

**PENGELOMPOKAN KABUPATEN/ KOTA DI PROVINSI SUMATERA
UTARA BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN
MANUSIA TAHUN 2022 DENGAN METODE
FUZZY C-MEANS DAN K- MEDOIDS**

Epridonta Sitepu^{1*}, Kris Suryowati², Noviana Pratiwi³

^{1,2,3}Program Studi Statistika, Fakultas Sains dan Teknologi Informasi, Universitas AKPRIND
Indonesia

Email : donta0100@gmail.com

*corresponding author

Abstract. *The development of human development indicators is important in an effort to measure and understand the social and economic progress of a region. North Sumatra Province, as one of the important regions in Indonesia, also needs to conduct an in-depth analysis related to human development at the Regency / City level. This study aims to group districts / cities in North Sumatra Province based on human development indicators using Fuzzy C-Means and K-Medoids methods. The data used in this study is secondary data that includes a number of human development indicators such as education, health, and income. Grouping is carried out using Fuzzy C-Means and K-Medoids algorithms to determine the characteristics of clusters formed based on HDI indicators, optimal clusters formed and the best method obtained. The conclusion obtained in this study obtained the best cluster results are the Fuzzy C-Means method at optimal $k = 6$ with a silhouette index value = 0.5850.*

Keywords: *Fuzzy C-Means, K-Medoids, Human Development Index*

Abstrak. Pengembangan indikator pembangunan manusia menjadi hal yang penting dalam upaya mengukur dan memahami kemajuan sosial dan ekonomi suatu wilayah. Provinsi Sumatera Utara, sebagai salah satu wilayah penting di Indonesia, juga perlu melakukan analisis yang mendalam terkait dengan pembangunan manusia di tingkat Kabupaten/Kota. Penelitian ini bertujuan untuk mengelompokkan Kabupaten/Kota di Provinsi Sumatera Utara berdasarkan indikator pembangunan manusia menggunakan metode Fuzzy C-Means dan K-Medoids. Data yang digunakan dalam penelitian ini adalah data sekunder yang mencakup sejumlah indikator pembangunan manusia seperti pendidikan, kesehatan, dan pendapatan. Pengelompokan dilakukan dengan menggunakan algoritma Fuzzy C-Means dan K-Medoids untuk mengetahui karakteristik *cluster* yang terbentuk berdasarkan Indikator IPM, *cluster* optimal yang terbentuk dan metode terbaik yang diperoleh. Kesimpulan yang didapat pada penelitian ini diperoleh hasil *cluster* terbaik adalah dengan metode *Fuzzy C-Means* pada k optimal=6 dengan nilai silhouette indeks = 0,5850.

Kata kunci : *Fuzzy C-Means, K-Medoids, Indeks Pembangunan Manusia*

1. Pendahuluan

Pembangunan manusia adalah salah satu parameter penting yang digunakan dalam mengukur kemajuan suatu wilayah atau negara. Baik itu pembangunan dibidang pendidikan, kesehatan, infrastruktur maupun ekonomi. Semua sasaran dari pembangunan tersebut adalah untuk masyarakat. Diharapkan dengan adanya pembangunan tersebut mampu mengatasi persoalan yang dialami oleh masyarakat seperti kemiskinan, putus sekolah, penganggura, kesehatan maupun kesenjangan sosial dapat teratasi dengan baik. Dalam proses pembangunan tersebut masyarakat juga dilibatkan dalam seluru proses pembangunan yang dilakukan. Guna mengetahui keberhasilan pembangunan itu sendiri, digunakanlah indeks yang dapat mengukurnya yaitu Indeks Pembangunan Manusia (IPM). IPM diperkenalkan oleh United Nation Development Programe (UNDP).

Indeks Pembangunan Manusia menjadi salah satu indikator yang penting dalam melihat sisi lain dari pembangunan. Setiap indikator komponen penghitungan IPM dapat dimanfaatkan untuk mengukur keberhasilan pembangunan kualitas hidup manusia (masyarakat/penduduk) (BPS Sumut, 2022). IPM dibangun melalui tiga aspek yakni kesehatan, pendidikan, dan taraf hidup layak. Dimensi pengetahuan diwakili oleh Indikator harapan lama sekolah dan rata-rata lama sekolah. Dimensi kesehatan diwakili oleh Indikator umur harapan hidup dan dimensi taraf hidup layak direpresentatif dengan Pengeluaran per Kapita.

