

# Pengelompokan Kabupaten/Kota Berdasarkan Indikator Rumah Layak Huni di Jawa Timur Tahun 2023 Menggunakan *K-Harmonic Means Clustering*

Azkiyatunufus Zulfani Nabila<sup>\*</sup>, Suwanda

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

azkinabila6@email.com, suwanda@unisba.ac.id

**Abstract.** Cluster analysis is one of the multivariate analysis techniques aimed at grouping objects based on their similarities. Clustering methods are generally categorized into hierarchical and non-hierarchical methods. One of the most commonly used non-hierarchical methods is K-Means Clustering; however, it has a notable weakness—it is sensitive to the initialization of cluster centers. To overcome this, the K-Harmonic Means (KHM) method was introduced as a more robust alternative. KHM utilizes the harmonic mean of squared distances from each data point to all cluster centers. In this study, the Davies-Bouldin Index is used as the validation metric to assess clustering quality. This research aims to cluster regencies/cities in East Java in 2023 based on indicators of decent housing, including type of flooring, wall type, roof type, access to proper sanitation, access to safe drinking water, and floor area per capita. The goal is to gain insight into housing disparities and provide a foundation for more targeted regional development policies. The analysis resulted in four optimal clusters: Cluster 1 consists of 7 regencies/cities, Cluster 2 includes 25, Cluster 3 consists of 5, and Cluster 4 comprises only 1 regency/city.

**Keywords:** *Cluster Analysis, K-Harmonic Means, Decent Housing.*

**Abstrak.** Analisis klaster yaitu salah satu teknik dari analisis multivariat yang bertujuan untuk mengelompokkan objek-objek berdasarkan kemiripan yang dimiliki. Metode klaster dibagi menjadi dua metode, yaitu metode klaster hierarki dan metode klaster nonhierarki. Salah satu metode yang populer dalam klaster nonhierarki adalah K-Means *Clustering*. Namun, metode ini memiliki kelemahan yaitu sensitif terhadap inisialisasi pusat klaster. Guna mengatasi hal ini, K-Harmonic Means dikembangkan sebagai alternatif yang lebih robust. Metode ini menggunakan rata-rata harmonik dari jarak kuadrat antara setiap titik data ke seluruh pusat klaster. Selain itu, untuk mengevaluasi hasil *clustering*, perlu adanya proses validasi. Indeks validasi yang digunakan pada penelitian ini adalah dengan Davies-Bouldin Index. Penelitian ini bertujuan mengelompokkan kabupaten/kota berdasarkan indikator rumah layak huni yaitu jenis lantai, jenis dinding, jenis atap, akses sanitasi layak, akses air minum layak, dan luas lantai per kapita di Jawa Timur tahun 2023 menggunakan K-Harmonic Means *Clustering* dengan validasi Davies-Bouldin Index untuk membantu dalam memahami kesenjangan perumahan dan menjadi dasar dalam perumusan kebijakan pembangunan daerah yang lebih tepat sasaran. Berdasarkan hasil analisis, terbentuk 4 klaster optimal, yaitu klaster 1 terdiri dari 7 kabupaten/kota, klaster 2 terdiri dari 25 kabupaten/kota, klaster 3 terdiri dari 5 kabupaten/kota, sementara klaster 4 hanya terdiri dari 1 kabupaten/kota.

**Kata Kunci:** *Analisis Klaster, K-Harmonic Means, Rumah Layak Huni.*

## A. Pendahuluan

Manusia memiliki berbagai kebutuhan mendasar, salah satunya adalah kebutuhan akan perlindungan. Perlindungan ini dikategorikan ke dalam dua aspek, yaitu perlindungan fisik dan perlindungan psikologis. Salah satu bentuk perlindungan fisik adalah tempat tinggal atau rumah yang berfungsi sebagai pelindung dari berbagai ancaman lingkungan eksternal. Menurut Maslow, kebutuhan akan rumah dan tempat berlindung termasuk dalam tingkat kebutuhan yang lebih tinggi dalam hierarki perkembangan kehidupan, yang umumnya baru dapat terpenuhi setelah kebutuhan dasar seperti sandang, pangan, dan kesehatan telah tercukupi (Sada, 2017). Rumah memiliki peran yang krusial dalam memenuhi kebutuhan hidup manusia, karena tidak hanya berfungsi sebagai tempat berteduh, tetapi juga sebagai lingkungan yang memberikan perlindungan dari berbagai kondisi alam yang tidak bersahabat, sehingga dapat menopang kehidupan yang lebih stabil (Setyobudi & Megawati, 2024).

