



# Penerapan K-Means Clustering dengan Evaluasi V-Measure untuk Pengelompokan Wilayah Berdasarkan Data Potensi Sumber Kesejahteraan Sosial

Riza Mar'atus Sholihah and Fachrur Rozi\*

Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim, Malang

## Abstrak

Kesejahteraan sosial merupakan dimensi penting pembangunan yang membutuhkan pemetaan wilayah berbasis data. Penelitian ini mengelompokkan kecamatan di Kabupaten/Kota Mojokerto berdasarkan data Potensi Sumber Kesejahteraan Sosial (PSKS) menggunakan *K-Means Clustering*. Data sekunder memuat 12 indikator dari BPS dan Dinas Sosial Mojokerto. Tahapan meliputi standardisasi *z-score*, deteksi *outlier*, penentuan jumlah klaster optimal dengan metode *Elbow*, penerapan *K-Means*, serta evaluasi menggunakan *V-Measure*. Hasil menunjukkan lima klaster dengan karakteristik spesifik: kawasan perkotaan, perhutanan, bersejarah, industri, dan pegunungan. Evaluasi menghasilkan *V-Measure* 0,756 (homogenitas 0,859; kelengkapan 0,672), menandakan kualitas pengelompokan yang baik dan representatif untuk mendukung penetapan prioritas program kesejahteraan sosial.

**Kata Kunci:** *K-Means Clustering*; *V-Measure*; Pengelompokan Wilayah; Kesejahteraan Sosial; Mojokerto

## Abstract

Social welfare is a key development dimension that requires evidence-based spatial targeting. This study clusters subdistricts in Mojokerto Regency/City using *K-Means Clustering* on Social Welfare Resource Potential (PSKS) data. The dataset comprises 12 indicators compiled from the Central Bureau of Statistics (BPS) and the Mojokerto Social Affairs Office. The workflow includes z-score standardization, outlier detection, Elbow-based selection of the optimal number of clusters, *K-Means* fitting, and evaluation with the *V-Measure*. Regions are grouped into five clusters with distinct profiles: urban, forestry-dominated, historical, industrial, and mountainous. The *V-Measure* is 0.756 (homogeneity 0.859; completeness 0.672), indicating accurate, well-separated clusters useful for program prioritization.

**Keywords:** *K-Means Clustering*; *V-Measure*; Regional Clustering; Social Welfare; Mojokerto

Copyright © 2025 by Authors, Published by JRMM Group. This is an open access article under the CC BY-SA License (<https://creativecommons.org/licenses/by-sa/4.0>)

\*Corresponding author. E-mail: [f.rozi@mat.uin-malang.ac.id](mailto:f.rozi@mat.uin-malang.ac.id)

## 1 Pendahuluan

Kesejahteraan sosial merupakan salah satu dimensi penting dalam pembangunan nasional untuk meningkatkan kualitas hidup masyarakat [1]. Upaya peningkatan kesejahteraan sosial tidak terlepas dari peran strategis pemerintah dalam merumuskan kebijakan dan program yang responsif terhadap kondisi sosial ekonomi masyarakat. Dalam implementasinya, permasalahan yang muncul berkaitan dengan ketidaktepatan sasaran penerima bantuan sosial yang disebabkan oleh ketidaksesuaian data masyarakat miskin, baik karena pendataan yang belum akurat maupun karena dinamika sosial yang terus berubah. Sehingga diperlukan langkah untuk mengidentifikasi dan mengelompokkan wilayah-wilayah yang memiliki karakteristik sosial ekonomi tertentu. Salah satu pendekatan yang dapat digunakan dalam menganalisis struktur data sosial adalah *clustering*, yaitu teknik analisis statistika multivariat yang mengelompokkan objek berdasarkan kemiripan karakteristik antar variabel [2].

Secara umum, metode *clustering* terbagi menjadi dua jenis, yaitu *clustering hierarki* dan *clustering non-hierarki*. *Clustering hierarki* membentuk struktur pengelompokan bertingkat, tetapi hasilnya cenderung kurang optimal [3]. Sebaliknya, *clustering non-hierarki*, seperti algoritma *K-Means Clustering*, memberikan hasil yang lebih stabil dan efisien, serta fleksibel dalam menentukan jumlah kelompok yang sesuai [4]. *K-Means* bekerja dengan membagi data ke dalam sejumlah kelompok ( $k$  klaster) berdasarkan jarak ke pusat kelompok (*centroid*) secara iteratif hingga konvergen. Evaluasi kualitas hasil pengelompokan penting dilakukan untuk menilai apakah klaster yang terbentuk representatif terhadap struktur data sebenarnya. Salah satu metode evaluasi yang digunakan dalam penelitian ini adalah *V-Measure*, yang merupakan ukuran berbasis rata-rata harmonik antara homogenitas (keseragaman anggota dalam satu kelompok) dan kelengkapan (kemampuan menangkap keseluruhan anggota dari suatu kategori) [5].

Penelitian ini memanfaatkan algoritma *K-Means Clustering* untuk mengelompokkan wilayah berdasarkan data Potensi Sumber Kesejahteraan Sosial (PSKS) pada Kabupaten dan Kota Mojokerto. Data ini mencakup 12 indikator sosial ekonomi yang relevan terhadap kondisi kemiskinan, potensi sumber daya, serta aktor-aktor kesejahteraan sosial seperti pekerja sosial profesional, pekerja sosial masyarakat, taruna siaga bencana, lembaga kesejahteraan sosial, karang taruna, dan lainnya. Permasalahan sosial seperti kemiskinan, urbanisasi, dan kawasan kumuh menjadi latar belakang utama pentingnya klasifikasi wilayah berbasis indikator data potensi sumber kesejahteraan sosial [6].

