

Comparison of K-Means and Fuzzy C-Means Algorithms for Clustering Based on Happiness Index Components Across Provinces in Indonesia

Inna Auliya, Fadhilah Fitri*, Nonong Amalita, dan Tessy Octavia Mukhti

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: fadhilahfitri@fmipa.unp.ac.id

Submitted : 30 Januari 2024

Revised : 23 Februari 2024

Accepted : 25 Februari 2024

ABSTRACT

Cluster analysis is a statistical technique used to group objects based on their shared characteristics. This research aims to assess how 34 provinces in Indonesia are clustered using happiness index indicators for the year 2021. The study compares two non-hierarchical cluster analysis methods, K-Means and Fuzzy C-Means. K-Means categorizes objects into clusters based on their proximity to the nearest cluster center, while Fuzzy C-Means employs a fuzzy grouping model assigning membership degrees from 0 to 1. The results indicate that both methods form three clusters. Evaluating standard deviation values and ratios, Fuzzy C-Means proves superior, displaying a larger standard deviation between groups and a smaller ratio (0.6680004) compared to K-Means. Consequently, the study concludes that the Fuzzy C-Means method is more optimal than K-Means.

Keywords: Cluster analysis, Fuzzy C-Means, K-Means, Happiness index



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Analisis multivariat dapat dipergunakan untuk menyederhanakan pengelompokan data, salah satu caranya adalah dengan memakai analisis *cluster* (Hair dkk, 2013). Analisis *cluster* ini mengelompokkan objek berdasarkan kesamaan karakteristik. Objek yang sangat mirip termasuk pada satu *cluster* serta objek yang kurang mirip termasuk dalam *cluster* lain atau tidak termasuk pada *cluster* yang sama (Mattjik dan Sumertajaya, 2011). Analisis *cluster* ini dibagi menjadi dua metode yaitu metode hirarki dan non-hierarki. Metode hirarki digunakan dalam mengelompokkan objek berdasarkan kecenderungan karakteristik, dan jumlah *cluster* yang diperlukan belum diketahui. Menurut Johnson dan Wichern (2002), metode non-hierarki mengelompokkan objek kepada kelompok, serta jumlah kelompok ditentukan lebih dahulu. Metode non-hierarki lebih efisien untuk memproses kumpulan data yang besar dibandingkan metode hirarki. Beberapa metode non-hierarki yang umum digunakan adalah K-Means dan Fuzzy C-Means.

K-Means melakukan analisis pengelompokan objek berdasarkan jaraknya dari pusat cluster terdekat (Johnson dan Wichern, 2002). Kelebihan K-Means meliputi sederhana, efektif, cepat, dan mampu menangani dataset besar dengan baik (Wu, 2012). Sedangkan, Fuzzy C-Means adalah teknik analisis pengelompokan yang memakai model *fuzzy clustering*, di mana pembentukan *cluster* berdasarkan pada tingkat keanggotaan yang bervariasi antara 0 hingga 1. Fuzzy C-Means mempunyai keunggulan, terutama dalam penanganan tumpang tindih cluster, sehingga menghasilkan informasi yang baik (Cebeci dan Yildiz, 2015). Metode ini juga memberikan detail tentang hubungan data dengan setiap *cluster* berdasarkan tingkat keanggotaan, memberikan informasi mendalam tentang ciri *cluster*. K-Means dan Fuzzy C-Means disebut cocok untuk data variabel kontinu (Luthifaturrahmah, 2014). Menurut Timothy J. Ross (2010) menggambarkan adanya dua jenis pengelompokan, yaitu pengelompokan tegas dan pengelompokan samar. K-Means dikategorikan sebagai pengelompokan tegas sebab setiap observasi ditempatkan secara tepat pada satu *cluster*, sedangkan Fuzzy C-Means dianggap sebagai pengelompokan samar sebab setiap observasi masuk ke suatu *cluster* dengan kemungkinan menjadi anggota parsial.