Kelayakan sebuah rumah tidak hanya dilihat dari keberadaannya sebagai tempat berteduh, tetapi juga dari aspek-aspek yang mendukung kesejahteraan penghuninya. Konsep rumah layak huni mencakup berbagai aspek, termasuk kelayakan struktur bangunan, kecukupan luas tempat tinggal, akses terhadap air bersih dan sanitasi, serta lingkungan yang sehat dan bebas dari risiko bencana (Bappenas, 2019). Mengacu pada definisi nasional dan global, hunian layak memiliki empat kriteria yang diwajibkan terpenuhi kelayakannya, yaitu: 1) Ketahanan bangunan yakni bahan bangunan atap, dinding, dan lantai rumah; 2) Kecukupan luas tempat tinggal yaitu luas lantai per kapita  $\geq 7,2$  m<sup>2</sup>; 3) Memiliki akses air minum layak; dan 4) Memiliki akses sanitasi layak. Dari sisi kesehatan, rumah yang tidak memenuhi standar kelayakan dapat menjadi sumber penyakit dan meningkatkan risiko gangguan kesehatan, baik fisik maupun mental. Penelitian menunjukkan bahwa individu yang tinggal di bangunan dengan kondisi yang tidak sehat cenderung mengalami tingkat stres yang lebih tinggi (Rizky & Zahrah, 2021). Selain itu, dengan melihat karakteristik fisik rumah juga dapat menunjukkan kemiskinan. Rumah yang tidak layak huni menjadi target utama dalam program rehabilitasi perumahan, yang penilaiannya didasarkan pada pemenuhan indikator karakteristik fisik rumah. Secara tidak langsung, dengan memanfaatkan data kondisi fisik rumah, upaya penanggulangan kemiskinan dapat dilakukan secara lebih efektif (Harahap, 2021).

Menurut BPS (2024), lebih dari sepertiga penduduk Indonesia, tinggal di rumah yang tidak layak huni. Tren persentase rumah tangga di Indonesia yang menempati rumah layak huni tergolong fluktuatif. Dalam tiga tahun terakhir, persentase rumah tangga yang tinggal di rumah layak huni berkisar sekitar 60%. Persentase rumah tangga yang tinggal di rumah layak huni lebih tinggi di daerah perkotaan dibandingkan dengan daerah pedesaan, dengan angka mencapai 65,47% di perkotaan dan 59,91% di pedesaan (Laksono, 2024). Sebagai salah provinsi dengan jumlah penduduk terbanyak di Indonesia yaitu sekitar 14,83% penduduk Indonesia, atau sekitar 41,64 juta jiwa yang tinggal di Provinsi Jawa Timur, masalah terkait perumahan layak masih menjadi tantangan di provinsi ini. Badan Pusat Statistik (BPS) mencatat bahwa pada tahun 2023 masih ada sekitar 29,26% rumah yang tidak layak huni di Jawa Timur (BPS, 2024).

Guna mengidentifikasi wilayah berdasarkan indikator rumah layak huni untuk membantu dalam memahami kesenjangan perumahan dan menjadi dasar dalam perumusan kebijakan pembangunan daerah yang lebih tepat sasaran, maka dilakukan analisis klaster. Analisis klaster yaitu salah satu teknik dari analisis multivariat yang bertujuan untuk mengelompokkan objek-objek berdasarkan kemiripan yang dimiliki (Rencher, 2002). Metode klaster dibagi menjadi dua metode, yaitu metode klaster hierarki dan metode klaster nonhierarki. Salah satu metode yang populer dalam klaster nonhierarki adalah K-Means. Namun, metode ini memiliki kelemahan yaitu sensitif terhadap inisialisasi pusat klaster. Guna mengatasi hal ini, K-Harmonic Means dikembangkan sebagai alternatif yang lebih robust. Sebagaimana dijelaskan dalam penelitian oleh Zhang, et. al. (1999) yang menunjukkan bahwa K-Harmonic Means tidak sensitif terhadap pemilihan pusat awal dan lebih cepat mencapai konvergensi dibandingkan K-Means. Penelitian oleh (Hamerly & Elkan, 2002) juga menegaskan bahwa KHM dapat meningkatkan kualitas pembentukan klaster melalui penggunaan fungsi keanggotaan dan pembobotan dalam proses klasterisasi. Selain itu, dalam analisis klaster, validasi diperlukan untuk menilai kualitas hasil pengelompokan. Penelitian ini menggunakan Davies-Bouldin Index sebagai metode validasi hasil *clustering*. Oleh karena itu, penulis akan menerapkan K-Harmonic Means untuk mengelompokkan kabupaten/kota berdasarkan indikator rumah layak huni di Jawa Timur tahun 2023 dengan validasi Davies-Bouldin Index.

Berdasarkan uraian masalah di atas, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana pengelompokan kabupaten/kota berdasarkan indikator rumah layak huni yaitu jenis lantai, jenis dinding, jenis atap, akses sanitasi layak, akses air minum layak, dan luas lantai per kapita di Jawa Timur tahun 2023 menggunakan K-Harmonic Means *Clustering*?” Adapun tujuan dalam penelitian ini adalah sebagai berikut: “Mengelompokan kabupaten/kota berdasarkan indikator rumah layak huni yaitu jenis lantai, jenis dinding, jenis atap, akses sanitasi layak, akses air minum layak, dan luas lantai per kapita di Jawa Timur tahun 2023 menggunakan K-Harmonic Means *Clustering*”.