Wilayah Kabupaten dan Kota Mojokerto merupakan daerah yang menghadapi tantangan terkait isu kesejahteraan sosial. Fenomena urbanisasi, keberadaan pemukiman kumuh, serta meningkatnya jumlah penduduk miskin menjadi permasalahan yang saling berkaitan. Berdasarkan data dari Badan Pusat Statistik menunjukkan bahwa tingkat kemiskinan di Kabupaten Mojokerto mengalami peningkatan sebesar 0,09% pada tahun 2023 dibandingkan tahun sebelumnya, dengan jumlah penduduk miskin mencapai sekitar 112,86 ribu jiwa [6]. Permasalahan ini meningkat salah satunya dipengaruhi oleh adanya pengelolaan sektor potensial, seperti pariwisata lokal yang belum optimal, yang seharusnya dapat mendorong pertumbuhan ekonomi masyarakat. Berbagai program pemerintah seperti Program Keluarga Harapan (PKH), bantuan pendidikan, dan subsidi telah digulirkan, namun permasalahan seperti ketidaktepatan data sasaran masih sering terjadi [1]. Oleh karena itu, diperlukan analisis berbasis data yang kuat untuk mendukung pengambilan kebijakan yang lebih tepat sasaran.

Penelitian sebelumnya telah menggunakan *K-Means* untuk berbagai konteks, seperti penentuan calon penerima bantuan [2], pengelompokan karakteristik warga [7], dan evaluasi kualitas kelompok dengan indeks seperti *Davies Bouldin Index* maupun *V-Measure* [8]. Namun, penelitian ini menambahkan dimensi baru dengan fokus pada data PSKS dan evaluasi menyeluruh terhadap kualitas klaster dengan *V-Measure*.

Dengan demikian, penelitian ini bertujuan untuk mengelompokkan wilayah di Mojokerto berdasarkan potensi sumber kesejahteraan sosial menggunakan metode *K-Means Clustering*,

serta mengevaluasi hasil pengelompokan tersebut dengan *V-Measure* untuk mengetahui sejauh mana kelompok yang terbentuk merepresentasikan struktur data yang sebenarnya. Melalui pendekatan ini, diharapkan diperoleh pemetaan wilayah yang lebih representatif terhadap kondisi sosial masyarakat sehingga intervensi terhadap program pemerintah yang bersangkutan dapat dilakukan secara lebih tepat.

## 2 Metode

Bagian ini menguraikan alur analisis yang digunakan untuk membentuk klaster wilayah secara andal dan terukur. Secara berurutan, kami melakukan standardisasi data agar setiap variabel berada pada skala yang sebanding, diikuti deteksi *outlier* untuk menjaga kekokohan hasil. Selanjutnya, kami memeriksa multikolinieritas antarfaktor, menentukan jumlah klaster optimal dengan metode *Elbow*, kemudian menerapkan *K-Means Clustering* dan mengevaluasi kualitas hasilnya menggunakan *V-Measure*. Rangkaian langkah tersebut dirangkum pada [Gambar 1](#) untuk memudahkan penelusuran metodologi. Tahap pertama dijelaskan pada subbagian berikut.

### 2.1 Standardisasi Data

Standardisasi data dilakukan agar variabel dalam penelitian tidak memiliki perbedaan satuan ukuran yang dapat menyebabkan ketidakvalidan perhitungan pada analisis [9]. Standardisasi data dilakukan dengan transformasi data ke dalam nilai *z-score*. Menurut [10], nilai *z-score* diperoleh dari pengurangan nilai data dengan rata-rata, kemudian dibagi dengan standar deviasi, sebagaimana dirumuskan sebagai berikut:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (1)$$

### 2.2 Deteksi *Outlier*

*Outlier* adalah nilai pencilan yang berbeda secara signifikan dalam distribusi data yang dmenyebabkan hasil analisis menjadi bias, meskipun dalam beberapa kasus juga mengandung informasi yang bernilai [11]. Menurut [12], deteksi *outlier* yang dapat dilakukan pada data dengan nilai *z-score*, jika data memiliki nilai kurang dari -3.00 dan lebih dari 3.00.

### 2.3 Uji Multikolinieritas

Multikolinieritas adalah kondisi yang terjadi karena adanya kemungkinan hubungan atau korelasi antar variabel [13]. Multikolinieritas dapat diketahui dengan menentukan Hipotesis uji multikolinieritas adalah:

$$\begin{aligned} H_0 &: \text{Tidak terdapat multikolinieritas} \\ H_1 &: \text{Terdapat multikolinieritas} \end{aligned}$$

Statistik uji multikolinieritas yang digunakan adalah nilai *Variance Inflation Factor* (VIF). Nilai VIF dapat diperoleh dengan rumus berikut:

$$\text{VIF}_j = \frac{1}{1 - R_{x_j|(x_1, x_2, \dots, x_{j-1}, x_{j+1}, \dots, x_m)}^2} \quad (2)$$

Pengambilan keputusan dalam uji multikolinieritas didasarkan pada batas nilai VIF. Jika nilai VIF  $\leq 10$ , maka hipotesis nol ( $H_0$ ) diterima, yang berarti tidak terdapat multikolinieritas. Sebaliknya, jika nilai VIF  $> 10$ , maka  $H_0$  ditolak, yang menunjukkan bahwa terjadi multikolinieritas di antara variabel independen dalam model [12].

