proses klasterisasi sampai akhirnya mencapai kondisi stabil, yaitu saat perubahan nilainya dari satu tahap ke tahap berikutnya sudah sangat kecil atau hampir tidak berubah sama sekali. Berikut hasil iterasi terakhir yaitu iterasi ke-14 pada Tabel 3.

Tabel 3. Hasil Metode Fuzzy C-Means

FCM: Iterasi 1	error: 1.7324e-06+0.98752i						
	X1	X2	X3	X4	X5	X6	Cluster
Data ke-1	98	1,39	14,23	12,6	75,36	11,3	2
Data ke-2	215	1,4	7,99	7,19	75,76	10,7	2
Data ke-3	139	1,43	5,97	5,42	76,43	11,5	2
Data ke-4	75	1,37	6,67	6,36	75,67	7,55	2
Data ke-5	76	1,3	7,1	7,26	74,36	8,93	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Data ke-34	15	1,45	17,26	18,1	73,83	12,3	2

Nilai *error* yang menurun dari iterasi ke iterasi dan pada iterasi ke 14 bernilai 1.7324e-06+0.98752i menunjukkan bahwa metode FCM berhasil memperbarui posisi pusat *cluster* dan keanggotaan dengan konsisten menuju pembentukan *cluster* yang stabil. Jika nilai *error* sudah sangat kecil atau tidak banyak berubah di setiap perulangan, maka proses perulangannya akan dihentikan.

Penerapan Metode *Fuzzy K-Nearest Neighbori*

Dalam penelitian ini, metode FKNN digunakan untuk melakukan proses klasterisasi terhadap data yang tidak berlabel, meskipuns secara umum FKNN lebih tepat digunakan untuk klasifikasi berbasis data berlabel. Penelitiaan ini dilakukan untuk melihat sejauh mana metode ini mampu mengelompokkan data secara akurat. Proses klasifikasi dilakukan dengan menentukan nilai keanggotaan *fuzzy* dari setiap data terhadap *cluster* berdasarkan mayoritas label dari *k* tetangga terdekatnya. Proses pengelompokkan dilakukan sepenuhnya mengguakan bantuan perangkat lunak MATLAB. Tabel 4. berikut ini menampilkan hasil dari penerapan metode *Fuzzy K-Nearest Neightbor*

Tabel 4. Hasil Metode Fuzzy K-Nearest Neighbor

	X1	X2	X3	X4	X5	X6	Cluster
Data ke-1	98	1,39	14,23	12,64	75,36	11,31	1
Data ke-2	215	1,4	7,99	7,19	75,76	10,7	2
Data ke-3	139	1,43	5,97	5,42	76,43	11,54	1
Data ke-4	75	1,37	6,67	6,36	75,67	7,55	2
Data ke-5	76	1,3	7,1	7,26	74,36	8,93	2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
Data ke-34	15	1,45	17,26	18,09	73,83	12,29	2

Pada Tabel 4 metode FKNN menghasilkan 2 *cluster*, yaitu *cluster* 1 dengan 18 provinsi dan *cluster* 2 dengan 16 provinsi. Perhitungan akhir pada masing-masing metode diperoleh nilai akhir *cluster* dilihat pada Tabel 5

Tabel 5. Hasil *Clustering* Metode FCM dan FKNN

Data ke-	*Cluster*	
	FCM	FKNN
1	2	1
2	2	2
3	2	1
4	2	2
5	2	2
⋮	⋮	⋮
34	2	1

Berdasarkan Tabel 5, metode FCM menghasilkan *cluster* 1 dengan 1 provinsi dan *cluster* 2 dengan 33 provinsi, sedangkan metode FKNN menghasilkan *cluster* 1 dengan 18 provinsi dan *cluster 2* dengan 16 provinsi.

Validasi Hasil Cluster

Setelah hasil *clustering* diperoleh, evaluasi dilakukan menggunakan *Silhouette Coefficient* (SC), dengan nilai masing-masing metode disajikan pada Tabel 6.

Tabel 6. Nilai *Silhouette Coefficient*

Metode *Clustering*	Nilai *Silhouette Coefficient*	Interpretasi
FCM	0,97656	Susunan Sangat Baik
FKNN	0,11982	Susunan Buruk

Nilai *Silhouette* untuk metode FCM sebesar 0,97656, termasuk dalam kategori sangat baik ($0{,}70 < SC \leq 1{,}00$), yang menunjukkan bahwa hasil klasterisasi optimal dan data terkelompokkan dengan baik. Sebaliknya, metode FKNN menghasilkan nilai *Silhouette* sebesar 0,1192, tergolong buruk ($SC \leq 0{,}25$), menandakan kelemahan dalam kedekatan antar objek dalam *cluster*. Dengan demikian, metode FCM memberikan hasil pengelompokan kesejahteraan rakyat yang lebih baik dibandingkan FKNN.

