

# Comparison K-Means and Fuzzy C-Means Methods to Grouping Human Development Index Indicators in Indonesia

Belia Mailien, Admi Salma\*, Syafriandi, Dina Fitria

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

\*Corresponding author: [admisalma@fmipa.unp.ac.id](mailto:admisalma@fmipa.unp.ac.id)

Submitted : 27 Juli 2022

Revised : 22 Desember 2022

Accepted : 27 Desember 2022

## ABSTRACT

*K-Means is a non-hierarchical cluster analysis that classifies items depending on how far they are from the nearest cluster center. Data that can be a member of a cluster is produced based on various memberships between 0 and 1 using the fuzzy grouping model used by Fuzzy C-Means, a non-hierarchical clustering method. The Human Development Index (HDI) is gathered to determine which technique is the best between the two. A key indicator of how well community initiatives to improve quality of life are doing is the HDI. There has been a growth in HDI in Indonesia, but it has not been distributed equally throughout the regencies and cities. Grouping is done to identify the features of each regency or city to ensure that HDI is distributed equally. The classification of the 2021 HDI indicators into 4 clusters using the K-means and Fuzzy C-means algorithms is covered in this work. The presence of the low, medium, and highest HDI indicators is the outcome of the grouping achieved. The K-Means method's C\_index and S\_Dbw index values are 0.108 and 2.444, respectively, while the Fuzzy C-Means method's C index and S\_Dbw index values are 0.105 and 2.312, respectively. Due to its lower C\_index and S\_Dbw index values, the Fuzzy C-Means approach produces the best grouping results.*

**Keywords:** Cluster Validity, Fuzzy C-Mean, HDI indicator, K-Means



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

Indeks Pembangunan Manusia (IPM) merupakan salah satu tolak ukur keberhasilan upaya peningkatan kualitas hidup masyarakat. Daerah dengan IPM tinggi dianggap berhasil dalam program pembangunannya. Untuk membangun IPM yang berkualitas, pemerintah harus memiliki program dan fasilitas yang memadai untuk meningkatkan IPM, serta dukungan masyarakat untuk memajukan program tersebut. Sejak 2014, angka IPM Indonesia telah dipublikasikan setiap tahun di tingkat nasional, negara bagian dan kabupaten/kota (BPS, 2021). Jika diperhatikan sebaran nilai IPM di setiap kabupaten/kota di Indonesia, kita dapat melihat nilai IPM yang beragam. Hal ini menunjukkan bahwa tingkat kemajuannya masih belum merata di seluruh Indonesia. IPM rendah atau tinggi mempengaruhi tingkat produktivitas penduduk. Semakin rendah IPM, semakin rendah tingkat produktivitas penduduk, dan semakin rendah produktivitas, semakin rendah pendapatan. Sebaliknya, semakin tinggi IPM, semakin tinggi produktivitas penduduk dan semakin tinggi tingkat pendapatan (Zusanti dkk, 2018). Nilai IPM yang tinggi, jika tidak diikuti dengan pemerataan pendapatan, akan menurunkan kesejahteraan masyarakat secara keseluruhan dan menyebabkan ketimpangan antar wilayah.

Untuk memudahkan pemerintah merumuskan strategi dan rencana mengatasi ketimpangan IPM di Indonesia, kabupaten/kota di Indonesia harus dikelompokkan berdasarkan tingkat IPM. Berdasarkan hasil pengelompokan yang diperoleh, ditentukan karakteristik kabupaten/kota dan mendeskripsikan karakteristik masing-masing kelompok berdasarkan indikator penentu nilai IPM. Analisis statistik yang dapat digunakan adalah analisis klaster. Analisis klaster merupakan mengelompokkan sekumpulan objek data ke dalam kelompok-kelompok sedemikian rupa sehingga objek dalam satu kelompok sangat mirip tetapi berbeda secara signifikan dari objek dalam kelompok lain (Han dkk, 2012).

*K-Means* adalah analisis klaster non hirarki yang mengklasifikasikan objek ke dalam kelompok klaster berdasarkan jaraknya ke pusat klaster terdekat (Jonhson dan Winchern, 2002). *K-Means* sangat sederhana, cepat, dan dapat memproses kumpulan data besar dengan sangat efisien (Wu, 2012). *Fuzzy C-Means* merupakan analisis klaster

yang menggunakan model pengelompokan fuzzy dimana data merupakan anggota dari kluster yang dibentuk berdasarkan derajat keanggotaan yang berbeda dari 0 sampai 1. *Fuzzy C-Means* memberikan hasil yang baik ketika kluster tumpang tindih (Cebeci dan Yildiz, 2015).