## 2.4 Metode *Elbow*

Metode Elbow digunakan untuk menentukan banyak kelompok optimal yang dapat dibentuk dalam proses pengelompokan pada *K-Means*. Metode *Elbow* dapat dilakukan dengan menganalisis grafik hubungan antara banyak kelompok ( $k$ ) dan nilai *Sum of Square Error* (SSE) yang diperoleh dengan rumus berikut [14]:

$$\text{SSE} = \sum_{h=1}^k \sum_{i=1}^{c_h} \|\vec{x}_i - \vec{v}_h\|^2 \quad (3)$$

Titik optimal sebagai  $k$ , banyak kelompok yang dipilih, ditentukan pada bagian grafik yang membentuk sudut siku atau yang memiliki besar sudut paling kecil [15].

## 2.5 *K-Means Clustering*

*K-Means* merupakan metode pengelompokan data ke dalam  $k$  kelompok berdasarkan kemiripan karakteristik antar objek. *K-Means Clustering* dilakukan dengan meminimalkan variasi dalam kelompok dan memaksimalkan perbedaan antar kelompok [16]. Dalam *K-Means Clustering*, pusat kelompok, *centroid* awal, ditentukan secara acak, kemudian diperbarui secara iteratif hingga konvergen. Jarak antara data dengan *centroid* dihitung menggunakan *Euclidean Distance*, kemudian data akan dikelompokkan berdasarkan nilai jarak minimum yang diperoleh [17]. Perhitungan *Euclidean Distance* dapat dilakukan dengan rumus berikut:

$$d_{ih} = \sqrt{\sum_{j=1}^m (x_{ij} - v_{hj})^2} \quad (4)$$

Sedangkan perhitungan *centroid* untuk setiap iterasi yang dilakukan dilakukan secara berulang hingga tidak ada perubahan signifikan dalam pengelompokan menggunakan rumus berikut [18]:

$$\vec{v}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \vec{x}_i \quad (5)$$

## 2.6 *V-Measure*

*V-Measure* adalah metode evaluasi eksternal yang berbasis entropi dengan menggabungkan aspek homogenitas dan kelengkapan [5]. Homogenitas mengukur sejauh mana anggota dalam suatu kelompok berasal dari kelompok yang sama, sedangkan kelengkapan menilai apakah seluruh anggota dari suatu kelompok terkumpul dalam satu kelompok. Nilai *V-Measure* berada pada rentang 0 hingga 1, dengan nilai 1 menunjukkan hasil pengelompokan yang sempurna, artinya hasil pengelompokan yang terbentuk sepenuhnya mencerminkan struktur kelompok yang sebenarnya. Nilai yang lebih rendah atau mendekati 0 mengindikasikan pengelompokan yang kurang akurat karena adanya pencampuran antar anggota kelompok [5]. Perhitungan evaluasi *V-Measure* melibatkan nilai entropi bersama dan entropi bersyarat dari hasil pengelompokan dan pembandingnya, kemudian dilakukan perhitungan nilai homogenitas ( $H$ ) yang dirumuskan sebagai berikut:

$$H = 1 - \frac{\frac{H(C | Y)}{n}}{H(C)} \quad (6)$$

Selanjutnya dilakukan perhitungan nilai kelengkapan ( $K$ ) yang dirumuskan sebagai berikut:

$$K = 1 - \frac{\frac{H(Y | C)}{n}}{H(Y)} \quad (7)$$

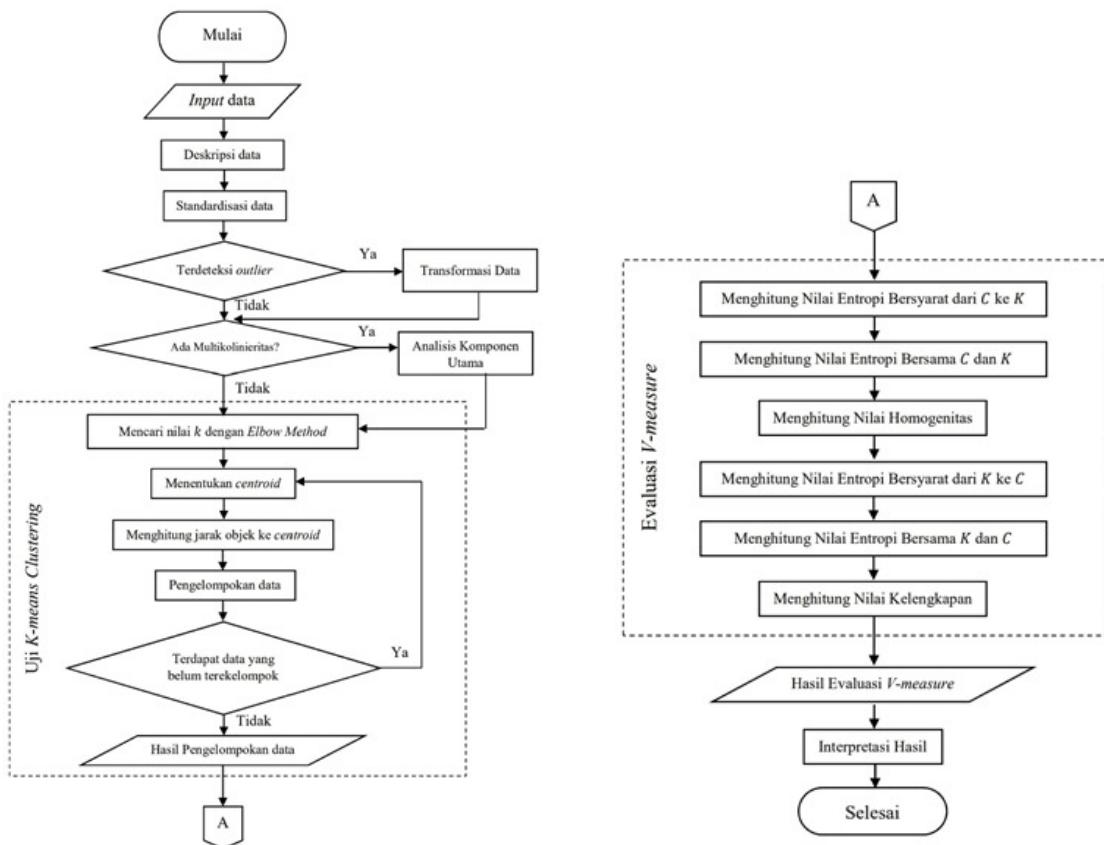
Setelah diperoleh nilai homogenitas dan kelengkapan, maka langkah selanjutnya dilakukan perhitungan nilai *V-Measure* dengan rumus sebagai berikut:

$$V = \frac{2HK}{H + K} \quad (8)$$

Interpretasi nilai *V-Measure* diklasifikasikan ke dalam tiga kategori utama, yaitu: tidak akurat ( $0 < v \leq 0.5$ ), akurat ( $0.5 < v \leq 0.9$ ), dan sempurna ( $v = 1$ ) [19]. Metode ini bersifat independen terhadap jumlah data, banyaknya kelompok, maupun ukuran sampel, sehingga dianggap sebagai pendekatan evaluasi yang objektif dan konsisten dalam menilai hasil pengelompokan [8]. Dengan demikian, *V-Measure* memberikan gambaran secara kuantitatif mengenai sejauh mana suatu model pengelompokan mampu merepresentasikan struktur distribusi kelas yang sebenarnya dalam data.

## 2.7 Diagram Alir Tahapan Penelitian

Diagram alir berikut menggambarkan langkah-langkah sistematis dalam analisis data implementasi metode *K-Means Clustering* kemudian evaluasi hasil pengelompokan dengan menggunakan metode *V-Measure* yang diawali dari proses *input* data hingga evaluasi hasil pengelompokan.



Gambar 1: Diagram Alir Penelitian

## 3 Hasil dan Pembahasan

Bagian ini menyajikan hasil penerapan metode *K-Means Clustering* pada data Potensi Sumber Kesejahteraan Sosial (PSKS) serta interpretasi setiap tahapan analisisnya. Pembahasan disusun secara berurutan, dimulai dari analisis deskriptif untuk memberikan gambaran umum data, kemudian dilanjutkan dengan hasil standardisasi, deteksi *outlier*, pengujian multikolinieritas,

penentuan jumlah klaster optimal, hingga evaluasi hasil pengelompokan menggunakan *V-Measure*. Uraian setiap tahap dijelaskan pada subbagian berikut.

### 3.1 Analisis Deskriptif

Analisis deskriptif dilakukan untuk memberikan gambaran umum mengenai 12 variabel Potensi dan Sumber Kesejahteraan Sosial (PSKS) di 21 kecamatan di Kabupaten dan Kota Mojokerto sebagaimana disajikan pada [Tabel 1](#).

**Tabel 1:** Statistik Deskriptif Variabel PSKS

Variabel	Rata-rata	Standar Deviasi	Min	Max
Pekerja Sosial Profesional ( $x_1$ )	2.333	4.305	0	18
Pekerja Sosial Masyarakat ( $x_2$ )	17.285	18.420	0	68
Taruna Siaga Bencana ( $x_3$ )	2.523	2.960	0	9
Lembaga Kesejahteraan Sosial ( $x_4$ )	127.952	111.003	0	361
Karang Taruna ( $x_5$ )	15.380	4.165	7	20
Lembaga Konsultasi Kesejahteraan Keluarga ( $x_6$ )	0.238	0.538	0	2
Keluarga Pioner ( $x_7$ )	5.380	8.529	0	28
Wahana Kesejahteraan Sosial Berbasis Masyarakat ( $x_8$ )	114.714	128.244	0	563
Wanita Pemimpin Kesejahteraan Sosial ( $x_9$ )	29.904	23.058	0	88
Penyuluhan Sosial ( $x_{10}$ )	1.809	2.040	0	6
Tenaga Kesejahteraan Sosial Kecamatan ( $x_{11}$ )	0.714	0.462	0	1
Dunia Usaha ( $x_{12}$ )	5.238	7.582	0	25

Berdasarkan [Tabel 1](#), hasil analisis statistik terhadap indikator data Potensi Sumber Kesejahteraan Sosial (PSKS) menunjukkan bahwa nilai minimum pada variabel  $x_1$  hingga  $x_{12}$ , adalah sebesar nol, terkecuali pada variabel  $x_5$  (Karang Taruna), yang memiliki nilai minimum sebesar 7, pada Kecamatan Kranggan, Magersari, dan Prajurit Kulon. Nilai nol ini mengindikasikan bahwa tidak semua kecamatan memiliki indikator PSKS secara lengkap. Sementara itu, nilai minimum pada  $x_5$  dipengaruhi oleh jumlah desa dalam satu kecamatan yang cukup banyak, sehingga menunjukkan persebaran Karang Taruna yang konsisten di setiap desa.