Secara umum penelitian ini dilakukan untuk mengetahui kabupaten/kota yang memiliki nilai Indeks Pembangunan Manusia (IPM) yang rendah, sehingga menjadi perhatian pemerintah setempat untuk memajukan Pembangunan Manusia. Proses pengelompokan dapat dilakukan dengan mengimplementasikan metode *clustering*, yang dimana metode *clustering* merupakan pengelompokan objek ke dalam suatu kelompok yang memiliki kesamaan tinggi dibandingkan dengan objek yang berada dalam satu kelompok lain. Beberapa metode yang ada dalam *clustering* antara lain metode *Fuzzy C-Means* dan *K-Medoids*. Dari penelitian ini diharapkan dapat memberikan gambaran dan hasil pengelompokan kabupaten/kota berdasarkan dimensi Indeks Pembangunan Manusia agar pemerintah di provinsi Sumatera Utara dapat melihat kabupaten/kota mana yang harus lebih diperhatikan.

2. Metode Penelitian

Data yang digunakan pada penelitian ini adalah data sekunder yang diperoleh dari publikasi Indeks Pembangunan Manusia pada website resmi BPS Provinsi Sumatera Utara. Data <https://sumut.bps.go.id/subject/26/indeks-pembangunan-manusia.html#subjekViewTab4>. Data yang diambil merupakan dimensi Indeks Pembangunan Manusia untuk 33 kabupaten/kota di Provinsi Sumatera Utara tahun 2022.

a. Analisis Deskriptif

Statistika deskriptif adalah statistika yang tingkat pengerjaannya adalah untuk menghimpun, mengatur, dan mengolah data untuk dapat disajikan dan memberikan gambaran yang jelas mengenai suatu kondisi atau peristiwa tertentu dimana data diambil. Atau dengan kata lain, tugas statistika deskriptif adalah untuk menyajikan data dengan jelas agar dapat diambil pengertian atau makna tertentu berdasarkan penggambaran yang disajikan (Martias, 2021). Statistika deskriptif sesuai definisinya merupakan salah satu metode statistika yang berkaitan dengan pemberian dan penyajian sehingga dapat memberikan informasi yang berguna. Memberikan informasi yang berguna disini dimaksudkan bahwa penyajiannya akan memudahkan pembaca atau pengguna informasi untuk dapat membaca dan memanfaatkan data secara lebih mudah.

b. Standarisasi Data

Proses standarisasi dilakukan apabila diantara variabel-variabel yang diteliti terdapat perbedaan ukuran satuan yang besar. Perbedaan satuan yang mencolok dapat mengakibatkan perhitungan pada analisis *cluster* menjadi tidak valid. Oleh karena itu, perlu dilakukan proses standarisasi dengan melakukan transformasi pada data asli sebelum dianalisis lebih lanjut (Silvi, 2018). Z score merupakan transformasi sebuah distribusi data mengikuti distribusi standard. Pengertian distribusi normal adalah sebuah distribusi yang mempunyai rata-rata nol dan simpangan baku 1. Standarisasi dilakukan terhadap variabel yang relevan ke dalam bentuk *z-score*. Rumus perhitungan Z adalah sebagai berikut:

$$Z_i = \frac{x_i - \bar{x}}{\sigma} \quad (1)$$

Dimana:

X_i : Nilai data ke-i, untuk $i = 1, 2, 3, \dots, n$

\bar{x} : Nilai rata rata

σ : Standar deviasi

Z_i : Nilai Standarisasi Data ke-i, untuk $i = 1, 2, 3, \dots, n$

c. Uji Multikolinearitas

Dalam uji multikolinearitas jika terdapat nilai yang mencapai atau melebihi 0.8 maka terjadi multikolinearitas. Untuk mendeteksi ada atau tidaknya multikolinearitas dapat dilakukan dengan menghitung koefisien korelasi sederhana (korelasi *Pearson*) dengan persamaan sebagai berikut:

$$r_{x_1.x_2} = \frac{n \sum_{i=1}^n X_{i1} X_{i2} - (\sum_{i=1}^n X_{i1})(\sum_{i=1}^n X_{i2})}{\sqrt{(n \sum_{i=1}^n X_{i1}^2 - (\sum_{i=1}^n X_{i1})^2)(n \sum_{i=1}^n X_{i2}^2 - (\sum_{i=1}^n X_{i2})^2)}} \tag{2}$$

keterangan:

$r_{x_1.x_2}$: Koefisien korelasi pearson

n : Banyaknya pengamatan

X_{i1} : Nilai pengamatan ke-i pada variabel X_1

X_{i2} : Nilai pengamatan ke-i pada variabel X_2

i : 1. 2. 3. n

d. Jarak Euclidean

Dalam menjelaskan atau mengukur kemiripan/kesamaan antar objek digunakan pendekatan ukuran jarak (*distance*). Semakin besar nilai jarak antara objek berarti semakin besar pula perbedaan karakteristik antara objek. Sedangkan semakin kecil nilai jarak antara objek berarti semakin tinggi pula kemiripan/kesamaan karakteristik antara objek. Jarak Euclidean adalah besar suatu jarak pada garis lurus yang menghubungkan antar objek (Iis dkk.2022). Dalam analisis kluster, jarak Euclidean sering digunakan untuk mengukur seberapa dekat atau jauhnya dua titik data. Jarak euclidean diperoleh dengan rumus berikut :

$$d_{i,j} = \sqrt{\sum_{i=1}^n (x_{ik} - x_{jk})^2} \tag{3}$$

Dimana:

$d_{i,j}$: Jarak *Euclidean* dari objek ke-i dan objek ke-j

x_{ik} : Data ke-i pada variable k. dengan $i = 1.2. \dots n$

x_{jk} : Data ke-j pada variable k . dengan $i = 1.2. \dots n$

n : Jumlah data

e. Fuzzy C-Means

Fuzzy C-Means (FCM) adalah proses pengelompokan data yang ditentukan oleh tingkat keanggotaan (Firdaus dkk.2021). *Fuzzy C-Means* adalah salah satu algoritma yang banyak digunakan dalam metode *Fuzzy Clustering* (Andika.2019). *Fuzzy C-Means* adalah algoritma yang umum dan sering digunakan dalam pengelompokan data karena memberikan estimasi yang efisien dan tidak memerlukan banyak parameter (Lestari dkk, 2019). *Fuzzy C-Means* adalah salah satu teknik dari *Fuzzy Clustering* yang mengintegrasikan metode *Subtractive Clustering* dan *Fuzzy C-Means*. Hal ini dilakukan untuk mengatasi kekurangan *Fuzzy C-Means* dalam menentukan jumlah kluster, menginisialisasi matriks keanggotaan awal, dan menangani masalah inkonsistensi (Deana dkk,2022).

Algoritma *Fuzzy C-Means* (FCM) adalah sebagai berikut :

- 1) Input data yang akan di *cluster*. X . berupa matriks berukuran $n \times m$
 n : jumlah sampel data.
 m : variabel setiap data.
 X_{ij} : data sampel ke-i ($i = 1.2. \dots n$). atribut ke-j ($j = 1.2. \dots m$)
- 2) Menentukan:

- a. Jumlah *cluster* (c)
 - b. Pangkat pembobot (w): $w > 1$
 - c. Maksimum iterasi (*Max iter*)
 - d. *Error* terkecil (ϵ)
 - e. Fungsi objektif awal ($P_0 = 0$)
 - f. Iterasi awal ($t = 1$)
- 3) Membangkitkan bilangan random μ_{ik} ; $i = 1.2. \dots n$; $k = 1.2. \dots c$; sebagai elemen-elemen matriks partisi awal U . Hitung jumlah setiap kolom.

$$Q_j = \sum_{k=1}^c \mu_{ik} \text{ dengan } j = 1.2. \dots n$$

- 4) Menghitung pusat *cluster* ke- k : V_{kj} dengan $k = 1.2. \dots c$ dan $j = 1.2. \dots m$

$$V_{kj} = \frac{\sum_{i=1}^n ((\mu_{ik})^w * X_{ij})}{\sum_{i=1}^n (\mu_{ik})^w}$$

Keterangan:

V_{kj} : Pusat *cluster* ke- k dan atribut ke- j

μ_{ik} : Derajat keanggotaan sampel ke- i dan *cluster* ke- k

X_{ij} : Data pada sampel ke- i dan atribut ke- j

- 5) Menghitung fungsi objektif pada iterasi ke- t . P_t

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left(\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right] (\mu_{ik})^w \right) \quad (5)$$

- 6) Hitung perubahan matriks partisi:

$$\mu_{ik} = \frac{\left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}}{\sum_{k=1}^c \left[\sum_{j=1}^m (X_{ij} - V_{kj})^2 \right]^{\frac{-1}{w-1}}} \quad (6)$$

dengan $i : 1.2. \dots n$ dan $k = 1.2. \dots c$

- 7) Mengecek kondisi berhenti

- a. Jika $|P_t - P_{t-1}| < \xi$ atau $t > \text{max Iter}$ maka berhenti;
- b. Jika tidak: $= t + 1$. mengulangi langkah ke-4

f. *K-Medoids*

K-Medoids adalah penyempurnaan dari algoritma K-means pada *clustering* yang lebih efektif dalam menangani dataset yang mengandung outliers. Algoritma *K-Medoids* menetapkan pusat *cluster* berdasarkan representasi objek *cluster* yang disebut sebagai *medoid*. Keunggulan dari metode ini adalah kemampuannya untuk menangani kelemahan metode k-means yang sensitif terhadap outlier. Selain itu, kelebihan lainnya adalah hasil dari proses *clustering* tidak tergantung pada urutan masukan data dalam dataset (Agustini dan Kirana. 2022).

