

## Perbandingan Metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means* dalam Pengelompokan Wilayah di Sumatera Utara

Andre June Agri Saragih<sup>1</sup>, Rosman Siregar<sup>2</sup>

<sup>1</sup>Prodi Sarjana Matematika, FMIPA, Universitas Sumatera Utara, Medan-Indonesia 20155

Email: <sup>1</sup>[andrejune06@gmail.com](mailto:andrejune06@gmail.com), <sup>2</sup>[rosmansiregar@yahoo.com](mailto:rosmansiregar@yahoo.com)

### ABSTRAK

Pada bulan Maret 2023, persentase penduduk miskin di daerah perkotaan sebesar 8,23% dan di daerah perdesaan sebesar 8,03%, dengan penurunan 0,40 poin di daerah perkotaan dan kenaikan 0,07 poin di daerah perdesaan dibandingkan bulan September 2022. Berdasarkan fakta tersebut dapat disimpulkan bahwa belum terjadinya pembangunan yang merata di desa dan di kota. Analisis *cluster* merupakan salah satu alat yang dapat dipakai untuk mengelompokkan Kota dan Kabupaten di Provinsi Sumatera Utara. Penelitian ini menganalisis perbandingan metode *K-Means*, *K-Medoids* dan *Fuzzy C-Means* dalam melakukan analisis *cluster*. Dimana pada metode *K-Means*, *cluster* 1 berjumlah 4 kota/kabupaten, *cluster* 2 berjumlah 8 kota/kabupaten dan *cluster* 3 berjumlah 21 kota/kabupaten. Pada metode *K-Medoids*, hasil *cluster* sama dengan hasil *cluster* metode *K-Means*. Sedangkan pada metode *Fuzzy C-Means*, *Cluster* 1 berjumlah 9 Kota/Kabupaten, *cluster* 2 berjumlah 16 Kota/Kabupaten dan *cluster* 3 berjumlah 8 Kota/Kabupaten. Kemudian berdasarkan nilai *Davies Bouldin Index* (DBI) dari ketiga metode, disimpulkan bahwa hasil *cluster* metode *K-Means* lebih optimal dibandingkan dengan metode *K-Medoids* dan *Fuzzy C-Means* karena memiliki nilai DBI yang terkecil.

**Kata kunci:** Analisis *Cluster*, Metode *Fuzzy C-Means*, Metode *K-Means*, Metode *K-Medoids*.

### ABSTRACT

In March 2023, the percentage of poor people in urban areas was 8.23% and in rural areas was 8.03%, with a decrease of 0.40 points in urban areas and an increase of 0.07 points in rural areas compared to September 2022. Based on these facts, it can be concluded that there has not been even development between villages and city. Cluster analysis is a tool that can be used to group cities and regencies in North Sumatra Province. This research analyzes the comparison of the *K-Means*, *K-Medoids* and *Fuzzy C-Means* methods in conducting cluster analysis. Where in the *K-Means* method, cluster 1 has 4 cities/regencies, cluster 2 has 8 cities/regencies and cluster 3 has 21 cities/regencies. In the *K-Medoids* method, the cluster results are the same as the cluster results of the *K-Means* method. Meanwhile, in the *Fuzzy C-Means* method, Cluster 1 has 9 cities/regencies, cluster 2 has 16 cities/regencies and cluster 3 has 8 cities/regencies. Then, based on the *Davies Bouldin Index* (DBI) values from the three methods, it was concluded that the cluster results of the *K-Means* method were more optimal compared to the *K-Medoids* and *Fuzzy C-Means* methods because they had the smallest DBI value.

**Keywords:** Cluster Analysis, *Fuzzy C-Means* Method, *K-Means* Method, *K-Medoids* Method.

### A. Pendahuluan

Analisis *cluster* merupakan salah satu teknik dalam statistika yang digunakan dalam mengelompokkan sekumpulan data menjadi beberapa *cluster* dimana masing-masing *cluster* dengan karakteristik tersendiri. Beberapa metode yang dapat digunakan dalam melakukan analisis *cluster* yaitu metode *K-Means*, *K-Medoids*, dan *Fuzzy C Means* yang sama-sama melakukan pendekatan iteratif.

Yang termasuk metode non hierarki dalam melakukan analisis klaster adalah metode *K-Means*. Sekumpulan data dibagi menjadi beberapa *cluster* menggunakan pendekatan *K-*

*Means*, dengan gagasan bahwa data dengan kualitas serupa berada dalam *cluster* yang sama dan data dengan kualitas berbeda berada dalam *cluster* berbeda (Rohmawati et al., 2015). Kelebihan metode *K-Means* yaitu metodenya yang sederhana dan konvergensi yang cepat dicapai. Walaupun begitu, jumlah *cluster* yang akan digunakan perlu ditentukan terlebih dahulu dan pilihan jumlah *cluster* akan mempengaruhi kecepatan konvergensi (Yuan & Yang, 2019)

Metode PAM (*Partitioning Around Medoid*), juga dikenal sebagai metode *K-Medoids*, adalah salah satu metode non-hierarki dalam metode analisis *cluster*. Metode *K-*

*Medoids* memiliki kelebihan dalam penanganan outlier dibandingkan dengan metode *K-Means* karena objek – objek ini letaknya sangat jauh atau karakteristiknya sangat berbeda, sehingga akan mendistorsi nilai rata-rata *cluster* tersebut jika dimasukkan ke suatu *cluster* (Wira et al., 2019).

