

## Klasterisasi Rumah Tangga Miskin di Kota Samarinda Menggunakan Algoritma K-Modes dengan Validasi Klaster Davies Bouldin Index

### *Clustering of Poor Households in Samarinda City Using K-Modes Algorithm with Davies Bouldin Index Cluster Validation*

Nihayatul Khoiriyah<sup>1</sup>, Memi Nor Hayati<sup>2a)</sup>, Nariza Wanti Wulan Sari<sup>3</sup>, Aji Syarif Hidayatullah<sup>4</sup>

<sup>1,2,3</sup>Program Studi Statistika, Universitas Mulawarman, Indonesia

<sup>4</sup>Program Studi Administrasi Publik, Universitas Widya Gama Mahakam Samarinda, Indonesia

<sup>a)</sup>Corresponding author: [meminorhayati@fmipa.unmul.ac.id](mailto:meminorhayati@fmipa.unmul.ac.id)

#### ABSTRACT

Poverty alleviation is one of goals at Sustainable Development Goals (SDGs) that can be supported by grouping poor households based on similar characteristics, enabling social assistance programs to be delivered more effectively and accurately. This study aims to cluster poor households in Samarinda City using the K-Modes algorithm with validation through the Davies-Bouldin Index (DBI). The data were obtained from the 2023 Samarinda Poverty Survey and consist of 16 poor households indicators. The analysis was conducted by testing the number of clusters (K) from 2 to 10. The results indicate that the optimal number of clusters is 3, with a DBI value of 1.6997. Cluster 1 consists of 11,237 households, Cluster 2 consists of 2,327 households, and Cluster 3 consists of 1,451 households. The distinct characteristics of each cluster suggest that the clustering results can serve as a basis for designing more targeted social assistance programs. Future research is recommended to consider alternative clustering methods for categorical data, such as the ROCK algorithm, which utilizes link-based similarity by considering the number of common neighbors between objects, allowing it to better capture the inherent structure of categorical data compared to distance-based methods

**Keywords:** clustering, dbi, k-modes, poor households

#### 1. Pendahuluan

Sustainable Development Goals (SDGs) merupakan proyek Perserikatan Bangsa-Bangsa (PBB) yang bertujuan meningkatkan kesejahteraan dan kualitas hidup masyarakat dari satu generasi ke generasi berikutnya. SDGs terdiri dari 17 tujuan dengan tujuan pertama adalah tanpa kemiskinan (Bappenas, 2020). Kemiskinan menjadi masalah sosial di negara-negara berkembang seperti Indonesia. Tidak hanya pada tingkat nasional saja, melainkan telah menjadi masalah setiap daerah termasuk Kota Samarinda. Kota Samarinda merupakan Ibu Kota Provinsi Kalimantan Timur yang akan menjadi kota penyangga Ibu Kota Nusantara (IKN). Peran Kota Samarinda sebagai penyangga IKN membuka lapangan kerja baru guna menurunkan persentase kemiskinan (Daton & Utomo, 2022). Untuk mencapai tujuan tersebut, pengentasan kemiskinan menjadi program utama pemerintah Kota Samarinda yang tertulis dalam Rencana Pembangunan Jangka Menengah Daerah (RPJMD) Kota Samarinda tahun 2021 sampai 2026 (Hayati dkk., 2023).

Menurut BPS, tingkat kemiskinan Kota Samarinda pada tahun 2023 mencapai 4,81% yang menunjukkan penurunan jika dibandingkan tahun 2022 yang mencapai 4,85% dan tahun 2021 sebesar 4,99%. Penurunan tersebut menunjukkan adanya upaya pemerintah dalam memberikan program pengentasan kemiskinan namun belum efektif, karena persentase penurunan tingkat kemiskinan relatif kecil (BPS Kota Samarinda, 2023). Hasil survei kemiskinan Kota Samarinda tahun 2023 menunjukkan dari 43.049 rumah tangga yang terdata berhak menerima manfaat, diperoleh 15.849 rumah tangga tergolong sebagai rumah tangga miskin. Sedangkan, 27.200 rumah tangga yang seharusnya tidak menerima bantuan terdata sebagai penerima bantuan pemerintah (Hayati dkk., 2023). Berdasarkan informasi tersebut, data penerima bantuan yang tidak akurat menyebabkan distribusi bantuan dari pemerintah Kota Samarinda tidak tepat sasaran, sehingga tidak semua rumah tangga yang membutuhkan mendapatkan bantuan (Samarinda Pos, 2024).

Pengentasan kemiskinan tidak dapat dilakukan jika hanya berdasarkan informasi jumlah penduduk miskin saja, tetapi memerlukan nama dan alamat rumah tangga yang memenuhi kriteria miskin, agar bantuan yang diberikan sesuai dengan karakteristik dan kebutuhan setiap rumah tangga (Hayati dkk., 2023). Berdasarkan Peraturan Wali Kota Samarinda Nomor 66 Tahun 2023 membagi kriteria rumah tangga miskin berdasarkan aspek hunian dan sosial ekonomi. Kemudian berdasarkan kriteria tersebut dapat dikelompokkan rumah tangga miskin Kota Samarinda berdasarkan kemiripan aspek hunian dan sosial ekonomi agar dapat terlihat kelompok rumah tangga yang menjadi prioritas untuk dilakukan penanganan kemiskinan.