Algoritma *K-Medoids Clustering* adalah sebagai berikut:

- 1) Tentukan k (jumlah *cluster*) yang diinginkan
- 2) Pilih secara acak *medoid* awal sebanyak k dari n data
- 3) Hitung jarak masing-masing obyek ke *medoid* sementara menggunakan rumus jarak *euclidean*. kemudian tandai jarak terdekat obyek ke *medoid* dan hitung totalnya
- 4) Lakukan iterasi *medoid*.
- 5) Hitung total simpangan (S)
Jika a adalah jumlah jarak terdekat antara obyek ke *medoid* awal. dan b adalah jumlah jarak terdekat antara obyek ke *medoid* baru. maka total simpangan adalah $S = b - a$
Jika $S < 0$. maka tukar obyek dengan data untuk membentuk sekumpulan k baru sebagai *medoid*.
- 6) Ulangi langkah 3 sampai 5 dan hentikan jika sudah tidak terjadi perubahan anggota *medoid*.

g. Uji Silhouette Index

Silhouette Index merupakan salah satu metode yang digunakan untuk memvalidasi kebaikkan sebuah data. *cluster* tunggal (satu *cluster* dari sejumlah *cluster*). atau bahkan keseluruhan *cluster*. Metode ini paling banyak digunakan untuk memvalidasi *cluster* yang menggabungkan nilai kohesi dan separasi. Nilai *Silhouette Index* yang mendekati nilai 1 adalah nilai validitas terbaik (Qori'atunnadyah, 2023).

3. Hasil Dan Pembahasan

a. Analisis Deskriptif

Berdasarkan studi kasus dimensi Indeks Pembangunan Manusia (IPM) di Sumatera Utara tahun 2022 memiliki karakteristik yang disajikan dalam analisis deskriptif. Angka harapan hidup tertinggi terdapat pada Kota Pematang Siantar sebesar 74.25% dan nilai angka harapan hidup terendah terdapat pada Mandailing Natal Sebesar 63.05. Pada variabel harapan lama sekolah di Sumatera Utara tahun 2022 dengan nilai harapan lama sekolah tertinggi terdapat pada Kota Medan sebesar 14.77 tahun dan nilai harapan lama sekolah terendah terdapat pada Nias Selatan sebesar 12.48 tahun. Pada variabel rata-rata lama sekolah dengan nilai tertinggi terdapat pada Kota Medan sebesar 11.50 dan nilai rata-rata lama sekolah terendah terdapat pada Nias sebesar 5.88. Pada variabel pengeluaran per kapita dengan nilai pengeluaran per kapita tertinggi terdapat pada Kota Medan 15.503 dan nilai pengeluaran per kapita terendah terdapat pada Nias Barat sebesar 6.152.

b. Uji Multikolinearitas

Data dimensi Indeks Pembangunan Manusia di Sumatera Utara memiliki empat variabel, sehingga perlu dilakukan uji multikolinearitas untuk mengetahui adanya hubungan linear yang sempurna atau pasti diantara beberapa atau semua variabel. Adanya hubungan linear antara variabel dapat dilihat dari nilai koefisien korelasi antar variabel. Nilai koefisien Pearson yang melebihi 0,8 menandakan bahwa terjadi multikolinieritas. Berdasarkan hasil analisis menggunakan software R-Studio pada disajikan Tabel 1.

