

Klasterisasi Tingkat Kemiskinan Kabupaten/Kota di Indonesia Menggunakan Algoritma *K-Means* dan *K-Medoids*

Gita Risky Pratiwi^{1*}, Deden Wahiddin², Elsa Elvira Awal³, Ahmad Fauzi⁴

Fakultas Ilmu Komputer, Program Studi Teknik Informatika
Universitas Buana Perjuangan Karawang

**email*: If20.gitapратиwi@mhs.ubpkarawang.ac.id

Info Artikel

Dikirim: 12 August 2024

Diterima: 4 Desember 2024

Diterbitkan: 4 Desember 2024

Kata kunci:

Clustering;

Kemiskinan;

K-Means;

K-Medoids;

Sillhouette Score.

ABSTRAK

Kemiskinan adalah ketika sebuah masyarakat memiliki akses fisik yang terbatas terhadap lingkungan dasar. Kondisi permukiman miskin ini seringkali jauh di bawah standar kelayakan dan menyebabkan orang-orang di sana kesulitan mendapatkan uang untuk hidup. Fokus penelitian ini adalah untuk menentukan tingkat kemiskinan di Kab/Kota di Indonesia. Karena ada peningkatan angka kemiskinan di Indonesia, *clustering* diperlukan untuk pemerintah dapat memberikan bantuan yang tepat kepada mereka yang paling membutuhkan. Metode yang digunakan adalah algoritma *K-Means* dan *K-Medoids*. Hasil dari pengelompokan ini menghasilkan *cluster* 0 menunjukkan tingkat kemiskinan yang relatif rendah yaitu 250 kab/kota, *cluster* 1 menunjukkan tingkat kemiskinan yang tinggi yaitu 330 kab/kota, Sedangkan algoritma *K-Medoids* menghasilkan tiga klaster dengan tingkat kemiskinan rata-rata yang berbeda: *cluster* 0 menunjukkan tingkat kemiskinan yang relatif rendah yaitu 270 kab/kota, *cluster* 1 menunjukkan tingkat kemiskinan yang tinggi yaitu 310 kab/kota. Hal ini dapat menjadi referensi bagi pemerintah untuk meningkatkan perhatian wilayah dengan tingkat kemiskinan tinggi dalam upaya mengurangi tantangan ekonomi yang sedang berlangsung. Dengan menggunakan skor *sillhouette*, untuk membagi tingkat kemiskinan evaluasi algoritma *K-Means* dan *K-Medoids* dilakukan. Algoritma *k-means* menghasilkan nilai $K = 0.284$ sedangkan algoritma *K-Medoids* menghasilkan nilai $K = 0.224$.

1) PENDAHULUAN

Kemiskinan didefinisikan sebagai keadaan dimana seseorang baik laki-laki maupun perempuan, tidak dapat memenuhi hak-hak dasar mereka untuk menjalani hidup dan mengembangkan kehidupan dengan baik [1]. Seseorang dianggap miskin Pengeluaran rata-rata perbulan atau perkapita berada dibawah garis kemiskinan (GK) [2].

Seseorang dianggap miskin jika mereka tidak dapay kemampuan untuk memenuhi kebutuhan dasar mereka [3]. Jumlah penduduk meningkat di kota-kota menyebabkan kemacetan, angka pengangguran, kejahatan, dan kesulitan mendapatkan infrastruktur sosial dan perumahan yang layak [4]. Diantara factor-faktor yang harus dipertimbangkan dalam merencanakan penanggulangan kemiskinan adalah sosial, ekonomi, budaya, politik, dan waktu ruang. Untuk merancang strategi dalam mengurangi kemiskinan dapat diterapkan pada masyarakat di setiap wilayah, faktor-faktor penyebab kemiskinan harus diketahui terlebih dahulu [5]. Pada maret 2024, rata-rat rumah tangga yang mengalami kemiskinan adalah 4,78 sehingga menciptakan garis kemiskinan per

rumah tangga rata-rata 2.786.415,- perbulan, dengan komposisi 74,44% dari garis kemiskinan makanan dan 25.56% dari garis kemiskinan bukan makanan [6].

Tingkat kemiskinan di setiap Kabupaten atau Kota berbeda berdasarkan banyak indikator pendukung. Adanya penentuan dan pengukuran indikator kemiskinan diperlukan untuk menentukan dan mempermudah pengukuran tingkat kemiskinan di suatu wilayah. Dengan melihat pendapatan ekonomi dan jumlah penduduk, misalnya. Oleh karena itu, diperlukan gambaran tentang area mana yang perlu penanganan serius dan cepat. Sulit untuk mengidentifikasi wilayah mana yang memiliki tingkat kemiskinan tinggi, yang menyebabkan pemerintah tidak melakukan banyak hal untuk menangani masalah tersebut. Untuk mengatasi masalah ini, clustering harus dilakukan untuk menentukan provinsi mana yang termasuk kelompok kota dan kabupaten tingkat tinggi dan rendah. Ini akan memungkinkan pemerintah untuk mengambil kebijakan kemiskinan kota dan kabupaten dengan cepat [7].

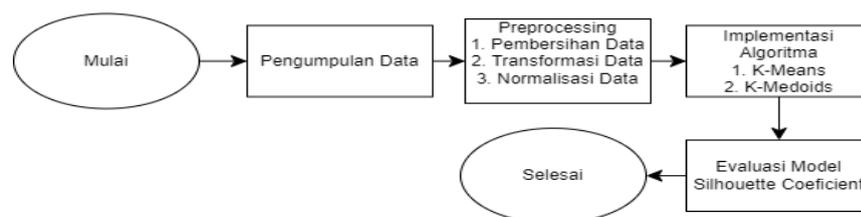
Clustering adalah pembagian objek yang mirip ke dalam kelompok yang berbeda, atau pembagaian sebuah set data menjadi subset yang berbeda sehingga setiap *subset* memiliki nilai datanya masing-masing. Dua jenis algoritma *clustering* dapat digunakan da sebuah *cluster* yang terdiri dari kumpulan objek yang serupa atau satu sama lain tetapi sama dengan objek di *cluster* lain. Algoritma hirarkis menemukan *cluster* secara berurutan dilokasi yang telah ditetapkan sebelumnya dan algoritma *partitionanl* menentukan lokasi spesifik dari setiap kelompok [8].

Menurut Freditasari (2022) Dalam hal pengelompokkan k-medoids berdasarkan faktor yang mempengaruhi kemiskinan di kab/kota aceh, hasil penelitian menunjukkan bahwa ada dua *cluster* yang ideal, *Cluster 1* terdiri dari 11 kabupaten sedangkan *cluster 2* terdiri dari 12 kabupaten. *Cluster 1* memiliki *persentase* penduduk yang lebih tinggi daripada *cluster 2*, tetapi *cluster 3* memiliki *Gini Ratio* AHH dan RLS yang lebih rendah daripada *cluster 1* [9]. Menurut penelitian yang dilakukan oleh Kumarahadi (2023) mengenai “Penerapan Metode *Hierarchical Clustering* Untuk Pengelompokkan Kota Atau Kabupaten Di Indonesia Berdasarkan Indikator Kemiskinan” hasilnya menunjukkan bahwa metode *agglomerative nesting average linkage* merupakan metode yang lebih baik karena memiliki *cophenetic value* dan *silhouette coefficient value* yang lebih besar yaitu 0,090 dan 0,71. hasil pengelompokkan terdiri dari dua *cluster*, *cluster 1* berisi 493 wilayah dengan kemiskinan rendah dan *cluster 2* berisi 21 wilayah dengan kemiskinan tinggi [7]. Menurut Fikri (2021) *clustering K-Means* digunakan untuk mengumpulkan data tentang tingkat kemiskinan di setiap kabupaten atau kota di Indonesia. Hasil adalah 3 dimensi baru PCA 0, PCA 1, PCA 2. Metode K-Means menggunakan proporsi kumulatif dan minimum persentase keragaman data sebesar 80% [10]

Berdasarkan latar belakang diatas, Penelitian sebelumnya menunjukkan bahwa penelitian ini hanya mengacu pada 1 atau 2 Kab/Kota saja. Dengan adanya penelitian ini, diharapkan bahwa pemerintah dapat mengawasi distribusi bantuan ke kab/kota dengan populasi miskin yang tinggi, yang berdampak pada pertumbuhan ekonomi Indonesia. Pada penelitian ini menggunakan teknik data *mining*.

2) METODE PENELITIAN

Pada tahap ini bertujuan untuk mengatur proses penelitian secara terstruktur. Metode penelitian memiliki manfaat untuk mengumpulkan data, menganalisis dang menginterpretasikan hasil penelitian. Adapun tahapan metode penelitian sebagai berikut:



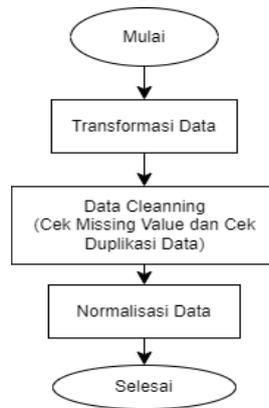
Gambar 1: Alur Penelitian

2.1 Pengumpulan Data

Penelitian menggunakan 580 data tingkat kemiskinan di Indonesia yang berasal dari *platform Badan Pusa Statistik* yang terdiri dari 4 atribut.

2.2 Preprocessing

Preprocessing dilakukan untuk menyelesaikan masalah, meningkatkan kualitas data, dan menyediakan data yang digunakan dalam proses analisis atau pelatihan model. Berikut ini adalah prosedur *preprocessing* yang digunakan dalam penelitian ini.:



Gambar 2. Tahapan *Preprocessing*

1) Transformasi Data

Transformasi adalah perubahan distribusi atau bentuk data. Beberapa jenis transformasi termasuk mengubah ukuran data Transformasi log untuk mengubah distribusi data dan transformasi akar kuadrat untuk mengubah data ke kumpulan data tertentu (*Scaling*)

2) *Data Cleaning*

Proses ini membersihkan data untuk memastikan data yang digunakan berkualitas tinggi adalah tujuan utama dari pembersihan data. Berikut adalah langkah-langkah dari pembersihan data:

- a. Menampilkan jumlah *missing value*,
- b. menghapus data *missing value*,
- c. menampilkan jumlah baris sebelum dan sesudah pengolahan data
- d. Cek duplikasi data

3) Normalisasi Data

Pada tahap ini untuk memastikan konsistensi, kualitas, dan efisiensi dalam pengolahan data. Adapun metode normalisasi yang digunakan sebagai berikut:

Min-Max Normalisasi:

$$\text{Nilai dinormalisasi} = \frac{\text{Nilai} - \text{Min}}{\text{Maks} - \text{min}} \quad (1)$$

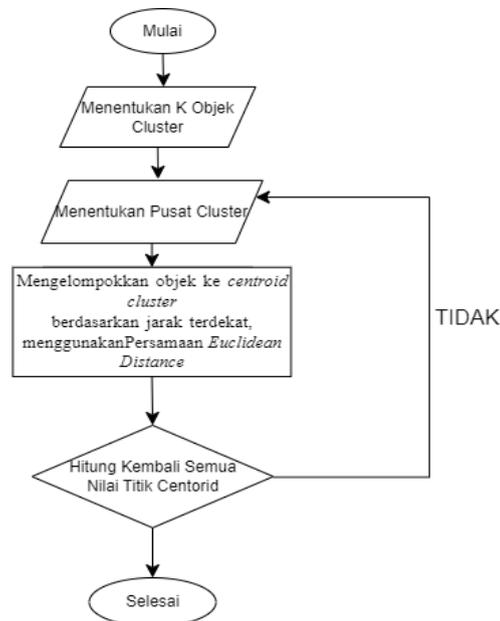
Metode ini mengubah nilai atribut menjadi rentang nilai 0 hingga 1. Nilai min yaitu 0 nilai maks yaitu 1.

2.3 Implementasi Algoritma

Setelah melakukan tahapan preprocessing langkah selanjutnya adalah menerapkan algoritma *K-Means* dan *K-Medoids*.

1) *K-Means*

Salah satu algoritma pengolahan data yang banyak digunakan yaitu *K-Means*, yang dirancang untuk membagi dataset menjadi *cluster* dan memasukkan data ke dalam kelompok yang paling dekat berdasarkan jarak *Euclidean* [11]. Kemampuan mengagregasi data dari sejumlah besar objek yang singkat dan efisien adalah salah satu kelebihan algoritma *K-means* [12]. Metode *K-Means* membutuhkan beberapa langkah untuk klasterisasi:



Gambar 3: Alur Algoritma *K-Means*

- a) Menentukan K objek *Cluster*. Rumus yang digunakan sebagai berikut:

$$SSE = \sum_{k=1}^K \sum_{x_i \in C_k} (X_i - \varphi_k)^2 \quad (3)$$

Keterangan:

C_k : k *cluster* yang terbentuk

K : Banyaknya *cluster*

X_i : Data x pada fitur ke - i

φ_k : Rata-rata K *cluster* pada nilai k (k= 1,2,3,...K)

- b) Menentukan pusat *cluster*

- c) Menghitung jarak Menggunakan persamaan *Euclidean Distance* dalam mengelompokkan objek ke kelompok centroid berdasarkan jarak terdekat:

$$D(i, j) = \sqrt{(X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (4)$$

Dimana:

$D(i, j)$: jarak data i ke pusat *cluster* j

X_{ki} : Data ke i pada atribut data ke k

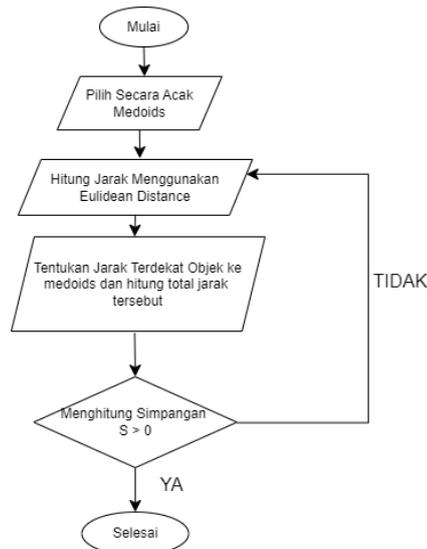
X_{kj} : Jarak antara titik pusat ke j pada atribut k

- d) Menghitung ulang semua nilai titik *centroid*

- e) Ulangi langkah b, c, d hingga titik centroid tidak berubah lagi.

2) *K-Medoids*

Kumpulan n objek dimasukkan ke dalam *cluster* k dengan metode partisi *clustering K-Medoids*. Objek yang dekat dengan pusat klaster dikumpulkan untuk membentuk *cluster* baru. Meskipun data diacak, hasil pengelompokan *K-Medoids* tidak berubah. [13]. Metode ini memulai dengan menghitung jarak antara semua kombinasi *medoids* dalam setiap kelompok yang mungkin untuk menemukan titik representatif pada dataset. [14]. Berikut adalah tahapan-tahapan dari algoritma *K-Medoids*:

Gambar 4: Alur Algoritma *K-Medoids*

- Dari n data, pilih medoid awal sebanyak k secara acak.
- Tentukan jarak antara masing-masing objek dan *medoid* sementara, menggunakan persamaan yang digunakan.:

$$c) \quad D(i, j) = \sqrt{(X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (5)$$

Dimana:

$D(i, j)$: jarak data i ke pusat *cluster* j

X_{ki} : Data ke i pada atribut data ke k

X_{kj} : Jarak antara titik pusat ke j pada atribut k

- Tentukan jarak objek terdekat ke *medoid* dan hitung jarak totalnya.
- Untuk menghitung total simpangan (S) hitung total jarak lama dan hitung nilai total jarak baru. Untuk membuat *medoid* dari sekumpulan k objek baru tukar objek dengan data *cluster* jika S kurang dari 0.

2.4 Evaluasi

Sillhouette coefficient adalah metrik evaluasi *cluster* yang digunakan untuk menentukan jumlah *cluster* yang ideal pada dataset. Nilai rata-rata positifnya berkisar antara -1 dan 1 nilai rata-rata negatif menunjukkan bahwa objek lebih cocok pada dataset [15]. Berikut persamaan yang digunakan:

$$Sill(c) = Sill(k) \frac{1}{|k|} \sum_{i=1}^k Sill(c_i) \quad (6)$$

Dimana :

$Sill(k)$: Nilai *Sillhouette* semua *cluster*

$|k|$: Banyak *cluster*

$Sill(C_i)$: Rata-rata nilai *sillhouette*

3) HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang dianalisis dalam penelitian ini merupakan data yang di ambil dari *website BPS.go.id* https://www.bps.go.id/id/statistics-table/2/NjIxIzI=/persentase-penduduk-miskin--p0--menurut_kabupaten-kota.html pada bulan Januari tahun 2024 dataset yang diperoleh berjumlah 578 dengan 3 kolom. Data yang digunakan sebagai berikut Nama Wilayah, Persentase penduduk miskin menurut kabupaten/kota dari tahun 2022-2024 Dataset yang digunakan dapat dilihat pada gambar berikut.

```
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Nama Wilayah    578 non-null    object
1   Unnamed: 1      580 non-null    object
2   Unnamed: 2      579 non-null    object
3   Unnamed: 3      579 non-null    object
dtypes: object(4)
```

Gambar: 5 Dataset yang gunakan

Gambar 5 menunjukkan dataset yang digunakan, data diambil dari *platform bps.go.id* kemudian di proses dengan melakukan tnsformasi data, *cleanning data*, dan menerapkan algoritma *K-Means* dan *K-Medoids*

3.1 Preprocessing

Tahapan *preprocessing* mencakup, transformasi data, data *cleaning* (duplikasi data dan memeriksa *missing value*) normalisasi data.

1) Transformasi Data

Pada langkah ini, nilai x1 hingga x10 diubah pada label atribut atau *variable*, yang dapat dilihat sebagai berikut:

Tabel 1. Tabel Transformasi

Atribut	Transformasi
Nama Wilayah	Provinsi
Unamed (2022)	X1
Unamed (2023)	X2
Unamed (2024)	X3

2) Data *Cleanning*

Pada tahap ini untukmemastikan bahwa data yang akan klasterisasi berkualitas tinggi dan relevan. Ini dimulai dengan menghitung berapa banyak nilai yang hilang pada dataset yang akan digunakan:

```
Index: 522 entries, 2 to 556
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Nama Wilayah    522 non-null    object
1   Unnamed: 1      522 non-null    float64
2   Unnamed: 2      522 non-null    float64
3   Unnamed: 3      522 non-null    float64
dtypes: float64(3), object(1)
```

Gambar 6. Hasil Cek *Missing Value*

Selanjutnya, menghapus kelebihan data atau duplikasi data dalam satu atau lebih entitas data dalam upaya mengurangi jumlah data yang berlebihan.

```

Jumlah missing values per kolom:
Nama Wilayah      2
Unnamed: 1        31
Unnamed: 2        31
Unnamed: 3        27
dtype: int64

```

Gambar 7. Hasil Cek Duplikasi Data

Pada gambar 7 menunjukkan bahwa dataset tidak memiliki nilai yang hilang. Hasil cek duplikasi data menunjukkan bahwa tidak ada baris duplikat dalam dataset.

3) Normalisasi Data

Tahap preprocessing berikutnya, normalisasi data. Tahap ini menggunakan normalisasi min-max untuk memastikan bahwa data disimpan secara efisien dan konsisten, seperti yang ditunjukkan pada tabel 2 di bawah ini.

Tabel 2. Hasil Normalisasi *Min-Max*

Kab/Kota	X1	X2	X3
Simuele	0.4226	0.7063	0.1593
Aceh Barat	0.4577	0.6362	0.2410
Aceh Barat D	0.2749	0.6538	0.2111
Aceh jaya	0.2808	0.7230	0.2035
Aceh selatan	0.3072	0.5950	0.2310
Aceh	0.3279	0.7397	0.3417
....
Yalimo	0.7858	0.1393	0.0345

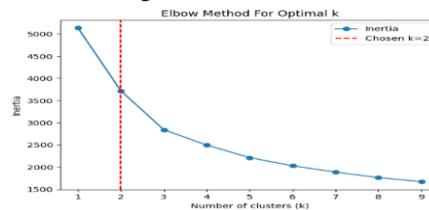
3.2 Implementasi Algoritma

1) Algoritma *K-Means*

Metode *K-Means* digunakan untuk menghitung jumlah *cluster*. Salah satu kelebihan algoritma *k-means* adalah kemampuan untuk mengelompokkan dan pencilaan objek dengan cepat. Salah satu kelemahan algoritma *k-means* adalah bahwa itu sangat sensitif terhadap pembangkitan acak titik pusat awal *cluster* dan hasil pengelompokan yang selalu berubah.

a) Tentukan Jumlah K

Jarak kuadrat jarak antara pusat *cluster* terdekat dan titik data disebut ukuran inerti. Jumlah *cluster* (k) ditampilkan pada sumbu x grafik, dan nilai inerti ditampilkan pada sumbu y.

Gambar 8. Hasil K optimal dengan Metode *Elbow*

Pada gambar 8 menunjukkan bahwa kuadran yang menurun sehingga membentuk sudut yang lebih dekat dengan siku-siku dibandingkan kuadran lainnya, nilai $k = 2$ adalah nilai yang paling cocok untuk klusterisasi data. Tabel 4 menunjukkan langkah selanjutnya untuk menentukan pusat kelompok.

b) Melakukan Inisialisasi Pusat *Cluster*

Hasil klusterisasi yang dihasilkan dapat dipengaruhi oleh titik awal yang dipilih secara acak. Tabel 3 berikut menunjukkan hasil pusat *cluster*.

Tabel 3. Pusat *Cluster*

<i>Cluster</i>	Kab/Kota	X1	X2	X3
0	Sampang	21.61	21.76	20.83
1	Wonosobo	16.17	15.58	15.28

Pada tabel 3 menunjukkan hasil dari pusat *cluster*, yang merupakan representasi rata-rata dari data dari masing-masing *cluster*. Dalam algoritma *clustering* seperti *K-Means* inialisasi pusat *cluster* adalah langkah awal, setelah itu algoritma akan mengiterasi untuk memperbaiki posisi pusat *cluster* hingga mencapai konvergensi.

c) Menentukan Jarak Setiap Data Ke Setiap *Centroid*

Hasil perhitungan diulang tiga kali sampai titik *centroid* tidak berubah lagi ini dilakukan dengan menghitung jarak menggunakan persamaan (4). Setelah mengetahui jarak, titik *centroid* baru dihitung. Tabel 4 dan 5 menunjukkan hasil iterasi.

Tabel 4. Hasil Perhitungan Jarak

No	Kab/Kota	X1	X2	X3	...	C 1	C 2	Jarak Terdekat	<i>Cluster</i>
1	Simuele	18.37	17.92	17.69	...	0.132	41.47	0.13	0
2	Aceh Singkil	0.355	0.635	19.06	...	0.156	41.44	0.156	1
...
580	Pegunungan Bintang	0.785	0.139	0.001	...	9.900	9.421	9.421	1

Setelah mengetahui hasil pengelompokkan pada tabel 7, langkah berikutnya adalah menentukan kategori tingkat kemiskinan rendah, sedang, dan tinggi. Dengan demikian hasil akhir pengelompokkan didapatkan yang terdapat pada tabel 8.

d) Memperbarui Nilai *Centroid*

Hasil total jumlah atribut dibagi dengan total jumlah anggota setiap *cluster* untuk menemukan nilai *centroid* baru. Tabel 11 menunjukkan pencarian nilai *centroid* baru.

Tabel 5. Nilai *Centroid* Baru

Centroid	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Kota Banjarmasin	0.12	0.76	0.47	0.81	0.76	0.89	0.96	0.60	0.22	0.11
Ogan Komering Ulu	0.25	0.58	0.28	0.65	0.60	0.76	0.83	0.31	0.32	0.25
Timor Tengah	0.73	0.34	0.08	0.41	0.42	0.32	0.56	0.14	0.59	0.003

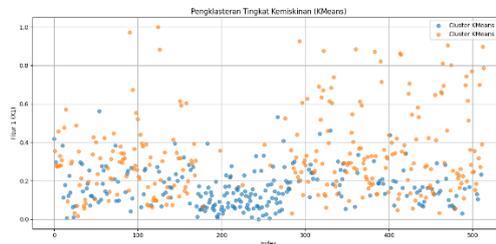
Ulangi langkah 3 dan 4 dengan pusat klaster yang baru digunakan. Karena tidak ada perubahan dalam hasil pengelompokkan pada iterasi kedua dan ketiga, proses dihentikan karena hasil dari *Cluster* 0 (Rendah) sebanyak 250 Kab/Kota, *Cluster* 1 (Tinggi) sebanyak 330 Kab/Kota. Hasil diperoleh dengan melakukan tiga kali perhitungan sampai titik *centroid* tidak berubah lagi.

Tabel 6. Hasil Cluster

<i>Cluster</i> 0	250 Kab/Kota
<i>Cluster</i> 1	330 Kab/Kota

2) Visualisasi Cluster K-Means

Gambar 9 berikut menunjukkan hasil visualisasi *cluster* yang dilakukan menggunakan algoritma *K-Means*.

Gambar 9: Visualisasi *Cluster K-Means*

Dari 580 data, *cluster_0* memiliki 250 Kab/Kota, *cluster_1* memiliki 330 Kab/Kota. Dengan menggunakan pengujian *python*, *clustering* tingkat kemiskinan Kab/Kota di Indonesia menghasilkan hasil *cluster* yang menunjukkan tingkat kemiskinan rendah dan tinggi.

3) Algoritma *K-Medoids*

Algoritma *K-Medoids* dibuat untuk mengatasi kelemahan algoritma *K-Means* yang sensitif terhadap *outlier*.

a) Inisialisasi *Medoid* Secara Acak

Inisialisasi pusat *cluster* secara acak dalam algoritma *K-Medoids* merupakan tahapan awal yang penting dalam proses pengelompokan data. Tabel berikut menunjukkan inisialisasi pusat *cluster* secara acak dalam algoritma *K-Medoids*.

Tabel 7. Inisialisasi medoid Acak

	Kabupaten_Kota	X1	X2	X3	Rata-Rata
0	Lampung Barat	0.265	0.582	0.300	0.623
1	Belitung	0.121	0.617	0.481	0.443

Pada tahap inisialisasi *k* medoid acak dari dataset dipilih. Dalam tabel 7 ini, $k = 2$ Inisialisasi *Medoid (Centroid)* dipilih secara acak dari tiga *medoid* awal dari baris-baris ini. *Medoid* Lampung Barat memiliki nilai rata-rata 0,623, Belitung memiliki nilai rata-rata 0,443.

b) Hitung Jarak Masing-Masing Objek Ke *Medoids*

Menghitung nilai jarak terdekat (*cost*) menggunakan persamaan *Euclidean*. Seperti tabel 8 berikut.

Tabel 8. Hasil Perhitungan Jarak *Euclidean*

No	Kab/Kota	Distance_Medoids_1	Distance_Medoids_2	Terdekat	Cluster
1	Aceh Barat	1.382	1.529	1.382	0
2	Aceh Barat D	0.759	1.467	0.759	1
3	Aceh Besar	1.570	0.852	0.852	0
...
580	Yalimo	1.811	1.309	2.45	1
Total Cost				452.25	

Dapat dilihat pada tabel 8 hasil iterasi 1 dari perhitungan menggunakan *Euclidean* yang memperoleh total *cost* yaitu 452.25

c) Memilih sebuah objek non *medoids* baru

Setelah diperoleh hasil jarak (*cost*) pada perulangan ke – 1 selanjutnya ke perulangan ke – 2, terdapat pada tabel 9 berikut.

Tabel 9. *Medoids* Baru

No	Kab/Kota	<i>Distance_Medoids_1</i>	<i>Distance_Medoids_2</i>	Terdekat	Cluster
1	Aceh Barat	0.443	0.451	0.443	0
2	Aceh Barat D	0.328	0.598	0.328	1
3	Aceh Besar	0.475	0.338	0.338	0
...
580	Yalimo	4.300	5.786	2.45	1
Total Cost				201.38	

Tabel 9 berikut menunjukkan hasil jarak (*cost*) pada perulangan ke-1, kemudian perulangan ke-2, dan medoid terkini (*non-medoid*) pada perulangan ke-2. Pada tabel 10 menunjukkan hasil *cluster* dari hasil proses iterasi.

Tabel 10. Nilai Medoid Baru

Cluster	Pusat Cluster	Jumlah Anggota
0	Musi Rawas	270
1	Berau	310

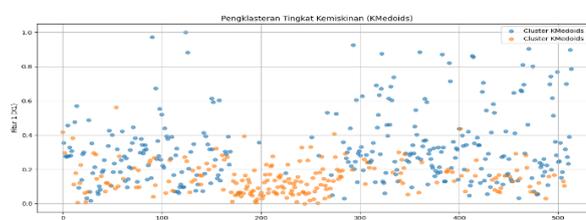
- d) Hitung Total Simpangan (*S*)
Setelah diperoleh hasil dari perulangan ke - 1 dan perulangan ke -2. Jumlah simpangan keseluruhan (*S*) dapat dihitung dengan membandingkan angka *cost* keseluruhan baru dan *cost* keseluruhan lama, baru dapat dilihat sebagai berikut:

$$S = Total\ cost\ baru - Total\ Cost\ Lama \quad (6)$$

$$s = 452.25 - 187.41 = 250.87 \quad (7)$$

Periksa konvergensi: $\Delta D \geq 0$, karena $= 250.87 \geq 0$, maka konvergensi tercapai dan proses dihentikan.

- e) Visualisasi *Clustering K-Medoids*
Gambar 9 di bawah ini menunjukkan hasil visualisasi *cluster* yang dilakukan dengan algoritma *K-Medoids*.

Gambar 10: Visualisasi Hasil *Clustering K-Medoids*

Dari 514 data, *cluster_0* memiliki 301 Kab/Kota, *cluster_1* memiliki 213 Kab/Kota *Clustering* tingkat kemiskinan Kab/Kota di Indonesia yang dilakukan dengan pengujian *python* algoritma *k-medoids* mendapatkan hasil *cluster* nya setelah proses iterasi ketiga yaitu terdapat tingkat rendah dan tinggi.

3.3 Evaluasi

Hasil evaluasi menggunakan nilai *silhouette score* yang menunjukkan seberapa dekat objek dalam *cluster* dengan satu sama lain. Nilai *Silhouette Score* berkisar antara satu dan satu. Seperti pada tabel 11.

Tabel 11. Hasil Evaluasi *Silhouette Score*

Nilai K	<i>Silhouette K-Means</i>	<i>Silhouette K-Medoids</i>
---------	---------------------------	-----------------------------

2	0.22999045128268317	0.22404116016451076
3	0.28469332735307884	0.09237929677125913
4	0.1804929620986072	0.08132443791282348
5	0.16623980770818705	0.09461287942257436
6	0.1854580533158082	0.06784886340815283
7	0.18266466879555834	0.06934871701260083
8	0.18152933130414145	0.040967471961965825
9	0.17247384198001847	0.06151935428656182
10	0.1680059213088525	0.04076191125656086

Pada tabel 11. Merupakan hasil evaluasi *clustering* antara algoritma *K-Means* dan *K-Medoids*. Hasil yang diperoleh *cluster* terbaik berdasarkan *silhouette coefficient* adalah evaluasi menggunakan algoritma *K-Means* dengan $K=3$ memiliki skor tertinggi yaitu 0.284, *K-Medoids* memiliki skor tertinggi pada $k=2$ dengan nilai 0.224, tetapi skor ini lebih rendah daripada *K-Means* pada $k=3$.

4. KESIMPULAN

Clustering Tingkat Kemiskinan menunjukkan tingkat kemiskinan rendah dan tinggi, penelitian yang dilakukan untuk *clustering* tingkat kemiskinan menggunakan algoritma *K-Means* dan *K-Medoids*. Algoritma *K-Means* dan *K-Medoids* menghasilkan temuan baru yang terdiri informasi tingkat kemiskinan berdasarkan dua kelompok. Tingkat kemiskinan yang lebih cenderung memiliki keuangan yang stabil disebut sebagai kemiskinan rendah. Hasil *clustering K-Means* 250 kab/kota dan *K-Medoids* 330 kab/kota. Tingkat kemiskinan tinggi, yang didefinisikan sebagai tingkat kemiskinan yang memiliki karakteristik berbeda. Hasil *clustering K-Means* 270 kab/kota dan *K-Medoids* 310 kab/kota. Hal ini dapat menjadi contoh bagi pemerintah untuk meningkatkan perhatian mereka terhadap daerah dengan tingkat kemiskinan tinggi dalam upaya mengurangi tantangan ekonomi yang sedang berlangsung. Dengan menggunakan skor *silhouette*, untuk membagi tingkat kemiskinan di kab/kota di Indonesia algoritma *K-means* dan *K-medoids* diuji. Sehingga menghasilkan nilai *K-Means* sebesar = 0.284, Nilai *K-Medoids* sebesar = 0.224.

REFERENSI

- [1] M. Rasyida, "Naïve Bayes Classification untuk Penentuan Status Penduduk Miskin," vol. 4, no. 2, 2020.
- [2] A. S. Rahmawati, D. Ispriyanti, and B. Warsito, "Pemodelan Kasus Kemiskinan Di Jawa Tengah Menggunakan Regresi Nonparametrik Metode B-Spline," *J. Gaussian*, vol. 6, no. 1, pp. 11–20, 2017.
- [3] K. Di and P. Riau, "PERBANDINGAN ALGORITMA K-MEANS CLUSTERING DAN K-MEDOIDS DALAM MENGELOMPOKKAN TINGKAT," vol. 8, no. 1, pp. 114–125, 2024.
- [4] F. Zahra, A. Khalif, and B. N. Sari, "Pengelompokan Tingkat Kemiskinan di Setiap Provinsi di Indonesia Menggunakan Algoritma K-Medoids," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 2830–7062, 2024.
- [5] S. Wahyuni and Y. A. Jatmiko, "Pengelompokan Kabupaten/Kota di Pulau Jawa Berdasarkan Faktor-Faktor Kemiskinan dengan Pendekatan Average Linkage Hierarchical Clustering," *J. Apl. Stat. Komputasi Stat.*, vol. 10, no. 1, p. 1, 2019.
- [6] "https://www.bps.go.id/".
- [7] I. Nasution, A. P. Windarto, and M. Fauzan, "Penerapan Algoritma K-Means Dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi," *Build. Informatics, Technol. Sci.*, vol. 2, no. 2, pp. 76–83, 2020.
- [8] K. Aprilia and F. Sembiring, "Analisis Garis Kemiskinan Makanan Menggunakan Metode Algoritma K-Means Clustering," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.)*, pp. 1–10, 2021.
- [9] F. P. Hidayat, R. P. Putra, M. D. Alfitrah, and E. Widodo, "Implementasi Clustering K-Medoids dalam Pengelompokan Kabupaten di Provinsi Aceh Berdasarkan Faktor yang Mempengaruhi Kemiskinan," *Indones. J. Appl. Stat.*, vol. 5, no. 2, p. 121, 2023.
- [10] R. Fikri, A. Mushardiyanto, M. N. Laudza'Banin, K. Maureen, and H. Patria, "Pengelompokan

- Kabupaten/Kota di Indonesia Berdasarkan Informasi Kemiskinan Tahun 2020 Menggunakan Metode K-Means Clustering Analysis,” *Semin. Nas. Tek. dan Manaj. Ind.*, vol. 1, no. 1, pp. 190–199, 2021.
- [11] M. N. Mursidin, A. Ridhalla Azhan, K. Aryasa, and R. A. Djamro, “Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Menggunakan Orange Data Mining (Studi Kasus: Kabupaten/Kota Provinsi Sulawesi Selatan),” *Edisi. Desember*, vol. 277, no. 2, pp. 277–289, 2022.
- [12] Y. Nurohmah, R. Mayasari, and B. Nurina Sari, “Optimalisasi Performa K-Means Clustering Dengan Pca Dalam Analisis Tingkat Kemiskinan Di Jawa Barat,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 3, pp. 1657–1665, 2023.
- [13] N. R. Tangke and A. A. D. Abdullah Ahmad Dzikrullah, “Implementasi K-Medoids Clustering Pada Kemiskinan Ekstrem di Provinsi Maluku,” *Emerg. Stat. Data Sci. J.*, vol. 1, no. 3, pp. 415–424, 2023.
- [14] F. Amelia, I. Iskandar, S. Kurnia Gusti, and E. Haerani, “Clustering Keluarga Miskin Desa Bina Baru Dengan Metode K-Medoids,” *Krea-Tif J. Tek. Inform.*, vol. 11, no. 1, pp. 1–13, 2023.
- [15] Y. I. Kurniawan, “Pengelompokan Prioritas Negara Yang Membutuhkan Bantuan Menggunakan Clustering K-Means dengan Elbow dan Silhouette,” *J. Pendidik. dan Teknol. Indones.*, vol. 3, no. 10, pp. 455–463, 2023.