Penelitian ini memberikan kontribusi teoritis dengan memperluas pemahaman mengenai penerapan metode K-Harmonic Means dalam proses klusterisasi. Dari sisi praktis, hasil pengelompokan kabupaten/kota di Provinsi Jawa Timur berdasarkan indikator rumah layak huni tahun 2023 menyajikan informasi yang dapat dimanfaatkan oleh pemerintah daerah, lembaga perumahan, dan pihak terkait lainnya dalam perencanaan pembangunan, penyaluran bantuan sosial, serta pengambilan kebijakan strategis. Selain itu, penelitian ini juga dapat dijadikan referensi awal bagi studi lanjutan yang ingin mengeksplorasi tema serupa dengan pendekatan atau wilayah yang berbeda.

## B. Metode

Penelitian ini menggunakan data sekunder yang diperoleh dari *website* resmi Badan Pusat Statistik Provinsi Jawa Timur yang berjumlah 38 data sesuai dengan jumlah kabupaten/kota di Jawa Timur. Data ini berisi persentase rumah tangga berdasarkan indikator rumah layak huni per kabupaten/kota di Jawa Timur pada tahun 2023. Adapun variabel yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

**Tabel 1.** Variabel Penelitian

| Variabel | Keterangan                                                                                                                                             | Skala | Satuan |
|----------|--------------------------------------------------------------------------------------------------------------------------------------------------------|-------|--------|
| $X_1$    | Jenis lantai terluas kategori layak: jenis bahan bangunan lantai bukan tanah                                                                           | Rasio | Persen |
| $X_2$    | Jenis dinding terluas kategori layak: jenis bahan bangunan dinding rumah yakni berupa tembok/plesteran/anyaman bambu-kawat, dan kayu/papan/batang kayu | Rasio | Persen |
| $X_3$    | Jenis atap terluas kategori layak: jenis bahan bangunan atap rumah yakni berupa beton dan genteng                                                      | Rasio | Persen |
| $X_4$    | Akses Sanitasi Layak                                                                                                                                   | Rasio | Persen |
| $X_5$    | Akses Air Minum Layak                                                                                                                                  | Rasio | Persen |
| $X_6$    | Luas Lantai per Kapita $\geq 7,2 \text{ m}^2$                                                                                                          | Rasio | Persen |

Dalam penelitian ini, data yang diperoleh akan diolah dengan menggunakan analisis kluster metode K-Harmonic Means. Dalam algoritma K-Harmonic Means, rata-rata harmonik digunakan sebagai fungsi pengganti minimum yang digunakan dalam K-Means. Notasi yang digunakan untuk merumuskan algoritma KHM adalah sebagai  $X = \{x_i | i = 1, \dots, n\}$   $n$  titik data dan pusat kluster  $M = \{m_c | c = 1, \dots, K\}$  (Zhang et al., 1999). Sehingga, fungsi K-Harmonic Means adalah sebagai berikut.

$$KHM(X, M) = \sum_{i=1}^n \frac{K}{\sum_{c=1}^K \frac{1}{\|x_i - m_c\|^q}} \quad \dots (1)$$

di mana  $q$  merupakan input parameter. Nilai  $q$  biasanya  $\geq 2$ .

Selanjutnya, menghitung nilai keanggotaan  $mem(m_c | x_i)$  untuk setiap data  $x_i$  pada setiap titik pusat kluster  $m_c$  dengan rumus sebagai berikut.

$$mem(m_c | x_i) = \frac{\|x_i - m_c\|^{-q-2}}{\sum_{c=1}^K \|x_i - m_c\|^{-q-2}} \quad \dots (2)$$

Kemudian dilakukan perhitungan nilai bobot  $w(x_i)$  untuk setiap data  $x_i$  menggunakan rumus sebagai berikut.

$$w(x_i) = \frac{\sum_{c=1}^K \|x_i - m_c\|^{-q-2}}{(\sum_{c=1}^K \|x_i - m_c\|^{-q})^2} \quad \dots (3)$$

Setelah didapat nilai keanggotaan dan nilai bobot pada setiap data maka selanjutnya dapat dilakukan perhitungan kembali setiap titik pusat  $m_c$  dengan rumus sebagai berikut.

$$m_c = \frac{\sum_{i=1}^n mem(m_c | x_i).w(x_i).x_i}{\sum_{i=1}^n mem(m_c | x_i).w(x_i)} \quad \dots (4)$$