Salah satu metode analisis data yang bertujuan mengelompokkan objek berdasarkan kesamaan karakteristik yang dimilikinya adalah analisis klaster (Han dkk., 2012). Objek yang bergabung dalam satu klaster saling berhubungan antara satu sama lain dan memiliki karakteristik yang berbeda dengan objek yang

berada pada kluster lain (Prasetyo, 2012). Salah satu metode analisis kluster adalah algoritma *K-Modes*. Algoritma *K-Modes* merupakan bentuk pengembangan dari algoritma *K-Means* yang dapat mengelompokkan data bertipe kategorik dengan cara menghitung fungsi jarak *mismatching* (Az-zahra dkk., 2021). Jarak *mismatching* didefinisikan sebagai ukuran ketidakmiripan antara 2 kategori dalam satu variabel, semakin kecil jumlah ketidakmiripan maka semakin identik kedua objek yang diamati (Jain & Dubes, 1988). Menurut Prasetyo (2014), pusat kluster dalam algoritma *K-Modes* berasal dari kategori yang paling banyak muncul dalam variabel kategorik yang disebut modus. Hal ini yang membedakan algoritma *K-Modes* dengan algoritma *K-Means*, algoritma *K-Means* menggunakan nilai rata-rata untuk memperbarui pusat klusternya.

Penerapan algoritma *K-Modes* dengan validasi kluster telah banyak dilakukan di berbagai bidang. Santika dkk. (2025) menggunakan algoritma *K-Modes* untuk mengelompokkan desa berdasarkan indeks ketahanan sosial dengan menggunakan validasi *Silhouette Coefficient* (SC). Hasil penelitian menunjukkan terbentuk empat kluster dengan nilai SC sebesar 0,09 yang menunjukkan kualitas kluster kurang baik. Temuan tersebut menunjukkan bahwa penggunaan SC pada data kategorik memiliki keterbatasan dalam mengevaluasi kualitas kluster secara optimal. Oleh karena itu, diperlukan metode validasi yang lebih sesuai untuk data kategorik, salah satunya *Davies-Bouldin Index* (DBI) yang mempertimbangkan kohesi dan separasi kluster secara simultan.

Penggunaan algoritma *K-Modes* dengan validasi DBI telah diterapkan oleh Riska dkk. (2024) untuk mengelompokkan kasus kebakaran di Provinsi DKI Jakarta dan menghasilkan enam kluster optimal dengan nilai DBI sebesar 6,22. Selain itu, Azahra dkk. (2024) menerapkan metode yang sama untuk mengelompokkan provinsi di Indonesia berdasarkan indikator rumah tangga miskin dan memperoleh dua kluster optimal dengan nilai DBI sebesar 1,94. Meskipun penelitian tersebut berhasil mengidentifikasi kelompok provinsi prioritas penanganan kemiskinan, hasil pengelompokan masih berada pada tingkat wilayah yang luas sehingga belum mampu menggambarkan karakteristik kemiskinan pada level rumah tangga. Namun, karakteristik kemiskinan dapat berbeda antar daerah sehingga memerlukan analisis yang lebih spesifik pada tingkat local atau kota.

Berdasarkan kajian tersebut, masih terdapat kesenjangan penelitian berupa terbatasnya penerapan algoritma *K-Modes* dengan validasi DBI untuk mengelompokkan rumah tangga miskin pada tingkat kota berdasarkan karakteristik sosial ekonomi dan hunian. Selain itu, belum ditemukan penelitian yang secara khusus mengidentifikasi karakteristik rumah tangga miskin di Kota Samarinda menggunakan kombinasi metode tersebut. Oleh karena itu, penelitian ini menerapkan algoritma *K-Modes* dan validasi DBI untuk mengelompokkan rumah tangga miskin di Kota Samarinda berdasarkan 16 indikator kemiskinan guna mengidentifikasi karakteristik setiap kluster sebagai dasar penyusunan program bantuan yang lebih tepat sasaran sesuai kebutuhan masing-masing kelompok rumah tangga.

## 2. Metodologi

### 2.1 Sumber Data

Data penelitian menggunakan data sekunder yang berasal dari Dinas Komunikasi dan Informatika Kota Samarinda yakni data Hasil Survei Kemiskinan Kota Samarinda tahun 2023.

### 2.2 Variabel Penelitian

Variabel yang digunakan merupakan indikator rumah tangga miskin Kota Samarinda dengan rincian yang dapat dilihat pada Tabel 1 sebagai berikut.

Tabel 1. Variabel Penelitian

Notasi	Variabel	Notasi	Variabel
$X_1$	Status penguasaan bangunan tempat tinggal	$X_9$	Tempat pembuangan akhir tinja
$X_2$	Status penguasaan lahan tempat tinggal;	$X_{10}$	Persepsi ekonomi
$X_3$	Jenis lantai terluas	$X_{11}$	Kepemilikan aset tidak bergerak
$X_4$	Jenis dinding terluas	$X_{12}$	Kepemilikan usaha
$X_5$	Jenis atap terluas	$X_{13}$	Pengeluaran rumah tangga per bulan
$X_6$	Sumber penerangan utama	$X_{14}$	Pendapatan rumah tangga per bulan
$X_7$	Bahan bakar/energi utama memasak	$X_{15}$	Status pekerjaan kepala keluarga
$X_8$	Penggunaan fasilitas tempat buang air besar	$X_{16}$	Keikutsertaan program bantuan pemerintah