Penelitian sebelumnya, metode K-Means dan Fuzzy C-Means telah dibandingkan. Contohnya, Ghosh dan Dubey (2013) menggunakan dataset Iris dan menemukan bahwa K-Means memberikan hasil yang lebih baik dalam hal waktu komputasi dibandingkan dengan Fuzzy C-Means. Sivarathri dan Govardhan (2014) juga menemukan bahwa Fuzzy C-Means memerlukan lebih banyak iterasi dan waktu, tetapi memberikan akurasi dan kualitas *cluster* yang lebih baik. Analisis *cluster* dapat diterapkan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan indeks

kebahagiaan, yang mencerminkan tingkat kepuasan masyarakat terhadap kehidupan sosial sehari-hari. Menurut laporan *World Happiness Report 2021*, Indonesia berada di peringkat 87 dari 146 negara dengan nilai kebahagiaan 71,49, menunjukkan tingkat kebahagiaan yang relatif lebih rendah dibandingkan dengan beberapa negara di dunia (Helliwell dkk, 2022).

Berdasarkan permasalahan ini, penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia dengan membandingkan metode K-Means dan metode Fuzzy C-Means, dengan fokus pada indeks kebahagiaan Indonesia tahun 2021. Harapannya, hasil analisis ini dapat memberikan gambaran yang bermanfaat bagi masyarakat dan pemerintah, sehingga mereka dapat bekerja sama untuk meningkatkan indeks kebahagiaan Indonesia dan mengatasi kesenjangan antar provinsi.

II. METODE PENELITIAN

Penelitian ini termasuk dalam kategori penelitian terapan yang menggunakan data sekunder dari publikasi Badan Pusat Statistik (BPS) mengenai Indeks Kebahagiaan Indonesia di 34 provinsi di Indonesia pada tahun 2021. Studi ini menganalisis 19 variabel seperti Kesehatan (X_1), Pendidikan dan keterampilan (X_2), Keharmonisan keluarga (X_3), Ketersediaan waktu luang (X_4), Hubungan sosial (X_5), Keadaan lingkungan (X_6), Kondisi keamanan (X_7), Pekerjaan usaha kegiatan utama (X_8), Pendapatan rumah tangga (X_9), Kondisi rumah dan fasilitas rumah (X_{10}), Perasaan/senang/riang/gembira (X_{11}), Perasaan/tidak khawatir/cemas (X_{12}), Perasaan tidak tertekan (X_{13}), Kemandirian (X_{14}), Penguasaan lingkungan (X_{15}), Pengembangan diri (X_{16}), Hubungan positif dengan orang lain (X_{17}), Tujuan hidup (X_{18}), dan Penerimaan diri (X_{19}).

Analisis dilakukan dengan memanfaatkan *software Rstudio*. Tahap pengelompokan dilakukan dengan menggunakan metode K-Means dan analisis Fuzzy C-Means dijalankan sesuai dengan langkah-langkah berikut.

1. Melakukan Standardisasi

Standardisasi terjadi ketika variabel memiliki ukuran unit yang berbeda (Alwi dan Hasrul, 2018). Standardisasi dapat dilakukan dengan menggunakan persamaan berikut.

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1)$$

Keterangan:

- z_{ij} : *z score* objek ke-*i* variabel ke-*j*
- x_{ij} : Objek ke-*i* variabel ke-*j*
- \bar{x}_j : Rata-rata variabel ke-*j*
- s_j : Simpangan baku variabel ke-*j*

2. Menentukan Cluster Optimal

Metode Elbow digunakan untuk menentukan jumlah cluster optimal dengan melihat penurunan paling besar antara jumlah cluster yang diuji. Jika terdapat siku pada grafik antara nilai cluster pertama dan kedua, jumlah cluster tersebut dianggap yang terbaik untuk analisis data.

$$SSE = \sum_{i=1}^g \sum_{x_{ij} \in V_{kj}} \|x_{ij} - v_{kj}\|^2 \quad (2)$$

Keterangan:

- g : Jumlah Cluster
- x_{ij} : Objek ke-*i* pada variabel ke-*j*
- v_{kj} : Merupakan pusat cluster objek ke-*k* pada variabel ke-*j*
- $x_{ij} \in V_{kj}$: Nilai keanggotaan titik data x_{ij} ke pusat cluster V_{kj}
- g : Jumlah Cluster

3. Algoritma K-Means

K-Means menganalisis objek menjadi cluster berdasarkan jarak ke pusat cluster terdekat objek tersebut (Johnson dan Wichern, 2002). Berikut algoritma metode K-Means:

- a. Menetapkan terlebih dahulu banyaknya cluster (g).
- b. Menetapkan titik pusat cluster (v_{ij}) yang dipilih secara acak.
- c. Hitung jarak objek ke titik pusat (d_{ik}).
- d. Mengelompokkan objek yang memiliki jarak terdekat objek ke titik pusat cluster.
- e. Untuk pusat cluster pada iterasi selanjutnya dapat dihitung dengan persamaan berikut.

$$d_{ik} = \left[\sum_{j=1}^p (x_{ij} - v_{kj})^2 \right]^{\frac{1}{2}} \quad (3)$$

Keterangan:

v_{ij} : Pusat *cluster* ke- i pada variabel ke- j

x_{kj} : Objek ke- k pada variabel ke- j

n_i : Banyaknya objek pada *cluster* ke- i

Proses iterasi berhenti jika tidak terjadi perpindahan objek antar *cluster*.

4. Algoritma Fuzzy C-Means

Menurut Everitt dkk, (2011), dalam *Fuzzy Clustering*, setiap objek memiliki tingkat keanggotaan yang mencerminkan intensitasnya dalam satu atau lebih *cluster*. Ini memungkinkan interpretasi probabilitas karena tingkat keanggotaan berkisar antara nol dan satu. Berikut adalah algoritma Fuzzy C-Means.

- Menetapkan banyaknya *cluster* (g), nilai pembobot keanggotaan w , dan error terkecil yang diharapkan ξ .
- Membuat matriks awal $U^{(0)}$ secara acak yang terdiri dari nilai derajat keanggotaan yang dapat dilihat pada persamaan berikut.

$$U_{g \times n} = \begin{bmatrix} \mu_{11}(x_1) & \mu_{21}(x_2) & \dots & \mu_{1n}(x_n) \\ \mu_{21}(x_1) & \mu_{22}(x_2) & \dots & \mu_{2n}(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{g1}(x_1) & \mu_{g2}(x_2) & \dots & \mu_{gn}(x_n) \end{bmatrix} \quad (4)$$

Keterangan:

$U_{g \times n}$: Matriks partisi berukuran $g \times n$

g : Ukuran *cluster* yang digunakan

n : Banyaknya Objek

μ_{ik} : adalah derajat keanggotaan objek ke- k pada *cluster* ke- i

- Menghitung pusat *cluster* ke- i pada variabel ke- j menggunakan persamaan berikut.

$$v_{ij} = \frac{\sum_{k=1}^n (\mu_{ik})^w \cdot x_{kj}}{\sum_{k=1}^n (\mu_{ik})^w} \quad (5)$$

Keterangan:

v_{ij} : Pusat *cluster* ke- i variabel ke- j

μ_{ik} : Derajat keanggotaan objek ke- k pada *cluster* ke- i

r : Iterasi

w : Parameter Pembobot keanggotaan

- Hitung jarak setiap objek ke titik pusat (d_{ik}).
- Lakukan update nilai matriks partisi U^t menggunakan persamaan berikut.

$$\mu_{ik} = \left[\sum_{l=1}^g \left(\frac{d_{il}}{d_{lk}} \right)^{\frac{2}{w-1}} \right]^{-1} \quad (6)$$

d_{ik} dengan merupakan jarak objek ke- k pada pusat *cluster* ke- i .

- Pengecakan kondisi berhenti.

Jika $(|U^t - U^{t-1}| < \xi)$, maka itersi berhenti

Jika tidak : maka $t = t + 1$ kemudian ulangi langkah ke-3

Dengan ε_L merupakan tingkat akurasi yang ditentukan.