Salah satu teknik analisis *cluster* yang dikenal sebagai *Fuzzy C-Means* menggunakan pengelompokan fuzzy, yang memungkinkan data dianggap sebagai anggota *cluster* dengan menunjukkan derajat keanggotaan yang berbeda dari 0 hingga 1. Metode ini menghasilkan kumpulan pusat *cluster* dengan derajat keanggotaan yang berbeda untuk setiap titik data.

Indeks Pembangunan Manusia (IPM) merupakan suatu tolak ukur untuk menguji kualitas Pembangunan manusia di suatu daerah. Provinsi Sumatera Utara, sebagai salah satu provinsi yang berkembang di Indonesia, membutuhkan pemahaman yang mendalam tentang distribusi pembangunan manusia di berbagai kota dan kabupaten di wilayah tersebut. Jika dilihat data penduduk miskin di Badan Pusat Statistik (BPS) dapat disimpulkan bahwa belum terjadi pembangunan yang merata di kabbbupaten dengan di kota.

Dalam konteks ini, analisis *cluster* dapat menjadi alat untuk mengelompokkan kota/kabupaten berdasarkan indikator indeks pembangunan manusia. Metode pengelompokan yang optimal dapat membantu pemerintah daerah dalam mengidentifikasi daerah-daerah yang memerlukan perhatian lebih dalam upaya peningkatan kesejahteraan masyarakat. Oleh karena itu, penelitian ini akan membandingkan metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means* dalam pengelompokan Kota dan Kabupaten di Provinsi Sumatera Utara berdasarkan indikator Indeks Pembangunan Manusia (IPM).

### 1. Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) adalah sebuah tolak ukur penting untuk menilai tingkat pembangunan manusia suatu wilayah. Pada Tahun 1990, United Nations Development (UNDP) pertama kali memperkenalkan Indeks Pembangunan Manusia (IPM). Menurut (Lengfelder, 1990), Indeks Pembangunan Manusia didefinisikan sebagai “a process of enlarging people’s choice” yang berarti suatu proses dimana manusia memiliki banyak pilihan.

Terdapat tiga dimensi yang digunakan dalam mengukur tingkat Indeks pembangunan manusia yaitu ekonomi, pendidikan dan Kesehatan

### 2. Analisis Cluster

Analisis *cluster* merupakan teknik analisis yang digunakan untuk membaca pola dari sekumpulan objek yang kemudian akan dikelompokkan ke dalam beberapa *cluster* berdasarkan persamaan karakteristiknya (Muhartini et al., 2022). Manfaat yang didapat dalam pemakaian analisis *cluster* adalah mempermudah peneliti dalam mengidentifikasi objek-objek ke dalam *cluster* berdasarkan karakteristiknya, membantu peneliti mengolah data mentah dan berperan dalam beberapa bidang. Menurut (Johnson & Wichern, 2007), alasan mengapa analisis klaster merupakan metode yang ketinggalan jaman adalah karena metode ini tidak membuat asumsi mengenai jumlah pengelompokan. Kesamaan atau jarak menjadi dasar dilakukannya pengelompokan.

### 3. K-Means

Metode *K-Means* ialah metode dalam melakukan analisis *cluster* yang termasuk ke dalam pendekatan non-hierarki. Dimana metode *K-Means* akan mempartisi sekumpulan objek menjadi beberapa *cluster* dimana objek yang sama karakteristiknya akan digabungkan menjadi satu *cluster*, sedangkan data yang tidak memiliki kesamaan karakteristik akan dipartisi menjadi beberapa *cluster* (Amalina et al., 2022).

Menurut (Sarwono, 2010), langkah-langkah metode *K-Means* adalah sebagai berikut :

- a. Jumlah *K-cluster* yang ingin dibentuk dapat ditentukan terlebih dahulu.
- b. Menentukan nilai secara random sebanyak k sebagai *centroid* awal.
- c. Dengan menggunakan rumus jarak Euclidean, carilah jarak setiap objek dengan setiap *centroid* hingga ditentukan jarak terdekat antara setiap objek dengan *centroid* tersebut.

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

Dimana:

$x_i$  = data kriteria,

$\mu_j$  = *centroid* pada *cluster* ke-j

- d. Klasifikasikan masing-masing objek dengan melihat kedekatannya atau jarak terkecil dengan *centroid*.
- e. Perbaharui nilai *centroid* dengan menghitung rata-rata *cluster* yang telah diklasifikasikan sebelumnya.
- f. Ulangi proses analisis dari langkah 3 sampai 5, sampai semua anggota tiap *cluster* tidak ada perubahan.