### 2.3 Teknik Analisis Data

Tahapan analisis data dalam penelitian ini menggunakan bantuan *software* R Studio sebagai berikut:

1. Melakukan proses *data cleaning* dengan menghapus objek pengamatan yang memiliki *missing value* seperti *noise* dan data yang tidak valid.
2. Mengubah variabel numerik ke dalam bentuk kategorik
3. Melakukan pengelompokan menggunakan algoritma *K-Modes* dengan langkah sebagai berikut:

- a. Penentuan jumlah kluster ( $K$ ). Pengujian dilakukan menggunakan nilai  $K$  yakni 2, 3, 4, 5, 6, 7, 8, 9, dan 10 jumlah optimal kelompok rumah tangga miskin di Kota Samarinda belum diketahui. Rentang ini dipilih untuk memperoleh hasil pengelompokan terbaik berdasarkan nilai DBI, sekaligus menjaga agar karakteristik kluster tetap mudah diinterpretasikan.
- b. Menentukan pusat kluster awal secara acak.
- c. Menghitung jarak setiap objek pengamatan terhadap pusat kluster menggunakan *simple matching dissimilarity* (Riska dkk., 2024). Secara matematis dapat dituliskan dengan persamaan (1) dan (2) berikut:

$$d(x_i, c_k) = \sum_{l=1}^m \delta(x_{il}, c_{kl}) \tag{1}$$

dengan,

$$\delta(x_{il}, c_{kl}) = \begin{cases} 0, & x_{il} = c_{kl} \\ 1, & x_{il} \neq c_{kl} \end{cases} \tag{2}$$

dimana,

$d(x_i, c_k)$  : jarak antara objek pengamatan ke- $i$  terhadap pusat kluster ke- $k$

$m$  : banyaknya variabel.

$x_{il}$  : objek pengamatan ke- $i$  pada variabel ke- $l$ .

$c_{kl}$  : pusat kluster ke- $k$  pada variabel ke- $l$ .

$i$  : 1, 2, 3, ...,  $n$ .

$k$  : 1, 2, 3, ...,  $K$

$\delta(x_{il}, c_{kl})$  : nilai ketidakmiripan antara objek ke- $i$  terhadap pusat kluster ke- $k$  pada variabel ke- $l$

- d. Mengalokasikan setiap objek pengamatan ke pusat kluster terdekat berdasarkan kedekatan jarak yakni yang memiliki nilai jarak terkecil.
  - e. Memperbarui pusat kluster menggunakan nilai modus dalam setiap kluster.
  - f. Mengulangi langkah ke- $c$  sampai ke- $e$  hingga tidak ada perubahan pusat kluster atau tidak ada lagi objek pengamatan yang berpindah kluster.
  - g. Mengulangi langkah ke- $c$  sampai ke- $f$  untuk setiap jumlah kluster yang berbeda.
4. Menentukan kluster optimal menggunakan metode *Davies Bouldin Index* (DBI). Metode DBI dipilih karena mampu mengevaluasi kualitas kluster berdasarkan kohesi (kedekatan objek dalam kluster yang sama) dan separasi (jarak antar kluster). Semakin kecil nilai DBI, semakin baik kualitas kluster yang terbentuk sehingga metode ini sesuai digunakan untuk menentukan jumlah kluster optimal pada hasil pengelompokan *K-Modes*. Menurut Riska dkk (2024), tahapan menghitung DBI dapat dilakukan sebagai berikut:

- a. Menghitung nilai *sum square of within* menggunakan persamaan berikut:

$$SSW_k = \frac{1}{n_k} \sum_{i=1}^n \sum_{l=1}^m d(x_{il}, c_k) \tag{3}$$

dengan,

$$d(x_{il}, c_k) = \begin{cases} 0, & x_{il} = c_k \\ 1, & x_{il} \neq c_k \end{cases} \tag{4}$$

dimana,

$SSW_k$  : nilai kohesi dari kluster ke- $k$ .

$n_k$  : jumlah objek pengamatan pada kluster ke- $k$ .

$x_{il}$  : objek pengamatan ke- $i$  pada variabel ke- $l$  dengan  $i = 1, 2, 3, \dots, n$  dan  $l = 1, 2, 3, \dots, m$ .

$c_k$  : pusat kluster ke- $k$  dengan  $k = 1, 2, 3, \dots, K$ .

$d(x_{il}, c_k)$  : jarak antara objek pengamatan ke- $i$  pada variabel- $l$  terhadap pusat kluster ke- $k$ .

- b. Menghitung nilai *sum square of between* menggunakan persamaan berikut:

$$SSB_{k,p} = d(c_k, c_p) \tag{5}$$

dengan,

$$d(c_k, c_p) = \begin{cases} 0, & c_k = c_p \\ 1, & c_k \neq c_p \end{cases} \tag{6}$$

dimana,

- $SSB_{k,p}$  : nilai separasi dari kluster ke- $k$  dan kluster ke- $p$ .
- $d(c_k, c_p)$  : jarak antara pusat kluster ke- $k$  terhadap pusat kluster ke- $p$  dengan  $k = 1, 2, 3, \dots, K$  dan  $p = 1, 2, 3, \dots, K$ .

c. Menghitung nilai rasio menggunakan persamaan berikut:

$$R_{k,p} = \frac{SSW_k + SSW_p}{SSB_{k,p}} \tag{7}$$

keterangan:

- $R_{k,p}$  : rasio antara kluster ke- $k$  dengan kluster ke- $p$ .
- $SSW_k$  : nilai kohesi pada kluster ke- $k$ .
- $SSW_p$  : nilai kohesi pada kluster ke- $p$ .
- $SSB_{k,p}$  : nilai separasi dari kluster ke- $k$  dan kluster ke- $p$ .

d. Menghitung nilai DBI menggunakan persamaan berikut:

$$DBI = \frac{1}{K} \sum_{k=1}^K \max_{k \neq p} (R_{k,p}) \tag{8}$$

keterangan:

- $DBI$  : nilai *Davies Bouldin Index*.
- $K$  : banyaknya kluster.
- $R_{k,p}$  : rasio antara kluster ke- $k$  dengan kluster ke- $p$ .
- $\max$  : maksimum nilai rasio antar kluster.