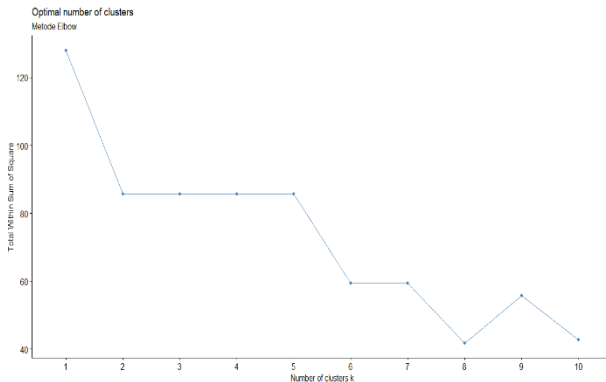
Tabel 1. Nilai Korelasi Antar Variabel

	AHH	HLS	RLS	Pengeluaran per Kapita
AHH	1	0.2589437	0.3328216	0.2530053
HLS	0.2589437	1	0.5968362	0.3056181
RLS	0.3328216	0.5968362	1	0.7616745
Pengeluaran per Kapita	0.2530053	0.3056181	0.7616745	1

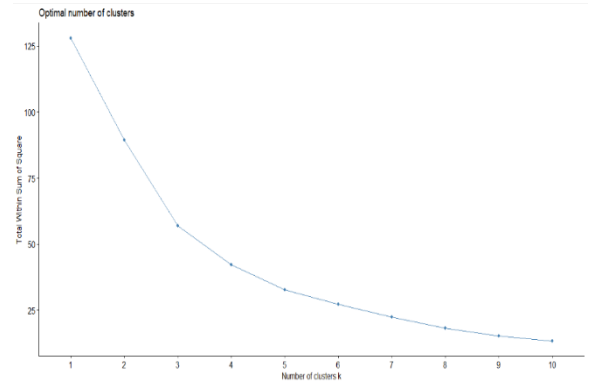
Berdasarkan Tabel 1 dapat dilihat nilai korelasi antar variabel penelitian tidak ada yang melebihi 0.8 yang artinya tidak terdapat multikolinearitas pada variabel-variabel penelitian yang akan digunakan dalam proses analisis. dengan demikian maka analisis selanjutnya memenuhi syarat untuk dilakukan.

c. *K-Optimum* Metode Elbow

Pada penelitian ini, penentuan *cluster* menggunakan metode elbow, dimana penentuan banyaknya *cluster* yang terbentuk dengan cara melihat hasil grafik yang membentuk siku. Berdasarkan hasil output software R-Studio, berikut adalah grafik metode elbow untuk menentukan jumlah *cluster* optimal untuk metode *Fuzzy C-Means* dan *K-Medoids* yang disajikan pada Gambar 1 dan 2.



Gambar 1. Grafik metode *Fuzzy C-Means*

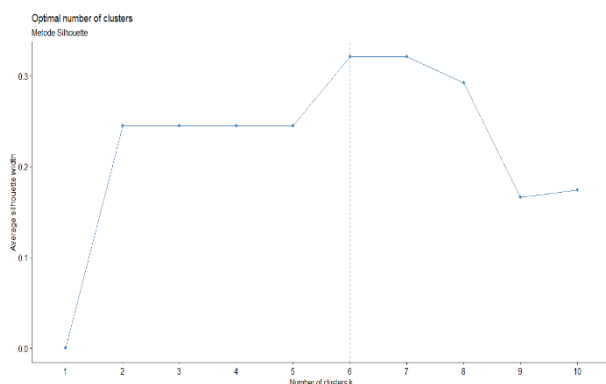


Gambar 2. Grafik metode *K-Medoids*

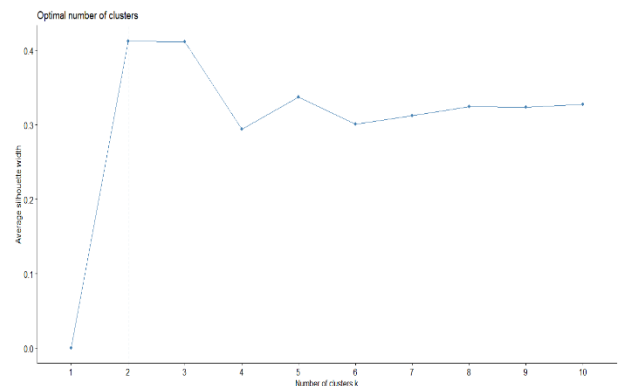
Untuk penentuan nilai k dengan metode elbow dengan grafik yang membentuk siku. Dari Gambar 1 dapat dilihat grafik membentuk siku pada angka 2, jadi nilai k optimum pada metode elbow untuk metode *Fuzzy C-Means* adalah 2 dan juga untuk metode *K-Medoids* pada Gambar 2 dapat dilihat grafik membentuk siku pada angka 3.