5. Validasi Analisis *Cluster*

Sebuah pengelompokan yang baik memiliki keragaman besar antar kelompok dan keragaman kecil di dalam kelompok (Johnson dan Wichern, 2002). Menurut Bunkerrs dan James (1996), simpangan baku di dalam suatu kelompok dapat dirumuskan sebagai berikut:

$$S_w = \frac{1}{g} \sum_{k=1}^g S_k \quad (7)$$

dengan $S_k = \sqrt{\frac{1}{n_k - 1} \sum_{i=1}^{n_k} (x_{ik} - \bar{x}_k)^2}$ (8)

dimana $k = 1, 2, \dots, g$ dan

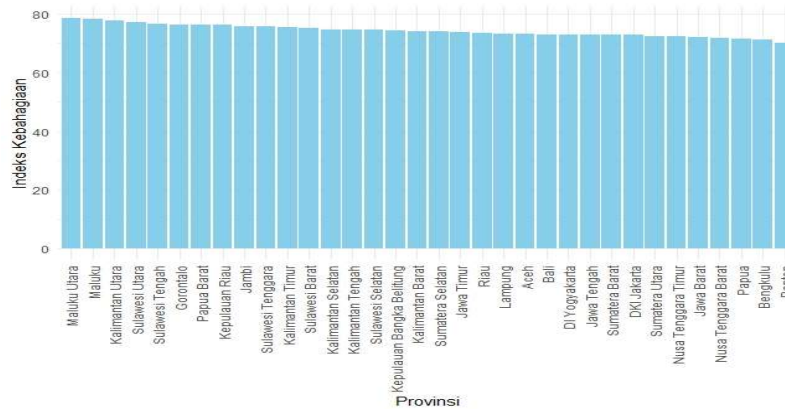
$$S_B = \sqrt{\frac{1}{g-1} \sum_{k=1}^g (\bar{x}_k - \bar{x})^2} \tag{9}$$

Keterangan:

- S_w : Simpangan Baku dalam kelompok
- S_B : Simpangan Baku antar kelompok
- \bar{x}_k : Nilai rata-rata pengamatan di dalam kelompok ke- k
- \bar{x} : Nilai rata-rata seluruh pengamatan
- n_k : Banyaknya pengamatan di dalam kelompok ke- k
- g : Banyaknya kelompok

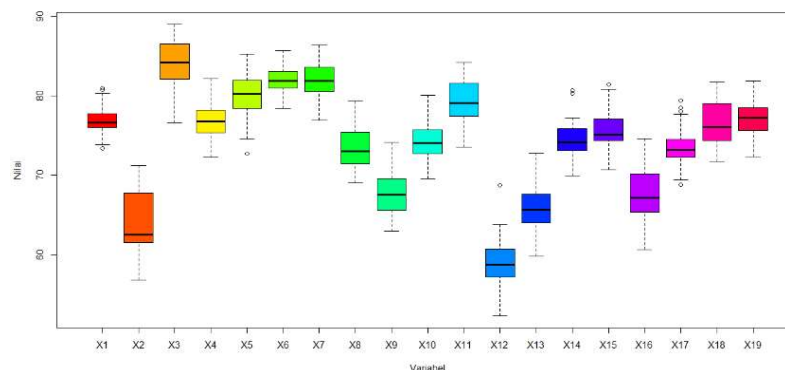
III. HASIL DAN PEMBAHASAN

Statistik deskriptif bertujuan untuk menggambarkan, menyajikan, dan merangkum karakteristik dasar dari variabel yang sedang diuji. Perbedaan indikator indeks penyusun kebahagiaan tahun 2021 dapat ditemukan dalam ilustrasi pada Gambar 1.



Gambar 1. Skor indikator indeks penyusun kebahagiaan menurut provinsi di Indonesia tahun 2021

Gambar 1 menunjukkan Provinsi Maluku Utara sebagai provinsi dengan indeks kebahagiaan tertinggi di Indonesia. Menurut antropolog Universitas Brawijaya, Hatib Abdul Kadir, ada enam faktor utama yang membuat penduduk Maluku Utara begitu bahagia, termasuk kondisi kesejahteraan sosial dan emosional yang baik, budaya yang terbuka, banyaknya pulau, hubungan keluarga erat, serta faktor sosial lainnya seperti tingkat perceraian yang rendah (Yudhistira, 2022). Sebaliknya, Provinsi Banten memiliki indeks kebahagiaan terendah di Indonesia. Gubernur Banten, Wahidin Halim, menyebutkan bahwa perbedaan regional antara perkotaan dan pedesaan menjadi faktor utama ketidakbahagiaan penduduk Banten (Ridho, 2022). Indeks kebahagiaan ini terdiri dari 3 dimensi dengan total 19 indikator.