5. Membuat profilisasi dan interpretasi hasil kluster

### 3. Hasil dan Pembahasan

#### 3.1 Data Cleaning

Penelitian ini menangani masalah *missing value* dengan melakukan penghapusan objek pengamatan yang memiliki nilai kosong. Berdasarkan hasil eksplorasi awal terhadap data hasil Survei Kemiskinan Kota Samarinda tahun 2023 diperoleh 15.849 rumah tangga yang masuk ke dalam kategori miskin. Namun, ditemukan 834 *missing value* yang perlu ditangani. Sebagai contoh data yang memiliki *missing value* dapat dilihat pada Tabel 2 berikut:

Tabel 2. Data Terindikasi Terdapat *Missing Value*

Objek ke- $i$	Variabel							
	$X_1$	$X_2$	...	$X_{10}$	$X_{11}$	...	$X_{15}$	$X_{16}$
1	Milik orang lain	Bebas sewa	...	Tidak khawatir	Tidak memiliki	...	Tidak terdapat	Bekerja
2	Milik orang lain	Kontrak/sewa	...	Tidak khawatir	Tidak memiliki	...	Terdapat	Bekerja
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
42	Milik orang lain	Bebas sewa	...	Tidak khawatir	-	...	Terdapat	Bekerja
43	Milik orang lain	Milik keluarga	...	Tidak khawatir	Tidak memiliki	...	Terdapat	Bekerja
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
15.849	Milik sendiri	Milik sendiri	...	Tidak khawatir	Tidak memiliki	...	Terdapat	Bekerja

Setelah dilakukan penghapusan objek pengamatan yang memiliki *missing value*, maka diperoleh data yang telah siap untuk dianalisis dan dapat lanjut ke proses pengelompokan yakni sebanyak 15.015 rumah tangga miskin.

#### 3.2 Konversi Data

Data yang diubah ke bentuk kategorik adalah variabel pengeluaran rumah tangga per bulan ( $X_{13}$ ), serta variabel pendapatan yang diperoleh per bulan ( $X_{14}$ ). Kedua variabel dikonversi menjadi variabel kategorik berdasarkan Peraturan Wali Kota Samarinda Nomor 66 Tahun 2023 tentang kriteria rumah tangga miskin. Rumah tangga dikatakan miskin apabila pengeluaran rumah tangga dalam 1 bulan berkisar antara Rp0 sampai dengan Rp1.000.000, sehingga variabel pengeluaran rumah tangga per bulan akan dibagi menjadi dua kategori

yakni pengeluaran < Rp1.000.000, pengeluaran > Rp1.000.000. Kemudian rumah tangga termasuk ke dalam kategori miskin apabila rata-rata pendapatan per bulan yang dihasilkan di bawah garis kemiskinan, sehingga variabel pendapatan yang dihasilkan per bulan akan dikonversi menjadi 3 kategori yakni pendapatan < Rp877.945, pendapatan > Rp877.945, dan “tidak memiliki penghasilan”.

**3.3 Analisis Pengelompokan dengan Algoritma K-Modes**

Analisis pengelompokan menggunakan algoritma *K-Modes* dilakukan setelah data melalui proses *data cleaning* sehingga siap untuk dianalisis. Adapun tahapan algoritma *K-Modes* dalam penelitian ini yakni sebagai berikut:

- a. Menentukan jumlah kluster (*K*) sebelum proses pengelompokan dilakukan  
 Penelitian ini akan menguji sebanyak 2, 3, 4, 5, 6, 7, 8, 9, dan 10 kluster. Sebagai contoh, dalam penelitian ini akan dilakukan perhitungan dengan menggunakan *K* = 3.
- b. Menentukan pusat kluster awal secara acak.  
 Penentuan acak ini bertujuan untuk memilih modus awal yang berfungsi sebagai pusat kluster dalam algoritma *K-Modes*. Penentuan pusat kluster awal dilakukan secara *trial* dan *error* menggunakan *software* dan terpilih objek ke-2.463, objek ke-2.511 dan objek ke-10.419 yang dapat dilihat pada Tabel 3 berikut:

**Tabel 3.** Pusat kluster (*c<sub>k</sub>*) untuk *K* = 3

Pusat kluster ( <i>c<sub>k</sub></i> )	<i>X<sub>1</sub></i>	<i>X<sub>2</sub></i>	<i>X<sub>3</sub></i>	...	<i>X<sub>14</sub></i>	<i>X<sub>15</sub></i>	<i>X<sub>16</sub></i>
<i>c<sub>1</sub></i>	Milik sendiri	Milik sendiri	Kayu kualitas rendah	...	≤ Rp877.945	Bekerja	Terdapat
<i>c<sub>2</sub></i>	Milik orang lain	Bebas sewa	Semen	...	≤ Rp877.945	Tidak bekerja	Terdapat
<i>c<sub>3</sub></i>	Milik orang lain	Rumah milik keluarga	Kayu kualitas rendah	...	Tidak memiliki penghasilan	Tidak bekerja	Tidak terdapat