d. *K-Optimum* Metode Silhouette

Pada penelitian ini, penentuan *cluster* menggunakan metode Silhouette dimana penentuan *cluster* optimum pada metode *silhouette*. dengan mengambil *cluster* optimal pada titik *cluster* yang paling tinggi. Berdasarkan hasil output software R-Studio, berikut adalah grafik metode elbow untuk menentukan jumlah *cluster* optimal untuk metode *Fuzzy C-Means* dan *K-Medoids* yang disajikan pada Gambar 3 dan 4



Gambar 3. Grafik metode *Fuzzy C-Means*



Gambar 4. Grafik metode *K-Medoids*

Berdasarkan Gambar 3 menunjukkan bahwa jumlah *cluster* optimum yang terbentuk adalah $k=6$ pada metode *Fuzzy C-Means* dan $k=2$ pada metode *K-Medoids*.

e. Metode *Fuzzy C-Means*

Berdasarkan data penelitian berikut adalah hasil *Fuzzy C-Means Clustering* dengan bantuan software R-Studio dan dikelompokkan berdasarkan nilai k optimum yang terbentuk pada Tabel 2 dan 3.

Tabel 2. Kelompok *Cluster* Metode *Fuzzy C-Means* $k=2$

<i>Cluster</i>	Objek
<i>Cluster 1</i>	Tapanuli Utara, Toba Samosir, Labuhan Batu, Simalungun, Karo, Deli Serdang, Samosir, Kota Sibolga, Kota P, Siantar, Kota Tebing Tinggi, Kota Medan, Kota Binjai dan Kota P, Sidempuan.
<i>Cluster 2</i>	Nias, Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Asahan, Dairi, Langkat, Nias Selatan, Humbahas, Pakpak Barat, Serang Berdagai, Batu Bara, Paluta, Padang Lawas, Labusel, Labura, Nias Utara, Nias Barat, Kota Tanjung Balai dan Kota Gunung Sitoli.

Berdasarkan Tabel 2 *cluster* kabupaten/kota yang terbentuk di atas maka dapat diketahui bahwa terdapat 13 kabupaten/kota pada *cluster* 1 dan terdapat 20 kabupaten/kota pada *cluster* 2.

Tabel 3. Kelompok *Cluster* Metode *Fuzzy C-Means* $k=6$

<i>Cluster</i>	Objek
<i>Cluster 1</i>	Tapanuli Utara, Toba Samosir, Karo, Deli Serdang, Kota Sibolga, Kota Tebing Tinggi dan Kota Binjai,
<i>Cluster 2</i>	Mandailing Natal, Tapanuli Selatan dan Kota Tanjung Balai,
<i>Cluster 3</i>	Kota P, Siantar, Kota Medan, dan Kota P, Sidempuan,
<i>Cluster 4</i>	Nias, Nias Selatan, Nias Utara dan Nias Barat,
<i>Cluster 5</i>	Labuhan Batu, Asahan, Simalungun, Dairi, Langkat, Serdang Berdagai, Batu Bara, Labusel dan Labura,
<i>Cluster 6</i>	Tapanuli Tengah, Humbahas, Pakpak Barat, Samosir, Paluta, Padang Lawas dan Kota Gunung Sitoli,

Berdasarkan *cluster* kabupaten/kota yang terbentuk di atas maka dapat diketahui bahwa terdapat 7 kabupaten/kota pada *cluster* 1, terdapat 3 kabupaten/kota pada *cluster* 2, terdapat 3 kabupaten/kota pada *cluster* 3, terdapat 4 kabupaten/kota pada *cluster* 4, terdapat 9 kabupaten/kota pada *cluster* 5 dan terdapat 7 kabupaten/kota pada *cluster* 6.

f. Metode *K-Medoids*

Berdasarkan data penelitian berikut adalah hasil *Fuzzy C-Means Clustering* dengan bantuan software R-Studio dan dikelompokkan berdasarkan nilai *k optimum* yang terbentuk pada Tabel 4 dan 5.

Tabel 4. Kelompok *Cluster* Metode *K-Medoids* $k=3$

<i>Cluster</i>	Objek
<i>Cluster 1</i>	Nias, Nias Selatan, Nias Utara dan Nias Barat.
<i>Cluster 2</i>	Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Tapanuli Utara, Toba Samosir, Labuhan Batu, Asahan, Simalungun, Dairi, Karo, Deli Serdang, Langkat, Humbahas, Pakpak Barat, Samosir, Serdang Berdagai, Batu Bara, Paluta, Padang Lawas, Labusel, Labura, Kota Sibolaga, Kota Tanjung Balai, Kota Tebing Tinggi dan Kota Gunung Sitoli.
<i>Cluster 3</i>	Kota P, Siantar, Kota Medan, Kota Binjai dan Kota P, Sidempuan.

