

Penerapan Algoritma *K-Medoids* pada Pengelompokan Kabupaten/Kota di Provinsi Kalimantan Timur Berdasarkan Indikator Kemiskinan

Saderisa^{a)}, Syaripuddin^{b)}, Indriasri Raming^{c)}

Program Studi Matematika, Universitas Mulawarman, Jln. Barong Tongkok, Samarinda, 75123, Indonesia

Email: ^{a)} saderisabaun12@gmail.com, ^{b)} syarifrahman2014@gmail.com, ^{c)} indriasriraming@fmipa.unmul.ac.id

Abstract

The K-Medoids algorithm is a partition-based method that uses representative objects called medoids as the center or centroid points. The K-Medoids algorithm is chosen because it is capable of clustering data with outliers and is more stable compared to the K-Means algorithm. This study aims to cluster regencies/cities in East Kalimantan Province based on poverty indicators using the K-Medoids algorithm. The data used are poverty indicator data from the year 2021. The variables include the number of people living in poverty, the number of pharmacies, the number of villages with income from agriculture, plantations, and fisheries, the number of villages with educational activities for packages A/B/C, the number of villages still using firewood, and the number of villages using wells as a water source. The clustering process was conducted using the K-Medoids algorithm to form 2 and 3 clusters. The research results indicate that effective clustering consists of 2 clusters, where the first cluster comprises 7 districts/cities and the second cluster consists of 3 districts/cities. The validation of the clustering results is conducted using the Silhouette Coefficient value, which shows the highest value for clustering quality, namely the 2 cluster grouping with a Silhouette Coefficient value of 0.6354268, indicating that the resulting cluster structure is a Medium Structure (a good structure). This study is expected to provide valuable information for the government in formulating policies to address poverty in East Kalimantan Province.

Keywords: *Clustering, K-Medoids Algorithm, Silhouette Coefficients, Poverty Indicators, and Data Mining.*

Abstrak

Algoritma *K-Medoids* merupakan metode berbasis partisi yang menggunakan objek representatif yang disebut *Medoids* sebagai titik pusat atau *centroid*. Algoritma *K-Medoids* dipilih karena mampu mengelompokkan data dengan *outlier* dan lebih stabil dibandingkan algoritma *K-Means*. Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Provinsi Kalimantan Timur berdasarkan indikator kemiskinan menggunakan algoritma *K-Medoids*. Data yang digunakan adalah data indikator kemiskinan tahun 2021. Variabel yang digunakan yaitu jumlah penduduk miskin, jumlah apotek, jumlah desa yang berpenghasilan dari sektor pertanian, perkebunan, dan perikanan, jumlah desa yang memiliki kegiatan pendidikan paket A/B/C, jumlah desa yang masih menggunakan kayu bakar, dan jumlah desa yang memiliki sumur sebagai sumber air. Proses pengelompokan dilakukan dengan menggunakan algoritma *K-Medoids* untuk membentuk 2 dan 3 klaster. Hasil penelitian menunjukkan bahwa pengelompokan yang baik adalah 2 klaster, di mana klaster pertama terdiri dari 7 kabupaten/kota dan klaster kedua terdiri dari 3 kabupaten/kota. Validasi hasil pengelompokan dilakukan dengan menggunakan nilai *Silhouette Coefficient* yang menunjukkan nilai terbesar pada kualitas pengelompokan yaitu

pengelompokan 2 klaster dengan nilai *Silhouette Coefficient* sebesar 0,6354268 yang menyatakan bahwa struktur klaster yang dihasilkan adalah *Medium Structure* (struktur yang baik). Penelitian ini diharapkan dapat memberikan informasi yang berguna bagi pemerintah dalam menyusun kebijakan untuk mengatasi kemiskinan di Provinsi Kalimantan Timur.

Kata-kata kunci: *Clustering*, Algoritma *K-Medoids*, *Silhouette Coefficients*, Indikator Kemiskinan, dan Data Mining.

PENDAHULUAN

Kemiskinan merupakan masalah global yang dihadapi hampir semua negara, termasuk Indonesia. Kemiskinan dapat didefinisikan sebagai kondisi ketika seseorang atau kelompok tidak mampu memenuhi kebutuhan dasar mereka (Hendayati & Nurhidayanti, 2021). Pandemi COVID-19 yang melanda Indonesia sejak awal 2020 memperburuk tingkat kemiskinan, termasuk di Provinsi Kalimantan Timur. Pandemi ini memengaruhi kondisi ekonomi dan sosial sejak akhir 2019, menyebabkan penurunan ekonomi daerah dan memperparah kondisi masyarakat. Untuk mengatasi penyebaran virus, pemerintah daerah menerapkan Pembatasan Sosial Berskala Besar (PSBB), yang membatasi aktivitas ekonomi, pendidikan, dan sosial masyarakat.

