

# Comparison the Performance of K-Means and K-Medoids Algorithms in Grouping Regencies/Cities in Sumatera Based on Poverty Indicators

Mardhiatul Azmi, Atus Amadi Putra\*, Dodi Vionanda, Admi Salma

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

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

Submitted : 09 November 2023  
Revised : 07 Desember 2023  
Accepted : 13 Februari 2023

## ABSTRACT

*K-Means is a non-hierarchical approach that separates data into a number of groups according on how far an object is from the closest centroid. K-Medoids is a non-hierarchical clustering technique that separates data into a number of groups according on how far away an object is from the closest medoid. The two approaches were put to the test using data on poverty in Sumatera in 2021, when the Covid-19 outbreak had caused the poverty rate to increase from the year before. This research is an applied research which begins by studying relevant theories. This study uses secondary data from the BPS website regarding poverty indicators. This research aims to determine regional groups and compare the results of grouping using the k-means and k-medoids algorithm. To find out the best performance of both algorithm, that is by looking at the lowest Davies Bouldin Index (DBI). This research findings show that the k-means algorithm can create up to 34 regencies/cities that are included in cluster 1, 52 regencies/cities in cluster 2, 23 regencies/cities in cluster 3, and 45 regencies/cities in cluster 4. K-Medoids algorithm, namely in clusters 1, 2, 3, and 4, respectively, as many as 53, 40, 37, and 24 regencies/cities. Based on the clustering results, the DBI was obtained as 1,584 for the k-means algorithm, and 2,359 for the k-medoids algorithm. The DBI of k-means is smaller than the DBI of k-medoids, which indicates that is a better than algorithm of k-medoids.*

**Keywords:** *Davies Bouldin Index, K-Means, K-Medoids, Poverty*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

Analisis multivariat adalah analisis yang mencakup dua variabel atau lebih. Analisis multivariat dibagi menjadi dua kategori yaitu analisis dependensi dan analisis interdependensi. Tujuan analisis dependensi adalah untuk melihat hubungan antara variabel dependen dengan variabel independen. Sedangkan analisis interdependensi bertujuan untuk melihat keterikatan antara semua variabel yang ada, dalam arti semua variabel merupakan variabel independen. Analisis interdependensi dibagi menjadi beberapa kategori, salah satunya adalah analisis *cluster*. Analisis *cluster* adalah analisis yang mengelompokkan objek ke dalam kelompok yang relatif homogen, dimana objek pada setiap kelompok mirip satu sama lain dan sangat berbeda dengan objek pada kelompok lain (Supranto, 2017).

Analisis *cluster* dibagi dalam dua kategori, yaitu hirarki dan non hirarki (Hair *et al.*, 2014). Hirarki adalah metode dimana objek dikelompokkan secara struktural berdasarkan kesamaannya dan jumlah *cluster* yang diinginkan tidak diketahui. Sedangkan non hirarki adalah metode pengelompokan objek dimana jumlah *cluster* yang akan dibentuk dapat ditentukan sebelumnya (Johnson dan Wichern, 2002). Algoritma yang paling populer pada metode non hirarki adalah algoritma *K-Means* dan *K-Medoids*. Selain itu, kedua algoritma tersebut cukup akurat dan fleksibel karena peneliti dapat menentukan berapa *cluster* yang ingin dibentuk (Farissa, 2021).

*K-Means* merupakan bagian dari metode non hirarki yang membagi objek data menjadi dua atau lebih kelompok. Dalam algoritma ini, data dikelompokkan dalam satu kelompok berkarakteristik sama, dan objek pada kelompok lain berbeda (Prasetyo, 2012). Sedangkan *K-Medoids* merupakan varian dari algoritma *K-Means*, dimana setiap *cluster* direpresentasikan oleh objek sebagai pusat *cluster*. Pusat *cluster* tersebut dipilih dari data sebanyak K objek secara acak sebagai *medoid* (Pujari, 2010). Untuk menguji kinerja dari kedua metode tersebut, dapat dilihat dari nilai *Davies*

*Bouldin Index* (DBI). DBI pertama kali ditemukan pada tahun 1979 oleh David L. Davies dan Donald W. Bouldin. Hasil pengelompokan dikatakan baik apabila nilai DBI yang diperoleh semakin rendah, dan tidak bernilai negatif.

Pada penelitian ini akan dibahas mengenai perbandingan algoritma *K-Means* dan *K-Medoids* dengan permasalahan kemiskinan di Sumatera pada tahun 2021. Kemiskinan merupakan ketidakmampuan suatu masyarakat untuk memenuhi kebutuhan pokok dan kebutuhan lainnya yang diukur dari sisi pengeluaran (BPS, 2022). Masalah kemiskinan dalam masyarakat mempengaruhi ekonomi, budaya, politik dan masyarakat. Secara luas, kemiskinan juga mengacu pada aspek-aspek selain pendapatan, seperti akses terhadap pelayanan kesehatan, pendidikan, air bersih dan sanitasi (Bappenas, 2017). BPS menyebutkan bahwa terdapat beberapa permasalahan yang menyebabkan meningkatnya angka kemiskinan pada tahun 2021 yaitu karena masih terjadinya pandemi covid-19 yang mengakibatkan perubahan aktivitas ekonomi masyarakat sehingga angka kemiskinan terpengaruhi, dan pengeluaran konsumsi rumah tangga mengalami peningkatan.

Perbandingan mengenai *K-Means* dan *K-Medoids* telah dilakukan terdahulu oleh Kamila, dkk (2019). Penelitian tersebut menghasilkan algoritma *K-Means* lebih baik daripada algoritma *K-Medoids*, karena DBI yang diperoleh pada algoritma *K-Means* sebesar 0.112 lebih kecil dibandingkan DBI pada algoritma *K-Medoids* sebesar 0.119, dengan jumlah *cluster* adalah 3. Selain itu, penelitian Harahap (2021) juga membandingkan algoritma *K-Means* dan *K-Medoids* untuk pengelompokan data kelas siswa tunagrahita. Hasil dari penelitian ini adalah DBI *K-Means* sebesar 0.161 dan *K-Medoids* sebesar 0.291. Hal ini berarti bahwa algoritma *K-Means* memberikan hasil yang lebih baik daripada *K-Medoids*, karena DBI *K-Means* lebih kecil dari DBI *K-Medoids*. Selanjutnya perbandingan *K-Means* dan *K-Medoids* juga diteliti oleh Soni dan Patel (2017) pada data IRIS. Penelitian oleh Soni dan Patel ini menghasilkan keunggulan *k-medoids* daripada *k-means* dalam waktu eksekusi, kualitas, dan jumlah *record*.

## II. METODE PENELITIAN

### A. Analisis Multivariat

Analisis multivariat merupakan analisis statistika yang melibatkan dua variabel atau lebih. Beberapa kegunaan teknik multivariat (Johnson dan Wichern, 2002), diantaranya dapat digunakan sebagai reduksi data atau penyederhanaan struktural, dengan tujuan mempermudah interpretasi tanpa mengurangi informasi yang ada. Selanjutnya teknik multivariat dapat pula digunakan untuk mengurutkan dan mengelompokkan suatu gugus data, dimana kelompok objek atau variabel serupa dibuat berdasarkan karakteristik yang diukur. Teknik multivariat juga dapat digunakan dalam menyelidiki ketergantungan antar variabel, peramalan, serta pengujian hipotesis. Data pada analisis multivariat biasanya ditulis dalam bentuk matriks. Matriks merupakan sekumpulan bilangan yang disusun dalam  $n$  baris dan  $p$  kolom.

### B. Analisis Cluster

Analisis *cluster* adalah analisis yang digunakan untuk mengelompokkan objek data ke dalam kelompok yang relatif homogen berdasarkan beberapa variabel, dimana objek dalam satu kelompok cenderung mirip satu sama lain (*within-cluster*), sedangkan dengan objek dalam kelompok lain berbeda jauh (*between-cluster*). Analisis *cluster* memeriksa ketergantungan (hubungan) antara semua set variabel. Salah satu kegunaan analisis *cluster* yaitu dapat digunakan sebagai alat mereduksi data, sehingga menghasilkan informasi yang bermakna dan memudahkan menarik kesimpulan tanpa menghilangkan informasi yang terkandung di dalam data asli (Supranto, 2018).

Analisis *cluster* terbagi ke dalam dua kategori (Hair dkk, 2014), yaitu hirarki dan non hirarki. Hirarki *clustering* adalah metode dimana objek dikelompokkan secara struktural berdasarkan kesamaan sifatnya dan jumlah *cluster* yang diinginkan tidak diketahui (Mattjik dan Sumertajaya, 2011). Hirarki *clustering* dibagi menjadi dua metode yaitu metode *agglomerative* (penggabungan) dan metode *divisive* (pemisahan gerombol). Sebaliknya, non hirarki *clustering* adalah metode pengelompokan objek dimana jumlah *cluster* yang akan digunakan dapat ditentukan terlebih dahulu sebagai bagian dari proses pengelompokan (Johnson dan Wichern, 2002). Banyaknya *cluster* yang akan dibentuk dapat dilakukan melalui pertimbangan teoretis, konseptual, dan praktis. Metode non hirarki terdiri dari beberapa algoritma, dimana metode yang populer yaitu algoritma *K-Means* dan algoritma *K-Medoids*.

#### 1. *K-Means Clustering*

*K-Means* adalah salah satu metode paling sederhana dan paling umum digunakan. *K-Means* merupakan metode pengelompokan non hirarki yang bertujuan untuk membagi data yang ada menjadi dua atau lebih kelompok, dimana data berkarakteristik yang sama masuk ke dalam kelompok yang sama, dan data yang berbeda ke dalam kelompok lain (Prasetyo, 2012). *K-Means Clustering* bertujuan untuk menemukan titik data prototipe untuk setiap *cluster*,

dimana semua titik data dipetakan ke *centroid* terdekat, yang kemudian membentuk *cluster*. Pusat *cluster* dari *K-Means* adalah rata-rata dari semua objek data dalam *cluster*. Dalam algoritma *k-means*,  $n$  objek dikelompokkan menjadi  $K$  *cluster* atribut data, dimana  $K < n$ , dan  $K$  adalah bilangan bulat positif. Untuk membagi kumpulan data, perlu mendefinisikan ukuran kedekatan. Ukuran untuk atribut numerik yang paling umum digunakan adalah jarak *Euclidean*, yang terdapat pada Persamaan (2).

Adapun langkah-langkah analisis *K-Means Clustering* (Bhatia, 2019) yang dilakukan adalah :

- 1) Terlebih dahulu tentukan jumlah *cluster* yang ingin dibentuk, serta tentukan pusat *cluster* (*centroid*) secara acak dari data.
- 2) Hitung jarak antara titik data dengan *centroid*.
- 3) Lakukan pengelompokan sementara data dengan melihat nilai terkecil dari jarak titik data dan *centroid*.
- 4) Hitung *centroid* baru dengan menggunakan Persamaan (1).

$$v_i = \frac{1}{n} \sum_{j=1}^n x_j \quad (1)$$

- 5) Lakukan iterasi sampai kondisi konvergensi terpenuhi.

## 2. *K-Medoids Clustering*

*K-Medoids* atau *Partition Around Medoids* (PAM) adalah sebuah algoritma *cluster* non hirarki yang merupakan varian dari algoritma *K-Means*. *K-Medoids* adalah algoritma yang setiap *cluster*-nya diwakili objek sebagai pusat *cluster* (Pujari, 2010). Pusat *cluster* pada algoritma *K-Medoids* dinamakan dengan *medoid*. Algoritma *K-Medoids* dilakukan secara iteratif atau berulang hingga setiap objek yang mewakili benar-benar merupakan *medoid* dalam *cluster*. Adapun langkah-langkah yang dilakukan pada analisis *K-Medoids Clustering* (Aggarwal dan Reddy, 2014) yaitu sebagai berikut :

- 1) Jumlah *cluster* yang ingin dibentuk ditentukan terlebih dahulu, dan pilih sebanyak  $K$  objek dari  $n$  data secara acak yang digunakan sebagai *medoid*.
- 2) Hitung jarak masing-masing *medoid* ke titik data.
- 3) Tentukan anggota *cluster* terhadap *medoid* sementara dengan melihat nilai jarak terdekat objek ke *medoid*, serta hitung total kedekatannya (*cost*).
- 4) Tentukan *medoid* baru dengan memilih sebanyak  $K$  objek dari data.
- 5) Hitung jarak setiap objek di setiap *cluster* menggunakan *medoid* baru.
- 6) Cari selisih kedekatan (*cost*) untuk menentukan nilai total simpangan ( $S$ ). Jika  $S < 0$ , maka lakukan iterasi hingga tidak terjadi perubahan letak *medoid*.

## C. Ukuran Jarak

Analisis *cluster* membutuhkan suatu ukuran untuk menentukan seberapa mirip atau berbeda objek data, karena tujuan *clustering* itu sendiri adalah untuk mengelompokkan objek yang mirip ke dalam *cluster* yang sama. Ukuran yang paling umum digunakan adalah jarak *Euclidean* (Supranto, 2018), dengan rumus (Everitt, 2018) yang ditunjukkan pada Persamaan (2) :

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2)$$

dimana  $d(x_i, x_j)$  adalah jarak antara objek ke- $i$  dan objek ke- $j$ ,  $p$  adalah banyak variabel,  $x_{ik}$  adalah objek ke- $i$  pada variabel ke- $k$ , dan  $x_{jk}$  adalah objek ke- $j$  pada variabel ke- $k$ .

## D. Validitas Cluster

Validasi digunakan untuk mengukur kualitas hasil *cluster*. Ukuran validitas yang digunakan yaitu DBI. DBI merupakan ukuran keunikan *cluster* dengan mempertimbangkan kekompakan *cluster* dan pemisahan antara *cluster*. Terdapat beberapa rumus yang digunakan untuk mendapatkan DBI (Prasetyo, 2014) yaitu sebagai berikut:

### 1. *Sum of Square Within Cluster*

Tujuan dari perhitungan *Sum of Square Within Cluster* (SSW) adalah untuk mengetahui keterikatan antara anggota *cluster* dalam sebuah *cluster* (kohesi/homogenitas). SSW diharapkan bernilai kecil, yang berarti karakteristiknya semakin mirip. Perhitungan SSW ditunjukkan pada Persamaan (3).

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_{jk}, c_{ik}) \quad (3)$$

dimana  $m_i$  merupakan jumlah data yang berada dalam *cluster* ke- $i$ ,  $x_{ik}$  yaitu data pada objek ke- $i$  variabel ke- $k$ ,  $c_{ik}$  adalah titik pusat *cluster* ke- $i$  pada variabel ke- $k$ , dan  $d(x_{ik}, c_{ik})$  adalah jarak antara data dengan titik pusat *cluster* (*centroid/medoid*).

**2. Sum of Square Between Cluster (SSB)**

Perhitungan *Sum of Square Between Cluster* (SSB) dilakukan dengan tujuan untuk mengetahui perbedaan antara satu *cluster* dengan *cluster* lainnya (separasi/heterogenitas). SSB diharapkan bernilai besar. Perhitungan SSB ditunjukkan pada Persamaan (4).

$$SSB_{i,j} = d(c_{ik}, c_{jk}) \tag{4}$$

dimana  $c_{ik}$  adalah titik pusat *cluster* ke- $i$  pada variabel ke- $k$ ,  $c_{jk}$  adalah titik pusat *cluster* ke- $j$  pada variabel ke- $k$ , dan  $d(c_{ik}, c_{jk})$  adalah jarak antara titik pusat satu *cluster* dengan *cluster* lainnya.

**3. Rasio**

Tujuan perhitungan rasio ( $R_{ij}$ ) adalah untuk mengetahui seberapa baik nilai perbandingan antara satu *cluster* dengan *cluster* lainnya. Rasio dapat dihitung menggunakan Persamaan (5).

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \tag{5}$$

**4. Davies Bouldin Index (DBI)**

*Davies Bouldin Index* (DBI) dikatakan semakin bagus ketika mendekati nilai nol (*non* negatif). Perhitungan DBI ditunjukkan pada Persamaan (6).

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j}) \tag{6}$$

dimana  $K$  merupakan jumlah *cluster* yang digunakan.

**E. Jenis dan Sumber Data**

Data yang digunakan pada penelitian ini adalah data sekunder yang bersumber dari *website* BPS mengenai data indikator kemiskinan di Sumatera pada tahun 2021. Adapun variabel yang digunakan adalah persentase penduduk miskin ( $X_1$ ), pengeluaran per kapita ( $X_2$ ), angka partisipasi sekolah usia 13-15 tahun ( $X_3$ ), indeks pembangunan manusia ( $X_4$ ), penduduk yang bekerja di sektor pertanian ( $X_5$ ), fasilitas perumahan yang memiliki air layak ( $X_6$ ), dan fasilitas perumahan yang memiliki jamban sendiri ( $X_7$ ). Selanjutnya langkah-langkah yang dilakukan pada penelitian ini adalah sebagai berikut :

1. Menyiapkan data yang akan diolah yaitu data indikator kemiskinan di Sumatera tahun 2021, dan melakukan analisis deskriptif data.
2. Melakukan standardisasi data.
3. Melakukan analisis *clustering* menggunakan algoritma *K-Means* dan algoritma *K-Medoids* dengan bantuan *software* RStudio.
4. Melakukan evaluasi kinerja algoritma berdasarkan nilai DBI pada Persamaan (6).

**III. HASIL DAN PEMBAHASAN**

**1. Standardisasi Data**

Data yang digunakan adalah data kemiskinan di Pulau Sumatera pada tahun 2021 menurut kabupaten/kota. Sebelum dilakukan pengelompokan pada data, terlebih dahulu dilakukan standardisasi agar semua nilai variabel jatuh pada rentang yang sama dan agar proses *clustering* lebih efektif. Hasil standardisasi data ditampilkan seperti pada Tabel 1.

**Tabel 1.** Standardisasi Data

No	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>
1	1.64830	-1.67927	0.68587	-1.00568	0.63475	0.08772	-1.86877
2	1.93123	-0.88182	0.29401	-0.38270	-0.18386	-0.04806	-0.34538
3	0.45917	-1.17376	0.68587	-0.77733	-0.03299	0.10388	-1.30684
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
154	-0.28096	2.38198	0.68587	1.76999	-1.62118	0.50905	0.94844

Setelah dilakukan standardisasi data, langkah selanjutnya yaitu melakukan pengelompokan menggunakan analisis *cluster* non hirarki dengan algoritma *K-Means* dan algoritma *K-Medoids*.

## 2. *K-Means Clustering*

Langkah pertama yang dilakukan untuk mengelompokkan kabupaten/kota di Sumatera pada tahun 2021 berdasarkan indikator kemiskinan yaitu menentukan jumlah *cluster* yang ingin dibentuk. Indikator kemiskinan akan dikelompokkan menjadi empat *cluster*. Selanjutnya dipilih secara acak sebanyak empat titik data yang digunakan sebagai pusat *cluster* (*centroid*). ukuran jarak yang digunakan adalah jarak *Euclidean* yang digunakan untuk mengetahui jarak terpendek setiap data dengan *centroid* nya. Proses *k-means clustering* dilakukan secara iterasi sampai kondisi konvergensi terpenuhi. *Centroid* terakhir yang diperoleh ditampilkan seperti pada Tabel 2.

**Tabel 2.** *Centroid* Akhir *K-Means Clustering*

No	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>
1	-0.673	1.209	0.363	1.459	-1.355	0.781	0.764
2	-0.303	0.058	0.363	-0.053	0.093	0.087	0.338
3	-0.107	-0.289	-1.688	-0.610	0.457	-0.604	0.059
4	0.941	-0.853	0.164	-0.744	0.696	-0.393	-1.029

Berdasarkan Tabel 2, diperoleh *centroid* akhir dari proses *K-Means*. Hasil pengelompokan kabupaten/kota berdasarkan indikator kemiskinan di Sumatera menggunakan algoritma *K-Means* disajikan seperti pada Tabel 3.

**Tabel 3.** Hasil Pengelompokan Algoritma *K-Means*

Cluster	Banyak Anggota (Kabupaten/Kota)
1	34
2	52
3	23
4	45

Berdasarkan Tabel 3, diketahui bahwa mayoritas kabupaten/kota di Sumatera masuk dalam *cluster* 2 yaitu sebesar 52 kabupaten kota, disusul dengan *cluster* 4 sebanyak 45 kabupaten/kota. Pada *cluster* 1 terdapat sebanyak 34 kabupaten/kota, serta sebanyak 23 kabupaten/kota tergabung dalam *cluster* 3. Setelah hasil pengelompokan diketahui, langkah selanjutnya yaitu melihat karakteristik dari tiap *cluster* yang hasilnya ditampilkan seperti pada Tabel 4.

**Tabel 4.** Karakteristik *K-Means Clustering* Berdasarkan Nilai Rata-Rata

Cluster	Rata-Rata						
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>
1	7.657	13044.059	97.494	77.529	7.964	90.021	93.359
2	9.463	10695.321	97.497	70.706	34.552	77.144	86.055
3	10.417	9986.652	81.586	68.193	41.241	64.320	81.299
4	15.531	8833.977	95.954	67.588	45.628	68.220	62.659

Berdasarkan Tabel 4, diketahui bahwa kabupaten/kota yang masuk pada *cluster* 1 merupakan kabupaten/kota yang memiliki rata-rata persentase penduduk miskin terendah diantara *cluster* lainnya, yaitu sebesar 7.657%. Namun demikian, pada *cluster* 1 ini memiliki nilai rata-rata pendapatan per kapita, indeks pembangunan manusia, fasilitas perumahan yang memiliki air layak dan jamban sendiri, paling tinggi dibandingkan *cluster* lainnya. Pada *cluster* 2, kabupaten/kota memiliki rata-rata persentase penduduk miskin sedang. Selanjutnya rata-rata kabupaten/kota pada *cluster* 3 menunjukkan persentase penduduk miskin tinggi, yaitu sebesar 10.417%. Selain itu, kabupaten/kota dengan rata-rata persentase penduduk miskin tertinggi masuk dalam *cluster* 4, yaitu sebesar 15.531%. Namun angka pengeluaran per kapita, indeks pembangunan manusia, dan fasilitas perumahan yang memiliki jamban sendiri pada *cluster* 4 ini merupakan angka terendah diantara empat *cluster* lainnya.

### 3. *K-Medoids Clustering*

Sama halnya dengan *K-Means clustering*, untuk mengelompokkan kabupaten/kota di Sumatera pada tahun 2021 berdasarkan indikator kemiskinan menggunakan algoritma *K-Medoids*, jumlah *cluster* yang ingin dibentuk terlebih dahulu ditentukan. Data akan dikelompokkan menjadi 4 *cluster*. Selanjutnya dipilih secara acak sebanyak empat titik data yang digunakan sebagai pusat *cluster*. Pusat *cluster* pada *K-Medoids* disebut sebagai *medoid*. Ukuran jarak *Euclidean* digunakan untuk mengetahui jarak terpendek dari setiap titik data dengan *medoid* nya. Proses *K-Medoids clustering* dilakukan secara iterasi sampai tidak terjadi perubahan letak *medoid*. *Medoid* terakhir yang diperoleh ditampilkan seperti pada Tabel 5.

**Tabel 5.** *Medoid* Akhir Algoritma *K-Medoids*

No	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>
1	1.005	-0.214	-0.094	-0.631	0.311	-0.352	-0.019
2	-0.375	-0.096	-0.143	-0.314	0.006	-0.500	-0.564
3	-0.640	0.138	0.686	0.460	-0.579	0.741	0.392
4	-0.496	0.911	0.686	1.823	-1.600	1.319	0.823

Berdasarkan Tabel 5, diperoleh *medoid* akhir dari proses *K-Medoids*. Hasil pengelompokan kabupaten/kota di Sumatera berdasarkan indikator kemiskinan menggunakan algoritma *K-Medoids* ditampilkan seperti pada Tabel 6.

**Tabel 6.** Hasil Pengelompokan Algoritma *K-Medoids*

Cluster	Banyak Anggota (Kabupaten/Kota)
1	53
2	40
3	37
4	24

Berdasarkan Tabel 6, *cluster* 1 memiliki anggota paling banyak yaitu mencakup sebanyak 53 kabupaten/kota, dan jumlah anggota paling sedikit berada pada *cluster* 4 sebesar 24 kabupaten/kota. Selanjutnya pada urutan kedua dan ketiga berturut-turut berada pada *cluster* 2 dan 3 dengan jumlah kabupaten/kota sebanyak 37 dan 24. Setelah hasil *cluster* diperoleh, selanjutnya dapat dilihat karakteristik dari masing-masing *cluster* berdasarkan nilai rata-rata (*mean*) nya, ditampilkan seperti pada Tabel 7.

**Tabel 7.** Karakteristik *K-Medoids Clustering* Berdasarkan Nilai Rata-Rata

Cluster	Rata-Rata						
	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>
1	15.936	9019.321	94.255	67.755	44.836	70.361	77.009
2	8.583	10268.150	90.627	69.376	39.100	62.504	67.932
3	8.040	11293.676	98.394	72.064	26.642	85.257	89.616
4	8.309	13421.750	96.643	78.888	5.507	94.607	93.644

Berdasarkan Tabel 7, diketahui bahwa kabupaten/kota pada *cluster* 1 merupakan kabupaten/kota dengan rata-rata persentase penduduk miskin paling tinggi diantara ketiga *cluster* lainnya yaitu 15.936%. Namun demikian, rata-rata persentase partisipasi sekolah anak usia 13-15 tahun dan IPM paling rendah diantara *cluster* lainnya. Di sisi lain, *cluster* 3 mencakup kabupaten/kota dengan rata-rata persentase penduduk miskin terendah sebesar 8.040%, dengan tingkat partisipasi sekolah usia 13-15 tahun tertinggi dibandingkan dengan *cluster* lainnya. Selanjutnya, kabupaten/kota yang memiliki nilai rata-rata persentase penduduk miskin tinggi berada pada *cluster* 2, dengan nilai sebesar 8.583%, dimana *cluster* 2 ini memiliki rata-rata pengeluaran per kapita sedang, angka partisipasi sekolah usia 13-15 tahun rendah diantara tiga *cluster* lainnya, IPM rendah, penduduk yang bekerja di sektor pertanian termasuk tinggi, dan fasilitas perumahan yang memiliki air layak dan jamban sendiri termasuk rendah dibandingkan tiga *cluster* lainnya. Selanjutnya pada *cluster* 4, kabupaten/kota yang berada di dalamnya memiliki nilai rata-rata persentase penduduk miskin sedang diantara ketiga *cluster* lainnya, dimana rata-rata angka partisipasi sekolah usia 13-15 tahun,

IPM, dan fasilitas perumahan yang memiliki air layak dan jamban sendiri termasuk paling tinggi dibandingkan tiga *cluster* lainnya.

#### 4. Uji Validasi

Uji validasi digunakan untuk mengukur kualitas hasil *cluster*. Ukuran validasi yang digunakan yaitu DBI, dimana hasil DBI untuk algoritma *K-Means* dan algoritma *K-Medoids* dapat dilihat seperti pada Tabel 8.

**Tabel 8.** Hasil DBI Algoritma *K-Means* dan *K-Medoids*

Algoritma	DBI
<i>K-Means</i>	1.584
<i>K-Medoids</i>	2.359

Berdasarkan Tabel 8, diperoleh hasil DBI untuk algoritma *K-Means* lebih kecil dibandingkan nilai DBI pada algoritma *K-Medoids*, dengan nilai berturut-turut adalah 1.584 dan 2.359.

#### IV. KESIMPULAN

Hasil pengelompokan kabupaten/kota di Sumatera berdasarkan indikator kemiskinan pada tahun 2021 menggunakan algoritma *K-Means* dan *K-Medoids* memberikan hasil yang berbeda. Hasil pengelompokan menggunakan algoritma *K-Means* yaitu sebanyak 34 kabupaten/kota masuk pada *cluster* 1, selanjutnya kabupaten/kota yang masuk pada *cluster* 2, 3, dan 4 berturut-turut adalah 52, 23, dan 45 kabupaten/kota. Pengelompokan menggunakan algoritma *K-Medoids* memberikan hasil yaitu sebanyak 53 kabupaten/kota pada *cluster* 1, 40 kabupaten/kota pada *cluster* 2, 37 kabupaten/kota pada *cluster* 3, dan 24 kabupaten/kota pada *cluster* 4. Berdasarkan hasil pengelompokan dari kedua algoritma tersebut, diperoleh DBI untuk masing-masing algoritma yaitu sebesar 1.584 untuk algoritma *K-Means* dan 2.359 untuk algoritma *K-Medoids*. Hal ini berarti bahwa algoritma *K-Means* lebih cocok digunakan untuk mengelompokkan kabupaten/kota di Sumatera berdasarkan indikator kemiskinan, dibandingkan algoritma *K-Medoids*, karena DBI yang dihasilkan pada algoritma *K-Means* lebih kecil daripada DBI pada algoritma *K-Medoids*.

#### UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah membantu penulis dalam menyelesaikan penulisan jurnal ini hingga dapat terselesaikan.

#### DAFTAR PUSTAKA

- Aggarwal, C. C., dan Reddy, C.K. (2014). *Data Clustering Algorithms and Applications*, Chapman and Hall/CRC.
- Bappenas. (2007). *Pemantauan dan Evaluasi Program-Program Penanggulangan Kemiskinan. Kumpulan Bahan Latihan*, 1–242.
- Badan Pusat Statistik. (2022). *Indeks Pembangunan Manusia*. Diakses pada 22 Mei 2022, dari <https://www.bps.go.id/subject/26/indeks-pembangunan-manusia.html>.
- \_\_\_\_\_. (2022a). *Kemiskinan dan Ketimpangan*. Diakses pada 30 Mei 2022, dari <https://www.bps.go.id/subject/23/kemiskinan-dan-ketimpangan.html>.
- Bhatia, Parteek. (2019). *Data Mining and Data Warehousing: Principal and Practical Techniques*. Cambridge: Cambridge University Press.
- Everitt, B. S., dkk. (2011). *Cluster Analysis. Fifth Edition*, London: John Wiley & Sons.
- Farissa, R. A., Mayasari, R., & Umaidah, Y. (2021). “Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung”, *Journal of Applied Informatics and Computing*, 5(2), 109–116.
- Hair, J., dkk. (2014). *Multivariate Data Analysis Seventh Edition*. Edinburgh: Pearson.

- Harahap, Fitriana. (2021). “Perbandingan Algoritma K-Means dan K-Medoids untuk Clustering Kelas Siswa Tunagrahita”, *Jurnal Terapan Informatika Nusantara*, 2(4), 191-197.
- Johnson, Richard A., dan Wichern, Dean W. (2002). *Applied Multivariate Statistical Analysis Fifth Edition*. New Jersey: Pearson.
- Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). “Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau”, *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119.
- Mattjik, A. A dan Sumertajaya, I Made. (2011). *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor: IPB Press.
- Prasetyo, Eko. (2012). *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi Offset.
- . (2014). *Data Mining*. Yogyakarta: Andi Offset.
- Pujari, Arun K. (2008). *Data Mining*, pp. 1-303.
- Soni, K. G, and Patel, A. (2017). “Comparative Analysis of K-Means and K-Medoids Algorithm on IRIS Data”, *International Journal of Computational Intelligence Research*. 13(5), 899-906.
- Supranto, J. (2018). *Analisis Multivariat: Arti & Interpretasi*. Jakarta: PT Rineka Cipta.