Berdasarkan *cluster* Kabupaten/Kota yang terbentuk di atas maka dapat diketahui bahwa terdapat 4 Kabupaten/Kota pada *cluster 1*, kemudian terdapat 25 Kabupaten/Kota pada *cluster 2* dan terdapat 4 Kabupaten/Kota pada *cluster 3*.

Tabel 5. Kelompok *Cluster* Metode *K-Medoids* $k=2$

<i>Cluster</i>	Objek
<i>Cluster 1</i>	Nias, Nias Selatan, Nias Utara dan Nias Barat.
<i>Cluster 2</i>	Mandailing Natal, Tapanuli Selatan, Tapanuli Tengah, Tapanuli Utara, Toba Samosir, Labuhan Batu, Asahan, Simalungun, Dairi, Karo, Deli Serdang, Langkat, Humbahas, Pakpak Barat, Samosir, Serdang Berdagai, Batu Bara, Paluta, Padang Lawas, Labusel, Labura, Kota Sibolga, Kota Tanjung Balai, Kota P, Siantar, Kota Tebing Tinggi, Kota Medan, Kota Binjai, Kota P, Sidempuan dan Kota Gunung Sitoli.

Berdasarkan *cluster* kabupaten/kota yang terbentuk di atas maka dapat diketahui bahwa terdapat 4 kabupaten/kota pada *cluster 1* dan terdapat 29 kabupaten/kota pada *cluster 2*.

g. Penentuan Metode Terbaik

Dari proses clustering kedua metode yaitu metode *Fuzzy C-Means Clustering* dan metode *K-Medoids Clustering* kemudian dilakukan penilaian metode mana yang terbaik untuk studi kasus dimensi Indeks Pembangunan Manusia (IPM) tahun 2022 dengan menggunakan silhouette index. Berikut adalah perhitungan silhouette index dengan bantuan software R-Studio pada Tabel 6:

Tabel 6. Nilai *Silhouette Index*

Metode	Metode k optimum	Jumlah Cluster	<i>Silhouette Index</i>
<i>Fuzzy c-means</i>	<i>Elbow</i>	2	0.504966507015425
	<i>Silhouette</i>	6	0.585058841961838
<i>K-Medoids</i>	<i>Elbow</i>	3	0.41
	<i>Silhouette</i>	2	0.37

Berdasarkan Tabel di atas dapat disimpulkan bahwa hasil *cluster* terbaik adalah dengan metode *Fuzzy C-Means* pada $k=6$ karena nilai *Silhouette Index* nya lebih besar daripada hasil *cluster* dengan metode *K-Medoids*.

Setelah mengetahui pengelompokan terbaik maka akan dilakukan *profiling cluster* untuk mengetahui karakteristik dari masing-masing *cluster* yang terbentuk. Karakteristik atau ciri *cluster* dilihat dengan nilai rata-rata setiap variabel pada masing-masing *cluster*. Berikut adalah *profiling cluster* dimensi IPM tahun 2022 di Sumatera Utara :

Tabel 7. Nilai *Profiling Cluster*

<i>Cluster</i>	AHH	HLS	RLS	Pengeluaran per Kapita
1	71.1	13.4	10.4	12316
2	64.1	13.4	9.22	11126
3	72.6	14.6	11.3	13195
4	69.6	12.9	6.46	6780
5	69.5	13.0	8.98	11431
6	68.9	13.5	9.30	9197

Dari 6 *cluster* yang terbentuk, akan dikelompokkan atau dikategorikan menjadi wilayah yang sangat tinggi, tinggi, cukup tinggi, cukup rendah, rendah dan sangat rendah dalam mempengaruhi Indeks Pembangunan Manusia pada tahun 2022 di Sumatera Utara berdasarkan nilai rata rata dari indikator Indeks Pembangunan Manusia. Kabupaten/kota dengan kelompok IPM sangat tinggi terdapat pada *cluster* 3, hal ini dikarenakan nilai rata-rata pada untuk variabel AHH, HLS, RLS dan pengeluaran per kapita lebih besar dibandingkan *cluster* lainnya, kemudian kabupaten/kota dengan kelompok IPM tinggi terdapat pada *cluster* 1, hal ini dikarenakan secara keseluruhan nilai rata-rata pada pada setiap variabel lebih besar setelah *cluster* 3, kemudian kabupaten/kota dengan kelompok IPM cukup tinggi terdapat pada *cluster* 6. Kemudian kabupaten/kota dengan kelompok IPM cukup rendah terdapat pada *cluster* 5, hal ini dikarenakan nilai rata-rata pada setiap variabel secara keseluruhan lebih besar setelah *cluster* 6, kemudian kabupaten/Kota dengan kelompok IPM rendah terdapat pada *cluster* 2, kabupaten/kota dengan kelompok IPM sangat rendah terdapat pada *cluster* 4. Berdasarkan hasil penjelasan tersebut kabupaten/kota pada *cluster* 2 dan 4, perlu menjadi perhatian pemerintah guna meningkatkan pembagunan pada kabupaten/kota tersebut, terutama pada kabupaten Nias, Nias Selatan, Nias Utara dan Nias Barat agar terciptanya masyarakat yang sejahtera.