Perbandingan antara metode *K-Means* dan metode *Fuzzy C-Means* telah dilakukan oleh peneliti sebelumnya. Ramadhan dkk (2017) pada pengelompokan data *user knowledge modeling* menggunakan *Silhouette Index* dan *Partition Coefficient Index* diperoleh bahwa metode *Fuzzy C-Means* lebih baik. Wulandari dan Yogantara (2022) pada data ekonomi dan kesehatan menggunakan *Davies Bouldin Index* dan *Silhouette Index* diperoleh metode terbaik adalah metode *Fuzzy C-Means* berdasarkan *Davies Bouldin Index* dan metode *K-Means* adalah metode terbaik menggunakan *Silhouette Index*.

Penelitian ini berbeda dari penelitian sebelumnya karena dalam menentukan metode terbaik digunakan *S\_Dbw index* dan *C\_index* yang cocok digunakan jika kluster yang dihasilkan tumpang tindih sementara *index* yang lain tidak dapat mengidentifikasi hal tersebut. *S\_Dbw index* mengukur varians dalam kluster dan kepadatan antar kluster (Legany dkk, 2006). *C\_index* mengukur perbedaan antara jarak maksimum dan minimum di dalam dan antar kluster (Charrad dkk, 2014). *S\_Dbw index* dan *C\_index* digunakan agar kriteria pengelompokan terdefinisi dengan baik sehingga menghasilkan partisi terbaik dari kumpulan data.

Pada penelitian ini akan dibahas mengenai perbandingan metode *K-Means* dan *Fuzzy C-Means* dengan permasalahan IPM di Indonesia Tahun 2021. Secara umum, output penelitian berupa hasil kluster terbaik dari metode *K-Means* dan *Fuzzy C-Means* yang menggambarkan karakteristik dari masing-masing kabupaten/kota di Indonesia berdasarkan indikator IPM Tahun 2021.

## II. METODE PENELITIAN

### A. *K-Means*

Jumlah kluster ( $g$ ) dan pusat kluster yang dipilih secara acak harus ditentukan terlebih dahulu untuk algoritma *K-Means* (Singh dkk, 2013). Langkah selanjutnya adalah menempatkan objek ke dalam kluster berdasarkan jarak terdekat objek ke titik pusat kluster. Pusat kluster pada iterasi berikutnya dapat dihitung menggunakan persamaan (1),

$$v_{lj} = \left(\frac{1}{n_l}\right) \sum_{k=1}^{n_l} x_{kj} \quad (1)$$

dengan  $v_{lj}$  merupakan pusat kluster ke- $l$  pada variabel ke- $j$ ,  $x_{kj}$  merupakan objek data ke- $k$  pada variabel ke- $j$ , dan  $n_l$  adalah banyak objek pada kluster ke- $l$ . Proses iterasi dihentikan ketika sudah tidak terjadi lagi perpindahan objek antar kluster.

### B. *Fuzzy C-Means*

Setiap objek memiliki derajat keanggotaan yang menunjukkan kekuatan keanggotaan di semua atau beberapa kluster pada pengelompokan fuzzy (Everitt dkk, 2011). Derajat keanggotaan diskalakan antara 0 sampai 1 yang diartikan sebagai probabilitas. Selain diinterpretasikan sebagai probabilitas, derajat keanggotaan dalam *Fuzzy C-Means* merepresentasikan seberapa dekat objek dengan titik pusat. Derajat keanggotaan suatu objek ke- $k$  pada kluster ke- $l$  adalah  $\mu_{lk} = \mu_{A_l}(x_k) \in [0,1]$ , dengan  $\mu_{lk}$  merupakan derajat keanggotaan objek ke- $k$  pada kluster ke- $l$  dan  $A_l$  adalah himpunan anggota kluster ke- $l$ . Dalam analisis *Fuzzy C-Means* kita perlu menentukan banyaknya kluster ( $g$ ), parameter pembobot keanggotaan  $m'$ , matriks partisi awal  $\mathbf{U}^{(0)}$  dan tingkat akurasi  $\epsilon_L$ . Nilai  $m'$  memiliki rentang  $m' \in [1, \infty]$ . Jika nilai  $m' = 1$  maka algoritma *Fuzzy C-Means* mendekati algoritma *K-Means*.  $m'$  mengontrol tingkat pembagian keanggotaan kluster. Nilai  $m'$  disarankan bernilai 1,25-2 (Ross, 2010). Menurut Rao dkk (2010) langkah-langkah metode *Fuzzy C-Means* setelah menentukan banyaknya kluster ( $g$ ), parameter pembobot keanggotaan  $m'$ , matriks partisi awal  $\mathbf{U}^{(0)}$  dan tingkat akurasi  $\epsilon_L$  adalah menghitung pusat kluster ke- $l$  pada variabel ke- $j$  pada iterasi ke- $r$  dapat dihitung menggunakan persamaan (2),