#### 4. *K-Medoids*

Suatu metode dalam analisis *cluster* yang akan mempartisi sekumpulan data menjadi beberapa *cluster* yaitu Metode *K-Medoids*. Metode *K-Medoids* dan *K-Means* termasuk ke dalam metode partitional yang akan membagi objek menjadi beberapa *cluster* (Wira et al., 2019). Yang membedakan metode *K-Medoids* dengan *K-Means* dapat ditemukan dari cara penentuan pusat *cluster*, dimana pada proses pengelompokkan dengan metode *K-Medoids* akan digunakan perwakilan (medoid) berupa objek atau data sebagai *centroid*.

Menurut (Han et al., 2011), tahapan *K-Medoids clustering* salah sebagai berikut :

- a. Tentukan objek sejumlah  $k$  dari semua  $n$  objek sebagai *medoids* secara acak.
- b. Kemudian Klasifikasikan objek yang bukan *medoids* ke dalam sebuah *cluster* dengan melihat jarak terdekat dengan *medoids* menggunakan rumus Euclidean Distance seperti pada Persamaan (1).
- c. Lalu tentukan objek kembali sebagai perwakilan untuk *medoids* baru secara acak yang berasal dari hasil klasifikasi sebelumnya.
- d. Lalu hitung kedekatan objek dengan *medoids* baru lalu klasifikasikan objek tersebut.
- e. Hitunglah selisih antara total distance baru dikurang dengan total distance lama sebagai total simpangan ( $S$ ). Iterasi akan berlanjut jika  $S < 0$ .
- f. Lakukan langkah 3, 4 dan 5 hingga diperoleh nilai total simpangan  $S > 0$ .

#### 5. *Fuzzy C-Means*

Metode dalam analisis *cluster* dalam menentukan anggota tiap *cluster* dilihat berdasarkan derajat keanggotaannya merupakan metode *Fuzzy C-Means*. *Fuzzy C-Means* memungkinkan untuk mengelompokkan dengan kondisi kelompok data yang tidak terbagi dengan jelas (Sormin et al., 2015).

Algoritma *Fuzzy C-Means*, yaitu :

- a. Tentukan data yang akan diolah yaitu  $X$  yang berupa matriks berukuran  $n \times m$ , dimana  $n$  adalah jumlah sampel data dan  $m$  adalah fitur setiap data.  $X_{ij}$  merupakan data sampel ke-  $i$  ( $i = 1, 2, \dots, n$ ), fitur ke-  $j$  ( $j = 1, 2, \dots, m$ ).
- b. Menentukan :
  - a. Total *cluster* :  $c$
  - b. Derajat pembobot :  $w$
  - c. Iterasi Maksimum :  $MaxIter$
  - d. *Error* terkecil yang diharapkan :  $\xi$
  - e. Fungsi Objektif Pertama :  $P_0 = 0$
  - f. Iterasi awal :  $t = 1$
- c. Pilihlah nilai secara acak yaitu  $\mu_{ik}$ ,  $i = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, c$ ; sebagai elemen matriks partisi awal  $U$ . Dengan nilai  $\mu_{ik}$  diantara 0 hingga 1 dan jumlah setiap barisnya sama dengan 1.
- d. Tentukan nilai pusat *cluster* ke- $k$ ,  $V_{kj}$  dengan  $k = 1, 2, \dots, c$  dan  $j = 1, 2, \dots, m$ .
 
$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w \times X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$$
- e. Pada iterasi ke-  $t$ , tentukan nilai fungsi objektif.
 
$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right)$$
- f. Hitunglah perubahan matriks partisi
 
$$\mu_{ik} = \frac{[\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}{\sum_{k=1}^c [\sum_{j=1}^m (X_{ij} - V_{kj})^2]^{-\frac{1}{w-1}}}$$

Dengan  $i = 1, 2, \dots, n$  dan  $k = 1, 2, \dots, c$
- g. Iterasi dapat berhenti Jika ( $|P_t - P_{t-1}| < \xi$ ) atau ( $t >$  iterasi maksimal ) maka berhenti.

#### 6. Pengujian Hasil *Cluster*

Pengujian atau validasi hasil dari analisis *cluster* merupakan suatu prosedur untuk mengetahui baik atau buruknya hasil analisis *cluster* yang dilakukan secara kuantitatif dan objektif. Salah satu metode yang dapat digunakan dalam mengevaluasi hasil *cluster* yaitu *Davies Bouldin Index* (DBI).

Metode *Davies Bouldin Index* (DBI) bertujuan untuk mencapai jarak maksimal suatu *cluster* dengan *cluster* lainnya dan juga akan memperkecil jarak setiap objek yang berada pada satu *cluster*. Hasil *cluster* yang optimal adalah hasil *cluster* yang memiliki nilai DBI terkecil (Hilmi et al., 2015).