Dari sisi nilai rata-rata, variabel dengan nilai rata-rata paling rendah adalah  $x_6$  (Lembaga Konsultasi Kesejahteraan Keluarga) sebesar 0.238. Hal ini menunjukkan bahwa ketersediaan lembaga tersebut di tiap kecamatan masih sangat terbatas. Lembaga ini hanya ditemukan di Kecamatan Jatirejo, Trowulan, Kemlagi, dan Sooko dengan jumlah sangat kecil. Ketersediaan lembaga tersebut umumnya berkaitan dengan tingkat kunjungan masyarakat, seperti yang terjadi di Kecamatan Trowulan yang memiliki daya tarik wisata sejarah. Sebaliknya, variabel dengan rata-rata tertinggi adalah  $x_4$  (Lembaga Kesejahteraan Sosial) sebesar 127.952, yang mencerminkan cakupan layanan yang lebih luas dan menyentuh aspek-aspek vital kehidupan masyarakat seperti pendidikan, kesehatan, dan ekonomi.

### 3.2 Standardisasi Data

Standardisasi data potensi sumber kesejahteraan sosial wilayah Mojokerto menggunakan transformasi *z-score* untuk menyamakan satuan setiap variabel, sehingga masing-masing variabel memiliki nilai rata-rata nol dan nilai standar deviasi satu. Hasil transformasi *z-score* secara disajikan pada [Tabel 2](#).

**Tabel 2:** Hasil Standardisasi Data PSKS dengan *Z-score*

No	Kecamatan	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$
1	Kranggan	-0.542	-0.938	-0.853	-1.153	-2.012	-0.442	-0.631	-0.894	-1.297	-0.887	-1.543	-0.691
2	Magersari	-0.542	-0.938	-0.853	-1.153	-2.012	-0.442	-0.631	-0.894	-1.297	-0.887	-1.543	-0.691
3	Prajurit Kulon	-0.542	-0.938	-0.853	-1.153	-2.012	-0.442	-0.631	-0.894	-1.297	-0.887	-1.543	-0.691
4	Dawarblandong	3.639	1.287	0.161	-0.432	0.629	-0.442	-0.631	0.104	2.519	0.584	0.617	2.606
5	Gedeg	-0.542	1.287	-0.853	0.496	-0.332	-0.442	2.652	-0.567	1.435	1.074	0.617	0.496
6	Jetis	0.155	0.473	-0.853	-0.909	-0.332	-0.442	0.307	-0.138	1.305	0.584	0.617	1.287
7	Pungging	-0.542	-0.721	0.837	0.829	0.869	-0.442	-0.279	1.655	-0.473	1.074	0.617	-0.163
8	Jatirejo	1.084	-0.721	0.161	2.009	0.869	1.414	-0.631	0.392	-0.039	2.054	-1.543	0.232
9	Trowulan	0.155	-0.016	0.161	0.361	0.149	1.414	1.714	-0.388	0.568	1.074	0.617	1.683
10	Ngoro	-0.542	-0.776	-0.515	-0.441	0.869	-0.442	-0.631	-0.567	-0.169	0.093	0.617	-0.691
11	Kemlagi	-0.542	-0.450	-0.515	0.415	1.109	1.414	-0.631	-0.419	-0.560	-0.887	0.617	-0.691
12	Pacet	-0.542	-0.721	0.499	1.072	1.109	-0.442	-0.631	3.496	0.004	-0.397	-1.543	-0.691
13	Dlanggu	-0.542	0.690	-0.853	-0.936	0.149	-0.442	1.831	-0.185	0.264	2.054	0.617	-0.295
14	Bangsal	0.387	-0.287	2.188	-0.684	0.389	-0.442	-0.631	0.548	0.611	-0.887	0.617	-0.163
15	Sooko	-0.542	-0.124	1.850	0.253	-0.091	3.269	-0.631	-0.427	-0.560	-0.887	0.617	-0.691
16	Mojoanyar	-0.542	2.753	2.188	2.099	-0.812	-0.442	0.776	-0.013	1.262	0.093	-1.543	-0.691
17	Puri	1.316	1.016	-0.853	-0.396	0.149	-0.442	1.011	0.338	-0.560	-0.887	0.617	1.815
18	Gondang	-0.542	-0.830	-0.177	-1.153	0.629	-0.442	-0.631	-0.169	-0.516	-0.887	0.617	-0.691
19	Mojosari	-0.077	-0.233	-0.177	0.820	0.869	-0.442	-0.279	-0.894	-0.907	0.093	0.617	-0.163
20	Trawas	0.155	-0.776	-0.177	-0.423	-0.572	-0.442	-0.631	0.072	0.134	-0.887	0.617	-0.427
21	Kutorejo	-0.310	0.962	-0.515	0.478	0.389	-0.442	-0.162	-0.154	-0.430	-0.397	0.617	-0.691

**Tabel 2** menyajikan hasil standardisasi data Potensi Sumber Kesejahteraan Sosial (PSKS) dari 21 kecamatan di wilayah Mojokerto menggunakan metode *z-score*. Proses standardisasi ini dilakukan agar tidak ada variabel yang mendominasi dalam proses analisis yang diakibatkan adanya perbedaan satuan atau skala pengukuran.

Berdasarkan **Tabel 2**, nilai yang diperoleh berkisar pada rentang yang berbeda-beda untuk setiap kecamatan dan indikator. seperti pada Kecamatan Dawarblandong memiliki *z-score* tertinggi pada beberapa indikator seperti  $x_1$  dan  $x_9$ , yang menunjukkan bahwa nilai aktual dari indikator-indikator tersebut jauh di atas rata-rata dibanding kecamatan lainnya. Sebaliknya, Kecamatan Kranggan, Magersari, dan Prajurit Kulon menunjukkan *z-score* negatif ekstrem hampir pada seluruh indikator, mencerminkan keterbatasan signifikan dalam potensi dan sumber kesejahteraan sosial.

### 3.3 Deteksi *Outlier*

Deteksi *outlier* dilakukan dengan menggunakan nilai *z-score*, di mana data dikategorikan sebagai *outlier* jika memiliki nilai  $z < -3.00$  atau  $z > 3.00$ . Berdasarkan hasil standardisasi, ditemukan beberapa data *outlier* sebagai berikut:

- Variabel pekerja sosial profesional ( $x_1$ ) pada Kecamatan Dawarblandong memiliki nilai *z-score* sebesar 3.639. Hal ini menunjukkan tingginya kebutuhan pekerja sosial profesional akibat kondisi geografis wilayah yang didominasi hutan industri dan berjarak jauh dari pusat pemerintahan.
- Variabel wahana kesejahteraan sosial berbasis masyarakat ( $x_8$ ) pada Kecamatan Pacet memiliki nilai *z-score* sebesar 3.496. Nilai ini menggambarkan tingginya partisipasi masyarakat dalam pengelolaan kesejahteraan sosial yang didukung oleh pemerintah daerah.
- Variabel lembaga konsultasi kesejahteraan keluarga ( $x_6$ ) tidak digunakan dalam analisis selanjutnya karena kontribusinya telah tercermin dalam variabel lembaga kesejahteraan sosial ( $x_4$ ) dan keluarga pioner ( $x_7$ ), serta untuk menghindari redundansi dalam model.

Data *outlier* yang terdeteksi tetap dipertahankan dalam analisis karena merepresentasikan kondisi realitas yang signifikan dan informatif terhadap karakteristik wilayah.

### 3.4 Uji Multikolinieritas

Hipotesis yang digunakan pada uji multikolinieritas sebagai berikut:

$$H_0 : \text{Tidak terdapat multikolinieritas}$$

$$H_1 : \text{Terdapat multikolinieritas}$$

Nilai VIF yang diperoleh untuk seluruh variabel disajikan pada [Tabel 3](#).

**Tabel 3:** Hasil Nilai *Variance Inflation Factor* (VIF)

Variabel	SSR	SST	R <sup>2</sup>	VIF	Keputusan
x <sub>1</sub>	17,699	20	0,885	8,695	Tidak Terjadi Multikolinieritas
x <sub>2</sub>	15,489	20	0,774	4,434	Tidak Terjadi Multikolinieritas
x <sub>3</sub>	10,315	20	0,516	2,065	Tidak Terjadi Multikolinieritas
x <sub>4</sub>	14,775	20	0,739	3,827	Tidak Terjadi Multikolinieritas
x <sub>5</sub>	16,715	20	0,836	6,098	Tidak Terjadi Multikolinieritas
x <sub>7</sub>	16,948	20	0,847	6,536	Tidak Terjadi Multikolinieritas
x <sub>8</sub>	11,530	20	0,577	2,361	Tidak Terjadi Multikolinieritas
x <sub>9</sub>	14,611	20	0,731	3,711	Tidak Terjadi Multikolinieritas
x <sub>10</sub>	13,022	20	0,651	2,866	Tidak Terjadi Multikolinieritas
x <sub>11</sub>	15,769	20	0,788	4,728	Tidak Terjadi Multikolinieritas
x <sub>12</sub>	17,653	20	0,883	8,520	Tidak Terjadi Multikolinieritas

Berdasarkan [Tabel 3](#), nilai *Variance Inflation Factor* (VIF) dari seluruh variabel dalam data penelitian memiliki nilai kurang dari 10. Dengan demikian, dapat diperoleh kesputusan diterima  $H_0$  yang menyatakan bahwa tidak terjadi multikolinieritas dalam variabel-variabel tersebut. Oleh karena itu, seluruh variabel dinyatakan layak dan dapat digunakan dalam proses *K-Means Clustering*.

### 3.5 Metode *Elbow*

Banyak kelompok optimal  $k$  dengan metode *Elbow* ditentukan dari grafik nilai SSE yang membentuk sudut siku (*elbow*) atau sudut terkecil. Titik ini menunjukkan banyak kelompok yang memberikan keseimbangan terbaik antara kompleksitas model dengan tingkat kesalahan. Perhitungan nilai SSE dilakukan menggunakan [Pers. 3](#), diawali dengan menghitung centroid ( $\vec{v}_h$ ) dari setiap kelompok menggunakan [Pers. 5](#). Nilai SSE untuk  $k = 1$  diperoleh menggunakan [Pers. 3](#):

$$SSE = (-0,542 - 0)^2 + (-0,938 - 0)^2 + \dots + (-0,691 - 0)^2 = 220,000$$

Hasil perhitungan SSE untuk beberapa nilai  $k$  disajikan pada [Tabel 4](#).

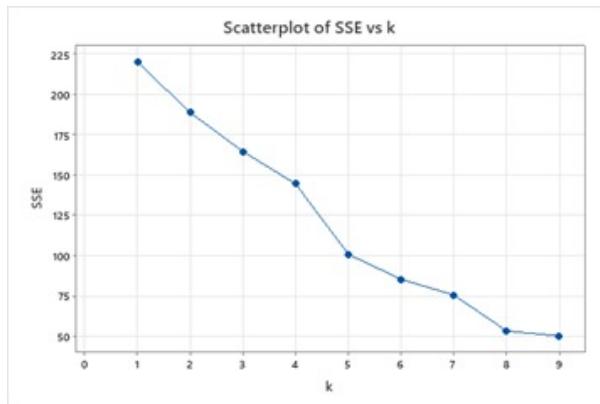
**Tabel 4:** Nilai *Sum of Square Error* (SSE)

k	SSE	Selisih
1	220.000	0
2	188.731	31.269
3	164.613	24.118
4	144.517	20.096
5	100.872	43.645
6	85.376	15.496
7	75.651	9.725
8	53.414	22.237

[Tabel 4](#) menunjukkan nilai *Sum of Square Error* (SSE) untuk masing-masing jumlah klaster  $k$  dari 1 hingga 8. SSE digunakan sebagai ukuran total jarak kuadrat antara titik data dan pusat klasternya. Semakin kecil nilai SSE, maka semakin baik data dalam satu klaster memiliki

kedekatan terhadap pusatnya. Namun demikian, penurunan nilai SSE tidak selalu signifikan seiring bertambahnya jumlah klaster.

Pada [Tabel 4](#) terlihat bahwa nilai SSE menurun secara bertahap seiring dengan meningkatnya nilai  $k$ , dengan penurunan paling signifikan terjadi pada saat  $k = 5$ , yaitu sebesar 43.645. Hal ini mengindikasikan adanya pembentukan struktur klaster yang lebih representatif pada jumlah klaster tersebut. Setelah  $k = 5$ , penurunan SSE menjadi relatif kecil dan tidak lagi memberikan peningkatan yang substansial terhadap kualitas pengelompokan. Oleh karena itu, titik  $k = 5$  dapat dianggap sebagai titik optimal dalam metode *Elbow*, yang ditandai dengan perubahan sudut yang tajam pada grafik dan selisih SSE yang paling besar. Nilai SSE pada [Tabel 4](#), yang merupakan jumlah kuadrat jarak antara titik dalam kelompok dan *centroid*, divisualisasikan dalam [Gambar 2](#).



**Gambar 2:** Grafik Metode *Elbow*

Grafik tersebut memiliki sumbu- $x$  sebagai nilai  $k$  dan sumbu- $y$  sebagai nilai SSE. Dari grafik terlihat bahwa titik siku atau sudut yang terkecil terbentuk pada  $k = 5$ . Pada  $k = 3$ , memang terjadi penurunan SSE yang signifikan ( $SSE = 24.118$ ), namun pola grafik belum membentuk sudut siku. Penurunan nilai SSE mulai membentuk sudut paling tajam dan landai pada  $k = 5$ , sehingga titik ini dipilih sebagai jumlah klaster optimal. Setelah  $k = 5$ , penurunan nilai SSE cenderung kecil dan tidak signifikan, yang menunjukkan bahwa menambah jumlah kelompok hanya akan meningkatkan kompleksitas tanpa meningkatkan kualitas pengelompokan secara substansial. Dengan demikian, jumlah klaster optimal yang digunakan dalam penelitian ini adalah  $k = 5$ .

### 3.6 Penerapan Metode *K-Means Clustering*

Data yang digunakan dalam proses pengelompokan dengan metode *K-Means Clustering* tersaji dalam [Tabel 2](#). Proses *K-Means Clustering* diawali dengan menentukan jumlah kelompok optimal ( $k$ ) menggunakan metode Elbow berdasarkan [Pers. 3](#). Hasil perhitungan SSE dari  $k = 1$  hingga  $k = 8$  ditunjukkan pada [Tabel 4](#). Berdasarkan titik siku yang terbentuk pada grafik, diperoleh nilai optimal  $k = 5$ . Langkah selanjutnya adalah penentuan *centroid* awal secara acak. *Centroid* awal diambil dari lima kecamatan, yakni Kecamatan Kranggan, Kecamatan Gedeg, Kecamatan Dlanggu, Kecamatan Gondang, dan Kecamatan Kutorejo seperti ditunjukkan pada [Tabel 5](#).

**Tabel 5:** Nilai *Centroid* Awal Iterasi ke-1

Titik Pusat	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_7$	$x_8$	$x_9$	$x_{10}$	$x_{11}$	$x_{12}$
C1	-0.542	-0.938	-0.853	-1.153	-2.012	-0.631	-0.894	-1.297	-0.887	-1.543	-0.691
C2	-0.542	1.287	-0.853	0.496	-0.332	2.652	-0.567	1.435	1.074	0.617	0.496
C3	-0.542	0.690	-0.853	-0.936	0.149	1.831	-0.185	0.264	2.054	0.617	-0.295
C4	-0.542	-0.830	-0.177	-1.153	0.629	-0.631	-0.169	-0.516	-0.887	0.617	-0.691
C5	-0.310	0.962	-0.515	0.478	0.389	-0.162	-0.154	-0.430	-0.397	0.617	-0.691

Setelah penentuan *centroid* awal, langkah selanjutnya adalah menghitung jarak *Euclidean* antara setiap data dan *centroid* menggunakan Pers. 4. Hasil selengkapnya dari perhitungan jarak dalam iterasi pertama disajikan dalam Tabel 6.