- c. Menghitung jarak setiap objek pengamatan terhadap pusat kluster menggunakan *simple matching dissimilarity* dengan persamaan (1) dan (2). Contoh perhitungan jarak antara objek pengamatan ke-1 terhadap pusat kluster ke-1 dalam dilakukan sebagai berikut.

$$\begin{aligned}
 d(x_1, c_1) &= \sum_{l=1}^{16} \delta(x_{1l}, c_{1l}) \\
 &= \delta(x_{11}, c_{11}) + \delta(x_{12}, c_{12}) + \dots + \delta(x_{116}, c_{116}) \\
 &= \delta(\text{Milik orang lain, Sendiri}) + (\text{Bebas Sewa, Sendiri}) \dots + \delta(\text{Bekerja, Bekerja}) \\
 &= 1 + 1 + 0 + 0 + 1 + \dots + 0 \\
 &= 6.
 \end{aligned}$$

Pada proses ini, perhitungan jarak dilakukan untuk seluruh objek terhadap pusat kluster 1, 2, dan 3 yakni untuk objek ke-2 hingga ke-15.015 seperti yang dapat disajikan pada Tabel 4 berikut:

**Tabel 4.** Hasil Perhitungan Jarak Setiap Objek Terhadap Pusat Kluster Awal Menggunakan *Simple Matching Dissimilarity* pada 3 Kluster

Objek ke- <i>i</i>	Nilai <i>Simple Matching Dissimilarity</i> Data Terhadap Pusat Kluster		
	<i>c<sub>1</sub></i>	<i>c<sub>2</sub></i>	<i>c<sub>3</sub></i>
1	6	5	5
2	4	4	5
3	6	5	5
⋮	⋮	⋮	⋮
15.015	3	6	7

- d. Mengalokasikan setiap objek pengamatan ke pusat kluster terdekat berdasarkan kedekatan jarak yakni yang memiliki nilai nilai *simple matching dissimilarity* terkecil. Hasil alokasi objek pengamatan ke pusat kluster terdekat dapat dilihat pada Tabel 5 berikut:

Tabel 5. Hasil Alokasi Kluster

Objek ke- <i>i</i>	Nilai <i>Simple Matching Dissimilarity</i> Data Terhadap Pusat Kluster			Alokasi kluster
	$c_1$	$c_2$	$c_3$	
1	6	5	5	2
2	4	4	5	1
3	6	5	5	2
⋮	⋮	⋮	⋮	⋮
15.015	3	6	7	1

Berdasarkan Tabel 5 menunjukkan bahwa jarak objek pengamatan ke-1 terhadap pusat kluster 2 dan 3 memiliki jarak paling minimum jika dibandingkan dengan jarak objek pengamatan ke-1 terhadap pusat kluster pertama. Sehingga, objek pengamatan ke-1 dialokasikan ke dalam kluster 2. Begitupun untuk objek pengamatan ke-2 hingga ke-15.015 dapat dialokasikan ke dalam kluster terdekat berdasarkan nilai *simple matching dissimilarity* yang paling kecil. Berdasarkan hasil alokasi kluster pada iterasi 1 diperoleh kluster 1 terdiri dari 10.078 rumah tangga miskin, kluster 2 terdiri dari 3.813 rumah tangga miskin, dan kluster 3 terdiri dari 1.124 rumah tangga miskin.

- e. Memperbarui pusat kluster menggunakan nilai modus atau nilai kategori yang paling banyak muncul pada setiap kluster. Pusat kluster baru dapat dilihat pada Tabel 6 berikut.

Tabel 6. Pusat Kluster Baru

Pusat kluster ( $c_k$ )	$X_1$	$X_2$	$X_3$	...	$X_{14}$	$X_{15}$	$X_{16}$
$c_1$	Milik sendiri	Milik sendiri	Kayu kualitas rendah	...	≤ Rp877.945	Terdaoat	Bekerja
$c_2$	Milik orang lain	Kontrak/ sewa	Semen	...	≤ Rp877.945	Terdapat	Tidak bekerja
$c_3$	Milik orang lain	Rumah milik keluarga	Kayu kualitas rendah	...	Tidak memiliki penghasilan	Tidak terdapat	Tidak bekerja

- f. Mengulangi langkah ke-c sampai ke-e hingga tidak ada perubahan pusat kluster atau tidak ada lagi objek pengamatan yang berpindah kluster.
- g. Mengulangi langkah ke-c sampai ke-f untuk setiap jumlah kluster yang berbeda.

**3.4 Validasi Kluster dengan Metode *Davies Bouldin Index***

Salah satu metode validasi kluster yang dapat digunakan adalah *Davies Bouldin Index* atau DBI. DBI merupakan suatu mencari ukuran jarak antara kluster yang satu dengan lainnya serta mengukur jarak antar objek dalam klasternya. Contoh perhitungan nilai DBI dilakukan pada  $K=3$  dengan rincian sebagai berikut.