Penarikan Kesimpulan

Untuk mengetahui karakteristik masing-masing *cluster* dilakukan analisis karakteristik *cluster* berdasarkan interpretasi *cluster* dari setiap *cluster* yang hasilnya dapat dilihat pada Tabel 7.

Tabel 7. Hasil Interpretasi *Cluster*

	X1	X2	X3	X4	X5	X6
Cluster 1	16165	0,31	4,3	4,14	84,15	12,24
Cluster 2	289,03	1,302	9,725	9,32	74,415	8,822

Berdasarkan hasil analisis *cluster*, wilayah terbagi ke dalam dua kelompok dengan karakteristik sosial-ekonomi yang berbeda, ditunjukkan pada Tabel 8 berikut:

Tabel 8. Jumlah Provinsi dari hasil *cluster*

Cluster	Jumlah Provinsi	Nama Provinsi
Cluster 1: wilayah sejahtera	1	DKI Jakarta

Cluster 2: wilayah kurang sejahtera	33	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua.

Adapun Tabel 8. Terlihat bahwa :

a. *Cluster* 1: wilayah Sejahtera

Cluster 1 dicirikan oleh nilai Indeks Pembangunan Manusia (IPM) yang sangat tinggi (84,15), tingkat kemiskinan yang rendah (4,3% pada semester 1 dan 4,14% pada semester 2), serta kepadatan penduduk yang sangat tinggi (16.165 jiwa/km²). Hal ini menunjukkan bahwa wilayah-wilayah dalam *cluster* ini secara umum tergolong sejahtera dan cenderung berada di kawasan perkotaan atau pusat ekonomi. Namun demikian, *cluster* ini juga menunjukkan tingkat pengangguran terbuka (TPT) yang cukup tinggi, yaitu sebesar 12,24%. Hal ini mengindikasikan bahwa meskipun masyarakatnya memiliki kualitas hidup yang baik, masih terdapat tantangan dalam aspek ketenagakerjaan yang perlu menjadi perhatian.

b. *Cluster* 2: Wilayah Kurang Sejahtera

Cluster 2 memiliki karakteristik yang berbeda, yaitu kepadatan penduduk yang sangat rendah (289,03 jiwa/km²), namun dengan laju pertumbuhan penduduk yang cukup tinggi (1,302%). Tingkat kemiskinan di wilayah ini juga relatif tinggi, yaitu 9,725% pada semester 1 dan 9,32% pada semester 2. Nilai IPM berada di kategori sedang (74,415), yang mencerminkan kualitas hidup yang belum sebaik *Cluster* 1. Meskipun demikian, tingkat pengangguran di *Cluster* 2 lebih rendah dibandingkan *Cluster* 1, yaitu sebesar 8,822%. Wilayah-wilayah dalam *cluster* ini kemungkinan besar merupakan daerah pedesaan atau sedang berkembang, dengan kebutuhan peningkatan kualitas hidup dan pengurangan angka kemiskinan sebagai fokus utama pembangunan.

KESIMPULAN

Berdasarkan hasil pengelompokkan dengan menggunakan metode *Fuzzy C-Means* dan *Fuzzy K-Nearest Neighbor* didapatkan nilai evaluasi terbaiknya ditunjukkan pada metode *Fuzzy C-Means* dengan nilai *Silhouette Coefficient* sebesar 0,97656 dibandingkan dengan nilai *Silhouette Coefficient* pada metode *Fuzzy K-Narest Neighbor* dengan nilai sebesar 0,11982. Hasil pengelompokkan menggunakan metode *Fuzzy C-Means* dengan banyaknya *cluster* yang terbentuk yaitu 2 *cluster*. *Cluster* 1 dikategorikan sebagai wilayah sejahtera dengan 1 provinsi yang berada di kawasan perkotaan dengan tantangan utama pada aspek ketenagakerjaan. Sementara itu *cluster* 2 dikategorikan sebagai wilayah yang kurang sejahtera dengan tantangan dalam meningkatkan kualitas hidup serta menurunkan angka kemiskinan .

DAF/.TAR PUSTAKA

[1] Alwi, W. & Hasrul, M., 2018. Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Propinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat. *Jurnal Matematika dan Statistika Serta Aplikasinya*, 6(1), pp. 1-8.

[2] Astria, D. & Suprayogi, 2017. Penerapan Algoritma Fuzzy C-Means Untuk Clustering Pelanggan Pada CV. Mataram Jaya Bawen. *Eksplora Informatika*, 6(2), pp. 169-178.