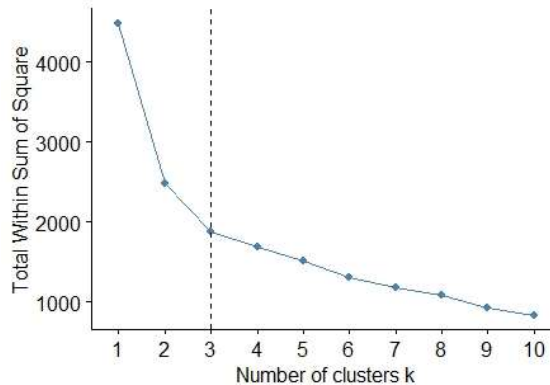
Gambar 2. Boxplot Indikator Indeks Penyusun Kebahagiaan Menurut Provinsi di Indonesia Tahun 2021

Nilai yang sangat bervariasi ditunjukkan oleh indikator indeks penyusun kebahagiaan dari Gambar 2. Hal ini ditunjukkan pada setiap boxplot di tiap variabel ada variabel cenderung lebih tinggi dan ada variabel yang cenderung lebih rendah, serta pada beberapa variabel ditemukannya nilai outlier yang mengindikasikan adanya nilai yang terlalu jauh dari nilai-nilai lainnya. Sehingga perlu dilakukan standarisasi data terlebih dahulu. Data distandarisasi dengan software *RStudio* menggunakan fungsi *scale()*. Output standarisasi dapat dilihat pada Tabel 1.

Tabel 1. Nilai Standarisasi Data

Amatan	X_1	X_2	X_3	...	X_{19}
1	-0,869	-0,289	0,047	...	0,005
2	-0,577	-0,681	-0,574	...	-0,725
3	-0,672	-0,408	-0,567	...	-0,543
⋮	⋮	⋮	⋮	⋮	-0,118
32	2,050	1,401	1,805	...	2,118
33	1,694	1,103	1,125	...	1,269
34	-0,486	-0,593	-2,790	...	-1,414

Sebelum analisis, diperlukan penentuan jumlah *cluster* dengan metode *elbow*. Pada Gambar 3, terlihat penurunan nilai pada $g = 3$ dengan SSE 251.1008, sehingga penelitian ini akan melanjutkan pengelompokan dengan nilai $g = 3$ menggunakan metode K-Means.



Gambar 3. Grafik Metode *Elbow* Metode K-Means

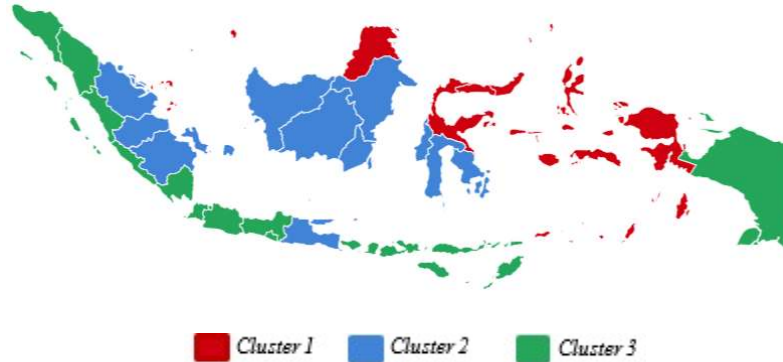
Setelah menentukan jumlah cluster, dilanjutkan dengan melakukan analisis Fuzzy C-Means didapatkan anggota untuk tiap *cluster* masing-masing adalah pada *cluster* 1 terdiri dari 8 provinsi, *cluster* 2 terdiri dari 12 provinsi, serta *cluster* 3 terdiri dari 14 provinsi. Dari hasil tersebut menunjukkan bahwa sebagian besar provinsi masuk ke dalam *cluster* 2. Dengan karakteristik *cluster* terlihat pada Tabel 2.