$$v_{lj}^{(r)} = \frac{\sum_{k=1}^n (\mu_{lk}^{(r)})^{m'} \cdot x_{kj}}{\sum_{k=1}^n (\mu_{lk}^{(r)})^{m'}} \quad (2)$$

Langkah selanjutnya adalah menghitung jarak objek ke pusat kluster. Setelah itu melakukan pembaruan elemen matriks partisi  $\mathbf{U}^{(r)}$  dengan persamaan (3),

$$\mu_{lk}^{(r+1)} = \left[ \sum_{i=1}^g \left( \frac{a_{lk}^{(r)}}{a_{ik}^{(r)}} \right)^{\frac{2}{(m^l-1)}} \right]^{-1} \quad (3)$$

$d_{lk}^{(r)}$  merupakan jarak objek ke- $k$  pada pusat kluster ke- $l$ . Pada metode *Fuzzy C-Means* konvergensi hasil pengelompokan ditentukan menggunakan persamaan (4),

$$\max_{l,k} |\mu_{lk}^{(r+1)} - \mu_{lk}^{(r)}| \leq \varepsilon_L \quad (4)$$

dimana  $\varepsilon_L$  merupakan tingkat akurasi yang ditentukan. Iterasi dihentikan ketika kriteria konvergensi terpenuhi. Namun jika belum terpenuhi maka hitung kembali pusat kluster sampai kriteria konvergensi terpenuhi.

### C. Validasi Analisis Kluster

Untuk mengetahui seberapa baik hasil pengelompokan, diperlukan beberapa ukuran validasi (Wu, 2012). *S\_Dbw index* memberikan keputusan yang tepat dalam mengidentifikasi hasil kluster (Legany dkk, 2006). *S\_Dbw index* mengukur varians dalam kluster dan kepadatan antar kluster. *S\_Dbw* dirumuskan pada persamaan (5),

$$S\_Dbw = Scatt + Dens\_bw \quad (5)$$

dengan

$$Scatt = \frac{1}{g} \sum_{i=1}^g \frac{|\sigma(v_l)|}{|\sigma(x_j)|} \quad Dens_{bw} = \frac{1}{g(g-1)} \sum_{i=1}^g \frac{density(u_{lm})}{\max\{density(v_l), density(v_m)\}}$$

*Scatt* merupakan varians intra kluster, *Dens\_bw* merupakan kepadatan antar kluster,  $g$  merupakan ukuran kluster yang digunakan,  $\sigma(v_l)$  merupakan varians pusat kluster ke- $l$ ,  $\sigma(x_j)$  merupakan varians variabel ke- $j$ ,  $u_{lm}$  adalah titik tengah *cluster* ke- $l$  dan kluster ke- $m$ .

Selain *S\_Dbw index*, *index* validasi yang biasa digunakan adalah *C\_index*. *C\_index* mengukur perbedaan jarak maksimum dan minimum dalam kluster dan antar kluster (Charrad dkk, 2014). *C\_index* dirumuskan pada persamaan (6),

$$C\_index = \frac{S_w - S_{min}}{S_{max} - S_{min}} \quad (6)$$

$S_w$  merupakan jumlah total jarak dalam kluster,  $S_{min}$  merupakan jumlah jarak minimum dalam kluster dan antar kluster,  $S_{max}$  jumlah jarak maksimum dalam kluster dan antar kluster.

Kriteria pengelompokan baik menggunakan *C\_index* adalah nilai yang lebih rendah karena menunjukkan jarak dalam kluster yang rendah dan jarak antar kluster yang berjauhan (Charrad dkk, 2014). Kriteria pengelompokan baik menggunakan *S\_Dbw index* adalah nilai yang lebih rendah karena menunjukkan keragaman dalam suatu kluster yang rendah dan kepadatan antar kluster yang rendah (Legany dkk, 2006).