Perhitungan nilai fungsi tujuan hingga perhitungan kembali setiap titik pusat terus dilakukan secara berulang hingga mendapatkan nilai fungsi tujuan tidak terdapat perubahan yang signifikan. Setelah iterasi selesai, setiap titik data  $x_i$  dimasukkan ke dalam *cluster* dengan nilai keanggotaan  $mem(m_c|x_i)$  terbesar. Setelah diperoleh hasil pengelompokan maka dilakukan validasi hasil *cluster* dengan menggunakan Davies-Bouldin Index. Davies-Bouldin Index mengukur kekompakan *cluster* dengan melihat seberapa kecil perbedaan antar objek dalam satu klaster, dan mengukur pemisahan antar *cluster* dengan mengevaluasi seberapa besar jarak antar klaster yang terbentuk (Hilmi et al., 2015). DB Index dihitung sebagai rata-rata dari nilai maksimum  $R_{i,j}$  untuk setiap *cluster* atau didefinisikan dengan persamaan sebagai berikut (Halkidi et al., 2001):

$$DB_K = \frac{1}{K} \sum_{i=1}^K R_i \quad \dots (5)$$

$$R_i = \max_{j=1, \dots, K, i \neq j} R_{ij}, \quad i = 1, \dots, K$$

di mana  $K$  adalah jumlah *cluster*.

$R_{i,j}$  adalah ukuran kesamaan antara dua *cluster*, yang dihitung menggunakan rumus:

$$R_{i,j} = \frac{s_i + s_j}{d_{i,j}} \quad \dots (6)$$

Di mana:

$s_i, s_j$  : Ukuran dispersi *cluster*, biasanya dihitung sebagai rata-rata jarak antara setiap titik dalam *cluster* ke *centroid*nya.

$d_{i,j}$  : Jarak antara pusat klaster ke- $i$  dan pusat klaster ke- $j$

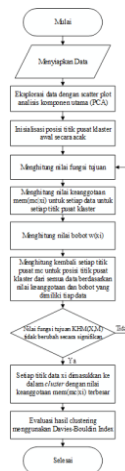
Dengan  $s_i$  dan  $s_j$  dapat diperoleh melalui persamaan berikut:

$$s_i = \frac{1}{n_i} \sum_{i=1}^{n_i} d(x_i, m_c) \quad \dots (7)$$

$$s_j = \frac{1}{n_j} \sum_{j=1}^{n_j} d(x_j, m_c) \quad \dots (8)$$

di mana  $n_i, n_j$  merupakan data dalam *cluster* ke- $i$  dan ke- $j$ , sementara  $d(x_i, m_c)$  merupakan jarak data ke- $i$  dan pusat klaster ke- $c$ . **Semakin kecil atau semakin mendekati nol nilai DB Index**, semakin baik kualitas *clustering*, karena ini menunjukkan bahwa *cluster* lebih terpisah dan memiliki dispersi dalam *cluster* yang lebih kecil. Sebaliknya, Semakin besar nilai DB Index, semakin buruk *clustering*, karena ini menunjukkan bahwa *cluster* terlalu dekat satu sama lain atau memiliki dispersi yang tinggi.

Telah digambarkan diagram alir pada Gambar 1 berikut yang dapat membantu memvisualisasikan alur kerja secara terstruktur dan efisien.



Gambar 1. Tahapan Penelitian.

### C. Hasil Penelitian dan Pembahasan

#### Analisis Deskriptif

Statistika deskriptif digunakan untuk menggambarkan dan meringkas karakteristik data yang diteliti yaitu data indikator rumah layak huni dari 38 kabupaten/kota di Jawa Timur. Tujuannya untuk memberikan gambaran umum tentang data dengan cara yang jelas. Analisis statistik deskriptif yang akan dibahas disini meliputi nilai rata-rata, maksimum, minimum, dan jumlah data yang dapat dilihat pada Tabel 2.

Tabel 2. Ringkasan Statistika Deskriptif

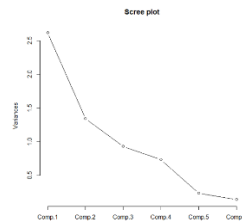
| Variabel | N  | Rata-Rata | Nilai Minimum | Nilai Maksimum |
|----------|----|-----------|---------------|----------------|
| $X_1$    | 38 | 94,27     | 68,59         | 100            |
| $X_2$    | 38 | 97,78     | 90,27         | 100            |
| $X_3$    | 38 | 93,97     | 65,82         | 98,88          |
| $X_4$    | 38 | 84,59     | 50,30         | 98,18          |
| $X_5$    | 38 | 96,12     | 79,26         | 100            |
| $X_6$    | 38 | 97,21     | 79,74         | 99,85          |

Berdasarkan Tabel 2, masing-masing indikator rumah layak huni menunjukkan variasi antar wilayah di Jawa Timur. Nilai terendah untuk jenis lantai bukan tanah ( $X_1$ ) terdapat di Kab. Bojonegoro (68,59%) dan tertinggi di Kota Kediri (100%). Jenis dinding layak ( $X_2$ ) terendah di Kab. Bondowoso (90,27%) dan tertinggi di Kota Malang (100%). Jenis atap layak ( $X_3$ ) terendah di Kota Surabaya (65,82%) dan tertinggi di Kab. Sampang (98,88%). Akses sanitasi layak ( $X_4$ ) terendah di Kab. Bangkalan (50,30%) dan tertinggi di Kota Madiun (98,18%). Akses air minum layak ( $X_5$ ) terendah di Kab. Lamongan (79,26%) dan tertinggi di Kota Probolinggo (100%). Terakhir, luas lantai per kapita ( $X_6$ ) terendah di Kota Surabaya (79,74%) dan tertinggi di Kab. Ngawi (99,85%).