[3] Ayuni, A. P., Kusnandar, D. & Martha, S., 2024. Implementasi Algoritma K-Medoids Dan Clustering Large Applications (Clara) Dengan Optimasi Silhouette Coefficient (Studi Kasus: Pengelompokan Indeks Pembangunan Manusia Berdasarkan Kabupaten/Kota di Indonesia). *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*, 13(2), pp. 191-200.

[4] Christiani, C., Tedjo , P. & Martono, B., 2014. Analisis Dampak Kepadatan Penduduk Terhadap Kualitas Hidup Masyarakat Provinsi Jawa Tengah. *Serat Acitya*, 3(1), pp. 102-114.

[5] Dewi, D. A. I. C. & Dewa, P. A. K., 2019. Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *JURNAL MATRIX*, November, 9(3), pp. 102-109.

[6] Dwitiyanti, N., Selvia, N. & Andrari, F. R., 2019. Penerapan Fuzzy C-Means Cluster Dalam Pengelompokan Provinsi Indonesia Menurut Indikator Kesejahteraan Rakyat. *Fakt. Exacta*, 12(3), pp. 201-209.

[7] Isah, Estri, M. N. & Larasati, N., 2024. Perbandingan Metode Fuzzy C-Means Dan Fuzzy Probabilistic C-Means Dalam Pengelompokan Provinsi Di Indonesia Berdasarkan Indeks Pembangunan Manusia Tahun 2022. *Seminar Nasional Sains Data 2024 (SENADA 2024)*, pp. 832-841.

[8] Kusuma, J., Kurniati, . A. P. & Karo, I. M. K., 2022. Analysis of Expertise Group Using The Fuzzy K-NN Classification Algorithm (Case Study: School of Computing Telkom University). *JURIKOM (Jurnal Riset Komputer)*, Juni, 9(3), pp. 564-572.

[9] Laeli, S., 2014. *Analisis Cluster dengan Average Linkage Method dan Ward's Method untuk Data Responden Nasabah Asuransi Jiwa Unit Link.* Yogyakarta: Universitas Negeri Yogyakarta.

[10] Mas'udia, P. E., Rismanto, R. & Mas`ud, A., 2018. Analysis of Comparison of Fuzzy Knn, C4.5 Algorithm, and Naïve Bayes Classification Method for Diabetes Mellitus Diagnosis. *International Journal of Computer Applications Technology and Research*, 7(8), pp. 363-369.

[11] Pitaloka, G. F., Dwidayati, N. K. & Mulyono, 2019. Perbandingan Metode Dalam Analisis Cluster Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah. *Seminar Nasional Edusainstek*, pp. 543-552.

[12] Rahmawati, T., Wilandari, Y. & Kartikasari, . P., 2024. Analisis Perbandingan Silhouette Coefficientdan Metode Elbowpada Pengelompokkan Provinsi Di Indonesia Berdasarkan Indikator Ipmdengan K-Medoids. *JURNAL GAUSSIAN*, 13(1), pp. 13-24.

[13] Ramadhan, A., Efendi, . Z. & Mustakim, 2019. Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling. *Seminar Nasional Teknologi Informasi, Komunikasi dan Industri (SNTIKI) 9*, pp. 219-226.

[14] Ramadhani, F., Satria, A. & Sari, . I. P., 2023. Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah. *Hello Word Jurnal Ilmu komputer*, Mei, 2(2), pp. 58-62.

[15] Salsabila, T. A. & Wachidah, L., 2022. Analisis Multidimensional Scaling pada Pemetaan Kabupaten/Kota di Jawa Barat Berdasarkan Indikator Kesejahteraan Rakyat. *Bandung Conference Series: Statistics*, July, 2(2), pp. 173-179.

[16] Saputri, F. W. & Ariantob, D. B., 2023. Perbandingan Performa Algoritma K-Means, Kmedoids, Dan Dbscan Dalam Penggerombolan Provinsi Di Indonesia Berdasarkan Indikator Kesejahteraan Masyarakat. *Jurnal Teknologi Informasi* , 17(2), p. Agustus.

[17] Satria, D. & Harmaini, 2023. Implementasi Algoritme Fuzzy K-Nearest Neighbor Untuk Klasifikasi Urutan Metagenom. *ADIL*, 5(1), pp. 36-45.

[18] Siburian, N., Cholissodin, I. & Adikara, P. P., 2020. Penerapan Metode Fuzzy K-Nearest Neighbor Pada Klasifikasi Penyakit Menular Seksual Pria. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* November, 4(11), pp. 4096-4102.

[19] Yasir, A. & Firmansyah, A. U., 2024. Implementasi Metode Fuzzy C-Means Dan Metode Ahp Dalam Pemilihan Promosi Jabatan Karyawan Berbasis Web (Studi Kasus: PT. Tunas Dwipa Matra Sekampung). *Journal of Science and Social Research,* November, VII(4), pp. 1616-1619.