Pandemi COVID-19 berdampak signifikan pada berbagai sektor, seperti perdagangan, nilai tukar, dan aktivitas bisnis yang mengalami penurunan tajam. Dampaknya meliputi berkurangnya tenaga kerja, meningkatnya pengangguran, penurunan pendapatan, perubahan pola konsumsi, serta kerentanan masyarakat terhadap penyakit dan perubahan ekonomi. Kebijakan pembatasan sosial memengaruhi seluruh lapisan masyarakat, terutama kelompok berpenghasilan rendah dan pekerja harian, menyebabkan banyak individu yang sebelumnya tidak miskin menjadi miskin akibat pembatasan tersebut. Di Provinsi Kalimantan Timur, tingkat kemiskinan menunjukkan ketidakstabilan selama periode 2019–2021. Menurut Badan Pusat Statistik (BPS), persentase penduduk miskin pada September 2019 adalah 5,91%, meningkat menjadi 6,64% pada September 2020, dan sedikit turun menjadi 6,27% pada September 2021. Data BPS hanya menyajikan persentase kemiskinan per kabupaten/kota, sehingga diperlukan pengelompokan wilayah berdasarkan tingkat kemiskinan untuk menentukan prioritas penanggulangan (Bahauddin et al., 2021).

Untuk mengatasi masalah kemiskinan di Kalimantan Timur, pengelompokan wilayah dapat dilakukan menggunakan metode *clustering* dalam data mining. *Clustering* mengelompokkan data ke dalam klaster berdasarkan kemiripan atau perbedaan karakteristik. Data mining memungkinkan ekstraksi informasi penting untuk meningkatkan efisiensi dan mengurangi biaya (Rofiq et al., 2020). Algoritma *K-Medoids*, salah satu metode *clustering*, digunakan untuk mengelompokkan data dengan lebih baik dibandingkan *K-Means* karena mampu menangani data dengan pencilan (*outlier*) dan tidak bergantung pada nilai rata-rata klaster (Kamila et al., 2019). Data dengan karakteristik serupa dikelompokkan dalam satu klaster, sedangkan yang berbeda ditempatkan pada klaster lain (Sanusi et al., 2022).

Penelitian ini menerapkan algoritma *K-Medoids* untuk mengelompokkan kabupaten/kota di Kalimantan Timur berdasarkan indikator kemiskinan tahun 2021, dengan validasi pengelompokan menggunakan *Silhouette Coefficient* untuk 2 dan 3 klaster. Penelitian ini bertujuan memperluas pemahaman tentang algoritma *K-Medoids*, menjadi referensi untuk penelitian *clustering*, dan menyediakan data kemiskinan dari tingkat tertinggi hingga terendah.

Algoritma *K-Medoids*

Algoritma *K-Medoids* merupakan metode partisi yang menggunakan objek representatif, disebut *Medoid*, sebagai pusat klaster. Algoritma ini berfokus pada meminimalkan total ketidaksesuaian antara setiap objek dan *Medoid* terdekatnya, dengan pengelompokan objek berdasarkan tingkat kemiripan tertinggi yang dihitung melalui jarak antar data.

Berbeda dengan algoritma *K-Means* yang kurang efektif dalam menangani data dengan pencilan (*outlier*), *K-Medoids* menggunakan *Medoid* sebagai titik pusat klaster. Pendekatan ini memungkinkan proses partisi dilakukan dengan meminimalkan ketidaksamaan antara objek dan *Medoid* sebagai inti

algoritma. *K-Medoids* bekerja dengan membagi q objek ke dalam K kluster. Setelah menetapkan K -*Medoids* awal, algoritma melakukan iterasi untuk menemukan *Medoid* yang lebih optimal dengan mengevaluasi semua kemungkinan pasangan objek, di mana satu objek dipilih sebagai *Medoid* dan yang lainnya tidak. Kualitas pengelompokan diukur untuk setiap pasangan, dan *Medoid* terbaik dari satu iterasi digunakan untuk iterasi berikutnya. Adapun tahapan-tahapan dari algoritma *K-Medoids* adalah sebagai berikut:

1. Memilih secara acak objek sebanyak K sebagai objek representatif o_m (*Medoids*).
2. Menghitung jarak *Euclidean* untuk setiap objek terhadap masing-masing *Medoids* seperti dinyatakan oleh Persamaan (1) sebagai berikut:

$$d(x_{ij}, o_{mj}) = \sqrt{(x_{i1} - o_{m1})^2 + (x_{i2} - o_{m2})^2 + \dots + (x_{iq} - o_{mq})^2}, \quad (1)$$

dengan $d(x_{ij}, o_{mj})$ adalah jarak dari data ke- i pada variabel ke- j terhadap *Medoids* ke- m pada variabel ke- j dimana $m = 1, 2, \dots, K$ serta $j = 1, 2, \dots, q$.