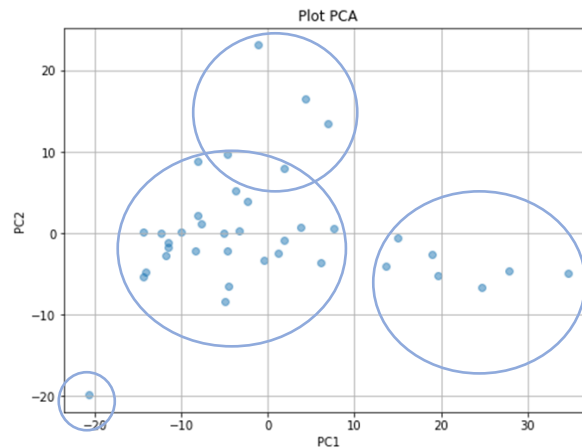
#### Eksplorasi Data

##### Principal Component Analysis (PCA)

Principal Component Analysis (PCA) digunakan untuk mereduksi dimensi data dengan mengubah variabel asli menjadi sejumlah komponen utama yang menjelaskan sebagian besar variasi dalam data. Teknik ini sangat berguna dalam eksplorasi data multivariat karena memungkinkan visualisasi pola, sehingga memudahkan identifikasi potensi kluster sebelum dilakukan analisis kluster yang lebih lanjut. Penentuan banyaknya komponen utama dapat dilihat dari scree plot. Scree plot merupakan plot antara  $i$  dengan  $\lambda_i$ . Penentuan banyaknya komponen utama dengan scree plot yaitu dengan melihat titik belok yang mulai melandai.

**Gambar 2.** Scree Plot PCA

Berdasarkan Gambar 2, terlihat bahwa kurva mulai landai pada komponen kedua ( $i = 2$ ), artinya bahwa dengan dua komponen utama saja sudah mencukupi untuk mewakili enam variabel awal.

**Gambar 3.** Scatter Plot 2 Komponen Utama

Berdasarkan Gambar 3, pada scatter plot PCA terdapat indikasi adanya 4 kluster dalam data, ditandai dengan kelompok titik yang berdekatan secara visual, meskipun jumlah pastinya masih perlu divalidasi lebih lanjut.

### Analisis dan Pembahasan

#### Analisis Kluster Menggunakan Metode K-Harmonic Means

Pada penelitian analisis kluster metode K-Harmonic Means (KHM) ini parameter ( $q$ ) yang digunakan yaitu bernilai 3,5 karena nilai tersebut merupakan yang terbaik menurut temuan Zhang, dan banyak kluster ( $K$ ) yang akan dianalisis adalah  $K = 2, 3$ , dan 4 kluster.

#### Inisialisasi Posisi Titik Pusat (*Centroid*) Kluster Awal

Penentuan *centroid* awal pada penelitian ini dilakukan secara acak dengan bantuan *Microsoft Excel*. Nilai *centroid* dari setiap titik hasil dari pemilihan acak pada masing-masing  $K = 2, 3$ , dan 4 disajikan pada tabel berikut.

**Tabel 3.** *Centroid* Awal

| $K = 2$          |       |       |       |       |       |       |
|------------------|-------|-------|-------|-------|-------|-------|
| Pusat Kluster    | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
| Kab. Madiun      | 87,68 | 98,88 | 98,06 | 91,99 | 96,39 | 99,17 |
| Kab. Gresik      | 99    | 99,14 | 90,59 | 95    | 92,54 | 96,26 |
| $K = 3$          |       |       |       |       |       |       |
| Pusat Kluster    | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
| Kab. Pacitan     | 95,4  | 97,62 | 95,86 | 77,45 | 80,19 | 98,64 |
| Kab. Tulungagung | 98,06 | 98,38 | 97,44 | 87,85 | 97,9  | 98,59 |
| Kota Mojokerto   | 99,71 | 99,37 | 85,86 | 96,48 | 99,95 | 96,11 |
| $K = 4$          |       |       |       |       |       |       |

| Pusat Klaster    | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ |
|------------------|-------|-------|-------|-------|-------|-------|
| Kab. Jember      | 96    | 95,55 | 98,62 | 64,47 | 97,51 | 97,36 |
| Kab. Mojokerto   | 97,05 | 98,81 | 95,25 | 89,38 | 98,9  | 98,12 |
| Kab. Probolinggo | 90,92 | 94,05 | 95,41 | 66,12 | 97,75 | 97,94 |
| Kota Malang      | 99,47 | 100   | 86,86 | 86,36 | 99,06 | 93,84 |

### Menghitung Nilai Fungsi Tujuan

Selanjutnya melakukan perhitungan fungsi tujuan dengan Persamaan 1, berikut merupakan hasil dari nilai fungsi tujuan pada masing-masing  $K = 2, 3$ , dan 4 saat sudah konvergen yang ditunjukkan pada Tabel 4.