Berikut tahapan melakukan perhitungan *Davies Bouldin Index* (DBI), yaitu :

- a. Menghitung nilai Sum of Square Within *cluster* (SSW) untuk menguji hubungan dalam sebuah *cluster* ke-i.
- b. Melakukan perhitungan nilai Sum of Squared Between *cluster* (SSB) untuk mencari perbedaan jarak antara dua *cluster* dengan cara menghitung jarak *centroid* suatu *cluster* dengan *cluster* lainnya.
- c. Menghitung nilai  $R_{ij}$  sebagai rasio untuk mengetahui nilai perbandingan yang baik antara *cluster* ke-i dan *cluster* ke-j.
- d. Hitung nilai *Davies Bouldin Index* (DBI) dengan persamaan:

$$DBI = \frac{1}{K} \sum_{j=1}^K \max(R_{i,j})$$

Dimana:

DBI = Nilai *Davies Bouldin Index* (DBI).

## B. Metode Penelitian

Pada penelitian ini dilakukan beberapa langkah sistematis dimulai dari studi literatur hingga pembuatan kesimpulan dan saran. Berikut adalah langkah sistematis yang dilakukan yaitu:

### 1. Studi Literatur

Membaca dan mempelajari referensi buku dan jurnal yang digunakan dalam metode ini, yaitu mengenai metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*.

### 2. Mengumpulkan data

Penelitian ini memanfaatkan data sekunder, yaitu data indikator Indeks Pembangunan Manusia (IPM) berupa data umur harapan hidup, data pengeluaran per Kapita, data rata-rata lama sekolah dan data harapan lama sekolah dari masing-masing Kota dan Kabupaten di Provinsi Sumatera Utara yang dapat diperoleh dari website Badan Pusat Statistik (BPS) Sumut.

### 3. Melakukan Pengolahan Data

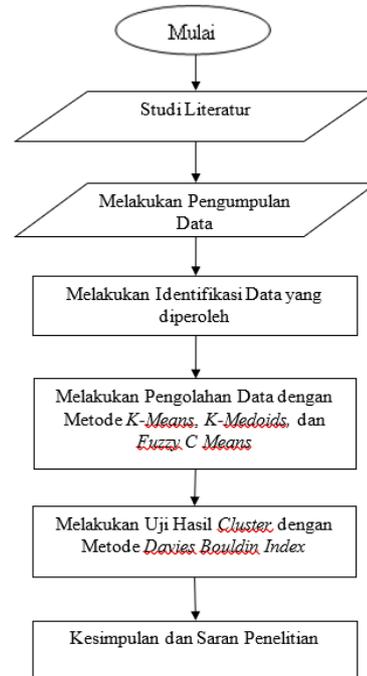
Melakukan Pengolahan data menggunakan metode *K-Means*, *K-Medoids*, dan *Fuzzy C-Means*. Kemudian melakukan analisis data dengan bantuan *software Matlab*.

### 4. Evaluasi Hasil Pengolahan Data

Menghitung skor metode *Davies Bouldin Index* (DBI) terhadap hasil *cluster* dari metode *K-Means*, *K-Medoids* dan *Fuzzy C-Means*.

### 5. Membuat Kesimpulan dan Saran

Membuat Kesimpulan dan saran mengenai perbandingan hasil pengolahan data yang telah didapatkan.



Gambar 1. Diagram alir penelitian

## C. Hasil dan Pembahasan

Data penelitian yang digunakan dalam penelitian ini adalah data sekunder yaitu data indikator Indeks Pembangunan Manusia (IPM) di provinsi Sumatera Utara yang diperoleh dari *website* BPS Sumut. Setelah data dikumpulkan, langkah selanjutnya adalah pengolahan data dengan metode *K-Means*, *K-Medoids* dan *Fuzzy C-Means*. Pada penelitian ini jumlah *cluster* yang akan digunakan adalah 3 *cluster*.

### 1. Pengolahan Data Metode K-Means

Pada metode *K-Means*, jumlah *cluster* yang digunakan yaitu 3 *cluster*. Pada metode *K-Means*, pengolahan data berhenti pada iterasi ke-3. Dimana didapatkan hasil *cluster* seperti pada Tabel 1 di bawah ini.

Tabel 1. Hasil *cluster K-Means*

No.	Kota dan Kabupaten	Hasil Cluster
1	Nias	1
2	Mandailing Natal	2
3	Tapanuli Selatan	3

4	Tapanuli Tengah	2
5	Tapanuli Utara	3
6	Toba	3
7	Labuhan Batu	3
8	Asahan	3
9	Simalungun	3
10	Dairi	3
11	Karo	3
12	Deli Serdang	3
13	Langkat	3
14	Nias Selatan	1
15	Humbang Hasundutan	2
16	Pakpak Bharat	2
17	Samosir	2
18	Serdang Bedagai	3
19	Batu Bara	3
20	Padang Lawas Utara	2
21	Padang Lawas	2
22	Labuhanbatu Selatan	3
23	Labuanbatu Utara	3
24	Nias Utara	1
25	Nias Barat	1
26	Sibolga	3
27	Tanjungbalai	3
28	Pematangsiantar	3
29	Tebing Tinggi	3
30	Medan	3
31	Binjai	3
32	Padangsidempuan	3
33	Gunungsitoli	2

Pada Tabel 1 dapat dilihat hasil *cluster* yaitu *cluster* 1 berjumlah 4 Kota/Kabupaten, *cluster* 2 berjumlah 8 Kota/Kabupaten dan *cluster* 3 berjumlah 21 Kota/Kabupaten. Setelah didapatkan hasil *cluster* seperti pada Tabel 1, kita dapat menentukan *cluster* baik, sedang dan buruk dengan melihat rata-rata indikator IPM masing-masing *cluster*.