Tabel 2. Karakteristik *Cluster* Berdasarkan Nilai Rata-Rata *Cluster* Data yang Distandarisasi

Cluster	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	Kategori
1	1,392	1,375	1,202	1,256	0,999	1,033	0,819	1,450	1,122	1,165	1,335	0,245	0,575	1,351	1,422	0,789	1,349	1,321	1,258	Tinggi
2	0,008	-0,471	0,179	0,138	0,153	-0,055	0,223	0,014	0,283	0,198	0,205	0,544	0,506	-0,090	-0,001	0,092	-0,035	0,133	0,206	Sedang
3	-0,803	-0,382	-0,840	-0,836	-0,702	-0,544	-0,659	-0,841	-0,884	-0,836	-0,938	-0,606	-0,762	-0,695	-0,812	-0,530	-0,741	-0,869	-0,895	Rendah

Hasil Tabel 2 menunjukkan karakteristik *cluster* berdasarkan rata-rata secara berurutan untuk cluster 1, 2, dan 3 yang terkategori sebagai tinggi, sedang, dan rendah dalam indeks penyusun kebahagiaan. Nilai negatif menunjukkan posisi di bawah rata-rata sebelum *clustering*, sementara yang nilai positif di atas rata-rata. *Cluster* kategori rendah memiliki rata-rata terendah pada setiap indikator indeks penyusun kebahagiaan, sementara *cluster* kategori sedang memiliki nilai

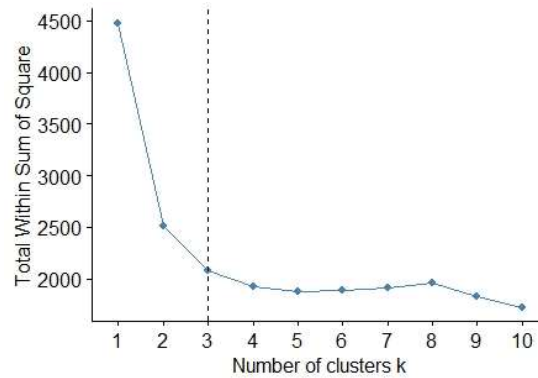
rata-rata lebih tinggi dan kategori rendah. *Cluster* kategori tinggi memiliki nilai rata-rata tertinggi untuk setiap indikator indeks. Gambar 4 menunjukkan sebaran hasil *clustering* dengan metode K-Means.



Gambar 4. Sebaran *Cluster* untuk Metode K-Means

Gambar 4 menunjukkan hasil pengelompokkan menggunakan metode K-Means berdasarkan indikator indeks kebahagiaan tahun 2021. Sebagian besar provinsi di Indonesia memiliki tingkat keberhasilan indikator indeks kebahagiaan rendah. Mayoritas berada di provinsi dekat ibu kota provinsi, seperti wilayah Jawa, dan Bali-Nusa. Sementara itu, indikator indeks kebahagiaan berada pada kategori kebahagiaan tinggi dan sedang, dengan mayoritas terdapat di provinsi yang jauh dari ibu kota, didominasi oleh wilayah Maluku dan sebagian Sumatera, Kalimantan, dan Sulawesi.

Pada analisis sebelumnya telah dilakukan pada metode K-Means. Sebelum analisis, diperlukan penentuan jumlah *cluster* dengan metode *elbow*. Pada Gambar 5, terlihat penurunan nilai pada $g = 3$ dengan SSE 272.0599, sehingga penelitian ini akan melanjutkan pengelompokkan dengan nilai $g = 3$ menggunakan metode Fuzzy C-Means.



Gambar 5. Grafik Metode *Elbow* Metode Fuzzy C-Means

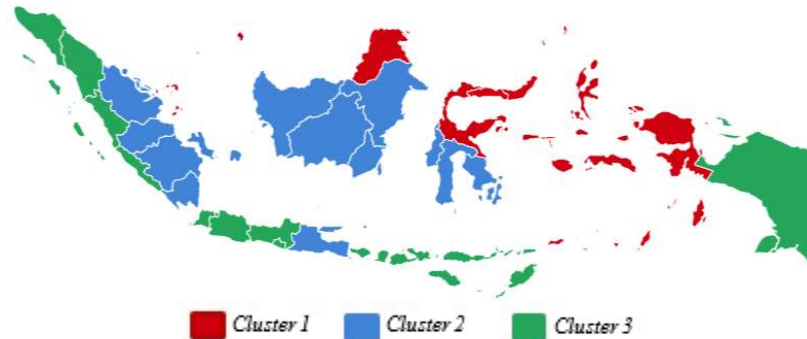
Setelah menentukan jumlah cluster, dilanjutkan dengan melakukan analisis Fuzzy C-Means didapatkan anggota untuk tiap *cluster* masing-masing adalah pada *cluster* 1 terdiri dari 8 provinsi, *cluster* 2 terdiri dari 13 provinsi, serta *cluster* 3 terdiri dari 13 provinsi. Dari hasil tersebut menunjukkan bahwa sebagian besar provinsi masuk ke dalam *cluster* 2 dan 3. Dengan karakteristik *cluster* dapat dilihat pada Tabel 3.