**Tabel 4.** Nilai Fungsi Tujuan

| Jumlah Klaster | Nilai Fungsi Tujuan | Iterasi |
|----------------|---------------------|---------|
| 2              | 778759,75655        | 21      |
| 3              | 607295,46728        | 39      |
| 4              | 386348,38137        | 27      |

### Menghitung Nilai Keanggotaan

Selanjutnya menghitung nilai keanggotaan menggunakan Persamaan 2, hasil akhir dari nilai keanggotaan masing-masing data pengamatan untuk  $K = 2$  disajikan pada Tabel 5.

**Tabel 5.** Nilai Keanggotaan  $K = 2$

| Objek    | $m_1$    | $m_2$    |
|----------|----------|----------|
| 1        | 0,74543  | 0,25457  |
| 2        | 0,00645  | 0,99355  |
| $\vdots$ | $\vdots$ | $\vdots$ |
| 38       | 0,00053  | 0,99947  |

Berikut merupakan nilai keanggotaan masing-masing data pengamatan untuk  $K = 3$  yang disajikan pada Tabel 6.

**Tabel 6.** Nilai Keanggotaan  $K = 3$

| Objek    | $m_1$    | $m_2$    | $m_3$    |
|----------|----------|----------|----------|
| 1        | 0,21072  | 0,71896  | 0,07032  |
| 2        | 0,00008  | 0,99844  | 0,00149  |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 38       | 0,00058  | 0,08077  | 0,91865  |

Berikut merupakan nilai keanggotaan masing-masing data pengamatan untuk  $K = 4$  yang disajikan pada Tabel 7.

**Tabel 7.** Nilai Keanggotaan  $K = 4$

| Objek    | $m_1$    | $m_2$    | $m_3$    | $m_4$    |
|----------|----------|----------|----------|----------|
| 1        | 0,2682   | 0,4481   | 0,2723   | 0,0114   |
| 2        | 0,0004   | 0,9698   | 0,0296   | 0,0001   |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| 38       | 0,0001   | 0,9982   | 0,0014   | 0,0004   |

### Menghitung Nilai Bobot

Langkah selanjutnya menghitung nilai bobot dengan Persamaan 3, hasil perhitungan nilai bobot ditunjukkan pada Tabel 8.

**Tabel 8.** Nilai Bobot

| Objek    | Nilai Bobot |          |          |
|----------|-------------|----------|----------|
|          | $K = 2$     | $K = 3$  | $K = 4$  |
| 1        | 45,0399     | 36,4508  | 31,7469  |
| 2        | 23,2828     | 10,5286  | 15,6797  |
| $\vdots$ | $\vdots$    | $\vdots$ | $\vdots$ |
| 38       | 18,8463     | 19,0117  | 15,4453  |

Setelah memperoleh nilai keanggotaan dan nilai bobot tiap objek, maka titik pusat kluster baru dapat dihitung dengan persamaan 4.

### Penetapan Anggota Kluster

Penentuan kluster akhir untuk setiap objek pengamatan dilakukan dengan mengacu pada nilai keanggotaan tertinggi pada saat iterasi berhenti yang dimiliki oleh data tersebut terhadap seluruh kluster. Hasil pengelompokan untuk tiap jumlah kluster disajikan pada Tabel 9.

**Tabel 9.** Hasil Pengelompokan Tiap Kluster

| $K = 2$       |                 |
|---------------|-----------------|
| Kluster       | Jumlah Kab/Kota |
| 1             | 10              |
| 2             | 28              |
| $K = 3$       |                 |
| Kluster       | Jumlah Kab/Kota |
| 1             | 7               |
| 2             | 19              |
| 3             | 12              |
| $K = 4$       |                 |
| Pusat Kluster | Jumlah Kab/Kota |
| 1             | 7               |
| 2             | 25              |
| 3             | 5               |
| 4             | 1               |

### Evaluasi Hasil *Clustering* Menggunakan *Davies-Bouldin Index*

Setelah pengelompokan dilakukan menggunakan algoritma *K-Harmonic Means* (KHM), langkah selanjutnya adalah menghitung nilai *Davies-Bouldin Index* (DBI) untuk mengevaluasi kualitas hasil kluster yang terbentuk. Perhitungan nilai *Davies-Bouldin Index* (DBI) pada hasil clustering dengan algoritma *K-Harmonic Means* dilakukan untuk tiga jumlah kluster yang telah dianalisis, yaitu  $K = 2$ , 3, dan 4 menggunakan persamaan 5. Perbandingan hasil validasi kluster tersebut dapat dilihat pada Tabel 10.