- a. Menghitung nilai *Sum Square of Within* (SSW)

Dalam menghitung nilai SSW, memerlukan informasi mengenai pusat kluster pada iterasi terakhir, Pusat kluster baru pada iterasi terakhir dapat dilihat pada Tabel 7 berikut.

Tabel 7. Pusat Kluster pada  $K=3$

Pusat kluster ( $c_k$ )	$X_1$	$X_2$	$X_3$	...	$X_{14}$	$X_{15}$	$X_{16}$
$c_1$	Milik sendiri	Milik sendiri	Kayu kualitas rendah	...	≤ Rp877.945	Terdaoat	Bekerja
$c_2$	Milik orang lain	Kontrak/ sewa	Semen	...	≤ Rp877.945	Terdapat	Tidak bekerja
$c_3$	Milik orang lain	Rumah milik keluarga	Kayu kualitas rendah	...	Tidak memiliki penghasilan	Tidak terdapat	Tidak bekerja

Kemudian nilai SSW dapat dihitung dengan menggunakan persamaan (3) dan (4) sebagai berikut.

$$\begin{aligned}
 SSW_1 &= \frac{1}{11.237} [d(\text{Milik orang lain, Milik sendiri}) + \dots + d(\text{Bekerja, Bekerja})] \\
 &= \frac{1}{11.237} [1+1+0+0+\dots+0] \\
 &= 2,928451
 \end{aligned}$$

Hasil perhitungan nilai SSW dapat dilihat pada Tabel 8 berikut.

**Tabel 8.** Nilai SSW pada 3 Klaster

Klaster ke-k	Nilai SSW
1	2,928451
2	2,842716
3	2,641626

- b. Menghitung nilai *Sum Square of Between* (SSB)  
 Nilai SSB diperoleh dengan menghitung jarak antara pusat klaster  $c_k$  dan  $c_p$  menggunakan persamaan (5) dan (6) berikut.

$$\begin{aligned}
 SSB_{1,2} &= d(c_1, c_2) \\
 &= \delta(\text{Milik sendiri, Milik orang lain}) + \dots + \delta(\text{Bekerja, Tidak bekerja}) \\
 &= 1+1+1+0+0+0+0+0+\dots+1 \\
 &= 4
 \end{aligned}$$

Nilai SSB dapat dilihat pada Tabel 9 berikut.

**Tabel 9.** Nilai SSB antara 3 Klaster

SSB	$c_1$	$c_2$	$c_3$
$c_1$	0	4	4
$c_2$	4	0	3
$c_3$	4	3	0

- c. Menghitung nilai rasio  
 Nilai rasio merupakan nilai perbandingan antara klaster ke-k dengan klaster ke-p. Nilai rasio dapat dihitung dengan menggunakan persamaan (7) sebagai berikut.

$$\begin{aligned}
 R_{1,2} &= \frac{SSW_1 + SSW_2}{SSB_{1,2}} \\
 &= \frac{2,928451 + 2,842716}{4} \\
 &= 1,442792
 \end{aligned}$$

Nilai rasio dapat dilihat pada Tabel 10 sebagai berikut.

**Tabel 10.** Nilai Rasio Antara 3 Klaster

SSB	$c_1$	$c_2$	$c_3$
$c_1$	-	1,442792	1,392519
$c_2$	1,442792	-	1,828114
$c_3$	1,392519	1,828114	-

- d. Menghitung nilai DBI  
 Nilai DBI dapat dihitung setelah memperoleh nilai rasio, berdasarkan persamaan (8) maka nilai DBI dihitung dengan menggunakan persamaan (8) sebagai berikut.

$$\begin{aligned}
 DBI &= \frac{1}{3} \max_{r \neq p} (1,442792 + 1,828114 + 1,828114) \\
 &= 1,69967330706517
 \end{aligned}$$

Nilai DBI dapat dilihat pada Tabel 11 berikut.

**Tabel 11.** Nilai DBI 3 Klaster

Jumlah Klaster	Nilai DBI
2	1,9022
3	1,6996

Jumlah Klaster	Nilai DBI
4	1,7008
5	2,2325
6	2,0619
7	2,0268
8	2,0320
9	2,0415
10	1,9696

Berdasarkan Tabel 11, pengelompokan rumah tangga miskin menggunakan algoritma *K-Modes* dengan 3 klaster menghasilkan nilai DBI terkecil yaitu 1,6996 jika dibandingkan jumlah klaster lainnya. Nilai DBI yang lebih kecil menunjukkan bahwa klaster yang terbentuk memiliki tingkat homogenitas yang tinggi dalam satu klaster (kohesi baik) serta memiliki pemisahan yang jelas antar klaster (separasi baik). Dengan demikian, jumlah klaster optimal untuk pengelompokan rumah tangga miskin di Kota Samarinda tahun 2023 adalah 3 klaster karena mampu merepresentasikan perbedaan karakteristik rumah tangga miskin secara lebih baik.

Hasil ini sejalan dengan penelitian yang dilakukan oleh Azahra dkk. (2024) yang menunjukkan bahwa algoritma *K-Modes* dengan validasi DBI dapat menghasilkan pengelompokan yang optimal pada data kategorik. Meskipun demikian, jumlah klaster optimal yang diperoleh berbeda, dimana penelitian ini menghasilkan 3 klaster optimal, sedangkan penelitian sebelumnya menghasilkan 2 klaster optimal. Perbedaan tersebut mengindikasikan bahwa karakteristik kemiskinan pada Tingkat kota cenderung lebih beragam dan kompleks dibandingkan pada Tingkat provinsi, sehingga memerlukan jumlah klaster yang lebih banyak untuk menggambarkan variasi karakteristik yang ada.