### D. Sumber Data dan Teknik Analisis Data

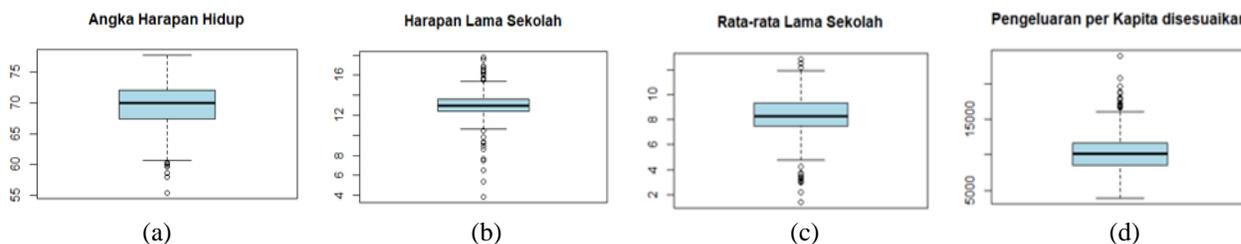
Jenis penelitian berdasarkan permasalahan dan tujuan penelitian ini adalah penelitian terapan. Pada penelitian ini data yang digunakan adalah data sekunder yang diperoleh dari *website* Badan Pusat Statistik (BPS) yaitu data indikator Indeks Pembangunan Manusia di Indonesia Tahun 2021 menurut kab/kota. Dari data tersebut akan dilakukan pengelompokan kabupaten/kota di Indonesia menurut indikator IPM. Variabel penelitian terdiri dari 4 variabel yaitu Angka Harapan Hidup (AHH), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), dan Pengeluaran per Kapita (PPP) yang disesuaikan.

Adapun langkah-langkah analisis dalam penelitian ini adalah:

1. Melakukan statistika deskriptif untuk mengetahui karakteristik dari masing-masing variabel yang digunakan.
2. Melakukan standardisasi data.
3. Melakukan analisis kluster *K-Means* dan *Fuzzy C-Means*.
4. Melakukan interpretasi hasil kluster yang diperoleh.
5. Menguji validitas hasil kluster untuk memperoleh metode terbaik antara *K-Means* dan *Fuzzy C-Means*.
6. Menarik kesimpulan dan saran.

### III. HASIL DAN PEMBAHASAN

Statistika deskriptif dilakukan untuk melihat ringkasan data secara umum mengenai variabel yang digunakan. Gambaran mengenai perbedaan capaian pada indikator penentu nilai IPM Tahun 2021 dapat dilihat seperti pada Gambar 1.



**Gambar 1.** (a) Boxplot Angka Harapan Hidup Menurut Kabupaten/Kota di Indonesia Tahun 2021, (b) Boxplot Harapan Lama Sekolah Menurut Kabupaten/Kota di Indonesia Tahun 2021, (c) Boxplot Rata-rata Lama Sekolah Menurut Kabupaten/Kota di Indonesia Tahun 2021, (d) Boxplot Pengeluaran per Kapita disesuaikan Menurut Kabupaten/Kota di Indonesia Tahun 2021

Berdasarkan Gambar 1, indikator AHH menunjukkan adanya nilai AHH yang sangat rendah. Hal ini menunjukkan masih belum tercapainya dimensi umur panjang dan hidup sehat pada kabupaten/kota tersebut. Kabupaten/kota dengan nilai AHH yang sangat rendah dibandingkan kabupaten/kota lain adalah Kabupaten Nduga yang terdapat di Provinsi Papua. Pada indikator HLS dan RLS terdapat nilai yang sangat rendah dan sangat tinggi. Hal ini berarti beberapa kabupaten/kota memiliki kondisi pendidikan yang baik namun masih ada beberapa kabupaten/kota yang perlu perhatian lebih agar tingkat kemajuan di bidang pendidikan tidak tertinggal dari kabupaten/kota lainnya. Kabupaten/kota yang memiliki nilai HLS dan RLS yang sangat rendah dibanding kabupaten/kota lain adalah Kabupaten Nduga yang terdapat di Provinsi Papua. Kabupaten/kota yang memiliki nilai HLS dan RLS tertinggi dibanding kabupaten/kota lainnya adalah Kota Banda Aceh. Selanjutnya indikator PPP disesuaikan yang memiliki nilai tertinggi dibanding kabupaten/kota lainnya. Hal ini menunjukkan tingginya kemampuan daya beli pada beberapa kabupaten/kota tersebut yang berarti bahwa masyarakat di kabupaten/kota tersebut telah mencapai dimensi IPM yang ketiga yaitu memiliki standar hidup yang layak. Kabupaten/kota dengan nilai PPP tertinggi dibandingkan kabupaten/kota lainnya adalah Kota Jakarta Selatan. Berdasarkan hasil eksplorasi data yang telah dilakukan, kabupaten Nduga memiliki nilai yang sangat rendah pada 3 indikator penentu nilai IPM.