Tabel 3. Karakteristik *Cluster* Berdasarkan Nilai Rata-Rata *Cluster* Data yang Distandardisasi

<i>Cluster</i>	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}	X_{17}	X_{18}	X_{19}	Kategori
1	1,324	1,210	1,086	1,232	0,925	1,013	0,825	1,352	1,168	1,065	1,255	0,177	0,392	1,266	1,342	0,864	1,366	1,242	1,258	Tinggi
2	-0,120	-0,349	-0,028	-0,089	-0,024	-0,134	0,038	-0,150	0,039	0,025	-0,017	0,354	0,314	-0,197	-0,180	-0,103	-0,188	-0,082	0,206	Sedang
3	-0,780	-0,464	-0,794	-0,805	-0,639	-0,517	-0,649	-0,823	-0,830	-0,779	-0,907	-0,611	-0,682	-0,644	-0,766	-0,537	-0,696	-0,850	-0,895	Rendah

Hasil Tabel 3 menunjukkan karakteristik *cluster* berdasarkan rata-rata secara berurutan untuk cluster 1, 2, dan 3 yang terkategori sebagai tinggi, sedang, dan rendah dalam indeks penyusun kebahagiaan. Nilai negatif menunjukkan posisi

di bawah rata-rata sebelum *clustering*, sementara yang nilai positif di atas rata-rata. *Cluster* kategori rendah memiliki rata-rata terendah pada setiap indikator indeks penyusun kebahagiaan, sementara *cluster* kategori sedang memiliki nilai rata-rata lebih tinggi dan kategori rendah. *Cluster* kategori tinggi memiliki nilai rata-rata tertinggi untuk setiap indikator indeks. Gambar 6 menunjukkan sebaran hasil *clustering* dengan metode Fuzzy C-Means.



Gambar 6. Sebaran *Cluster* untuk Metode Fuzzy C-Means

Gambar 6 menampilkan hasil pengelompokan menggunakan metode Fuzzy C-means berdasarkan indikator indeks kebahagiaan tahun 2021. Sebagian besar provinsi di Indonesia memiliki tingkat keberhasilan indikator indeks kebahagiaan sedang dan rendah. Mayoritas berada di provinsi dekat ibu kota provinsi, seperti wilayah Sumatera, Jawa, dan Bali-Nusa. Sementara itu, indikator indeks kebahagiaan berada pada kategori kebahagiaan tinggi, dengan mayoritas terdapat di provinsi yang jauh dari ibu kota, didominasi oleh wilayah Maluku dan sebagian Kalimantan dan Sulawesi.

Untuk melihat metode mana yang optimal antara K-Means dan Fuzzy C-Means, maka perlu dilakukan uji validasi *cluster* dengan menggunakan rasio simpangan baku dapat dilihat pada Tabel 4.

Tabel 4. Validasi *Cluster* K-Means dan Fuzzy C-Means

Nilai	K-Means	Fuzzy C-Means
Sw	251.1008	252.2168
Sb	375.8992	325.3501
Sw/Sb	0.6680004	0.7752166

Pada Tabel 4, validasi dapat dilihat dengan menggunakan simpangan baku antar kelompok dan rasio simpangan baku. Pemilihan nilai optimal untuk pengelompokan provinsi-provinsi Indonesia dilakukan berdasarkan nilai simpangan baku antar kelompok yang besar dan rasio simpangan baku yang kecil. Dengan demikian, metode terpilih adalah K-Means.

IV. KESIMPULAN

Metode K-Means lebih optimal dengan nilai rasio kecil dan simpangan baku besar. Pengelompokan provinsi Indonesia berdasarkan indeks kebahagiaan menggunakan K-Means dan Fuzzy C-Means menunjukkan tiga *cluster*, diukur dengan metode *elbow*. Provinsi dengan tingkat kebahagiaan rendah pada metode K-Means terdiri dari 14 provinsi dan metode Fuzzy C-Means terdiri dari 13 provinsi, untuk tingkat kebahagiaan sedang metode K-Means terdiri dari 12 provinsi dan metode Fuzzy C-Means terdiri dari 13 provinsi. Sementara, pada tingkat kebahagiaan tinggi pada metode K-Means dan Fuzzy C-Means masing-masing terdiri dari 8 provinsi. Hasil pengelompokan menunjukkan provinsi di kawasan dekat ibukota memiliki kebahagiaan rendah, sementara tingkat kebahagiaan tinggi terdistribusi di kawasan jauh seperti Maluku dan Sulawesi pada metode K-Means dan Fuzzy C-Means.

DAFTAR PUSTAKA

Alwi, Wahidah, dan Muh Hasrul. 2018. "Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Sulawesi Selatan Berdasarkan Indikator Kesejahteraan Rakyat". *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)* 6(1):35. doi: 10.24252/msa.v6i1.4782.

Badan Pusat Statistik Indonesia. 2021. *Indeks Kebahagiaan 2021*. Badan Pusat Statistik Indonesia.

- Bunkerrs, Matthew J., dan James R. Miller JR. 1996. "Definition of Climate Regions in the Northern Plains Using an Objective Cluster Modification Technique". *Journal of Climate* 9(1):130–46.
- Cebeci, Zeynel, dan Figen Yildiz. 2015. "Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures". *Journal of Agricultural Informatics* 6(3):13–23. doi: 10.17700/jai.2015.6.3.196.
- Everitt, B.S. Landau, S., Leese, dan M., Stahl, D. 2011. *Cluster Analysis*. 5th ed. edited by Isberg. London, UK: John Wiley & Sons.
- Ghosh, Soumi, dan Sanjay Kumar Dubey. 2013. "A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms". *International Journal of Computational Engineering Research* 4(4).
- Hair, J., Black, W., Babin, B., dan Anderson, R. 2013. *Preparation and Characterization of Nonwoven Fibrous Biocomposites for Footwear Components*. 7th ed.
- Helliwell, J. F., Huang, H., Wang, S., dan Norton, M. 2022. *World Happiness Report 2022*. John Helli. New York: Sustainable Development Solutions Network.
- Johnson, Richard A., dan Dean W. Wichern. 2002. *Applied Multivariate Statistical Analysis (5th Ed)*. 5th ed. New Jersey: Prentice Hall.
- Luthifaturrahmah. 2014. "Perbandingan Hasil Penggerombolan K-Means, Fuzzy K- Means, Dan Two Step Clustering". *Jurnal Pendidikan Matematika* 02(1):39–62.
- Mattjik, Ahmad Ansori, dan I. Made Sumertajaya. 2011. *Sidik Peubah Ganda Menggunakan SAS*. 1st ed. Bogor,: IPB PRESS.
- Ridho, Rasyid. 2022. Provinsi Banten Jadi Daerah Paling Tidak Bahagia, Ini Respons Gubernur Wahidin Halim. *Regional.Kompas*. diakses Januari 20, 2024 (<https://regional.kompas.com/read/2022/01/08/120349878/provinsi-banten-jadi-daerah-paling-tidak-bahagia-ini-respons-gubernur>).
- Sivarathri, Srinivas, dan Govardhan A. 2014. "Experiments on Hypothesis 'Fuzzy K-Means Is Better than K-Means for Clustering'". *International Journal of Data Mining & Knowledge Management Process* 4(5):21–34. doi: 10.5121/ijdkp.2014.4502.
- Timothy J. Ross. 2010. *Fuzzy Logic With Engineering Application*. Third Edit. NewYork(US).
- Wu, Junjie. 2012. *Advances in K-Means Cluster Ing: A Data Mining Thinking*. New York: Springer Science & Business Media.
- Yudhistira, Aria W. 2022. Mengapa Orang Maluku Utara Paling Bahagia Di Indonesia? *Katadata*. Retrieved January 20, 2024 (<https://katadata.co.id/ariayudhistira/analisisdata/61ee1b64b167e/mengapa-orang-maluku-utara-paling-bahagia-di-indonesia?page=2>).