### 3.5 Profilisasi

Hasil analisis klaster menggunakan algoritma *K-Modes* dengan menggunakan validasi DBI pada rumah tangga miskin di Kota Samarinda tahun 2023 menghasilkan 3 klaster optimal. Adapun hasil profilisasi klaster dapat dilihat pada Tabel 12 berikut.

**Tabel 12.** Hasil Pengelompokan Rumah Tangga Miskin Menggunakan Algoritma *K-Modes* pada  $K = 3$

Klaster	Banyak anggota	Kode rumah tangga
1	11.237	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, ...,15.015
2	2.327	17, 18, 19, 20, 23, 27, 45, 52, 59, 62, ..., 15.014
3	1.451	40, 50, 84, 114, 128, 152, 185, 197, ..., 15.004

Berdasarkan Tabel 12, klaster 1 merupakan kelompok terbesar dengan jumlah 11.237 rumah tangga atau sekitar 74,8% dari total data. Klaster 2 terdiri dari 2.327 rumah tangga (15,5%), sedangkan klaster 3 merupakan kelompok kecil dengan jumlah 1.452 rumah tangga (9,7%). Hal ini menunjukkan bahwa mayoritas rumah tangga miskin di Kota Samarinda berada pada klaster 1, sementara klaster 3 hanya mencakup sebagian kecil namun memiliki karakteristik yang lebih rentan. Selanjutnya untuk mengetahui karakteristik dari klaster yang terbentuk, maka perlu untuk mencari modus dari masing-masing klaster yang dapat disajikan pada Tabel 13 berikut.

**Tabel 13.** Karakteristik Masing-masing Klaster

Klaster	Karakteristik
1	Tinggal di rumah dan lahan milik sendiri dengan kondisi rumah sederhana seperti lantai dari kayu kualitas rendah, dinding kayu, serta atap dari seng. Sumber penerangan berasal dari listrik PLN dengan sumber energi memasak berasal dari gas 3 kg. Memiliki sanitasi yang layak seperti memiliki fasilitas buang air besar di rumah dengan tempat pembuangan akhir tinja berupa tangki septik. Tidak merasa khawatir dengan kondisi ekonomi rumah tangga dalam 1 tahun ke depan, tidak memiliki aset tidak bergerak dan tidak memiliki usaha sebagai sumber pendapatan di luar pendapatan tetap. Pengeluaran per bulan di atas Rp1.000.000 dengan pendapatan yang tidak sepadan yakni di bawah Rp877.945 per bulan. Kepala keluarga tidak bekerja dan terdapat anggota rumah tangga yang tercatat sebagai penerima bantuan.
2	Tinggal di rumah kontrakan yang dibangun di atas lahan milik orang lain dengan kondisi rumah sederhana seperti lantai dari semen, dinding kayu, serta atap dari seng. Sumber penerangan berasal dari listrik PLN dengan sumber energi memasak berasal dari gas 3 kg. Memiliki sanitasi yang layak seperti memiliki fasilitas buang air besar di rumah dengan tempat pembuangan akhir tinja berupa tangki septik. Tidak merasa khawatir dengan kondisi ekonomi rumah tangga dalam 1 tahun ke depan, tidak memiliki aset tidak bergerak dan tidak memiliki usaha sebagai sumber pendapatan di luar

Klaster	Karakteristik
	pendapatan tetap. Pengeluaran per bulan di atas Rp1.000.000 dengan pendapatan yang tidak sepadan yakni di bawah Rp877.945 per bulan. Kondisi diperparah dengan kepala keluarga tidak bekerja dan ditandai dengan adanya anggota rumah tangga yang tercatat sebagai penerima bantuan.
3	Tinggal di rumah milik keluarga dan dibangun di atas lahan milik orang lain dengan kondisi rumah sederhana seperti lantai dari kayu kualitas rendah, dinding kayu, serta atap dari seng. Sumber penerangan berasal dari listrik PLN dengan sumber energi memasak berasal dari gas 3 kg. Memiliki sanitasi yang layak seperti memiliki fasilitas buang air besar di rumah dengan tempat pembuangan akhir tinja berupa tangki septik. Tidak merasa khawatir dengan kondisi ekonomi rumah tangga dalam 1 tahun ke depan, tidak memiliki aset tidak bergerak dan tidak memiliki usaha sebagai sumber pendapatan di luar pendapatan tetap. Pengeluaran per bulan di atas Rp1.000.000 namun tidak memiliki pendapatan bulanan sedikitpun. Kondisi diperparah dengan kepala keluarga tidak bekerja dan ditandai adanya anggota rumah tangga yang tercatat sebagai penerima bantuan.

Tabel 13 menunjukkan bahwa perbedaan utama antar klaster terletak pada status kepemilikan tempat tinggal dan kondisi pendapatan. Klaster 1 relatif lebih stabil karena mayoritas memiliki tempat tinggal sendiri, sedangkan kluster 2 dan 3 menunjukkan ketergantungan terhadap pihak lain. Klaster 3 merupakan kelompok paling rentan dibandingkan klaster lainnya karena rumah tangga pada klaster ini tidak memiliki pendapatan sama sekali namun tetap memiliki pengeluaran di atas Rp1.000.000 per bulan. Kondisi ini mengindikasikan adanya ketidakseimbangan ekonomi yang tinggi dan potensi ketergantungan terhadap bantuan eksternal. Oleh karena itu, klaster ini menjadi prioritas utama dalam penyaluran bantuan sosial karena memiliki Tingkat kerentanan yang paling tinggi dibandingkan klaster lainnya. Berdasarkan karakteristik tersebut, kluster 3 menjadi prioritas utama dalam penyaluran bantuan sosial karena memiliki Tingkat kerentanan paling tinggi. Sementara itu, kluster 1 dan 2 memerlukan pendekatan kebijakan yang berbeda, seperti pemberdayaan ekonomi dan peningkatan akses terhadap pekerjaan.

#### 4. Kesimpulan

Berdasarkan hasil penelitian, penelitian ini berhasil mengelompokkan rumah tangga miskin di Kota Samarinda menggunakan algoritma *K-Modes* dengan validasi *Davies Bouldin Index*, dimana diperoleh jumlah klaster optimal sebanyak 3 klaster dengan nilai DBI terkecil yaitu 1,6996. Klaster 1 terdiri dari 11.237 rumah tangga miskin, klaster 2 sebanyak 2.327 rumah tangga miskin, dan klaster 3 sebanyak 1.451 rumah tangga miskin. Hasil analisis menunjukkan bahwa setiap klaster memiliki karakteristik yang berbeda, dengan klaster 3 sebagai kelompok paling rentan. sehingga perlu menjadi prioritas dalam program pengentasan kemiskinan di Kota Samarinda. Oleh karena itu, rumah tangga pada klaster 3 layak diprioritaskan sebagai penerima bantuan sosial. Penelitian ini memberikan informasi pengelompokan berbasis karakteristik yang dapat digunakan sebagai dasar dalam penentuan kebijakan bantuan sosial yang lebih tepat sasaran. Hasil penelitian ini diharapkan dapat membantu pemerintah dalam meningkatkan efektivitas program pengentasan kemiskinan.

#### 5. Daftar Pustaka

- Az-zahra, A. A., Marsaoly, A. F., Lestyani, I. P., Salsabila, R., & Madjida, W. O. Z. (2021). Penerapan Algoritma K-Modes Clustering Dengan Validasi Davies Bouldin Index Pada Pengelompokan Tingkat Minat Belanja Online Di Provinsi Daerah Istimewa Yogyakarta. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, 9(1), 24–36.
- Azahra, S., Zilrahmi, Vionanda, & Fitri Fadhilah. (2024). K-Modes Analysis with Validation of the DBI in Grouping Provinces in Indonesia based on Indicators of Poor Households. *UNP Journal of Statistics and Data Science*, 2(2), 173–178.
- Badan Pusat Statistik Kota Samarinda. (2023). *Statistik Daerah Kota Samarinda 2023*. Samarinda: Badan Pusat Statistik Kota Samarinda.
- Bappenas. (2020). *Pedoman Teknis Penyusunan Rencana Aksi Tujuan Pembangunan Berkelanjutan (TPB)/Sustainable Development Goals (SDGs)*. Jakarta: Badan Perencanaan Pembangunan Nasional.
- Daton, Z. D., & Utomo, A. P. (2022). *Fakta Seputar IKN, 4 Daerah Penyangga, Samarinda Jadi "Jantung", Balikpapan sebagai "Otot."*. Kompas.Com. <https://regional.kompas.com/read/2022/01/31/064050278/fakta-seputar-ikn-4-daerah-penyangga-samarinda-jadi-jantung-balikpapan?page=all>.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. Waltham: Morgan Kaufmann Publishers.
- Hayati, M. N., Agus, S., Lepu, A. A., Hidayatullah, A. S., Hartoyo, A. S., Ariyanto, B. D., & Udin. (2023). *Hasil Survei Kemiskinan Kota Samarinda Tahun 2023*. Samarinda: Dinas Komunikasi dan Informatika Kota Samarinda.

- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for Clustering Data*. Upper Saddle River: Prentice-Hall Inc.
- Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, 9(1), 95–100.
- Prasetyo, E. (2012). *Data Mining Konsep dan Aplikasinya Menggunakan MATLAB*. Yogyakarta: CV Andi Offset.
- Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan MATLAB*. Yogyakarta: CV Andi Offset.
- Riska, W. H., Permana, D., Putra, A. A., & Zilrahmi. (2024). *Categorical Data Clustering with K-Modes Method on Fire Cases in DKI Jakarta Province*. 2(1), 56–63.
- Samarinda Pos. (2024). *Angka Kemiskinan Ekstrem Berkurang, Penerima Bantuan Tetap Butuh Pendampingan*. Sapos.Co.Id. <https://www.sapos.co.id/metropolis/2455199631/angka-kemiskinan-ekstrem-berkurang-penerima-bantuan-tetap-butuh-pendampingan>.
- Santika, Perdana, H., & Satyahadewi, N. (2025). Analisis Pengelompokan Wilayah Desa di Kabupaten Landak Berdasarkan Indeks Ketahanan Sosial Menggunakan K-Modes. *Buletin Ilmiah Math. Stat. Dan Terapannya (Bimaster)*, 14(2), 175–182.
- Wali Kota Samarinda. (2023). *Peraturan Wali Kota Samarinda tentang Kriteria Rumah Tangga Miskin di Daerah (Perwali Nomor 66 Tahun 2023)*.