Standardisasi data dilakukan menggunakan *software* RStudio dengan fungsi *scale()*. Hasil standardisasi ditampilkan seperti pada Tabel 1.

**Tabel 1.** Hasil Standardisasi Data

Amatan	AHH	HLS	PPP	RLS
1	-1,270	0,666	-1,169	0,640
2	-0,646	0,984	-0,570	0,149
3	-1,525	1,196	-0,789	0,272
⋮	⋮	⋮	⋮	⋮
512	-1,104	-1,842	-1,807	-2,144
513	-1,151	-4,061	-1,839	-3,278
514	-1,246	-2,425	-2,080	-3,180

Sebelum masuk ke algoritma *K-Means* dan *Fuzzy C-Means* perlu ditentukan terlebih dahulu jumlah kluster yang akan dibentuk. Penetapan jumlah kluster disesuaikan dengan kategori IPM oleh BPS yaitu rendah, sedang, tinggi dan sangat tinggi. Sehingga jumlah kluster yang akan dibentuk adalah 4 kluster.

Setelah menentukan jumlah kluster, selanjutnya digunakan analisis *K-Means*. Ringkasan hasil analisis kluster dengan menggunakan metode *K-Means* dapat dilihat seperti pada Tabel 2.

**Tabel 2.** Hasil kluster metode *K-Means*

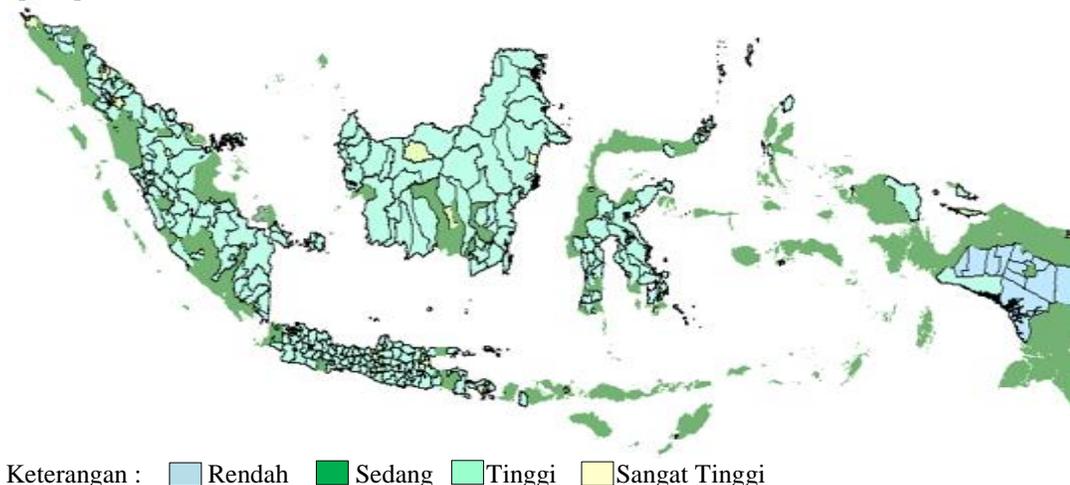
Kluster	1	2	3	4
Jumlah kab/kota	14	175	237	88

Berdasarkan Tabel 2, pengelompokan kabupaten/kota berdasarkan capaian IPM menunjukkan bahwa sebagian besar kabupaten/kota di Indonesia berada pada kluster 2 dan 3. Karakteristik dari masing-masing kluster berdasarkan nilai rata-rata kluster dapat dilihat seperti pada Tabel 3.

