

**Perbandingan Metode C-Means dan Fuzzy C-Means Pada Pengelompokan Kabupaten/Kota Di Kalimantan Berdasarkan Indikator IPM Tahun 2019***Comparison of C-Means and Fuzzy C-Means Methods in the Districts/Cities on the Island of Kalimantan Based on the 2019 HDI Indicators***Mahmudi<sup>1</sup>, Rito Goejantoro<sup>2</sup>, Fidia Deny Tisna Amijaya<sup>3</sup>**<sup>1,2</sup>Laboratorium Statistika Komputasi, FMIPA, Universitas Mulawarman<sup>3</sup>Laboratorium Matematika Komputasi, FMIPA, Universitas MulawarmanE-mail: [Mahmudjaza5@gmail.com](mailto:Mahmudjaza5@gmail.com)**Abstract**

The Human Development Index is an indicator used to measure one important aspect related to the quality of the results of economic development, namely the degree of human development. Data Mining is a technique or process for obtaining information from large database warehouses. Based on its function, one of the data mining tasks was to group data, where the method used in this study was the C-Means and Fuzzy C-Means grouping methods. The two classification methods were applied to the human development index indicator data. The purpose of this study was to determine the best method based on the ratio of the standard deviation in clusters to the standard deviation between clusters. Based on the results of the analysis, it was concluded that the best method is the C-Means method with the value of the standard deviation value in the cluster against the standard deviation between clusters of 0.434 which results in 5 clusters, namely cluster 1 consisting of 9 districts/cities, cluster 2 consisting of 7 districts/cities, cluster 3 consists of 10 districts/cities, cluster 4 consists of 15 districts/cities and cluster 5 consists of 15 districts/cities.

**Keywords:** C-Means, Fuzzy C-Means, Human Development Index, Standard Deviation

**Pendahuluan**

Indeks Pembangunan Manusia (IPM) atau dikenal dengan sebutan *Human Development Index* (HDI) merupakan indikator yang digunakan untuk mengukur salah satu aspek penting yang berkaitan dengan kualitas dari hasil pembangunan ekonomi (Tambunan 2003). Indeks Pembangunan Manusia merupakan tolak ukur keberhasilan pembangunan manusia yang mencakup tiga dimensi kehidupan manusia, yaitu kesehatan (diukur dengan indikator Angka Harapan Hidup), pendidikan (diukur dengan indikator Harapan Lama Sekolah (HLS) dalam tahun dan Rata-rata Lama Sekolah (RLS)), dan kualitas hidup layak (diukur dengan indikator Pengeluaran Per Kapita) (BPS, 2019).

Sejak tahun 2010, pembangunan manusia di Indonesia setiap tahunnya mengalami peningkatan yang cukup signifikan. Rata-rata pertumbuhan IPM di Indonesia adalah 0,89 persen per tahun. Dalam kurun waktu sembilan tahun terjadi kenaikan IPM sebesar 5,39 poin. Perkembangan ini menunjukkan semakin membaiknya pembangunan manusia secara umum di Indonesia. Pada tahun 2019, IPM Indonesia tumbuh 0,74 persen atau meningkat sebesar 0,53 poin dibanding tahun sebelumnya, sehingga mencapai 71,92. Meskipun mengalami pertumbuhan, tetapi capaian ini lebih lambat dibandingkan pertumbuhan tahun sebelumnya

yang mencapai 0,82 persen. Dengan capaian ini, status pembangunan manusia di Indonesia masih berada pada level tinggi (berada pada kisaran antara  $70 \leq \text{IPM} < 80$ ). Posisi ini belum berubah sejak tahun 2016 dan untuk meningkatkan status menjadi sangat tinggi memerlukan upaya yang keras dan dalam waktu yang cukup lama (BPS, 2019).

*Clustering* adalah pengelompokan data yang di dasarkan pada kesamaan karakteristik di antara objek-objek pengamatan yang akan diteliti. *Clustering* data mempertimbangkan sebuah pendekatan penting untuk mencari kesamaan dalam data dan menempatkan data tersebut ke dalam suatu kelompok. Sehingga, objek-objek yang berada dalam *cluster* tersebut akan mempunyai kesamaan karakteristik (Santoso 2015). Terdapat beberapa algoritma pengelompokan yang dapat digunakan di antaranya adalah Algoritma *C-Means* dan *Fuzzy C-Means*.

Metode *C-Means* merupakan metode *clustering* yang paling sederhana dan umum. Hal ini dikarenakan *C-Means* mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien (Alfina 2012). *C-Means* merupakan salah satu metode pengelompokan data yang berusaha mempartisi data yang ada ke dalam suatu kelompok sehingga

tingkat kemiripan antar anggota dalam satu *cluster* sangat tinggi sedangkan tingkat kemiripan dengan anggota pada *cluster* lain sangat rendah di mana keanggotaan sebuah data dinyatakan secara tegas yaitu dengan memberi nilai 1 (menjadi anggota) dan 0 (tidak menjadi anggota) (Prasetyo, 2014).