**Tabel 2.** Rata-rata indikator *K-Means*

<i>Cluster</i>	UHH	PpK	RLS	HLS
<i>Cluster</i> 1	71,94	6942	6,63	13,1
<i>Cluster</i> 2	72,67	9498	9,42	13,63
<i>Cluster</i> 3	73,71	12205	9,94	13,49

Dari Tabel 2 memberikan informasi berikut:

Untuk *Cluster* 1, nilai indikator Umur Harapan Hidup (UHH) sebesar 71,94, indikator Pengeluaran per Kapita (Ppk) sebesar 6942, indikator Rata-rata Lama Sekolah (RLS) sebesar

6,63 dan indikator Harapan Lama Sekolah (HLS) sebesar 13,1.

Untuk *Cluster* 2, nilai indikator Umur Harapan Hidup sebesar 72,67, indikator Pengeluaran per Kapita sebesar 9498, indikator Rata-rata Lama Sekolah sebesar 9,42 dan indikator Harapan Lama Sekolah sebesar 13,63.

Untuk *Cluster* 3, nilai indikator Umur Harapan Hidup sebesar 73,71, indikator Pengeluaran per Kapita sebesar 12205, indikator Rata-rata Lama Sekolah sebesar 9,94 dan indikator Harapan Lama Sekolah sebesar 13,49.

Sehingga dapat disimpulkan bahwa *cluster* 1 merupakan *cluster* yang buruk, *cluster* 2 merupakan *cluster* sedang, dan *cluster* 3 merupakan *cluster* yang baik.

## 2. Pengolahan Data Metode *K-Medoids*

Pengolahan data dengan metode *K-Medoids* dengan menggunakan banyak *cluster* yaitu 3 *cluster*. Pada metode ini, iterasi berhenti pada iterasi ke-3. Berikut adalah hasil *cluster* dengan metode *K-Means* seperti pada Tabel 3 di bawah ini.

**Tabel 3.** Hasil *cluster K-Medoids*

No.	Kota dan Kabupaten	Hasil <i>Cluster</i>
1	Nias	1
2	Mandailing Natal	2
3	Tapanuli Selatan	3
4	Tapanuli Tengah	2
5	Tapanuli Utara	3
6	Toba	3
7	Labuhan Batu	3
8	Asahan	3
9	Simalungun	3
10	Dairi	3
11	Karo	3
12	Deli Serdang	3
13	Langkat	3
14	Nias Selatan	1
15	Humbang Hasundutan	2
16	Pakpak Bharat	2
17	Samosir	2
18	Serdang Bedagai	3
19	Batu Bara	3
20	Padang Lawas Utara	2
21	Padang Lawas	2
22	Labuhanbatu Selatan	3
23	Labuanbatu Utara	3
24	Nias Utara	1
25	Nias Barat	1
26	Sibolga	3

27	Tanjungbalai	3
28	Pematangsiantar	3
29	Tebing Tinggi	3
30	Medan	3
31	Binjai	3
32	Padangsidempuan	3
33	Gunungsitoli	2

Pada Tabel 3 memberikan informasi yaitu *cluster* 1 berjumlah 4 Kota/Kabupaten, *cluster* 2 berjumlah 8 Kota/Kabupaten dan *cluster* 3 berjumlah 21 Kota/Kabupaten. Setelah didapatkan hasil *cluster* seperti pada Tabel 3, kita dapat menentukan *cluster* baik, sedang dan buruk dengan melihat rata-rata indikator IPM masing-masing *cluster*.

**Tabel 4.** Rata-rata indikator IPM *K-Medoids*

<i>Cluster</i>	UHH	PpK	RLS	HLS
<i>Cluster</i> 1	71,94	6942	6,63	13,1
<i>Cluster</i> 2	72,67	9498	9,42	13,63
<i>Cluster</i> 3	73,71	12205	9,94	13,49

Dari Tabel 4 memberikan informasi berikut:

Untuk *Cluster* 1, nilai indikator Umur Harapan Hidup (UHH) sebesar 71,94, indikator Pengeluaran per Kapita (PpK) sebesar 6942, indikator Rata-rata Lama Sekolah (RLS) sebesar 6,63 dan indikator Harapan Lama Sekolah (HLS) sebesar 13,1.

Untuk *Cluster* 2, nilai indikator Umur Harapan Hidup sebesar 72,67, indikator Pengeluaran per Kapita sebesar 9498, indikator Rata-rata Lama Sekolah sebesar 9,42 dan indikator Harapan Lama Sekolah sebesar 13,63.

Untuk *Cluster* 3, nilai indikator Umur Harapan Hidup sebesar 73,71, indikator Pengeluaran per Kapita sebesar 12205, indikator Rata-rata Lama Sekolah sebesar 9,94 dan indikator Harapan Lama Sekolah sebesar 13,49. Sehingga dapat disimpulkan bahwa *cluster* 1 merupakan *cluster* yang buruk, *cluster* 2 merupakan *cluster* sedang, dan *cluster* 3 merupakan *cluster* yang baik.