3. Menetapkan setiap objek ke gugus yang sesuai dengan *Medoids* terdekat dan menghitung total jarak minimum yang merupakan jumlah ketidakmiripan dari semua objek ke *Medoids* terdekat berdasarkan jarak antara objek terhadap setiap *Medoids* yang paling terdekat.
4. Memilih secara acak objek yang tidak representatif o_h (*Non-Medoids*).
5. Menghitung jarak *Euclidean* untuk setiap objek terhadap masing-masing *Non-Medoids* seperti dinyatakan oleh Persamaan (2) sebagai berikut:

$$d(x_{ij}, o_{hj}) = \sqrt{(x_{i1} - o_{h1})^2 + (x_{i2} - o_{h2})^2 + \dots + (x_{iq} - o_{hq})^2}, \quad (2)$$

dengan $d(x_{ij}, o_{hj})$ adalah jarak dari data ke- i pada variabel ke- j terhadap *Non-Medoids* ke- h pada variabel ke- j dimana $h = 1, 2, \dots, n - K$ serta $j = 1, 2, \dots, q$.

6. Menetapkan setiap objek ke gugus yang sesuai dengan *Non-Medoids* terdekat dan menghitung total jarak minimum yang merupakan jumlah *dissimilarity* dari semua objek ke *Non-Medoids* terdekat berdasarkan jarak antara objek terhadap setiap *Non-Medoids* yang paling terdekat.
7. Menghitung selisih dari total jarak minimum dengan cara mengurangi total jarak minimum *Non-Medoids* dengan total jarak minimum *Medoids* dan apabila hasil selisih total jarak minimum lebih dari nol maka perhitungan algoritma *K-Medoids* dihentikan.
8. Mengganti *Medoids* o_m dengan *Non-Medoids* o_h apabila pertukaran seperti mengurangi total jarak minimum.
9. Ulangi langkah (4-8) sehingga tidak ada lagi perubahan *Medoids*.

Silhouette Coefficients

Salah satu metode evaluasi yang dapat digunakan untuk melihat kualitas dan kekuatan kluster adalah metode *Silhouette Coefficient*. Metode ini merupakan metode validasi kluster yang menggabungkan metode *Cohesion* dan *Separation*. Tahapan perhitungan *Silhouette Coefficient* adalah sebagai berikut:

1. Menghitung rata-rata jarak dari suatu data ke- i dengan semua data yang berada pada satu kluster yang sama dengan menggunakan Persamaan (3),

$$a_i = \frac{1}{n_p - 1} \sum_{r=1}^{n_p - 1} d_{i,r}, r \neq i, \quad (3)$$

dengan p merupakan anggota kluster, $p = 1, 2, \dots, K$.

- Menghitung rata-rata jarak suatu data ke- i dengan semua data yang berada pada kluster yang berbeda dengan menggunakan Persamaan (5), kemudian ambil nilai terkecilnya berdasarkan Persamaan (4)

$$b_i = \min\{d_i(p)\}, r \neq i \quad (4)$$

dengan rumus jarak suatu data ke- i dengan semua data pada kluster yang berbeda adalah

$$d_i(p) = \frac{1}{n_p} \sum_{r=1}^{n_p} d_{i,r}, r \neq i, \quad (5)$$

dengan $p = 1, 2, \dots, K$.

- Menghitung nilai *Silhouette Coefficient* untuk setiap data ke- i

$$SC_1(i) = \frac{b_i - a_i}{\max\{a_i, b_i\}}, i = 1, 2, \dots, n, \quad (6)$$

Nilai SC dari sebuah kluster ($SC_2(p)$) diperoleh dengan menghitung rata-rata nilai $SC_1(i)$ semua data yang bergabung dalam kluster tersebut dengan menggunakan Persamaan (7),

$$SC_2(p) = \frac{1}{n_p} \sum_{x_i \in C_p}^{n_p} SC_1(i) \quad (7)$$

setelah itu nilai SC global diperoleh dengan menghitung rata-rata nilai $SC_2(p)$ dari semua kluster dengan menggunakan Persamaan (8).

$$SC = \frac{\sum_{p=1}^K (n_p \times SC_2(p))}{\sum_{p=1}^K n_p}, \quad (8)$$

METODE PENELITIAN

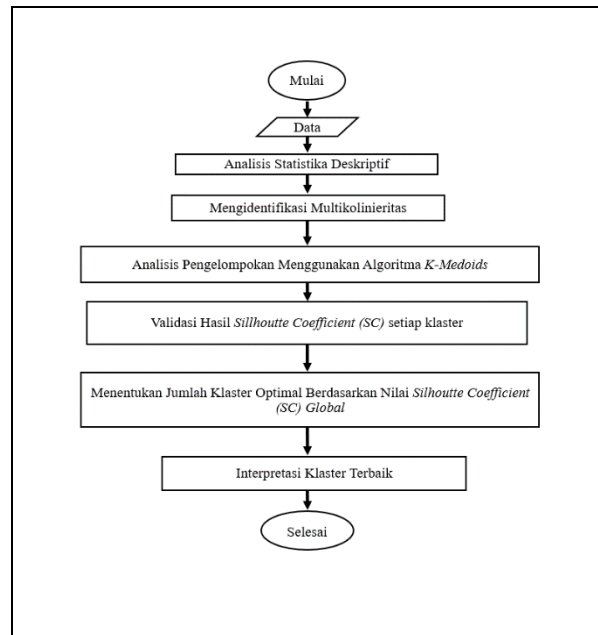
Teknik Analisis Data

Terkait dengan permasalahan dan tujuan ini, maka tahapan analisis data menggunakan algoritma *K-Medoids* adalah sebagai berikut:

- Melakukan analisis statistika deskriptif.
- Mengidentifikasi multikolinieritas dengan menggunakan metode korelasi *pearson*.
- Melakukan pengelompokan data dengan menerapkan algoritma *K-Medoids*.
- Menghitung nilai *Silhouette Coefficient* untuk mengetahui kualitas dari hasil pengelompokan dengan tahapan sebagai berikut:

Kerangka Penelitian

Berikut ini diberikan ilustrasi kerangka pada penelitian melalui diagram *flowchart* yang dapat dilihat Gambar 1.



GAMBAR 1. KERANGKA PENELITIAN

HASIL DAN PEMBAHASAN

Pada bagian ini, kita akan membahas proses perhitungan pengelompokan Kabupaten/Kota di Provinsi Kalimantan Timur dengan menggunakan algoritma *K-Medoids*. Data dalam penelitian ini menggunakan data indikator kemiskinan di Provinsi Kalimantan Timur Tahun 2021. Data tersebut adalah data berupa data sekunder, yaitu jumlah penduduk, miskin, jumlah banyaknya apotek, jumlah banyaknya desa berpenghasilan di sektor pertanian, perkebunan, dan perikanan., jumlah banyaknya desa yang memiliki kegiatan pendidikan paket A/B/C, jumlah banyaknya desa yang masih menggunakan kayu bakar, dan jumlah banyaknya desa yang memiliki sumur sebagai sumber air. Data tersebut disajikan pada Tabel 1.

TABEL 2. Data Penelitian

Data ke- <i>i</i>	Kabupaten/Kota	x_1	x_2	x_3	x_4	x_5	x_6
1	Paser	27.560	32	134	5	5	21
2	Kutai Barat	15.380	35	178	7	6	12
3	Kutai Kartanegara	62.360	114	199	30	1	3
4	Kutai Timur	37.780	48	131	24	2	9
5	Berau	13.620	47	94	17	2	2
6	Penajam Paser Utara	12.130	43	47	10	0	4
7	Mahakam Hulu	3.180	6	50	15	7	2
8	Kota Balikpapan	18.530	127	4	13	0	0
9	Kota Samarinda	42.840	187	12	12	0	0
10	Kota Bontang	8.410	27	3	8	0	0

Berdasarkan data pada Tabel 2, dilakukan analisis multikolinieritas. Salah satu metode untuk mendeteksi multikolinieritas adalah dengan memeriksa nilai koefisien korelasi antar variabel. Nilai koefisien korelasi ini dihitung menggunakan *Software R Studio* dan hasilnya dapat dilihat pada Tabel 3.

TABEL 3. Koefisien Korelasi Antar Variabel

Variabel	x_1	x_2	x_3	x_4	x_5	x_6
x_1	1	0,637	0,482	0,643	-0,308	0,080
x_2	0,637	1	-0,206	0,253	-0,210	0,566
x_3	0,482	-0,206	1	0,400	0,438	0,602
x_4	0,643	0,253	0,400	1	-0,210	-0,308
x_5	-0,308	-0,210	0,438	-0,210	1	0,566
x_6	0,080	0,566	0,602	-0,308	0,566	1

Berdasarkan hasil Tabel 3, terlihat bahwa nilai dari koefisien korelasi antar variabel penelitian yang berbeda bernilai dibawah 0,8 yang berarti tidak terdapat multikolinieritas antar variabel dalam penelitian dan dapat dilanjutkan ke dalam proses pengelompokan dengan algoritma *K-Medoids*.

Analisis pengelompokan menggunakan algoritma *K-Medoids* menentukan jumlah kelompok yang terbentuk dan memerlukan titik pusat awal untuk mengelompokkan objek penelitian. Dalam penelitian ini, pengelompokan dibentuk menjadi 2 dan 3 klaster.

Pertama, proses pengelompokan dengan menggunakan algoritma *K-Medoids* untuk 2 klaster adalah sebagai berikut:

Memilih objek representatif secara acak, sebagai *Medoids* (o_m). Pada bagian ini digunakan jumlah klaster $K = 2$. Pusat klaster dipilih secara acak sebanyak K dari data yang digunakan menggunakan *Software R Studio*. Pusat klaster yang terpilih yaitu *Medoids 1* menggunakan data ke-5 (Berau) dan *Medoids 2* menggunakan data ke-9 (Kota Samarinda). Pusat klaster yang terpilih dapat dilihat Tabel 4.

TABEL 4. *Medoids* (o_{mj}) untuk 2 Klaster

<i>Medoids</i> (o_{mj})	Variabel					
	o_{m1}	o_{m2}	o_{m3}	o_{m4}	o_{m5}	o_{m6}
5	13.620	47	94	17	2	2
9	42.840	187	12	12	0	0

Menghitung jarak *Euclidean* untuk setiap data pengamatan terhadap masing-masing *Medoids* menggunakan Persamaan (1). Hasil data dapat dilihat pada Tabel 5, dengan keseluruhan data hasil perhitungan jarak terhadap masing-masing *Medoids*.

TABEL 5. Data Hasil Perhitungan Jarak Terhadap Masing-Masing Objek *Medoids* (o_{mj}) untuk 2 Klaster

Data ke- i	<i>Medoids</i> (o_{mj})	
	1	2
1	13.940,08	15.281,29
2	1.762,10	27.460,92
3	48.740,16	19.521,04
4	24.160,03	5.063,33
5	0	29.220,45
6	1.490,76	30.710,35
7	10.440,17	39.660,43
8	4.911,47	24.310,07
9	29.220,45	0
10	5.210,84	34.430,37

Menentukan kluster dengan menetapkan setiap objek ke gugus yang sesuai berdasarkan nilai jarak minimum dari kedua jarak *Medoids* yang dihasilkan dan menghitung total jarak minimum yang merupakan jumlah jarak minimum dari semua objek terhadap *Medoids* yang paling terdekat.

TABEL 6. Data Hasil Penempatan Jarak terhadap masing-masing *Medoids* (o_{mj}) untuk 2 kluster

Data ke- i	<i>Medoids</i>		Min	Kluster yang diikuti
	1	2		
1	13.940,08	15.281,29	13.940,08	1
2	1.762,10	27.460,92	1.762,10	1
3	48.740,16	19.521,04	19.521,04	2
4	24.160,03	5.063,33	5.063,33	2
5	0	29.220,45	0	1
6	1.490,76	30.710,35	1.490,76	1
7	10.440,17	39.660,43	10.440,17	1
8	4.911,47	24.310,07	4.911,47	1
9	29.220,45	0	0	2
10	5.210,84	34.430,37	5.210,84	1
Total Jarak Minimum			62.339,82	

Berdasarkan Tabel 6, didapatkan data ke-1 masuk ke kluster 1 karena memiliki jarak terdekat pada *Medoids* 1. Begitu juga sebaliknya, data ke-3 masuk ke kluster 2 karena memiliki jarak terdekat pada *Medoids* 2. Hasil total jarak minimum *Medoids* yaitu sebesar 62.339,82.

Kemudian memilih secara acak *Non-Medoids* untuk 2 kluster menggunakan *Software R Studio*. Pusat kluster yang terpilih yaitu *Non-Medoids* 1 menggunakan data ke-7 (Mahakam Hulu) dan *Non-Medoids* 2 menggunakan data ke-3 (Kutai Kartanegara) dan hasil perhitungan jarak *Euclidean* untuk semua data pengamatan untuk setiap *Non-Medoids* dan kluster yang diikuti sesuai jarak terdekat dapat dilihat pada Tabel 7.

TABEL 7. Data Hasil Penempatan Jarak terhadap masing-masing *Non-Medoids* o_{mj} untuk 2 kluster

Data ke- i	<i>Non-Medoids</i>		Min	Kluster yang diikuti
	1	2		
1	24.380,17	34.800,17	24.380,17	1
2	12.200,71	46.980,08	12.200,71	1
3	59.180,29	0	0	2
4	34.600,12	24.580,18	24.580,18	2
5	10.440,17	48.740,16	10.440,17	1
6	8.950,08	50.230,28	8.950,08	1
7	0	59.180,29	0	1
8	15.350,55	43.830,44	15.350,55	1
9	39.660,43	19.521,04	19.521,04	2
10	5.230,26	53.950,43	5.230,26	1
Total Jarak Minimum			120.653,2	

Setelah memperoleh nilai total jarak minimum *Medoids* dan *Non-Medoids*. Selanjutnya menghitung selisih dari total jarak minimum dengan cara mengurangkan total jarak minimum *Non-Medoids* dengan total jarak minimum *Medoids* sebagai berikut:

$$\begin{aligned} \text{Selisih total jarak minimum} &= \text{Non-Medoids} - \text{Medoids} \\ &= 120.653,2 - 62.339,82 \\ &= 58.313,35 \end{aligned}$$

Hasil selisih total jarak minimum yang didapatkan lebih dari 0, maka perhitungan algoritma *K-Medoids* dihentikan. Berdasarkan hasil perhitungan yang telah dilakukan, maka diperoleh bahwa pengelompokan untuk 2 klaster yang baik adalah pengelompokan dengan *Medoids* pada data pengamatan ke-5 (berau) dan data pengamatan ke-9 (Kota Samarinda) karena memiliki total jarak minimum sebesar 62.339,82 lebih kecil dari total jarak minimum *Non-Medoids*.

Ulangi langkah-langkah diatas pada perhitungan selanjutnya hingga hasil yang diperoleh tidak terdapat perubahan pada *Medoids*. Untuk data hasil perhitungan *K-Medoids* untuk 2 dan 3 klaster dapat dilihat pada Tabel 8 dan 9.

TABEL 8. Data Hasil Perhitungan *K-Medoids* untuk 2 Klaster

Data ke- <i>i</i>	Kabupaten/Kota	Klaster
1	Paser	1
2	Kutai Barat	1
3	Kutai Kartanegara	2
4	Kutai Timur	2
5	Berau	1
6	Penajam Paser Utara	1
7	Mahakam Hulu	1
8	Kota Balikpapan	1
9	Kota Samarinda	2
10	Kota Bontang	1

Berdasarkan Tabel 8, hasil analisis menunjukkan bahwa pengelompokan Kabupaten/Kota di wilayah Provinsi Kalimantan Timur yang masuk ke dalam klaster 1 dengan tingkat kemiskinan terendah yaitu Kabupaten Paser, Kutai Barat, Berau, Penajam Paser Utara, Mahakam Hulu, Kota Balikpapan, dan Kota Bontang. Sedangkan untuk Kabupaten/Kota di wilayah Provinsi Kalimantan Timur yang termasuk klaster 2 dengan tingkat kemiskinan tertinggi yaitu Kabupaten Kutai Kartanegara, Kutai Timur, dan Kota Samarinda.

TABEL 9. Data Hasil Perhitungan *K-Medoids* untuk 3 Klaster

Data ke- <i>i</i>	Kabupaten/Kota	Klaster
1	Paser	1
2	Kutai Barat	2
3	Kutai Kartanegara	3
4	Kutai Timur	1
5	Berau	2
6	Penajam Paser Utara	2
7	Mahakam Hulu	2
8	Kota Balikpapan	2
9	Kota Samarinda	1
10	Kota Bontang	2

Berdasarkan Tabel 9, hasil analisis menunjukkan bahwa pengelompokan Kabupaten/Kota di wilayah Provinsi Kalimantan Timur yang masuk ke dalam kluster 1 dengan tingkat kemiskinan terendah yaitu Paser, Kutai Timur dan Kota Samarinda. Sedangkan untuk Kabupaten/Kota di wilayah Provinsi Kalimantan Timur yang termasuk kluster 2 dengan tingkat kemiskinan sedang yaitu Kabupaten Kutai Barat, Berau, Penajam Paser Utara, Mahakam Hulu, Kota Balikpapan, dan Kota Bontang. Untuk Kabupaten/Kota di wilayah Provinsi Kalimantan Timur yang termasuk kluster 3 dengan tingkat kemiskinan tertinggi yaitu Kabupaten Kutai Kartanegara.

Pada bagian ini, hasil validasi nilai *Sillhoutte Coefficient* (SC) setiap kluster dengan menggunakan algoritma *K-Medoids* untuk 2 dan 3 kluster dapat dilihat pada Tabel 10 dan 12.

TABEL 10. Hasil Validasi Kluster Algoritma K-Medoids Berdasarkan Nilai Silhouette Coefficient untuk setiap Data ke- i untuk 2 Kluster

Data ke- i	Kabupaten/Kota	Kluster	$SC(i)$
1	Paser	1	0.2196502
2	Kutai Barat	1	0.7959339
3	Kutai Kartanegara	2	0.5429409
4	Kutai Timur	2	0.3736704
5	Berau	1	0.8151429
6	Penajam Paser Utara	1	0.8159092
7	Mahakam Hulu	1	0.7131615
8	Kota Balikpapan	1	0.7198178
9	Kota Samarinda	2	0.5720713
10	Kota Bontang	1	0.7859698

Hasil validasi kluster berdasarkan nilai *Sillhoutte Coefficient* setiap kluster ke- p dapat dilihat pada Tabel 11 sebagai berikut:

TABEL 11. Hasil Validasi Kluster Berdasarkan Nilai Silhouette Coefficient setiap Kluster pada 2 Kluster

Kluster ke- p	$SC_2(p)$
1	0.69508361
2	0.49622753

Berdasarkan Tabel 11, dapat diketahui bahwa untuk kluster pertama memiliki nilai SC sebesar 0.69508361 yang artinya kluster tersebut dapat dikatakan *medium structure* (struktur kluster yang baik) dimana pengelompokan objek i berada dalam kluster tepat tetapi memerlukan metode tambahan lainnya untuk memperjelas keanggotaan tiap kluster. Kumpulan objek i yang bergabung dalam kluster yang sama memiliki karakteristik yang mirip. Kluster kedua memiliki nilai SC sebesar 0.49622753 yang artinya kluster tersebut dapat dikatakan *weak structure* (struktur kluster yang lemah) dimana sedikit sekali objek i yang berada dalam kluster yang tepat dan sisanya bergabung pada kluster yang tidak semestinya. Kumpulan objek i yang bergabung dalam kluster yang sama memiliki karakteristik yang berbeda.

Perhitungan nilai *Silhouette Coefficient* global yaitu dengan menghitung rata-rata nilai $SC_2(p)$ dengan Persamaan (8):

$$SC = \frac{n_1 \times SC_2(1) + n_2 \times SC_2(2)}{(n_1 + n_2)}$$

$$\begin{aligned}
 &= \frac{7 \times 0.69508361 + 3 \times 0.49622753}{(7 + 3)} \\
 &= 0.6354268
 \end{aligned}$$

Hasil validasi untuk 2 klaster berdasarkan nilai *SC* global adalah sebesar 0.6354268 artinya pengelompokan secara keseluruhan dapat dikatakan *medium structure* (struktur yang baik) artinya pengelompokan objek *i* berada dalam klaster tepat tetapi memerlukan metode tambahan lainnya untuk memperjelas keanggotaan tiap klaster. Kumpulan objek *i* yang bergabung dalam klaster yang sama memiliki karakteristik yang mirip.

Dilakukan perhitungan *Silhouette Coefficient* seperti langkah 1 sampai dengan langkah 3 untuk *clustering* dengan algoritma *K-Medoids* untuk 3 klaster. Hasil perhitungan *Silhouette Coefficient* dengan bantuan *Software R Studio* dapat dilihat Tabel 12.

TABEL 12. Hasil Validasi Klaster Algoritma *K-Medoids* Berdasarkan Nilai *Silhouette Coefficient* untuk setiap Data ke-*i* untuk 3 Klaster

Data ke- <i>i</i>	Kabupaten/Kota	Klaster	<i>SC(i)</i>
1	Paser	1	0.1871009
2	Kutai Barat	2	0.7355570
3	Kutai Kartanegara	3	0.0000000
4	Kutai Timur	1	0.6891120
5	Berau	2	0.7877436
6	Penajam Paser Utara	2	0.8009674
7	Mahakam Hulu	2	0.6826557
8	Kota Balikpapan	2	0.5443493
9	Kota Samarinda	1	0.4789053
10	Kota Bontang	2	0.7739335

Pada bagian ini dilakukan perbandingan dari 2 pengelompokan dengan jumlah *Medoids* yang berbeda berdasarkan *Silhouette Coefficient* global. Perbandingan hasil validasi klaster dan pengelompokan tersebut dapat dilihat pada Tabel 13.

TABEL 13. Perbandingan Hasil Validasi Klaster Berdasarkan Nilai *Silhouette Coefficient* Global

Jumlah Klaster	<i>SC Global</i>
2	0.6354268
3	0.5680325

Berdasarkan Tabel 13, dapat diketahui nilai *Silhouette Coefficient* untuk validasi data hasil pengelompokan Kabupaten/Kota di wilayah Provinsi Kalimantan Timur berdasarkan indikator kemiskinan dengan menggunakan algoritma *K-Medoids* memiliki nilai *Silhouette Coefficient* yang berbeda pada rentang nilai *SC* yang sama. Nilai *Silhouette Coefficient* terbesar adalah pengelompokan pada 2 klaster yaitu 0.6354268 yang artinya klaster tersebut dapat dikatakan *medium structure* (struktur klaster yang baik) dimana pengelompokan objek *i* berada dalam klaster yang tepat tetapi memerlukan metode tambahan lainnya untuk memperjelas keanggotaan tiap klaster. Kumpulan objek *i* yang bergabung dalam klaster yang sama memiliki karakteristik yang mirip. Oleh karena itu, dapat diputuskan bahwa pengelompokan yang baik pada pengelompokan Kabupaten/Kota di Provinsi Kalimantan Timur dengan menggunakan algoritma *K-Medoids* adalah 2 klaster.

KESIMPULAN

Klaster yang terbentuk pada pengelompokan kabupaten/kota di provinsi Kalimantan Timur berdasarkan kemiskinan dengan menggunakan algoritma *K-Medoids* adalah 2 klaster yaitu klaster 1 dan klaster 2. Klaster 1 adalah tingkat kemiskinan terendah yang beranggotakan 7 kabupaten/kota yaitu kabupaten Paser, Kutai Barat, Berau, Penajam Paser Utara, Mahakam Hulu, Kota Balikpapan, dan Kota Bontang. Sedangkan klaster 2 adalah tingkat kemiskinan tertinggi yang beranggotakan 3 kabupaten/kota yaitu kabupaten Kutai Kartanegara, Kutai Timur dan Kota Samarinda. Nilai *Silhouette Coefficient* untuk validasi data hasil pengelompokan kabupaten/kota di provinsi Kalimantan Timur berdasarkan indikator kemiskinan menggunakan algoritma *K-Medoids* yaitu pada 2 klaster dengan nilai sebesar 0.6354268 yang menyatakan bahwa struktur klaster yang dihasilkan pada pengelompokan ini adalah *Medium Structure* (struktur yang baik). Diharapkan dalam penelitian selanjutnya yaitu membandingkan algoritma *K-Medoids* dan *K-Means* dengan menggunakan pengukuran jarak *Euclidean* dan *Manhattan*.

UCAPAN TERIMAKASIH

Terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam penulisan skripsi ini, khususnya kepada para pembimbing dan penguji yang telah memberikan arahan dan masukan sehingga penelitian ini memperoleh hasil yang optimal dan penelitian ini dapat diselesaikan.

REFERENSI

- Hendayanti, N. P. N. & Nurhidayanti, M. 2021. "Klasifikasi Tingkat Keparahan Kemiskinan Provinsi di Indonesia dengan Analisis Diskriminan". *Jurnal Matematika dan Pendidikan Matematika*. 5(1). 14-21.
- BPS Kalimantan Timur, 2023. https://kaltim.bps.go.id/publication/2022/12/29/4e01eccc92dce7d016d66332/statisti_K-potensi-desa-kalimantan-timur-2021.html diakses Desember 2023.
- Han, J. & Kamber, M. 2006. *Data Mining, concepts and Techniques*. Morgan Kauffman, Walthan.
- Defiyanti, S., Jajuli, M., & Rohmawati, N. 2017. "Optimalisasi *K-Medoids* dalam Pengklasteran Mahasiswa Pelamar Beasiswa dengan Cubic Clustering Criterion". *Jurnal TEKNO SI*, 3(1).
- Kamila, I., Khairunnisa, U., & Mustakim. 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), 119-125.
- Wahyudi, D., Handrizal., Parlina, I., Windarto, A.P., Suhendro, D., & Wanto, A. 2019. "Mengelompokkan Garis Kemiskinan Menurut Provinsi Menggunakan Algoritma *K-Medoids*". *Prosiding Seminar Nasional Riset Informasi Sains (SENARIS)*:452-461.
- Septian, R. 2023. "Penerapan Algoritma *K-Medoids* Pada Pengelompokan Wilayah Provinsi di Indonesia Berdasarkan Indikator Pendidikan". *Skripsi*. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Mulawarman.
- Marlina, D., Putri, N. F., Fernando, A., & Ramadhan, A. 2018. "Implementasi Algoritma *K-Medoids* dan *K-Means* untuk Pengelompokan Wilayah Sebaran Cacat pada Anak". *Jurnal CoreIT*. 4(2), 64-71.