*Fuzzy C-Means* adalah suatu metode pengelompokan data yang ditentukan oleh derajat keanggotaan.. Metode *Fuzzy C-Means* adalah salah satu metode *clustering* yang mengalokasikan kembali data ke dalam masing-masing kelompok dengan memanfaatkan teori *Fuzzy*. Dalam teori *Fuzzy*, keanggotaan sebuah data tidak diberi nilai secara tegas dengan nilai 1 (menjadi anggota) dan 0 (tidak menjadi anggota), melainkan dengan suatu nilai derajat keanggotaan yang jangkauan nilainya 0 sampai 1 (Prasetyo, 2014). Metode *Fuzzy C-Means* sering digunakan dalam melakukan pengelompokan, karena metode ini memberikan hasil yang halus dan cukup efektif dan memiliki kemampuan untuk mengelompokkan data yang jumlahnya lebih besar dan lebih kokoh terhadap data *outlier* (Kusumadewi, 2006).

Berdasarkan uraian di atas, maka peneliti tertarik untuk melakukan penelitian dengan judul “Perbandingan metode *C-Means* dan *Fuzzy C-Means* pada pengelompokan Kabupaten/Kota di Kalimantan Berdasarkan Indikator IPM” yang bertujuan untuk melihat kemampuan dari setiap algoritma dalam *clustering* sehingga mendapatkan hasil algoritma terbaik.

**Data Mining**

Data mining merupakan bidang dari kumpulan beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, *database* dan visualisasi untuk penanganan masalah pengambilan informasi dari sejumlah *database* yang besar (Larose, 2005). Data mining bertujuan untuk mengambil intisari pengetahuan dari sekumpulan data sehingga didapatkan struktur yang dapat dimengerti manusia serta meliputi basis data dan pengaturan data, prapemrosesan data, pertimbangan model dan inferensi, ukuran ketertarikan, pertimbangan kompleksitas, pascapemrosesan terhadap struktur yang ditemukan, visualisasi dan *online updating* (Suyanto, 2017).

**Analisis Cluster**

Analisis *cluster* adalah analisis yang bertujuan untuk mengelompokkan data atau objek yang di dasarnya hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan di antaranya (Prasetyo, 2012). Terdapat beberapa

*cluster* yang dianggap baik yang di antaranya mempunyai kesamaan (homogenitas) yang tinggi antar anggota dalam suatu kelompok (*within cluster*) dan mempunyai perbedaan (heterogenitas) yang tinggi antara kelompok yang satu dengan kelompok yang lainnya (*between cluster*) (Santoso, 2015). Asumsi dalam analisis kelompok yaitu sampel yang diambil harus mewakili populasi dan tidak terdapat variabel penelitian yang memiliki hubungan linier yang kuat dengan variabel lainnya (nonmultikolinieritas) (Gujarati, 2003).

**C-Means**

*C-Means* merupakan salah satu metode data *clustering* nonhierarki yang memparsi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok lain (Agusta, 2007).

**Fuzzy C-Means**

*Fuzzy C-Means* adalah suatu teknik pengelompokan data di mana keberadaan tiap-tiap data dalam suatu kelompok ditentukan oleh nilai keanggotan. Konsep dasar dalam penggunaan metode *Fuzzy C-Means* adalah menentukan pusat kelompok yang akan menandai lokasi rata-rata untuk tiap-tiap kelompok. *Fuzzy C-Means* adalah suatu metode pengelompokan yang memungkinkan satu bagian data untuk memiliki dua atau lebih kelompok (Kusumadewi, 2006).

**Variabel Penelitian**

Variabel yang digunakan dalam penelitian ini adalah:

- X<sub>1</sub> : Rata-rata Lama Sekolah (Tahun)
- X<sub>2</sub> : Harapan Lama Sekolah (Tahun)
- X<sub>3</sub> : Umur Harapan Hidup (Tahun)
- X<sub>4</sub> : Pengeluaran Per Kapita (Ribu Rupiah/Tahun)

**Metode Penelitian**

Adapun langkah-langkah yang dilakukan dalam analisis data adalah sebagai berikut:

1. Langkah yaitu melakukan analisis statistika deskriptif untuk melihat gambaran umum pada data penelitian.
2. Standarisasi data dengan menggunakan persamaan berikut:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{1}$$

dengan

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \tag{2}$$

$$s_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \quad (3)$$

di mana

- $z_{ij}$  : standarisasi data ke- $i$  pada variabel ke- $j$
- $\bar{x}_j$  : rata-rata dari variabel ke- $j$
- $s_j$  : simpangan baku dari variabel ke- $j$
- $x_{ij}$  : data ke- $i$  dari variabel ke- $j$
- $\ell$  : banyak variabel
- $n$  : banyak data pengamatan

3. Mendeteksi hubungan antar variabel penelitian dengan cara melihat nilai koefisien korelasi berdasarkan persamaan berikut:

$$r = \frac{n(\sum_{i=1}^n x_i y_i) - (\sum_{i=1}^n x_i) \cdot (\sum_{i=1}^n y_i)}{\sqrt{n(\sum_{i=1}^n x_i^2) - (\sum_{i=1}^n x_i)^2} \cdot \sqrt{n(\sum_{i=1}^n y_i^2) - (\sum_{i=1}^n y_i)^2}} \quad (4)$$

dengan  $r$  adalah nilai koefisien korelasi antara variabel  $x$  dan  $y$  dan  $n$  adalah banyaknya data. Jika terdapat koefisien korelasi antar variabel bernilai di atas 0,8 maka perlu dilakukan reduksi variabel, jika tidak maka dilanjutkan ke tahap berikutnya.

4. Melakukan pengelompokan data dengan metode *C-Means*. Berikut adalah tahapan dalam algoritma *C-Means*:
  - a. Menentukan banyak *cluster* ( $k$ ) sesuai dengan keinginan peneliti.
  - b. Pilih data sebanyak  $k$  secara acak yang akan dijadikan sebagai pusat *cluster* awal.
  - c. Menghitung jarak antar data dengan pusat *cluster* yang telah ditentukan sebelumnya menggunakan jarak Euclid dengan persamaan berikut:

$$d(z_i, c_p) = \sqrt{\sum_{j=1}^{\ell} (z_{ij} - c_{pj})^2} \quad (5)$$

di mana

- $d(z_i, c_p)$  : jarak antara data pengamatan ke- $i$  yang telah distandarisasi dengan pusat *cluster* ke- $p$
- $z_{ij}$  : data pengamatan ke- $i$  yang telah distandarisasi pada variabel ke- $j$
- $c_{pj}$  : pusat *cluster* ke- $p$  pada variabel ke- $j$

- d. Menempatkan data ke pusat *cluster* yang ada, dengan melihat jarak terdekat berdasarkan persamaan berikut:

$$u_{ip} = \begin{cases} 1, & d(z_i, c_p) = \min\{d(z_i, c_1), d(z_i, c_2), \dots, d(z_i, c_k)\} \\ 0, & \text{lainnya} \end{cases} \quad (6)$$

- e. Menghitung pusat *cluster* baru dari data pada masing-masing *cluster* berdasarkan persamaan:

$$\bar{c}_{pj} = \frac{1}{M_p} \sum_{i=1}^{M_p} z_{ij} \quad (7)$$

di mana

- $\bar{c}_{pj}$  : rata-rata seluruh anggota *cluster* ke- $p$  dari variabel ke- $j$
- $z_{ij}$  : data pengamatan ke- $i$  yang telah distandarisasi pada variabel ke- $j$
- $M_p$  : banyaknya data dalam *cluster* ke- $p$

- f. Jika nilai pusat *cluster* dan pengalokasian anggota *cluster* terdapat perubahan maka kembali ke langkah c, d dan e sampai pusat *cluster* tidak berubah lagi dan tidak ada anggota yang berpindah ke *cluster* lainnya.

5. Pengelompokan data dengan menggunakan metode *Fuzzy C-Means*. Berikut adalah tahapan menggunakan metode *Fuzzy C-Means*:

- a. Menentukan jumlah kelompok yang akan dibentuk ( $k$ ) dengan syarat ( $k \geq 2$ ), nilai pangkat atau pembobot ( $w$ ) dengan syarat ( $w > 1$ ).
- b. Membangkitkan bilangan acak dengan syarat pada persamaan berikut:

$$\sum_{p=1}^k u_{ip} = 1 \quad (8)$$

di mana  $u_{ip}$  adalah bilangan acak ke- $i$  pada *cluster* ke- $p$

- c. Menghitung pusat *cluster* untuk setiap *cluster* dengan persamaan:

$$c_{pj} = \frac{\sum_{i=1}^n (u_{ip})^w \cdot z_{ij}}{\sum_{i=1}^n (u_{ip})^w} \quad (9)$$

di mana

- $c_{pj}$  : pusat *cluster* ke- $p$  pada variabel ke- $j$
- $u_{ip}$  : bilangan acak ke- $i$  pada *cluster* ke- $p$
- $z_{ij}$  : data pengamatan ke- $i$  yang telah distandarisasi pada variabel ke- $j$
- $w$  : pangkat pembobot

- d. Menghitung nilai derajat keanggotaan setiap data pada setiap *cluster* dengan persamaan berikut:

$$u_{ip} = \left[ \frac{d(z_i, c_p)^{\frac{-2}{w-1}}}{\sum_{p=1}^k d(z_i, c_p)^{\frac{-2}{w-1}}} \right] \quad (10)$$

dengan

$$d(z_i, c_p) = \left[ \sum_{j=1}^{\ell} (z_{ij} - c_{pj})^2 \right]^{\frac{1}{2}} \quad (11)$$

di mana

- $u_{ip}$  : nilai derajat keanggotaan data pengamatan ke- $i$  dari

- $d(z_i, c_p)$  : jarak antara data pengamatan dengan ke- $i$  dengan pusat *cluster* ke- $p$
- $w$  : pangkat pembobot
- $z_{ij}$  : data pengamatan ke- $i$  yang distandarisasikan pada variabel ke- $j$
- $c_{pj}$  : pusat *cluster* ke- $p$  dari variabel ke- $j$
- e. Menempatkan data ke pusat *cluster* yang ada dengan melihat jarak terdekat berdasarkan persamaan berikut:

$$w_{ip} = \begin{cases} 1, & u_{ip} = \max\{u_{i1}, u_{i2}, \dots, u_{in}\} \\ 0, & \text{lainnya} \end{cases} \quad (12)$$

- f. Kembali ke langkah c, d dan e sampai pusat *cluster* tidak berubah lagi dan tidak ada anggota yang berpindah ke *cluster* lainnya.
6. Penentuan metode terbaik dari hasil pengelompokan metode *C-Means* dan *Fuzzy C-Means* dengan membandingkan nilai rasio simpangan baku dalam *cluster* terhadap simpangan baku antar *cluster*. Simpangan baku dalam *cluster* ( $s_w$ ) dinyatakan dalam persamaan berikut:

$$s_w = \frac{1}{k} \sum_{p=1}^k s_p \quad (13)$$

dengan

$$s_p = \sqrt{\frac{\sum_{i=1}^{M_p} (z_{ip} - \bar{z}_p)^2}{M_p - 1}} \quad (14)$$

$$\bar{z}_p = \frac{1}{M_p} \sum_{i=1}^{M_p} z_{ip} \quad (15)$$

di mana

- $s_w$  : simpangan baku di dalam *cluster*
- $s_p$  : simpangan baku di dalam *cluster* ke- $p$
- $z_{ip}$  : data pengamatan ke- $i$  yang telah distandarisasi *cluster* ke- $p$
- $\bar{z}_p$  : rata-rata data pengamatan yang telah distandarisasi *cluster* ke- $p$
- $M_p$  : banyaknya data dalam *cluster* ke- $p$
- Simpangan baku antar *cluster* ( $s_b$ ) dinyatakan dalam persamaan berikut:

$$s_b = \sqrt{\frac{1}{k-1} \sum_{p=1}^k (\bar{z}_p - \bar{z})^2} \quad (16)$$

dengan

$$\bar{z} = \frac{1}{k} \sum_{p=1}^k \bar{z}_p \quad (17)$$

- $s_b$  : simpangan baku antar *cluster*
- $\bar{z}_p$  : nilai rata-rata *cluster* ke- $p$
- $\bar{z}$  : nilai rata-rata dari keseluruhan *cluster*
7. Interpretasi hasil pengelompokan terbaik berdasarkan nilai rasio  $s_w$  terhadap  $s_b$  terkecil.

**Hasil dan Pembahasan**

**1. Statistika Deskriptif**

**Tabel 1.** Statistika deskriptif data indikator IPM Kabupaten/Kota di Pulau Kalimantan tahun 2019

Variabel	Minimum	Maksimum	Rata-rata
X <sub>1</sub>	6	11,510	8,232
X <sub>2</sub>	11,150	14,990	12,670
X <sub>3</sub>	63,580	74,410	70,520
X <sub>4</sub>	7.206	16,843	10,834

Berdasarkan Tabel 1 diperoleh informasi sebagai berikut:

- a. Indikator pendidikan yang terdiri terdiri dari dua variabel yaitu variabel X<sub>1</sub> dan X<sub>2</sub>. Berdasarkan Tabel 1 dapat diketahui bahwa nilai minimum variabel X<sub>1</sub> selama 6 tahun dan nilai maksimum selama 11,510 tahun dengan rata-rata selama 8,232 tahun. Nilai minimum variabel X<sub>2</sub> adalah selama 11,150 tahun dan nilai maksimum selama 14,990 tahun dengan rata-rata selama 12,670 tahun.
- b. Indikator kesehatan yaitu indikator yang terdiri dari variabel X<sub>3</sub>. Berdasarkan Tabel 1 dapat diketahui bahwa nilai minimum variabel X<sub>3</sub> adalah selama 63,580 tahun dan nilai maksimum selama 74,410 tahun dengan rata-rata selama 70,520 tahun.
- c. Indikator standar hidup layak yaitu indikator yang terdiri dari variabel X<sub>4</sub>. Berdasarkan Tabel 1 dapat diketahui bahwa nilai minimum variabel X<sub>4</sub> adalah sebesar 7.206 Ribu Rupiah/Orang/Tahun dan nilai maksimum sebesar 16.843 Ribu Rupiah/Orang/Tahun dengan rata-rata sebesar 10.834 Ribu Rupiah/Orang/Tahun.

**2. Standarisasi data**

Standarisasi data dilakukan agar data memiliki rentang nilai yang tidak jauh berbeda, sehingga data yang lebih besar tidak akan mendominasi data yang lebih kecil sehingga diperoleh hasil yang semakin akurat.

**Tabel 2.** Standarisasi data indikator IPM Kabupaten/Kota di Pulau Kalimantan tahun 2019.

Data	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
1	-1,252	-0,086	-0,694	-0,422
2	-1,391	-0,699	1,296	-0,766
3	-0,925	-0,334	0,897	-1,591
⋮	⋮	⋮	⋮	⋮
56	1,411	1,245	1,399	0,313

**3. Koefisien Korelasi**

**Tabel 3.** Koefisien korelasi data indikator IPM Kabupaten/Kota di Pulau Kalimantan tahun 2019.

Variabel	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>
X <sub>1</sub>	1,000	0,784	0,392	0,730
X <sub>2</sub>	0,784	1,000	0,350	0,600
X <sub>3</sub>	0,392	0,350	1,000	0,080
X <sub>4</sub>	0,730	-0,600	0,080	1,000

Berdasarkan Tabel 3 terlihat bahwa nilai dari koefisien korelasi antar variabel penelitian tertinggi adalah sebesar 0,784 yang bernilai di bawah 0,8 yang artinya tidak terjadi multikolinieritas antar variabel dalam penelitian dan dapat dilanjutkan keproses pengelompokan dengan metode *C-Means* dan *Fuzzy C-Means*.

**4. Pengelompokan metode C-Means**

- a. Langkah awal dari analisis *cluster* adalah menentukan banyaknya *cluster*. Sebenarnya tidak ada aturan dalam menentukan berapa banyaknya *cluster*. Sehingga penelitian ini banyaknya *cluster* yang digunakan dipilih secara acak yaitu sebanyak 5 *cluster* di mana  $k = 2, 3, 4, 5, 6$ .
- b. Menentukan pusat *cluster* awal ( $k = 2$ )  
Pusat *cluster* awal ditentukan secara acak, yaitu data ke-47 dan data ke-51 sebagai berikut:

**Tabel 4.** Pusat *cluster* awal

Pusat <i>Cluster</i>	Variabel			
	$X_1$	$X_2$	$X_3$	$X_4$
$c_1$	-0,280	-0,204	0,568	-1,310
$c_2$	2,041	0,267	1,506	2,786

- c. Menghitung jarak antar data dengan pusat *cluster* awal menggunakan jarak Euclid.

**Tabel 5.** Jarak data pengamatan dengan pusat *cluster* awal menggunakan jarak Euclid

Data	Jarak Euclid	
	<i>Cluster 1</i>	<i>Cluster 2</i>
1	1,828	5,109
2	1,518	5,037
3	0,788	5,356
⋮	⋮	⋮
56	2,879	2,735

- d. Mengalokasikan data ke dalam salah satu pusat *cluster* terdekat.

**Tabel 6.** Alokasi data pengamatan ke pusat *cluster*

Data	Jarak Euclid		<i>Cluster</i> yang diikuti
	<i>Cluster 1</i>	<i>Cluster 2</i>	
1	1,828	5,109	1
2	1,518	5,037	1
3	0,788	5,356	1
⋮	⋮	⋮	⋮
56	2,879	2,735	2

Berdasarkan Tabel 6 diperoleh bahwa *cluster 1* beranggotakan 47 kabupaten/kota. Sedangkan *cluster 2* beranggotakan 9 kabupaten/kota.

- e. Menghitung pusat *cluster* baru berdasarkan hasil alokasi data pengamatan.

**Tabel 7.** Pusat *cluster* baru ( $c^{(1)}$ )

Pusat <i>Cluster</i>	Variabel			
	$X_1$	$X_2$	$X_3$	$X_4$
$c_1$	-0,339	-0,335	-0,199	-0,297
$c_2$	1,772	1,749	1,041	1,549

- f. Berdasarkan hasil pada Tabel 7 dapat dilihat bahwa pusat *cluster* baru memiliki perbedaan nilai dengan pusat *cluster* sebelumnya, maka akan dilanjutkan iterasi ke-2 dan kembali ke langkah c, d dan e sampai pusat *cluster* tidak berubah lagi dan tidak ada anggota yang berpindah ke *cluster* lainnya.
- Menghitung jarak data pengamatan dengan pusat *cluster* yang baru ( $c^{(1)}$ ) iterasi ke-2.

**Tabel 8.** Jarak data pengamatan dengan pusat *cluster* ( $c^{(1)}$ ) baru menggunakan jarak Euclid

Data	Jarak Euclid	
	<i>Cluster 1</i>	<i>Cluster 2</i>
1	1,075	4,406
2	1,922	4,629
3	1,794	4,636
⋮	⋮	⋮
56	2,913	1,428

- Mengalokasikan data ke dalam pusat *cluster* baru ( $c^{(1)}$ ) iterasi ke-2

**Tabel 9.** Alokasi data pengamatan ke pusat *cluster* baru iterasi ke-2

Data	Jarak Euclid		<i>Cluster</i> yang diikuti
	<i>Cluster 1</i>	<i>Cluster 2</i>	
1	1,075	4,406	1
2	1,922	4,629	1
3	1,794	4,636	1
⋮	⋮	⋮	⋮
56	2,913	1,428	2

Berdasarkan Tabel 9 diperoleh bahwa *cluster 1* beranggotakan 46 kabupaten/kota. Sedangkan *cluster 2* beranggotakan 10 kabupaten/kota.

- Melakukan perhitungan pusat *cluster* baru iterasi kedua  $c^{(2)}$

**Tabel 10.** Pusat *cluster* baru ( $c^{(2)}$ )

Pusat <i>Cluster</i>	Variabel			
	$X_1$	$X_2$	$X_3$	$X_4$
$c_1$	-0,362	-0,365	-0,219	-0,306
$c_2$	1,666	1,681	1,007	1,409

- Berdasarkan hasil pada Tabel 10 dapat dilihat bahwa pusat *cluster* baru iterasi ke-2 memiliki perbedaan nilai dengan pusat *cluster* sebelumnya, maka akan dilanjutkan iterasi ke-3.
- Menghitung jarak data pengamatan dengan pusat *cluster* yang baru ( $c^{(2)}$ ) iterasi ke-3

**Tabel 11.** Jarak data pengamatan dengan pusat *cluster* ( $c^{(2)}$ ) baru menggunakan jarak Euclid

Data	Jarak Euclid	
	Cluster 1	Cluster 2
1	1,075	4,406
2	1,922	4,629
3	1,794	4,636
⋮	⋮	⋮
56	2,913	1,428

Setelah dilakukan perhitungan jarak data pengamatan dengan pusat *cluster* baru ( $c^{(2)}$ ), nilai jarak Euclid yang didapatkan tidak terdapat perbedaan, sehingga hasil alokasi data pengamatan akan sama seperti Tabel 9 dan akan menghasilkan pusat *cluster* yang sama seperti Tabel 10 di mana hasil anggota *cluster* tidak berpindah ke *cluster* yang lain dan memiliki nilai pusat *cluster* yang sama.

**Tabel 12.** Hasil pengelompokan *C-Means* untuk  $k=2$

Cluster	Anggota Cluster (Kabupaten/Kota)	Jumlah Anggota
1	Sambas, Bengkayang, Landak, Mempawah, Sanggau, Ketapang, Sintang, Kapuas Hulu, Sekadau, Melawi, Kayong Utara, Kubu Raya, Kota Singkawang, Kotawaringin Barat, Kotawaringin Timur, Kapuas, Barito Selatan, Barito Utara, Sukamaru, Lamandau, Seruyan, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Tanah Laut, Kota Baru, Banjar, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Paser, Kutai Barat, Kutai Timur, Penajam Paser Utara, Mahakam Ulu, Malinau, Bulungan, Tana Tidung dan Nunukan	46
2	Kota Pontianak, Kota Palangkaraya, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang dan Kota Tarakan	10

**5. Pengelompokan Metode Fuzzy C-Means**

a. Menentukan jumlah *cluster* yang akan dibentuk dan menentukan nilai pangkat pembobot ( $w$ ) dengan syarat ( $w > 1$ ), di mana dalam penentuan jumlah *cluster* tidak ada aturan yang baku untuk penentuan nilai tersebut. Sehingga dalam penelitian ini banyaknya *cluster* yang digunakan dipilih secara acak yaitu sebanyak 5 *cluster* di mana  $k = 2, 3, 4, 5, 6$ . Berdasarkan penelitian Klawon dan Keller nilai dari  $w$  yang paling optimal dan sering digunakan adalah 2, sehingga dalam penelitian ini pangkat pembobot ( $w$ ) yang digunakan adalah 2.

b. Membangkitkan bilangan acak ( $u_{ip}$ ) dengan syarat yang tertulis pada Persamaan (2.12) dari  $k = 2$ . Dimana dalam penelitian ini, penentuan bilangan acak ( $u_{ip}$ ) dibangkitkan dengan menggunakan *software R*.

**Tabel 13.** Bilangan Acak ( $u_{ip}$ )

Data Pengamatan	Bilangan Acak	
	1	2
1	0,929	0,071
2	0,808	0,192
3	0,825	0,175
⋮	⋮	⋮
56	0,090	0,910

c. Menghitung pusat *cluster* ( $c$ ) berdasarkan bilangan acak yang telah didapatkan.

**Tabel 14.** Pusat Cluster ( $c$ )

Pusat Cluster	Variabel			
	$X_1$	$X_2$	$X_3$	$X_4$
$c_1$	-0,449	-0,451	-0,253	-0,353
$c_2$	1,292	1,307	0,777	1,047

d. Memperbaiki derajat keanggotaan ( $u_{ip}^{(1)}$ ) setiap data pada setiap kelompok

**Tabel 15.** Hasil Perhitungan Derajat Keanggotaan Data Pada Setiap Kelompok

Data Pengamatan	Bilangan Acak	
	1	2
1	0,929	0,071
2	0,808	0,192
3	0,825	0,175
⋮	⋮	⋮
56	0,090	0,910

e. Mengalokasikan data ke dalam salah satu *cluster* dengan melihat nilai derajat keanggotaan

**Tabel 16.** Hasil Pengalokasian Data Berdasarkan Nilai Derajat Keanggotaan

Data	Derajat Keanggotaan		Cluster yang diikuti
	Cluster 1	Cluster 2	
1	0,929	0,071	1
2	0,808	0,192	1
3	0,825	0,175	1
⋮	⋮	⋮	⋮
56	0,090	0,910	2

f. Kembali ke langkah c, d dan e sampai pusat *cluster* tidak berubah lagi dan tidak ada anggota yang berpindah ke *cluster* lainnya.

- Menghitung pusat *cluster* ( $c$ ) berdasarkan hasil derajat keanggotaan yang telah diperbaiki yang telah didapatkan.

**Tabel 17.** Pusat Cluster ( $c^1$ )

Pusat Cluster	Variabel			
	$X_1$	$X_2$	$X_3$	$X_4$
$c_1$	-0,449	-0,451	-0,253	-0,353
$c_2$	1,292	1,307	0,777	1,047

Setelah dilakukan perhitungan dapat dilihat bahwa pusat *cluster* tidak menunjukkan adanya perubahan nilai atau dengan kata lain nilai pusat *cluster* awal sama dengan nilai pusat *cluster* baru. Sehingga hasil pengelompokan Kabupaten/Kota menggunakan metode *Fuzzy C-Means* dapat dihentikan di mana hasil nilai pusat *cluster* baru sama dengan nilai pusat *cluster* awal.