**Tabel 3.** Karakteristik kluster berdasarkan nilai rata-rata kluster data yang standardisasi

Kluster	AHH	HLS	PPP	RLS	Kategori
1	-1,548	-3,654	-1,936	-3,019	Rendah
2	-0,963	-0,099	-0,643	-0,323	Sedang
3	0,457	-0,158	0,054	-0,133	Tinggi
4	0,931	1,202	1,442	1,479	Sangat Tinggi

Berdasarkan Tabel 3, karakteristik masing-masing kluster dari nilai rata-rata kluster secara berurutan adalah kluster 1,2,3 dan 4 diklasifikasikan berdasarkan indikator rendah, sedang, tinggi dan sangat tinggi. Nilai rata-rata kluster yang negatif menunjukkan bahwa nilai tersebut berada di bawah rata-rata data sebelum dibagi menjadi beberapa kluster. Nilai rata-rata kluster yang positif pada rata-rata kluster menandakan bahwa nilai tersebut berada di bawah rata-rata data sebelum dibagi menjadi beberapa kluster. Kluster dengan kategori rendah merupakan kluster dengan nilai rata-rata yang terendah di setiap indikator IPM dibandingkan kluster lain yaitu nilai yang berada jauh di bawah rata-rata, kluster dengan kategori sangat tinggi merupakan nilai rata-rata kluster yang tertinggi di setiap indikator IPM, kluster dengan kategori sedang memiliki nilai rata-rata kluster yang juga di bawah rata-rata tetapi lebih tinggi daripada kluster dengan kategori rendah, dan kluster dengan kategori tinggi memiliki nilai rata-rata kluster untuk indikator AHH dan PPP di atas nilai rata-rata. Sebaran hasil kluster yang diperoleh pada metode *K-Means* dapat dilihat seperti pada Gambar 2.



**Gambar 2.** Sebaran Hasil Kluster Metode *K-Means*

Berdasarkan Gambar 2, hasil pengelompokan kabupaten/kota berdasarkan capaian IPM menunjukkan bahwa banyak kabupaten/kota di Indonesia berada pada capaian indikator IPM dengan kategori sedang dan tinggi. Kabupaten/kota dengan kategori tersebut rata-rata tersebar di Pulau Sumatera, Kalimantan, Sulawesi dan Jawa.

Capaian indikator IPM dengan kategori sangat tinggi paling banyak terdapat di kota besar di Indonesia dan ibukota provinsi di Indonesia. Capaian indikator IPM kategori rendah banyak terdapat di Pulau Papua.

Setelah dilakukan analisis dengan metode *K-Means*, selanjutnya adalah menganalisis dengan metode *Fuzzy C-Means*. Ringkasan hasil kluster yang diperoleh sehingga kriteria konvergensi tercapai dapat dilihat seperti pada Tabel 4.

**Tabel 4.** Hasil kluster metode *Fuzzy C-Means*

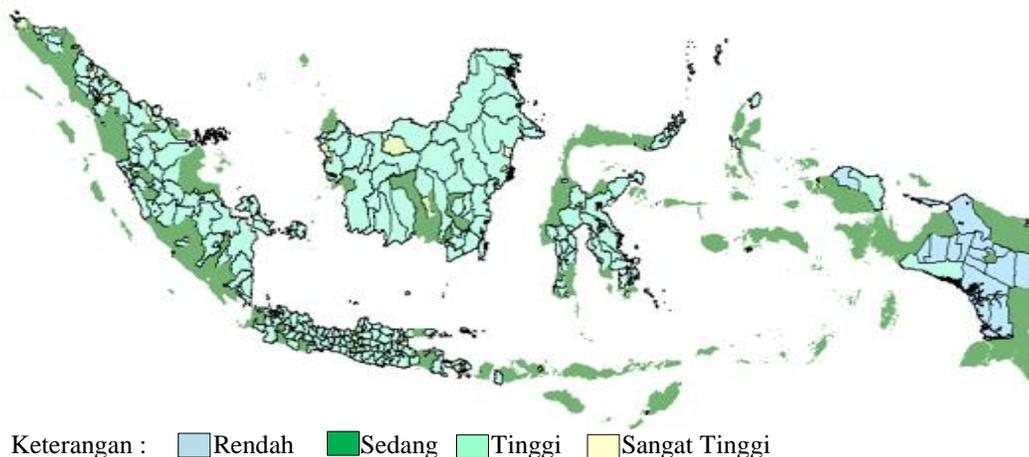
Kluster	1	2	3	4
<b>Jumlah kab/kota</b>	18	222	183	91

Berdasarkan Tabel 4, kluster pertama memiliki anggota kluster paling sedikit dibanding kluster lain. Anggota kluster terbanyak terdapat pada kluster 2. Karakteristik dari masing-masing kluster berdasarkan nilai rata-rata kluster dapat dilihat seperti pada Tabel 5.