4. Kesimpulan

Berdasarkan hasil uji validitas yang dihasilkan diketahui bahwa hasil *cluster* terbaik adalah dengan metode *Fuzzy C-Means* pada $k=6$ karena nilai *Silhouette Index* nya lebih besar daripada hasil *cluster* dengan metode *K-Medoids*. *cluster* 1 adalah sebanyak 7 kabupaten/kota, *cluster* 2 adalah sebanyak 3 kabupaten/kota, *cluster* 3 adalah sebanyak 3 kabupaten/kota, *cluster* 4 adalah sebanyak 4 kabupaten/kota, *cluster* 5 adalah sebanyak 9 kabupaten/kota, dan *cluster* 7 adalah sebanyak 7 kabupaten/kota.

Daftar Pustaka

- Agustini, L., Sumarno, & Kirana, I. O. (2022). Pengelompokan Data Janjang Panen Kelapa Sawit Menggunakan Algoritma K-Medoids Pada PT SIR MANDAU. *Journal of Machine Learning and Data Analytics (MALDA)*, 01(01), 36–44.
- Andika, T. H. (2024). Pengenalan Pola Berbasis Segmentasi Citra Menggunakan Algoritma Fuzzy C-Means Dan K-Means. *Aisyah Journal of Informatics and Electrical Engineering*, 1, 1–10
- Badan Pusat Statistik Provinsi Sumatera Utara. (2022). *Indeks Pembangunan Manusia Provinsi Sumatera Utara* (A. B. W. S. S. T. MM, Ed.). Badan Pusat Statistik Provinsi Sumatera Utara.
- Deana, R., Suhaedi, D., & Harahap, E. (2022). Konstruksi Sistem Inferensi Fuzzy Menggunakan Subtractive Fuzzy C-Means pada Data Parkinson. *Bandung Conference Series: Mathematics*, 2(1), 51–52. <https://doi.org/10.29313/bcsm.v2i1.1837>
- Firdaus, H. S., Nugraha, A. L., Sasmito, B., Awaluddin, M., & Nanda, C. A. (2021). Perbandingan Metode Fuzzy C-Means dan K-Means Untuk Pemetaan Daerah Rawan Kriminalitas di Kota Semarang. *Elipsoida*, 04(01), 58–64
- Iis*, Yahya, I., Wibawa, G. N. A., Baharuddin, Ruslan, & Laome, L. (2022). Penggunaan Korelasi Cophenetic untuk Pemilihan Metode Cluster Berhierarki pada Mengelompokkan Kabupaten/Kota Berdasarkan Jenis Penyakit di Provinsi Sulawesi Tenggara Tahun 2020. *Prosiding Seminar Nasional Sains dan Terapan (Sinta) VI 2022*, 1.
- Lestari, D., Fadillah, N., & Ihsan, A. (2019). Sistem Deteksi Kualitas Beras Berdasarkan Warna Menggunakan Fuzzy C-Means Clustering Guna Membantu Tingkat Pegetahuan Masyarakat. *InfoTekJar: Jurnal Nasional Informatika Dan Teknologi Jaringan*, 3(2), EDISI MARET. <https://doi.org/https://doi.org/10.30743/infotekjar.v3i2.920>
- Martias, L. D. (2021). Statistika Deskriptif Sebagai Kumpulan Informasi. *FIHRIS: Jurnal Ilmu Perpustakaan Dan Informasi*, 16(1), 40–59. <https://doi.org/10.14421/fhrs.2021.161.40-59>
- Qori'atunnadyah, M. (2023). Metode C-Means untuk Pengelompokkan Kabupaten/Kota Provinsi Jawa Timur berdasarkan Indikator Indeks Pembangunan Manusia (IPM). *Journal of Informatics Development*, 2(2), 51–58. <https://doi.org/10.30741/jid.v2i2.1013>
- Silvi, R. (2018). Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia.