

## PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI JAWA TENGAH BERDASARKAN INDIKATOR KESEHATAN BAYI DAN BALITA MENGUNAKAN ALGORITMA *FUZZY C-MEANS* DAN *K-MEDOIDS*

Risa Nur'aini<sup>1</sup>, Tatik Widiharih<sup>2</sup>, Bagus Arya Saputra<sup>3</sup>

<sup>1,2,3</sup>Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro

\*e-mail: [risanuraini80@gmail.com](mailto:risanuraini80@gmail.com)

DOI: 10.14710/j.gauss.13.1.189-198

### Article Info:

Received: 2023-05-26

Accepted: 2024-10-10

Available Online: 2024-10-14

### Keywords:

*Clustering; Fuzzy C-Means; K-Medoids; Silhouette ; Standar deviation ratio; Health*

**Abstract:** The infant and toddler mortality rate is an indicator of health in Indonesia. If the health level of infants and toddlers in an area is low, the health status in that area is low. Clustering of districts based on infant and toddler health indicators can be used as a guideline for the government in allocating funds and determining health services program. Clustering is a statistical data processing technique that is useful for grouping an area into a klaster based on certain characteristics. This clustering of infants and toddlers uses *Fuzzy C-Means* and *K-Medoids* algorithms. The optimal cluster of the two methods is selected using silhouette validation, while for selection of the best method for profiling using standard deviation ratio values. The optimal number of clusters in the *Fuzzy C-Means* and *K-Medoids* algorithms based on silhouette validation for each method is 2 clusters. Based on the value of the standard deviation ratio of each method, the value of *Fuzzy C-Means* (1.1062) is smaller than *K-Medoids* (1,1771), so the 2-klaster of *Fuzzy C-Means* method is used for the profiling step. Cluster 1 is a group of regions with infant and toddler mortality rates, malnutrition and low birth weight babies, so there must be an increase in the complete basic immunization program and health services. Cluster 2 is a regional group with very good health services, so it is better to maintain the quality of these health services.

## 1. PENDAHULUAN

Kondisi angka kematian bayi (0-12 bulan) dan balita (12-59 bulan) per 1000 kelahiran hidup dalam kurun waktu satu tahun menggambarkan tingkat permasalahan kesehatan masyarakat. Angka tersebut berkaitan dengan faktor penyebab kematian, tingkat pelayanan kesehatan, status gizi ibu hamil dan menyusui, status gizi anak, tingkat keberhasilan program KIA dan KB, serta kondisi lingkungan, sosial dan ekonomi. Tingkat kesehatan bayi dan balita menjadi salah satu indikator kesehatan di suatu wilayah (Dinas Kesehatan Jawa Tengah, 2021). Pengelompokan Kabupaten/kota berdasarkan indikator kesehatan bayi dan balita dapat dijadikan pedoman bagi pemerintah dalam hal alokasi dana dan peningkatan program pelayanan kesehatan khususnya terhadap bayi dan balita. Indikator tersebut meliputi angka kematian bayi dan balita, persentase bayi yang mendapatkan imunisasi dasar lengkap, persentase bayi dan balita yang mendapat pelayanan kesehatan, persentase balita kurang gizi dan persentase bayi berat lahir rendah.

Pengelompokan atau klasterisasi merupakan teknik pengolahan data secara statistika yang berguna untuk mengelompokan suatu wilayah dalam suatu klaster berdasarkan karakteristik tertentu. Pengelompokan kesehatan bayi dan balita ini menggunakan algoritma *Fuzzy C-Means* dan *K-Medoids* (*partitioning around medoids*) karena kedua algoritma tersebut memiliki kinerja yang lebih baik ditinjau dari validitas dibandingkan metode klaster konvensional Menurut Bezdek *et al* (1981), *Fuzzy C-Means* adalah metode klasterisasi di mana setiap data yang masuk ke dalam klaster ditentukan oleh derajat keanggotaannya.

Prinsip dasar dari *Fuzzy C-Means* ialah menentukan pusat kluster yang akan menandai lokasi rata-rata untuk masing-masing kluster. *K-Medoids* menyempurnakan algoritma *K-Means* dengan menggunakan nilai *medoids* sebagai pusat kluster. *Medoids* merupakan nilai yang letaknya terpusat pada suatu kluster, *medoids* dapat mewakili objek di suatu kluster. (Arora *et al.*, 2006).

Kluster yang terbentuk selanjutnya dilakukan validasi untuk memilih banyak kluster yang optimal. Validasi kluster dapat dilakukan dengan berbagai cara, salah satunya validasi kluster internal menggunakan *silhouette*. Menurut Rousseeuw (1987), *silhouette* dibangun berdasarkan kedekatan rata-rata pada setiap nilai di skala rasio. Jumlah kluster terbaik dipilih berdasarkan nilai *silhouette* yang mendekati 1. Tahap selanjutnya ialah pemilihan metode yang akan dipakai dalam profilisasi. Pemilihan metode pengelompokan terbaik menggunakan nilai rasio simpangan baku. Metode kluster yang dipilih adalah kluster optimal yang memiliki nilai rata-rata simpangan baku dalam kluster kecil tetapi nilai simpangan baku antar kluster besar.

## 2. TINJAUAN PUSTAKA

*Outlier* menggambarkan data atau nilai yang terletak sangat jauh atau ekstrim dari data yang lain, *outlier* dalam data multivariat adalah data ekstrim dalam dua atau lebih variabel. Pemeriksaan *outlier multivariate* dapat dilakukan dengan jarak mahalanobis ( $d^2$ ) yang berdistribusi *chi square* ( $\chi^2$ ) dengan derajat bebas (df) sebanyak  $p$  variabel. Jarak mahalanobis didefinisikan dengan persamaan berikut:

$$d^2(i) = (x_{ip} - \bar{x}_p)^T \Sigma^{-1} (x_{ip} - \bar{x}_p) \quad (1)$$

dengan:

$d^2(i)$  = jarak mahalanobis objek ke- $i$  dengan pusat data;  $x_i$  = objek pengamatan ke- $i$  untuk setiap variabel ke- $p$ ;  $\bar{x}_p$  = rata-rata dari tiap variabel ke- $p$ ;  $\Sigma$  = matriks kovarian dari variabel  $p$ .

Analisis kluster menggunakan ukuran jarak akan bermasalah apabila variabel yang digunakan memiliki skala yang berbeda-beda. Cara yang paling sering digunakan untuk standarisasi variabel adalah menggunakan *Z-score*. Proses ini mengubah data menjadi nilai standar, dengan rata-rata sebesar 0 dan simpangan baku sebesar 1 (Hair *et al.*, 2006)

Patro *et al* (2015) menyatakan bahwa formula *Z-score* sebagai berikut:

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (2)$$

dengan :

$Z_{ij}$  = objek ke- $i$  pada variabel ke- $j$  yang telah distandarkan;  $x_{ij}$  = objek ke- $i$  pada variabel ke- $j$ ;  $\bar{x}_j$  = nilai rata-rata variabel ke- $j$ ;  $s_j$  = simpangan baku variabel ke- $j$

### Analisis Kluster

Analisis kluster merupakan bagian dari teknik multivariat yang tujuan utamanya untuk mengelompokkan objek berdasarkan karakteristiknya. Analisis kluster mengelompokkan  $n$  objek berdasarkan  $p$  variabel yang relatif memiliki kemiripan karakteristik. Analisis kluster tidak terlalu memperhitungkan asumsi dasar statistik tetap memperhatikan asumsi sampel representatif dan multikolinearitas (Hair *et al.*, 2006).

#### a. Sampel Representatif

Sampel Representatif adalah sampel yang digunakan dapat mewakili populasi. Pengujian sampel representatif ini dilakukan dengan melihat syarat kecukupan sampel. Kecukupan

sampel diidentifikasi dengan nilai *Kaiser Meyer Olkin* (KMO). Sekelompok data dapat memenuhi syarat kecukupan sampel jika nilai KMO lebih besar dari 0,5 (Widarjono, 2015). Nilai KMO dapat dihitung menggunakan rumus berikut ini:

$$KMO = \frac{\sum_{j=1}^p \sum_{l=1, l \neq j}^p r_{Z_j Z_l}^2}{\sum_{j=1}^p \sum_{l=1, l \neq j}^p r_{Z_j Z_l}^2 + \sum_{j=1}^p \sum_{l=1, l \neq j}^p \rho_{Z_j Z_l, Z_m}^2} \quad (3)$$

dengan:

KMO = nilai Kaiser-Mayer-Olkin;  $p$  = banyak variabel data;  $n$  = banyaknya objek ;  $Z_{ij}$  = objek pengamatan ke- $i$  pada variabel ke- $j$  yang telah distandarisasi;  $Z_{il}$  = objek pengamatan ke- $i$  pada variabel ke- $l$  yang telah distandarisasi;  $Z_j$  = variabel ke- $j$  yang telah distandarisasi;  $Z_l$  = variabel ke- $l$  yang telah distandarisasi;  $\bar{Z}_j$  = rata-rata variabel ke- $j$  yang telah distandarisasi;  $\bar{Z}_l$  = rata-rata variabel ke- $l$  yang telah distandarisasi;  $r_{Z_j Z_l}$  = korelasi antara variabel  $Z_j$  dan  $Z_l$  ;  $\rho_{Z_j Z_l}$  = korelasi parsial antara variabel  $Z_j$  dan  $Z_l$  dengan menjaga agar  $Z_m$  tetap konstan.

#### b. Multikolinearitas

Uji multikolinearitas bertujuan untuk menguji adanya korelasi antar variabel independent dalam sebuah model regresi. Analisis kluster mengharapkan untuk tidak terjadi multikolinearitas. Salah satu kriteria untuk melihat adanya multikolinearitas ialah nilai VIF  $\geq 10$  (Ghozali, 2006). Nilai VIF dapat dicari menggunakan formula berikut:

$$VIF_j = \frac{1}{1 - R_j^2} \quad (4)$$

dengan:

$VIF_j$  = nilai VIF untuk variabel ke- $j$ ,  $j = 1, 2, \dots, p$ ;  $R_j^2$  = koefisien determinasi antara  $X_j$  dengan variabel bebas lainnya pada persamaan model regresi.

Apabila terjadi multikolinearitas antar variabel dapat dilakukan *Principal Component Analysis*. PCA merupakan teknik kombinasi linier yang mengurangi dimensi sekumpulan variabel yang mungkin berkorelasi menjadi kumpulan variabel baru yang tidak berkorelasi. Variabel baru yang tidak berkorelasi disebut *principal component* (PC). Kriteria pemilihan banyak PC didasarkan nilai eigen setiap variabel lebih harus dari 1 dan PC ke- $p$  yang dipilih telah mampu menjelaskan varians data dengan proporsi kumulatif varians sebesar 80%.

#### **Fuzzy C-Means Clustering**

Menurut Bezdek (1981), *Fuzzy C-Means* adalah metode pengelompokan dimana setiap data yang mengikuti aturan himpunan *fuzzy*. Setiap anggota dalam suatu kluster ditentukan oleh derajat keanggotaannya. Derajat keanggotaan *fuzzy* bernilai antara 0 hingga 1. Algoritma *Fuzzy C-Means clustering* ialah sebagai berikut:

- 1) Input data yang akan dikluster, berupa matriks berukuran  $n \times p$
- 2) Langkah selanjutnya ialah menentukan beberapa input yang dibutuhkan dalam perhitungan *Fuzzy C-Means*, yaitu:
  - a. Jumlah kluster ( $c$ )
  - b. Pangkat ( $w$ )
  - c. Maksimum iterasi (MaxIter)
  - d. Error terkecil ( $\epsilon > 0$ )
  - e. Fungsi objektif awal ( $P_0 = 0$ )
- 3) Membangkitkan bilangan random ( $\mu_{ik}^0$ ) sebagai elemen-elemen matriks partisi awal  $U$ . Jumlah setiap baris dalam  $\mu_{ik}^0$  sama dengan 1 (Khang et al., 2020).

4) Hitung pusat kluster ke-k ( $V_{kj}$ ) pada iterasi ke-1

$$V_{kj}^1 = \frac{\sum_{i=1}^n ((\mu_{ik}^0)^w C_{ij})}{\sum_{i=1}^n (\mu_{ik}^0)^w}, \forall k, \forall j \quad (16)$$

5) Menghitung fungsi objektif ( $P_t$ ) pada iterasi ke-1

$$P_1 = \sum_{i=1}^n \sum_{k=1}^c \left( \left[ \sum_{j=1}^p (C_{ij} - V_{kj}^1)^2 \right] (\mu_{ik}^0)^w \right) \quad (17)$$

6) Hitung perubahan derajat keanggotaan ( $\mu_{ik}$ ) pada iterasi ke-1

$$\mu_{ik}^1 = \frac{\left[ \sum_{j=1}^p (C_{ij} - V_{kj}^1)^2 \right]^{\frac{1}{w-1}}}{\sum_{k=1}^c \left[ \sum_{j=1}^p (C_{ij} - V_{kj}^1)^2 \right]^{\frac{1}{w-1}}} \forall i, \forall k \quad (18)$$

7) Anggota setiap kluster ditentukan dari derajat keanggotaan pada iterasi ke-1

$$a_{ik}^1 = \begin{cases} 1, & \mu_{ik}^1 = \max \{ \mu_{1k}^1, \mu_{2k}^1, \dots, \mu_{nk}^1 \} \\ 0, & \text{lainnya} \end{cases} \quad (19)$$

$a_{ik}$  adalah anggota kluster pada objek ke- $i$  kluster ke- $k$ . Keberadaan data dalam suatu kluster di *Fuzzy C-Means* ditentukan oleh nilai keanggotaan yang mendekati 1 atau yang paling besar.

Proses iterasi diulang untuk iterasi ke-2,3,4,...,  $t$  sampai mencapai kondisi ( $P_t - P_{t-1} < \epsilon$ ) atau ( $t > \text{MaxIter}$ ) yang mana ( $P_t - P_{t-1} < \epsilon$ ) ialah selisih pusat kluster kurang dari nilai *error* yang diharapkan atau jika  $t$  (jumlah iterasi) sudah lebih besar dari iterasi maksimum.

### ***K-Medoids Clustering***

*K-Medoids* menyempurnakan algoritma *K-Means* dengan menggunakan nilai *medoids* sebagai pusat kluster (Arora *et al.*, 2006). *K-Means* merupakan proses pengelompokan data menggunakan nilai *mean* sebagai pusat kluster. *K-Medoids* lebih baik dibanding *K-Means* dalam aspek waktu eksekusi, sensitivitas terhadap pencilan dan *noise* (Arora *et al.*, 2006). Algoritma *K-Medoids* ialah sebagai berikut:

- 1) Input data yang akan dikluster, berupa matriks berukuran  $n \times p$  ( $n$ = jumlah sampel data,  $p$  = atribut setiap data).
- 2) Memilih *medoids* awal ( $M_k^1$ ) secara random sebagai pusat kluster sebanyak  $c$  dengan ( $k = 2, 3, \dots, c$ ) pada iterasi ke-1.
- 3) Menghitung jarak terdekat antara objek dengan pusat kluster (*medoids*) menggunakan perhitungan jarak euclidean pada iterasi ke-1.

$$d_{euc}(C_i, c_k^1) = \sqrt{\sum_{j=1}^p (C_{ij} - c_{kj}^1)^2} \quad (20)$$

- 4) Menentukan anggota kluster berdasarkan jarak terdekat antara objek dengan *medoids* ( $s_i^1$ ) yang diambil dari nilai terkecil pada iterasi ke-1.

$$s_i^1 = \min(d_{euc}\{C_i, c_k^1\}) \quad (21)$$

$s_i^1$  adalah jarak terdekat antara objek ke- $i$  dengan *medoids* ke- $k$  pada iterasi ke-1

- 5) Menghitung total ( $s_i^1$ ).

$$S_1 = \sum_{i=1}^n s_i^1 \quad (22)$$

- 6) Memilih objek non *medoids* pada masing-masing kluster secara acak sebagai *medoids* baru untuk iterasi ke-2.

- 7) Ulangi langkah ke-3, 4 dan 5 untuk iterasi ke-2.  
 8) Hitung  $S_{total\ jarak}$  merupakan simpangan antara total jarak objek iterasi baru dengan iterasi sebelumnya.

$$S_{total\ jarak} = S_{(t+1)} - S_t \quad (23)$$

dengan :

$S_t$  = jumlahan jarak terdekat objek pada iterasi ke- $t$

$S_{(t+1)}$  = jumlahan jarak terdekat objek pada iterasi ke- $t + 1$

Iterasi akan dihentikan jika nilai  $S_{total\ jarak} \geq 0$  dan apabila nilai  $S_{total\ jarak} < 0$ , maka iterasi dilanjutkan dengan mengulang langkah ke-6 yaitu memilih *medoids* baru (Park dan Jun, 2009).

### Validasi *Silhouette*

Salah satu bentuk validasi kluster kriteria internal ialah validasi menggunakan *silhouette*. Menurut Rousseeuw (1987), *silhouette* dibangun berdasarkan kedekatan rata-rata pada setiap nilai di skala rasio seperti pada kasus jarak *euclidean*. Rentang nilai *silhouette* dirumuskan sebagai berikut:

$$s_i = \frac{b_i - a_i}{\max(b_i, a_i)} \quad (24)$$

dengan :

$s_i$  = nilai koefisien *silhouette* objek ke- $i$  dengan  $i=1,2,\dots, n$

$a_i$  = rata-rata jarak antara objek ke- $i$  dengan objek lainnya dalam satu kluster

$b_i$  = minimum rata-rata jarak anatar objek ke- $i$  dengan objek lainnya di masing-masing kluster.

Rentang objek secara keseluruhan dapat dirumuskan sebagai berikut:

$$SC = \frac{\sum_{i=1}^n s_i}{n} \quad (25)$$

dengan: SC = rentang nilai dari koefisien *silhouette*;  $n$  = banyaknya objek

Kluster yang memiliki struktur hasil pengelompokan baik apabila nilai SC-nya lebih dari 0. Nilai SC yang mendekati 1 menunjukkan struktur hasil pengelompokan yang sangat baik.

### Rasio Simpangan Baku

Metode terbaik dalam penelitian ini akan ditentukan dengan melihat rasio simpangan baku. Rasio simpangan baku merupakan perbandingan antara rata-rata simpangan baku dalam kluster dan simpangan baku antar kluster (Silvi, 2018). Rata-rata simpangan baku dalam kluster secara matematis sebagai berikut:

$$S_w = \frac{1}{c} \sum_{k=1}^c S_k \quad (26)$$

$$S_k = \sqrt{\frac{1}{n_k - 1} \sum_{i=1}^c (C_{ik} - \bar{C}_k)^2} \quad (27)$$

dengan:

$S_w$  = rata-rata simpangan baku di dalam kluster;  $S_k$  = simpangan baku pada kluster ke- $k$ ;  $n_k$  = jumlah objek pada kluster ke- $k$ ;  $c$  = banyaknya kluster;  $C_{ik}$  = objek ke- $i$  yang menjadi anggota kluster ke- $k$ ;  $\bar{C}_k$  = rata-rata objek pada kluster ke- $k$

Rata-rata simpangan baku antar kluster dapat dirumuskan sebagai berikut:

$$S_B = \sqrt{\frac{1}{c-1} \sum_{k=1}^c (\bar{C}_k - \bar{C})^2} \quad (28)$$

dengan :

$S_B$  = rata-rata simpangan baku antar kluster;  $\bar{C}$  = rata-rata keseluruhan kluster  
 Metode kluster terbaik merupakan kluster optimal yang memiliki nilai simpangan baku dalam kluster lebih kecil dari nilai simpangan baku antar kluster ( $S_w < S_B$ ).

### Profilisasi

Tahap terakhir ialah profilisasi, tahapan ini melibatkan pemeriksaan persamaan karakteristik variabel di dalam sebuah kluster. Nilai *centroid* mendefinisikan kemiripan karakteristik dari masing-masing variabel dalam sebuah kluster. Nilai *centroid* dapat digunakan untuk menguraikan setiap kluster dengan cara memberikan label.

## 3. METODE PENELITIAN

Data dalam penelitian ini adalah data sekunder yang diperoleh dari publikasi Dinas Kesehatan Provinsi Jawa Tengah melalui Buku Profil Kesehatan Provinsi Jawa Tengah tahun 2021. Variabel yang digunakan dalam penelitian ini berupa Angka kematian bayi (X1), Angka kematian balita (X2), Persentase bayi yang mendapat imunisasi dasar lengkap (X3), Persentase bayi yang mendapatkan pelayanan kesehatan (X4), Persentase balita yang mendapatkan pelayanan kesehatan (X5), Persentase balita kurang gizi (X6) dan Persentase berat bayi lahir rendah (X7).

1. Input data yang berukuran  $n \times p$
2. Deteksi *outlier* dengan jarak mahalanobis
3. Standarisasi data menggunakan *Z-score*
4. Uji asumsi analisis kluster
5. Principal Component Analysis (PCA), dilakukan apabila terjadi multikolinearitas antara variabel yang dianalisis.
6. Menentukan banyaknya kluster ( $c$ ) dalam penelitian, banyak kluster dalam penelitian ini adalah  $k = 2, 3, \dots, 7$
7. Analisis data menggunakan algoritma *Fuzzy C-Means*
8. Validasi *Silhouette* untuk algoritma *Fuzzy C-Means*
9. Analisis data menggunakan algoritma *K-Medoids*
10. Validasi *Silhouette* untuk algoritma *K-Medoids*
11. Membandingkan nilai rasio simpangan baku ( $S_w < S_b$ ) untuk memilih algoritma pengklasteran yang tepat dalam penelitian ini.
12. Profilisasi hasil kluster terbaik terhadap hasil kluster terbaik.

## 4. HASIL DAN PEMBAHASAN

Hasil deteksi *outlier* menggunakan jarak mahalanobis berdasarkan *output* R Studio. Jarak mahalanobis dibandingkan dengan nilai *chi square* ( $\chi_{\alpha;p}^2$ ) dengan taraf signifikansi 0,05 dan derajat bebas sebanyak variabel  $p = 7$  diperoleh nilai *chi square* ( $\chi_{0,05;7}^2$ ) sebesar 14,0671. Nilai jarak mahalanobis objek ke-24 sebesar 19,0265 lebih besar dari nilai *chi square* ( $\chi_{0,05;7}^2$ ), sehingga dapat disimpulkan bahwa objek ke-24 diidentifikasi sebagai *outlier*. Analisis kluster dengan *Fuzzy C-Means* dan *K-Medoids* tetap dilanjutkan karena algoritma ini bersifat *robust* terhadap *outlier*. Kemudian data distandarkan menggunakan standarisasi *Z-Score* dan diuji asumsi kluster.

a. Represetatif

Penelitian ini menggunakan seluruh data dari 35 kabupaten/kota yang ada di Jawa Tengah, sehingga tidak diperlukan uji sampel representatif.

b. Non Multikolinearitas

Nilai VIF untuk setiap variabel berdasarkan *output* uji multikolinearitas disajikan dalam tabel 1.

Tabel 1. Nilai VIF dari Setiap Variabel

Variabel	VIF
X1	44.038214
X2	47.683474
X3	1.419573
X4	1.746377
X5	1.494989
X6	1.494989
X7	2.310801

Berdasarkan tabel 1, diperoleh kesimpulan bahwa terjadi multikolinearitas diantara variabel bebas, sehingga asumsi non multikolinearitas tidak terpenuhi. Langkah selanjutnya ialah melakukan *Principal Component Analysis* untuk mengatasi multikolinearitas tersebut.

**Principal Component Analysis**

Berdasarkan output PCA, diperoleh nilai *eigen*, proporsi varian dan proporsi kumulatif varian disajikan dalam tabel 2.

Tabel 2. Nilai *Eigen*, Proporsi Varian dan Proporsi Kumulatif dari Setiap PC

<i>Principal Component</i>	Nilai <i>Eigen</i>	Proporsi Varian	Proporsi Kumulatif
<b>PC1</b>	3.01319	0.4305	0.4305
<b>PC2</b>	1.59043	0.2272	0.6577
<b>PC3</b>	0.87352	0.1248	0.7824
<b>PC4</b>	0.79829	0.1140	0.8965
<b>PC5</b>	0.36755	0.0525	0.9490
<b>PC6</b>	0.34602	0.0494	0.9984
<b>PC7</b>	0.01098	0.0016	1.0000

Tabel diatas memperlihatkan *eigenvalue* setiap variabel. Nilai *eigenvalue* pada PC1 dan PC2, masing-masing 3.01319 dan 1.59043 lebih besar dari 1, namun proporsi kumulatif dari 2 variabel tersebut ialah 65.77%, sehingga 2 variabel terpilih belum mewakili 80% variabel indikator kesehatan bayi dan balita di Jawa Tengah. Proporsi kumulatif komponen PC1, PC2, PC3 dan PC4 mewakili 89.65% variabel indikator kesehatan bayi dan balita di Jawa Tengah, sehingga 4 komponen tersebut dapat digunakan untuk analisis selanjutnya.

**Fuzzy C-Means**

Berdasarkan hasil analisis pengklasteran untuk  $k = 2, 3, \dots, 7$ ; menggunakan metode *Fuzzy C-Means* dan uji validasi *silhouette*, maka diperoleh perbandingan nilai *silhouette coefficient* untuk setiap klaster disajikan dalam tabel 3.

Tabel 3. Nilai *Silhouette Coefficient* Algoritma *Fuzzy C-Means Clustering*

<i>Cluster</i>	<i>Silhouette Coefficient</i>
<b>2 Cluster</b>	<b>0.43018</b>
3 Cluster	0.38008
4 Cluster	0.34664
5 Cluster	0.26033
6 Cluster	0.30692
7 Cluster	0.16725

Berdasarkan tabel 3 di atas, diperoleh kesimpulan bahwa pengelompokan dengan 2 kluster merupakan hasil kluster optimal pada algoritma *Fuzzy C-Means*. Hal tersebut disebabkan karena nilai *silhouette coefficient* 2 kluster paling mendekati 1 dibandingkan dengan kluster yang lain.

#### ***K-Medoids***

Berdasarkan hasil analisis pengklasteran untuk  $k = 2, 3, \dots, 7$ ; menggunakan metode *K-Medoids* dan uji validasi *silhouette*, maka diperoleh perbandingan nilai *silhouette coefficient* untuk setiap kluster yang disajikan dalam tabel 4.

Tabel 4. Nilai *Silhouette Coefficient* Algoritma *K-Medoids*

<i>Cluster</i>	<i>Silhouette Coefficient</i>
<b>2 Cluster</b>	<b>0.26594</b>
3 Cluster	0.17766
4 Cluster	0.20914
5 Cluster	0.23044
6 Cluster	0.19176
7 Cluster	0.20682

Berdasarkan tabel 4 di atas diperoleh kesimpulan bahwa pengelompokan dengan 2 kluster merupakan kluster optimal yang baik pada algoritma *K-Medoids*. Hal tersebut disebabkan karena nilai *silhouette coefficient* 2 kluster paling mendekati 1 dibandingkan dengan kluster yang lain.

#### **Rasio Simpangan Baku**

Metode terbaik dipilih berdasarkan nilai rasio simpangan baku terkecil. Berdasarkan tabel 5 di bawah ini, dapat disimpulkan bahwa metode terbaik untuk pengelompokan data indikator kesehatan bayi dan balita di Jawa Tengah tahun 2021 adalah algoritma *Fuzzy C-Means*.

Tabel 5. Perbandingan Nilai Rasio Simpangan Baku

Metode	Nilai Rasio Simpangan Baku
<b><i>Fuzzy C-Means</i></b>	<b>1.10619</b>
<i>K-Medoids</i>	1.17705

Anggota setiap kluster berdasarkan metode *Fuzzy C-Means* 2 kluster ialah sebagai berikut:

Kluster 1 : Kabupaten Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Klaten, Wonogiri, Grobogan, Blora, Rembang, Semarang, Temanggung, Kendal, Batang, Pekalongan, Brebes, Kota Magelang, Kota Salatiga, Kota Pekalongan dan Kota Tegal.

Kluster 2 : Kabupaten Cilacap, Magelang, Boyolali, Sukoharjo, Karanganyar, Sragen, Pati, Kudus, Jepara, Demak, Pemalang, Tegal, Kota Surakarta dan Kota Semarang.

#### **Profilisasi**

Tahap profilisasi melibatkan nilai *centroid* kluster dari setiap variabel, Rata-rata setiap variabel disajikan pada tabel 6.



Tabel 6. Rata-Rata Variabel pada Setiap Kluster

Cluster	X1	X2	X3	X4	X5	X6	X7
Cluster 1	<b>10.189</b>	<b>11.466</b>	84.133	92.671	86.429	<b>7.419</b>	<b>5.481</b>
Cluster 2	5.805	6.676	<b>94.421</b>	<b>98.300</b>	<b>93.171</b>	5.264	3.621

Berdasarkan tabel 6, diperoleh kesimpulan sebagai berikut:

- 1) Klaster 1 yang terdiri dari 21 Kabupaten/Kota dikategorikan sebagai kelompok yang memiliki indikator kesehatan bayi dan balita rendah karena angka kematian bayi dan balita, balita kurang gizi dan berat bayi lahir rendah sangat tinggi, sehingga sebaiknya meningkatkan kegiatan imunisasi dasar lengkap serta pelayanan kesehatan terhadap bayi dan balita.
- 2) Klaster 2 beranggotakan 14 Kabupaten/Kota dikategorikan sebagai kelompok yang memiliki indikator kesehatan bayi dan balita tinggi karena rata-rata tertinggi terdapat pada variabel pelayanan kesehatan bayi dan balita serta imunisasi dasar lengkap, sehingga sebaiknya tetap mempertahankan kegiatan imunisasi dan pelayanan kesehatan terhadap bayi dan balita.

## 5. KESIMPULAN

Berdasarkan hasil pembahasan dari kedua metode yang digunakan diperoleh kesimpulan bahwa:

1. Jumlah klaster optimum pada algoritma *Fuzzy C-Means* dan *K-Medoids* berdasarkan validasi *silhouette* masing-masing sebanyak 2 klaster.
2. Hasil pengelompokan Kabupaten/Kota di Jawa Tengah berdasarkan indikator kesehatan bayi dan balita tahun 2021 menggunakan algoritma *Fuzzy C-Means* dan *K-Medoids* diperoleh nilai rasio simpangan baku terkecil yaitu pada metode *Fuzzy C-Means* 2 klaster.
3. Karakteristik dari masing-masing klaster berdasarkan algoritma *Fuzzy C-Means* 2 klaster, yaitu: Klaster 1 merupakan Kabupaten/Kota yang memiliki indikator kesehatan bayi dan balita yang rendah, sehingga perlu meningkatkan kegiatan imunisasi dasar lengkap serta kegiatan pelayanan kesehatan terhadap bayi maupun balita. Klaster 2 merupakan Kabupaten/Kota yang memiliki indikator kesehatan bayi dan balita tinggi, sehingga hanya perlu mempertahankan pelayanan kesehatan terhadap bayi dan balita saja.

## DAFTAR PUSTAKA

- Arora, P., Varshney, S., dan Deepali. (2016). Analysis of *K-Means* and *K-Medoids* Algorithm For Big Data. *Procedia Computer Science*, 507-512. doi:10.1016/j.procs.2016.0
- Bezdek, J. C., Ehrlich, R., dan Full, W. (1983, Mei 1983). FCM : The *Fuzzy C-Means* Clustering Algorithm. *Computers and Geosciences*, 10, 191-203. Dipetik Oktober 10, 2022
- Dinas Kesehatan Jawa Tengah. (2021). *Profil Kesehatan Jawa Tengah*. Semarang: Dinas Kesehatan Jawa Tengah. Dipetik Desember 12, 2022
- Ghozali, I. (2006). *Aplikasi Analisis Multivariate Dengan Program SPSS* (4 ed.). Semarang: Badan Penerbit Universitas Diponegoro. Dipetik Oktober 26, 2022
- Hair, J., Anderson, R., Black, W., Babin, B., dan Tatham, R. (2006). *Multivariate Data Analysis* (6 ed.). New Jersey: Pearson Education, Inc. Dipetik Oktober 24, 2022

- Kaufman, L., & Rousseeuw, P. J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis (Wiley Series in Probability and Statistics)*. New York: Wiley.
- Khang, T. D., Vuong, N. D., Tran, M. K., dan Fowler, M. (2020, Juni 7). *Fuzzy C-Means Clustering Algorithm with Multiple Fuzzification Coefficient*. *Algorithms*, 1-11. doi:doi:10.3390/a13070158
- Park, H. -S., dan Jun, C. -H. (2009). A simple and fast algorithm for *K-Medoids* clustering. *Expert Systems with Applications*, 3336-3341. doi:doi:10.1016/j.eswa.2008.01.039
- Patro, S. K., dan Sahu, K. K. (2015, Maret 3). Normalization: A Preprocecing Stage. *International Advanced Research Journal in Science, Engineering and Technology*, 2(3), -. doi:10.17148/IARJSET.2015.2305
- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation cluster analysis. *Journal of Computational and Aplied Mathematics*, 20, 53-65. Dipetik Desember 14, 2022, dari <https://reader.elsevier.com/reader/sd/pii/0377042787901257?token=82842BD256DCCEEF29D754DCF268496C452FB696351703C680A8B1C6BF92D409B0B9C97E70485C3566E0276F4A5A9danoriginRegion=eu-west-1danoriginCreation=20221214145620>
- Silvi, R. (2018). Analisis Cluster dengan Data Outlier Menggunakan Metode Centroid Linkage dan *K-Means* Clusterig untuk Pengelompokan Indikator HIV/AIDS di Indonesia. *Jurnal Matematika MANTIK 4.1*, 22-31.
- Widarjono, A. (2015). *Analisis Multivariat Terapan : Dengan Program SPSS, AMOS Dan SMARTPLS* (2 ed.). Yogyakarta: STIM YKPN.