**Tabel 5.** Karakteristik kluster berdasarkan nilai rata-rata kluster data yang di standardisasi

Kluster	AHH	HLS	PPP	RLS	Kategori
1	-1,386	-3,096	-1,823	-2,717	Rendah
2	0,573	-0,157	0,057	-0,157	Tinggi
3	-0,691	-0,115	-0,545	-0,268	Sedang
4	0,872	1,120	1,289	1,365	Sangat Tinggi

Berdasarkan Tabel 5, karakteristik masing-masing kluster dari nilai rata-rata kluster mengikuti urutan kluster 1,3,2 dan 4 diklasifikasikan berdasarkan indikator rendah, sedang, tinggi dan sangat tinggi. Nilai rata-rata kluster yang negatif menandakan bahwa nilai tersebut berada di bawah rata-rata data sebelum dipecah ke dalam kluster. Nilai rata-rata kluster yang positif menunjukkan bahwa nilai tersebut berada di atas rata-rata data sebelum dipecah ke dalam kluster. Kluster dengan kategori rendah merupakan kluster dengan nilai rata-rata terendah di setiap indikator IPM yang nilainya jauh berada di bawah rata-rata dari data sebelum dibagi menjadi beberapa kluster, kluster dengan kategori sangat tinggi menempati nilai rata-rata kluster yang paling tinggi di setiap indikator IPM, kluster dengan kategori sedang memiliki nilai rata-rata kluster yang juga di bawah rata-rata seperti pada kategori rendah tetapi nilai rata-rata kluster pada kategori sedang lebih tinggi dan kluster dengan kategori tinggi memiliki nilai rata-rata kluster dengan indikator AHH dan PPP yang berada di atas nilai rata-rata dari data sebelum dipecah ke dalam bentuk kluster. Sebaran hasil kluster yang diperoleh dapat dilihat seperti pada Gambar 3.



**Gambar 3.** Sebaran Hasil Kluster Metode *Fuzzy C-Means*

Gambar 3 menampilkan sebaran pengelompokkan pada metode *Fuzzy C-Means* dengan hasil yang tidak jauh berbeda dibanding metode *K-Means*. Pengelompokkan kabupaten/kota berdasarkan capaian IPM menunjukkan bahwa

rata-rata kabupaten/kota di Indonesia memiliki capaian indikator IPM dengan kategori sedang dan tinggi yang paling banyak tersebar di Pulau Sumatera, Kalimantan, Sulawesi dan Jawa. Capaian indikator IPM dengan kategori sangat tinggi paling banyak terdapat di kota besar di Indonesia dan ibukota provinsi di Indonesia. Capaian indikator IPM dengan kategori rendah secara garis besar terdapat di Pulau Papua.

Untuk mengetahui seberapa baik hasil klaster yang terbentuk dari metode *K-Means* dan *Fuzzy C-Means* serta metode mana yang paling baik di antara kedua metode klaster tersebut maka perlu dilakukan uji validasi klaster menggunakan nilai *C\_index* dan *S\_Dbw index*. Hasil validasi metode *K-Means* dan *Fuzzy C-Means* dapat dilihat seperti pada Tabel 6.

**Tabel 6.** Validasi Klaster *K-Means* dan *Fuzzy C-Means*

Indeks Validasi	<i>K-Means</i>	<i>Fuzzy C_Means</i>
<i>C_index</i>	0,108	0,105
<i>S_Dbw index</i>	2,444	2,312

Tabel 6 menunjukkan nilai *C\_index* dan *S\_Dbw index* dari metode *K-Means* dan *Fuzzy C-Means* yang diperoleh bahwa metode *Fuzzy C-Means* lebih baik dibanding metode *K-Means* karena nilai *C\_index* dan *S\_Dbw index* dari metode *Fuzzy C-Means* lebih kecil dibanding metode *K-Means*.