**Tabel 10.** Nilai DBI Hasil Algoritma K-Harmonic Means

| Jumlah Kluster | Nilai DBI |
|----------------|-----------|
| $K = 2$        | 1,0224    |
| $K = 3$        | 1,0981    |
| $K = 4$        | 0,7719    |

Berdasarkan Tabel 10, diperoleh nilai DBI yang paling rendah sebesar 0,7719 yakni pada nilai  $K = 4$ . Hal ini dapat diartikan bahwa dari ketiga jumlah kluster yang telah digunakan dapat diperoleh bahwa jumlah  $K = 4$  merupakan jumlah kluster yang paling baik untuk menghasilkan hasil kluster yang optimal untuk digunakan dalam pengelompokan kabupaten/kota di Jawa Timur berdasarkan indikator rumah layak huni selaras dengan pola yang tampak pada scatter plot PCA, yang secara visual menunjukkan kecenderungan terbentuknya empat kelompok data berdasarkan komponen utama pertama dan kedua.

### Karakteristik Hasil Analisis *Clustering*

Setelah hasil klaster yang optimal sudah diketahui, yaitu  $K = 4$ , langkah berikutnya melakukan *profiling* untuk melihat karakteristik dari tiap klaster. Untuk mengidentifikasi karakteristik ini, rata-rata dari variabel-variabel pada tiap klaster dapat dihitung sebagai perwakilan atau gambaran umum dari klaster tersebut. Perhitungan rata-rata pada masing-masing klaster disajikan dalam Tabel 11.

**Tabel 11.** Karakteristik Hasil Klaster Algoritma K-Harmonic Means Berdasarkan Rata-Rata

| Klaster | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $X_6$ | Kategori      |
|---------|-------|-------|-------|-------|-------|-------|---------------|
| 1       | 91,81 | 93,53 | 95,61 | 62,72 | 97,06 | 97,84 | Rendah        |
| 2       | 97,67 | 99,05 | 93,83 | 89,58 | 95,76 | 97,36 | Sangat Tinggi |
| 3       | 79,66 | 97,07 | 97,99 | 87,61 | 96,2  | 99,06 | Tinggi        |
| 4       | 99,68 | 99,48 | 65,82 | 97,81 | 98,15 | 79,74 | Sedang        |

Berdasarkan Tabel 11, maka diperoleh informasi bahwa daerah pada klaster 1 didominasi oleh kabupaten yang masih memiliki banyak rumah tidak layak huni, terutama pada rata-rata indikator akses terhadap sanitasi layak masih di bawah 80 persen. Klaster ini memiliki rata-rata persentase indikator rumah layak huni berkategori rendah. Daerah pada klaster 2 didominasi oleh kabupaten/kota yang telah memiliki banyak rumah layak huni, dengan semua indikator di atas 80 persen. Klaster ini memiliki rata-rata persentase indikator rumah layak huni berkategori sangat tinggi. Daerah pada klaster 3 didominasi oleh kabupaten yang telah memiliki banyak rumah layak huni, dengan indikator yang unggul ialah jenis atap terluas, dan luas lantai per kapita. Klaster ini memiliki rata-rata persentase indikator rumah layak huni berkategori tinggi. Sementara daerah pada klaster 4 hanya beranggotakan Kota Surabaya yang di mana satu-satunya kota di Jawa Timur yang indikator jenis atap terluas dan luas lantai per kapitanya belum memenuhi kelayakan dibanding kota lainnya. Namun, empat indikator lainnya lebih unggul dibanding klaster lainnya. Klaster ini memiliki rata-rata persentase indikator rumah layak huni berkategori sedang.

### D. Kesimpulan

Berdasarkan nilai Davies-Bouldin Index, hasil pengelompokan kabupaten/kota di Jawa Timur tahun 2023 yang optimal dengan menggunakan metode K-Harmonic Means dengan  $q = 3,5$  adalah sebanyak 4 klaster karena diperoleh nilai DBI yang paling rendah yaitu sebesar 0,7719. Pada hasil pengelompokan 4 klaster diperoleh bahwa Klaster 1 terdiri dari 7 kabupaten, Klaster 2 terdiri dari 25 kabupaten/kota, dan Klaster 3 terdiri dari 5 kabupaten. Sementara Klaster 4 hanya beranggotakan Kota Surabaya. Dengan hasil tersebut, kabupaten/kota yang tergolong dalam klaster dengan indikator rumah layak huni rendah, khususnya terkait akses sanitasi, perlu mendapatkan perhatian khusus melalui program bantuan terpadu, seperti peningkatan infrastruktur sanitasi dan edukasi masyarakat. Penelitian selanjutnya disarankan untuk menggunakan pendekatan *clustering* dan metode validasi yang berbeda guna membandingkan serta meningkatkan kualitas hasil pengelompokan.