### 3. Pengolahan Data Metode *Fuzzy C-Means*

Pada metode *Fuzzy C-Means*, pengolahan data menggunakan 3 *cluster*. Pada metode ini, iterasi berhenti pada Iterasi ke-36.

```

Command Window
>> fuzzycmeans
Iteration count = 1, obj. fcn = 55931002.678053
Iteration count = 2, obj. fcn = 34049630.135658
Iteration count = 3, obj. fcn = 22735359.635184
Iteration count = 4, obj. fcn = 19508717.482115
Iteration count = 5, obj. fcn = 18522317.895891
Iteration count = 6, obj. fcn = 18135461.977766
Iteration count = 7, obj. fcn = 17908686.739065
Iteration count = 8, obj. fcn = 17692396.456548
Iteration count = 9, obj. fcn = 17448875.496972
Iteration count = 10, obj. fcn = 17208722.964014
Iteration count = 11, obj. fcn = 17026821.336215
Iteration count = 12, obj. fcn = 16921639.413229
Iteration count = 13, obj. fcn = 16871831.979224
Iteration count = 14, obj. fcn = 16850884.627838
Iteration count = 15, obj. fcn = 16842612.262393
Iteration count = 16, obj. fcn = 16839450.641560
Iteration count = 17, obj. fcn = 16838263.372841
Iteration count = 18, obj. fcn = 16837821.908245
Iteration count = 19, obj. fcn = 16837658.699254
Iteration count = 20, obj. fcn = 16837598.567336
Iteration count = 21, obj. fcn = 16837576.458281
Iteration count = 22, obj. fcn = 16837568.339466
Iteration count = 23, obj. fcn = 16837565.360363
Iteration count = 24, obj. fcn = 16837564.267720
Iteration count = 25, obj. fcn = 16837563.867085
Iteration count = 26, obj. fcn = 16837563.720210
Iteration count = 27, obj. fcn = 16837563.666371
Iteration count = 28, obj. fcn = 16837563.646637
Iteration count = 29, obj. fcn = 16837563.639404
Iteration count = 30, obj. fcn = 16837563.636752
Iteration count = 31, obj. fcn = 16837563.635781
Iteration count = 32, obj. fcn = 16837563.635425
Iteration count = 33, obj. fcn = 16837563.635294
Iteration count = 34, obj. fcn = 16837563.635246
Iteration count = 35, obj. fcn = 16837563.635229
Iteration count = 36, obj. fcn = 16837563.635222
fx >>

```

**Gambar 2.** Output matlab *Fuzzy C-Means*

Berdasarkan hasil output pada Gambar 2, iterasi berhenti pada iterasi ke-36 dengan nilai fungsi objektif yaitu sebesar 16837563,63522 dikarenakan sudah memenuhi kondisi  $|P_t - P_{t-1}| < 0,00001$ .

Kemudian diperoleh hasil *cluster* seperti pada Tabel 5 dibawah ini.

**Tabel 5.** Hasil *cluster Fuzzy C-Means*

No.	Kota dan Kabupaten	Hasil <i>Cluster</i>
1	Nias	2
2	Mandailing Natal	1
3	Tapanuli Selatan	1
4	Tapanuli Tengah	1
5	Tapanuli Utara	1
6	Toba	3
7	Labuhan Batu	1
8	Asahan	1
9	Simalungun	1
10	Dairi	1
11	Karo	3
12	Deli Serdang	3
13	Langkat	1

14	Nias Selatan	2
15	Humbang Hasundutan	2
16	Pakpak Bharat	2
17	Samosir	2
18	Serdang Bedagai	1
19	Batu Bara	1
20	Padang Lawas Utara	1
21	Padang Lawas	2
22	Labuhanbatu Selatan	1
23	Labuanbatu Utara	3
24	Nias Utara	2
25	Nias Barat	2
26	Sibolga	3
27	Tanjungbalai	1
28	Pematangsiantar	3
29	Tebing Tinggi	3
30	Medan	3
31	Binjai	1
32	Padangsidempuan	1
33	Gunungsitoli	2

Dari Tabel 5 diperoleh hasil *cluster* yaitu *cluster* 1 berjumlah 16 kota/kabupaten, *cluster* 2 berjumlah 9 Kota/Kabupaten, dan *cluster* 3 berjumlah 8 Kota/Kabupaten. Setelah didapatkan hasil *cluster* seperti pada Tabel 5, kita dapat menentukan *cluster* baik, sedang dan buruk dengan melihat rata-rata indikator IPM masing-masing *cluster*.

**Tabel 6.** Rata-rata indikator IPM *Fuzzy C-Means*

<i>Cluster</i>	UHH	PpK	RLS	HLS
<i>Cluster</i> 1	73,08	11422	9,49	13,41
<i>Cluster</i> 2	72,68	8022	8,19	13,40
<i>Cluster</i> 3	74,21	13137	10,53	13,72

Pada Tabel 6 memberikan informasi berikut yaitu untuk *Cluster* 1, nilai indikator Umur Harapan Hidup (UHH) sebesar 73,08, indikator Pengeluaran per Kapita (PpK) sebesar 11422, indikator Rata-rata Lama Sekolah (RLS) sebesar 9,49 dan indikator Harapan Lama Sekolah (HLS) sebesar 13,41.

Untuk *Cluster* 2, nilai indikator Umur Harapan Hidup sebesar 72,68, indikator Pengeluaran per Kapita sebesar 8022, indikator Rata-rata Lama Sekolah sebesar 8,19 dan indikator Harapan Lama Sekolah sebesar 13,40.

Untuk *Cluster* 3, nilai indikator Umur Harapan Hidup sebesar 74,21, indikator Pengeluaran per Kapita sebesar 13137, indikator

Rata-rata Lama Sekolah sebesar 10,53 dan indikator Harapan Lama Sekolah sebesar 13,71. Sehingga dapat disimpulkan bahwa *cluster* 3 merupakan *cluster* baik, *cluster* 2 merupakan *cluster* buruk, dan *cluster* 1 merupakan *cluster* sedang.

#### 4. Pengujian Hasil *Cluster*

Setelah dilakukan pengelompokan Kota dan Kabupaten di Provinsi Sumatera Utara dengan metode *K-Means*, *K-Medoids* dan *Fuzzy C-Means*, selanjutnya kan dilakukan pengujian hasil *cluster* dengan menggunakan metode *Davies Bouldin Indeks* (DBI).

Berikut adalah perbandingan nilai DBI dari metode *K-Means*, *K-Medoids* dan *Fuzzy C-Means* seperti pada Tabel 7 dibawah ini.

**Tabel 7.** Perbandingan nilai DBI

Metode	DBI
<i>K-Means</i>	0,50997
<i>K-Medoids</i>	0,70149
<i>Fuzzy C-Means</i>	0,59536

Pada Tabel 7 dapat dilihat bahwa nilai DBI untuk hasil *cluster* metode *K-Means* yaitu sebesar 0,50997, nilai DBI untuk hasil *cluster* metode *K-Medoids* yaitu sebesar 0,70149 dan nilai DBI untuk hasil *cluster* metode *Fuzzy C-Means* yaitu sebesar 0,59536.

#### 5. Analisis Perbandingan

Dalam melakukan analisis *cluster* pada data penelitian, ketiga metode *cluster* yaitu *K-Means*, *K-Medoids* dan *Fuzzy C-Means* memiliki pendekatan yang berbeda dalam menghasilkan pengelompokan Kota dan Kabupaten.

Pada metode *K-Means*, pusat *cluster* atau *centroid* ditentukan secara acak, yang kemudian pada iterasi selanjutnya pusat *cluster* atau *centroid* ditentukan dari nilai rata-rata *cluster* hasil klasifikasi yang telah dilakukan sebelumnya. Iterasi berhenti jika tidak terdapat perubahan anggota *cluster*. Pada pengerjaan analisis *cluster* dengan metode *K-Means* diperoleh hasil *cluster* yaitu *cluster* 1 berjumlah 4 kota/kabupaten, *cluster* 2 berjumlah 8 kota/kabupaten dan *cluster* 3 berjumlah 21 kota/kabupaten.

Dalam metode *K-Medoids*, penentuan pusat *cluster* yaitu dengan menggunakan *medoids* secara acak. Kemudian dengan pusat *cluster* tersebut dihitung jarak serta dilakukan klasifikasi objek berdasarkan kedekatan jarak dengan pusat *cluster*. Kemudian untuk iterasi selanjutnya akan digunakan *medoids* baru yang ditentukan secara acak. Proses iterasi berhenti dengan melihat nilai total simpangan(S).

Metode *Fuzzy C-Means*. Pada metode ini pengelompokan dilakukan dengan melihat derajat keanggotaan setiap objek di setiap *cluster* yang berkisar antara 0 hingga 1. Untuk melihat objek masuk ke suatu *cluster* dengan melihat nilai derajat keanggotaan tertinggi di antara *cluster* tersebut. Pada metode ini didapatkan *cluster* 1 berjumlah 16 kota/kabupaten, *cluster* 2 berjumlah 9 Kota/Kabupaten, dan *cluster* 3 berjumlah 8 Kota/Kabupaten.

Kemudian untuk melihat metode yang optimal digunakan nilai *Davies Bouldin Index* (DBI) terhadap hasil *cluster* ketiga metode yaitu metode *K-Means*, *K-Medoids* dan *Fuzzy C-Means*.

Untuk melihat hasil *cluster* yang optimal adalah dengan cara melihat nilai DBI yang terkecil (Hilmi et al., 2015). Jika dilihat dari nilai DBI pada Tabel 4.29, maka metode *K-Means* memiliki hasil *cluster* yang lebih optimal karena nilai memiliki nilai DBI yang paling mendekati nilai 0 dibandingkan dengan metode *K-Medoids* dan *Fuzzy C-Means*. Artinya metode *K-Means* lebih baik dalam melakukan pengelompokan Kota dan Kabupaten di Provinsi Sumatera Utara berdasarkan indikator Indeks Pembangunan Manusia.

Oleh karena itu, untuk menyelesaikan masalah dalam penelitian ini yaitu pengelompokan Kota dan Kabupaten berdasarkan indikator Indeks Pembangunan Manusia (IPM) akan digunakan hasil *cluster* yang diperoleh dengan metode *K-Means*.

## D. Kesimpulan dan Saran

### 1. Kesimpulan:

Berdasarkan hasil penelitian yang sudah dilakukan, diperoleh kesimpulan bahwa hasil pengelompokan antara metode *K-Means* dan *K-Medoids* adalah sama, sedangkan metode *Fuzzy C-Means* berbeda. Dimana diperoleh hasil pengelompokannya yaitu *cluster* 1 berjumlah 4 Kota/Kabupaten, *cluster* 2 berjumlah 8 Kota/Kabupaten dan *cluster* 3

berjumlah 21 Kota/Kabupaten. Sedangkan pada pengelompokan dengan metode *Fuzzy C-Means* diperoleh hasil yaitu *cluster* 1 berjumlah 9 Kota/Kabupaten, *cluster* 2 berjumlah 16 Kota/Kabupaten dan *cluster* 3 berjumlah 8 Kota/Kabupaten. Kemudian berdasarkan nilai *Davies Bouldin Index* (DBI), menunjukkan bahwa nilai DBI pada metode *K-Means* lebih kecil dibandingkan dengan metode *K-Medoids* dan *Fuzzy C-Means* yaitu 0,50997. Maka pada penelitian ini dapat dilakukan dengan metode *K-Means* karena metode *K-Means* lebih optimal dalam mengelompokkan Kota dan Kabupaten di Provinsi Sumatera Utara berdasarkan indikator Indeks Pembangunan Manusia.

## 2. Saran

Berdasarkan hasil pengelompokan yang didapat, pemerintah Provinsi Sumatera Utara dapat mengetahui Kota dan Kabupaten yang perlu pengawasan lebih dalam upaya meningkatkan kesejahteraan rakyat. Kemudian penelitian ini diharapkan dapat dilakukan dengan menggunakan metode-metode analisis *cluster* lainnya dalam melakukan pengelompokan Kota dan Kabupaten berdasarkan Indeks Pembangunan Manusia berdasarkan indikator Indeks Pembangunan Manusia (IPM).

## E. Daftar Pustaka

- Amalina, T., Pramana, D. B., & Sari, B. N. (2022). Metode *K-Means* Clustering Dalam Pengelompokan Penjualan Produk Frozen Food. *Jurnal Ilmiah Wahana Pendidikan*, 8(15), 574–583. <https://doi.org/10.5281/zenodo.7052276>
- Han, J., Kambe, M., & Pe, J. (2011). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- Hilmi, M. N., Wilandari, Y., & Yasin, H. (2015). Pemetaan Preferensi Mahasiswa Baru Dalam Memilih Jurusan Menggunakan Artificial Neural Network (ANN) Dengan Algoritma Self Organizing Maps (SOM). *Jurnal Gaussian*, 4(1), 53–60. <http://ejournal-s1.undip.ac.id/index.php/gaussian>

- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. In *Pearson Prentice Hall*.
- Lengfelder, C. (1990). *Policies for human development*.
- Muhartini, A. A., Febriati, T., & Sukmawati, S. (2022). Analisis Cluster Untuk Mengelompokkan Penggunaan Kartu Perdana Seluler di Universitas Bina Bangsa (Survei Mahasiswa Jurusan Pendidikan Matematika T.1 2021-2022). *Jurnal Bayesian : Jurnal Ilmiah Statistika Dan Ekonometrika*, 2(1), 15–30. <https://doi.org/10.46306/bay.v2i1.25>
- Rohmawati, N., Defiyanti, S., & Jajuli, M. (2015). Implementasi Algoritma K-Means Dalam Pengklasteran Mahasiswa Pelamar Beasiswa. *Jurnal Ilmiah Teknologi Terapan*, 1(2), 62–68.
- Sarwono, Y. T. (2010). Aplikasi Model Jaringan Syaraf Tiruan Dengan Radial Basis Function Untuk Mendeteksi Kelainan Otak ( Stroke Infark ). *Jurnal Sistem Informasi*, 1–10.
- Sormin, R. P. A., Rumlawang, F. Y., & Sinay, L. J. (2015). Aplikasi Metode Fuzzy C-Means Untuk Pengklasteran Kelayakan Rumah Di Desa Wayame. *Jurnal Ilmu Matematika Dan Terapan*, 9(2), 135–146. <https://doi.org/10.30598/barekengvol9iss2pp135-146>
- Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. *Jurnal Terapan Sains & Teknologi*, 1(3), 53–68. <https://doi.org/10.21067/jtst.v1i3.3046>
- Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *Multidisciplinary Scientific Journal*, 2(2), 226–235. <https://doi.org/10.3390/j2020016>