#### IV. KESIMPULAN

Metode terbaik adalah metode *Fuzzy C-Means* karena memiliki nilai *C\_index* dan *S\_Dbw index* yang lebih rendah dibandingkan metode *K-Means*. Pengelompokan kabupaten/kota di Indonesia berdasarkan indikator penentu nilai IPM menggunakan metode *K-Means* dan *Fuzzy C-Means* dengan ukuran klaster yaitu 4 dengan capaian indikator IPM rendah, sedang, tinggi dan sangat tinggi. Capaian indikator IPM rendah terdiri dari 14 kabupaten/kota pada metode *K-Means*, pada metode *Fuzzy C-Means* terdiri dari 18 kabupaten/kota. Capaian indikator IPM sedang terdiri dari 175 kabupaten/kota pada metode *K-Means*, pada metode *Fuzzy C-Means* terdiri dari 183 kabupaten/kota. Capaian indikator IPM tinggi terdiri dari 237 kabupaten/kota pada metode *K-Means*, pada metode *Fuzzy C-Means* terdiri dari 222 kabupaten/kota. Capaian indikator IPM sangat tinggi terdiri dari 88 kabupaten/kota pada metode *K-Means*, pada metode *Fuzzy C-Means* terdiri dari 91 kabupaten/kota. Hasil pengelompokan pada metode *K-Means* dan *Fuzzy C-Means* memperlihatkan bahwa banyak kabupaten/kota di Pulau Papua memiliki capaian indikator IPM yang lebih rendah dibandingkan kabupaten/kota di pulau lain di Indonesia. Hal ini menunjukkan bahwa kabupaten/kota di Pulau Papua masih belum memenuhi kondisi yang diharapkan pada dimensi IPM. Capaian indikator IPM kategori sedang pada metode *K-Means* dan *Fuzzy C-Means*, tersebar diseluruh pulau besar di Indonesia. Capaian indikator IPM dengan kategori tinggi pada metode *K-Means* dan *Fuzzy C-Means* mayoritas terdapat pada pulau Sumatera, Jawa, Kalimantan dan Sulawesi. Capaian indikator IPM dengan kategori sangat tinggi pada metode *K-Means* dan *Fuzzy C-Means* terdapat pada ibukota provinsi di beberapa provinsi di Indonesia serta kota besar di Indonesia.

#### DAFTAR PUSTAKA

- Badan Pusat Statistik (BPS). "Indikator Indeks Pembangunan Manusia Tahun 2021". Publikasi BPS diakses dari <http://www.sumbar.bps.go.id/>, tanggal 10 Januari 2022.
- Cebeci, Z., & Yildiz, F. (2015). Comparison of *K-Means* and *Fuzzy C-Means* algorithms on different cluster structures. *Journal of Agricultural Informatics*, 6(3), 13-23. <https://doi.org/10.17700/jai.2015.6.3.196>
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: an R package for determining the relevant number of clusters in a data set. *Journal of statistical software*, 61, 1-36. <https://doi.org/10.18637/jss.v061.i06>
- Everitt, B. S. Landau, S., Leese, & M., Stahl, D. (2011). *Cluster analysis*. John Wiley & Sons.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Johnson, R. A., & Wichern, D. W., (2002). *Applied Multivariate Statistical Analysis (5th Ed)*. New Jersey: Prentice Hall.

- Legany, C., Juhasz, S., & Babos, A. (2006, February). Cluster validity measurement techniques. *In Proceedings of the 5th WSEAS international conference on artificial intelligence, knowledge engineering and databases* (pp. 388-393). World Scientific and Engineering Academy and Society (WSEAS) Stevens Point, Wisconsin, USA.
- Ramadhan, A., Efendi, Z., & Mustakim, M. (2017). Perbandingan K-Means dan Fuzzy C-Means untuk Pengelompokan Data User Knowledge Modeling. In *Seminar Nasional Teknologi Informasi Komunikasi dan Industri* (pp. 219-226).
- Rao, V. S., & Vidyavathi, D. S. (2010). Comparative investigations and performance analysis of FCM and MFPCM algorithms on iris data. *Indian Journal of Computer Science and Engineering*, 1(2), 145-151.
- Ross, T. J. (2005). *Fuzzy logic with engineering applications*. John Wiley & Sons.
- Singh, A., Yadav, A., & Rana, A. (2013). K-Means with Three different Distance Metrics. *International Journal of Computer Applications*, 67(10). <https://doi.org/10.5120/11430-6785>
- Wu, J. (2012). *Advances in K-Means clustering: a data mining thinking*. New York: Springer Science & Business Media.
- Wulandari, L., & Yogantara, B. O. (2022). Algorithm analysis of K-means and fuzzy C-means for clustering countries based on economy and health. *Faktor Exacta*, 15(2).
- Zusanti, R. D., Sasana, H., & Rusmijati, R. (2020). Analisis Pengaruh IPM, Per-tumbuhan Ekonomi dan TPT Terhadap Ketimpangan Wilayah di Pulau Jawa 2010-2018. *Journal of Economic*, 2(3), 602-615.