### Ucapan Terimakasih

Tentunya dalam proses penelitian ini tidak terlepas dari dukungan dan bantuan dari berbagai pihak. Penulis menyampaikan terima kasih kepada Bapak Dr. Suwanda, M. S., selaku dosen pembimbing penulis yang dengan sabar membimbing dan menyumbangkan waktu, pengetahuan, dan tenaganya dalam membantu penulis menyelesaikan penelitian ini. Serta kedua orang tua atas do'a, dukungan, motivasi, dan juga dukungan materi kepada penulis.

### Daftar Pustaka

Ajeng Mega Pratiwi, & Mutaqin, A. K. (2021). Penerapan Algoritma Naïve Bayes Classifier dalam Memprediksi Status Keberlanjutan Polis Nasabah Asuransi PT.X. *Jurnal Riset Statistika*, 1(2), 117–126. <https://doi.org/10.29313/jrs.v1i2.435>

- Bappenas. (2019). *Daftar Indikator-Indikator Tingkat Pembangunan Berkelanjutan (TPB) dan Standar Pelayanan Minimum (SPM) yang Menjadi Prioritas*. Sepakat Bappenas. [https://sepakat.bappenas.go.id/wiki/index.php/Daftar\\_Indikator\\_Indikator\\_Tingkat\\_Pembangunan\\_Berkelanjutan\\_%28TPB%29\\_dan\\_Standar\\_Pelayanan\\_Minimum\\_%28SPM%29\\_yang\\_Menjadi\\_Prioritas](https://sepakat.bappenas.go.id/wiki/index.php/Daftar_Indikator_Indikator_Tingkat_Pembangunan_Berkelanjutan_%28TPB%29_dan_Standar_Pelayanan_Minimum_%28SPM%29_yang_Menjadi_Prioritas)
- BPS. (2024). Statistika Perumahan dan Pemukiman Provinsi Jawa Timur. In *BPS Provinsi Jawa Timur*. <https://doi.org/https://doi.org/3302001.35>
- Devila Mustika Prancisca, & Darwis, S. (2021). Prediksi Sisa Umur Bearing Menggunakan Regresi Eksponensial. *Jurnal Riset Statistika*, 1(2), 107–116. <https://doi.org/10.29313/jrs.v1i2.434>
- Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). Clustering Validation Techniques. *Journal of Intelligent Information Systems*. <https://doi.org/10.1023/A:1012801612483>
- Hamerly, G., & Elkan, C. (2002). Alternatives to the k-means algorithm that find better clusterings. *International Conference on Information and Knowledge Management, Proceedings*, 600–607. <https://doi.org/10.1145/584792.584890>
- Harahap, T. (2021). Komparasi Indikator Rumah Layak Huni dan Permukiman Kumuh Indonesia. *Journal of Science and Applicative Technology*, 5(1), 163. <https://doi.org/10.35472/jsat.v5i1.426>
- Hilmi, M. N., Wilandari, Y., & Yasin, H. (2015). PEMETAAN PREFERENSI MAHASISWA BARU DALAM MEMILIH JURUSAN MENGGUNAKAN ARTIFICIAL NEURAL NETWORK (ANN) DENGAN ALGORITMA SELF ORGANIZING MAPS (SOM). *Jurnal Gaussian*, 4. <https://doi.org/https://doi.org/10.14710/j.gauss.4.1.53-60>
- Laksono, M. Y. (2024). 63 dari 100 Keluarga di Indonesia Tempati Rumah Layak Huni. Kompas.Com. <https://www.kompas.com/properti/read/2024/02/11/120000921/63-dari-100-keluarga-di-indonesia-tempati-rumah-layak-huni>
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*. John Wiley & Sons.
- Rizky, S. F., & Zahrah, A. (2021). Arsitektur Hunian untuk Penderita Gangguan Kesehatan Mental. *Jurnal Sosiologi USK (Media Pemikiran & Aplikasi)*, 15(2), 155–173. <https://doi.org/10.24815/jsu.v15i2.21452>
- Sada, H. J. (2017). Kebutuhan Dasar Manusia dalam Perspektif Pendidikan Islam. *Al-Tadzkiyyah: Jurnal Pendidikan Islam*, 8(2), 213. <https://doi.org/10.24042/atjpi.v8i2.2126>
- Sanusi, W., Syam, R., Adawiyah, D. R., Matematika, J., Matematika, F., Pengetahuan, I., Universitas, A., & Makassar, N. (2019). Model Regresi Nonparametrik dengan Pendekatan Spline. *Journal of Mathematics*, 2(1), 70–81. <http://www.ojs.unm.ac.id/jmathcos>
- Setyobudi, G. A., & Megawati, S. (2024). Implementasi Program Rehabilitasi Rumah Tidak Layak Huni (Rutilahu) Di Kelurahan Genteng Kota Surabaya. *Publika*, 268–277. <https://doi.org/10.26740/publika.v12n1.p253-263>
- Zhang, B., Hsu, M., & Dayal, U. (1999). K-Harmonic means - A data clustering algorithm. *HP Laboratories Technical Report*, 124.