**Tabel 18.** Hasil Pengelompokan *Fuzzy C-Means* Untuk  $k = 2$

Cluster	Anggota Cluster (Kabupaten/Kota)	Jumlah Anggota
1	Sambas, Bengkayang, Landak, Mempawah, Sanggau, Ketapang, Sintang, Kapuas Hulu, Sekadau, Melawi, Kayong Utara, Kubu Raya, Kota Singkawang, Kotawaringin Barat, Kotawaringin Timur, Kapuas, Barito Selatan, Barito Utara, Sukamaru, Lamandau, Seruyan, Katingan, Pulang Pisau, Gunung Mas, Barito Timur, Murung Raya, Tanah Laut, Kota Baru, Banjar, Barito Kuala, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu, Balangan, Paser, Kutai Barat, Penajam Paser Utara, Mahakam Ulu, Bulungan, Tana Tidung dan Nunukan	44
2	Kota Pontianak, Kota Palangkaraya, Kota Banjarmasin, Kota Banjar Baru, Kutai Kartanegara, Kutai Timur, Berau, Kota Balikpapan, Kota Samarinda, Kota Bontang, Malinau dan Kota Tarakan	12

**6. Penentuan Metode Terbaik**

a. Metode *C-Means*

Perhitungan *rasio* simpangan baku dalam kelompok ( $s_w$ ) dan simpangan baku antar kelompok ( $s_b$ ) metode *C-Means* menggunakan  $k = 2, 3, 4, 5, 6$ .

**Tabel 19.** Hasil Perhitungan Nilai  $s_w/s_b$  Untuk  $k = 2, 3, 4, 5, 6$

Jumlah Cluster (k)	$X_1$	$X_2$	$X_3$	$X_4$	Mean
2	0,439	0,498	0,836	0,637	0,603
3	0,530	0,564	0,602	0,541	0,559
4	0,397	0,518	0,624	0,340	0,470
5	<b>0,427</b>	<b>0,509</b>	<b>0,395</b>	<b>0,404</b>	<b>0,434</b>
6	0,369	0,615	0,401	0,412	0,449

Berdasarkan nilai *rasio* simpangan baku menunjukkan bahwa metode *C-Means* dengan jumlah  $k = 5$  memiliki hasil *cluster* yang lebih baik dibandingkan dengan hasil *cluster* yg lainnya. Hal ini dapat dilihat dari nilai *rasio*  $s_w$  terhadap  $s_b$  pada jumlah *cluster* 5 lebih kecil dibandingkan hasil *cluster* yang lain, dengan anggota *cluster* 1 sebanyak 9 Kabupaten/Kota, anggota *cluster* 2 sebanyak

7 Kabupaten/Kota, anggota *cluster* 3 sebanyak 10 Kabupaten/Kota, anggota *cluster* 4 sebanyak 15 Kabupaten/Kota dan anggota *cluster* 5 sebanyak 15 Kabupaten/Kota.

b. Metode *Fuzzy C-Means*

Perhitungan *rasio* simpangan baku dalam kelompok ( $s_w$ ) dan simpangan baku antar kelompok ( $s_b$ ) metode *Fuzzy C-Means* menggunakan  $k = 2, 3, 4, 5, 6$ .

**Tabel 20.** Hasil Perhitungan Nilai  $s_w/s_b$  Untuk  $k = 2, 3, 4, 5, 6$

Jumlah Cluster (k)	$X_1$	$X_2$	$X_3$	$X_4$	Mean
2	0,472	0,562	0,836	0,813	0,671
3	0,510	0,562	0,737	0,464	0,568
4	0,384	0,540	0,590	0,413	0,482
5	<b>0,452</b>	<b>0,518</b>	<b>0,402</b>	<b>0,402</b>	<b>0,444</b>
6	0,434	0,508	0,414	0,428	0,446

Berdasarkan nilai *rasio* simpangan baku menunjukkan bahwa metode *Fuzzy C-Means* dengan jumlah  $k = 5$  memiliki hasil *cluster* yang lebih baik dibandingkan dengan hasil *cluster* yg lainnya. Hal ini dapat dilihat dari nilai *rasio*  $s_w$  terhadap  $s_b$  pada jumlah *cluster* 5 lebih kecil dibandingkan dengan hasil *cluster* yang lain, dengan anggota *cluster* 1 sebanyak 9 Kabupaten/Kota, anggota *cluster* 2 sebanyak 7 Kabupaten/Kota, anggota *cluster* 3 sebanyak 10 Kabupaten/Kota, anggota *cluster* 4 sebanyak 15 Kabupaten/Kota dan anggota *cluster* 5 sebanyak 15 Kabupaten/Kota.

**7. Interpretasi**

Berdasarkan hasil yang telah didapatkan maka dapat disimpulkan untuk kedua metode tersebut jumlah *cluster* yang paling tepat digunakan yaitu sebanyak 5 *cluster*. Dengan nilai *rasio*  $s_w$  terhadap  $s_b$  metode *C-Means* lebih kecil dibandingkan dengan metode *Fuzzy C-Means*, hal tersebut menandakan bahwa metode *C-Means* lebih tepat digunakan pada pengelompokan Kabupaten/Kota di Kalimantan berdasarkan indikator IPM tahun 2019 dengan jumlah *cluster*  $k = 5$ .

**Tabel 21.** Hasil Pengelompokan Terbaik Menggunakan Metode *C-Means* Dengan Jumlah  $k = 5$

Cluster	Anggota Cluster (Kabupaten/Kota)	Banyak Anggota
1	Barito Selatan, Katingan, Pulang Pisau, Banjar, Barito Kuala, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara dan Balangan	9
2	Kota Pontianak, Kota Palangkaraya, Kota Banjarmasin, Kota Banjar Baru, Kota Balikpapan, Kota Samarinda dan Kota Bontang	7
3	Kota Singkawang, Barito Utara, Paser, Kutai Barat, Kutai Kartanegara, Kutai	10

Timur, Berau, Malinau, Bulungan dan Kota Tarakan

**Tabel 21.** Hasil Pengelompokan Terbaik Menggunakan Metode *C-Means* Dengan Jumlah  $k = 5$  (Lanjutan)

Cluster	Anggota Cluster (Kabupaten/Kota)	Banyak Anggota
4	Bengkayang, Landak, Mempawah, Sanggau, Ketapang, Sintang, Kapuas Hulu, Sekadai, Melawi, Kayong Utara, Kubu Raya, Sukamara, Mahakam Ulu, Tana Tidung dan Nunukan	15
5	Sambas, Kotawaringin Barat, Kotawaringin Timur, Kapuas, Lamandau, Seruyan, Gunung Mas, Barito Timur, Murung Raya, Tanah Laut, Kota Baru, Tapin, Tabalong, Tanah Bumbu dan Penajam Paser Utara	15

**Kesimpulan**

1. Pengelompokan metode *Fuzzy C-Means* dengan menggunakan jumlah  $k = 2$  sampai dengan  $k = 6$  berdasarkan nilai rasio simpangan baku menunjukkan bahwa metode *Fuzzy C-Means* dengan jumlah  $k = 5$  memiliki hasil pengelompokan yang lebih tepat digunakan dibandingkan dengan jumlah *cluster* yg lainnya dengan *cluster* 1 beranggotakan 9 Kabupaten/Kota, *cluster* 2 beranggotakan 7 Kabupaten/Kota, *cluster* 3 beranggotakan 10 Kabupaten/Kota, *cluster* 4 beranggotakan 15 Kabupaten/Kota dan *cluster* 5 beranggotakan 15 Kabupaten/Kota.
2. Pengelompokan metode *C-Means* dengan menggunakan jumlah  $k = 2$  sampai dengan  $k = 6$  berdasarkan nilai rasio simpangan baku menunjukkan bahwa metode *C-Means* dengan jumlah  $k = 5$  memiliki hasil pengelompokan yang lebih tepat digunakan dibandingkan dengan jumlah *cluster* yg lainnya dengan *cluster* 1 beranggotakan 9 Kabupaten/Kota, *cluster* 2 beranggotakan 7 Kabupaten/Kota, *cluster* 3 beranggotakan 10 Kabupaten/Kota, *cluster* 4 beranggotakan 15 Kabupaten/Kota dan *cluster* 5 beranggotakan 15 Kabupaten/Kota.
3. Berdasarkan perhitungan nilai rasio  $s_w$  terhadap  $s_b$ , metode *C-Means* adalah sebesar 0,434 sedangkan nilai rasio  $s_w$  terhadap  $s_b$  metode *Fuzzy C-Means* adalah sebesar 0,444. Dari hasil nilai tersebut dapat disimpulkan bahwa metode pengelompokan yang paling tepat digunakan di antara

metode *C-Means* dan *Fuzzy C-Means* pada pengelompokan Kabupaten/Kota di Kalimantan berdasarkan indikator IPM tahun 2019 adalah metode *C-Means* yang menghasilkan 5 *cluster*.

**Daftar Pustaka**

Agusta, Y. 2007. *C-Means – Penerapan, Permasalahan dan Metode Terkait. Jurnal Sistem dan Informatika*, 47-60.

Alfina, T., Santosa, B., & Barakbah, A. R. 2012. Analisa Perbandingan Metode Hierarchical Clustering, *C-Means*, dan Gabungan Keduanya dalam Cluster Data. *Jurnal Teknik ITS*, 521-525.

Badan Pusat Statistik Kalimantan Timur. 2019. *Kalimantan Timur Dalam Angka 2019*. Samarinda: Badan Pusat Statistika Provinsi Kalimantan Timur.

Gujarati, D. 2003. *Ekonometrika Dasar*. Jakarta: Erlangga.

Kusumadewi, S., Hartati, Sri., Harjoko, Agus., dan Wardoyo, Retantyo. 2006. *Fuzzy Multi-Attribute Decision Making (FUZZY MADM)*. Yogyakarta: Graha Ilmu.

Larose, D.T. 2005. *Discovering Knowledge in Data*. New York: John Wiley and Sons.

Prasetyo, E. 2012. *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.

Prasetyo, E. 2014. *Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Andi Offset.

Santoso, S. 2015. *Menguasai Statistik Multivariat*. Jakarta: PT Elex Media Komputindo.

Suyanto. 2017. *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika.

Tambunan, T. H. 2003. *Perekonomian Indonesia*. Jakarta: Ghalia.