Tabel 6: Jarak Setiap Data terhadap *Centroid*

No	C1	C2	C3	C4	C5	Jarak Minimum	Kelompok
1	<b>0.000</b>	6.226	5.463	3.639	4.316	<b>0.000</b>	1
2	<b>0.000</b>	6.226	5.463	3.639	4.316	<b>0.000</b>	1
3	<b>0.000</b>	6.226	5.463	3.639	4.316	<b>0.000</b>	1
4	8.014	<b>6.115</b>	6.416	6.702	6.153	<b>6.115</b>	2
5	6.226	<b>0.000</b>	2.533	5.346	3.991	<b>0.000</b>	2
6	4.932	3.113	<b>2.973</b>	3.728	3.334	<b>2.973</b>	3
7	5.588	5.104	4.224	3.550	<b>3.286</b>	<b>3.286</b>	5
8	5.896	5.568	5.128	5.248	<b>4.472</b>	<b>4.472</b>	5
9	5.719	<b>2.556</b>	2.952	4.471	3.750	<b>2.556</b>	2
10	3.994	4.734	3.664	<b>1.386</b>	2.201	<b>1.386</b>	4
11	4.241	5.025	4.446	<b>1.735</b>	1.765	<b>1.735</b>	4
12	6.146	6.770	6.268	4.925	<b>4.814</b>	<b>4.814</b>	5
13	5.463	2.533	<b>0.000</b>	4.304	3.594	<b>0.000</b>	3
14	5.214	5.610	5.155	<b>3.015</b>	3.607	<b>3.015</b>	4
15	4.366	5.426	5.005	<b>2.677</b>	2.767	<b>2.677</b>	4
16	6.720	5.009	5.869	6.471	<b>4.828</b>	<b>4.828</b>	5
17	5.312	4.199	4.310	<b>4.166</b>	3.426	<b>4.166</b>	4
18	3.639	5.346	4.304	<b>0.000</b>	2.562	<b>0.000</b>	4
19	4.421	4.484	3.914	2.562	<b>1.803</b>	<b>1.803</b>	5
20	3.361	4.866	4.323	<b>1.736</b>	2.446	<b>1.736</b>	4
21	4.316	3.991	3.594	2.562	<b>0.000</b>	<b>0.000</b>	5

Iterasi selanjutnya perlu dilakukan dengan memperbarui nilai *centroid* menggunakan Pers. 5 hingga tidak terjadi perubahan kelompok. Dalam penelitian ini proses iterasi dilakukan sampai iterasi ke-4, di mana pada iterasi ke-3 dan ke-4 tidak lagi terjadi perubahan anggota kelompok. Hasil akhir pengelompokan wilayah disajikan dalam Tabel 7.

Tabel 7 menunjukkan hasil pengelompokan wilayah berdasarkan data Potensi Sumber Kesejahteraan Sosial (PSKS) dengan menggunakan metode *K-Means Clustering*. Dari hasil tersebut, diperoleh anggota untuk lima kelompok yang terbentuk dengan karakteristik nilai variabel yang berbeda-beda pada masing-masing kecamatan. Kelompok 1 terdiri dari Kecamatan Kranggan, Magersari, dan Prajurit Kulon, yang ditandai dengan nilai variabel yang relatif rendah secara konsisten. Hal ini menunjukkan bahwa ketiga kecamatan tersebut memiliki potensi sumber kesejahteraan sosial yang terbatas. Sementara itu, Kelompok 2 hanya terdiri dari Kecamatan Dawarbandong dan Puri yang memiliki nilai variabel tinggi, mengindikasikan keberadaan potensi sosial yang besar pada wilayah tersebut.

Kelompok 3 mencakup kecamatan-kecamatan seperti Gedeg, Jetis, Trowulan, Dlanggu, dan Mojoanyar, yang menunjukkan nilai variabel yang bervariasi namun cenderung sedang. Kelompok ini mencerminkan wilayah dengan potensi yang mulai berkembang dan memerlukan perhatian dalam perencanaan intervensi sosial. Kelompok 4 merupakan kelompok dengan jumlah kecamatan terbanyak, yaitu terdiri dari tujuh kecamatan, yang cenderung memiliki nilai variabel yang rendah hingga sedang. Wilayah dalam kelompok ini menunjukkan kondisi sosial yang memerlukan penguatan dalam aspek-aspek tertentu untuk menunjang kesejahteraan masyarakat secara menyeluruh.

Adapun Kelompok 5 terdiri dari Kecamatan Pungging, Jatirejo, dan Pacet, yang memiliki ciri khas tertentu pada aspek-aspek seperti peran karang taruna dan sektor wisata. Kelompok ini mencerminkan wilayah yang memiliki potensi sosial berbasis komunitas dan lingkungan, namun tetap memerlukan pendampingan dan pemberdayaan secara berkelanjutan. Hasil pengelompokan

**Tabel 7:** Hasil Pengelompokan dengan Metode *K-Means Clustering*

No	Kecamatan	$x_1$	$x_2$	...	Kelompok
1	Kranggan	-0.542	-0.938	...	1
2	Magersari	-0.542	-0.938	...	1
3	Prajurit Kulon	-0.542	-0.938	...	1
4	Dawarblandong	3.639	1.287	...	2
5	Gedeg	-0.542	1.287	...	3
6	Jetis	0.155	0.473	...	3
7	Pungging	-0.542	-0.721	...	5
8	Jatirejo	1.084	-0.721	...	5
9	Trowulan	0.155	-0.016	...	3
10	Ngoro	-0.542	-0.776	...	4
11	Kemlagi	-0.542	-0.450	...	4
12	Pacet	-0.542	-0.721	...	5
13	Dlanggu	-0.542	0.690	...	3
14	Bangsal	0.387	-0.287	...	4
15	Sooko	-0.542	-0.124	...	4
16	Mojoanyar	-0.542	2.753	...	3
17	Puri	1.316	1.016	...	2
18	Gondang	-0.542	-0.830	...	4
19	Mojosari	-0.077	-0.233	...	4
20	Trawas	0.155	-0.776	...	4
21	Kutorejo	-0.310	0.962	...	4

ini memberikan gambaran menyeluruh tentang distribusi potensi kesejahteraan sosial antar wilayah dan dapat dijadikan dasar dalam penyusunan kebijakan serta program intervensi sosial yang tepat sasaran dan berbasis data.

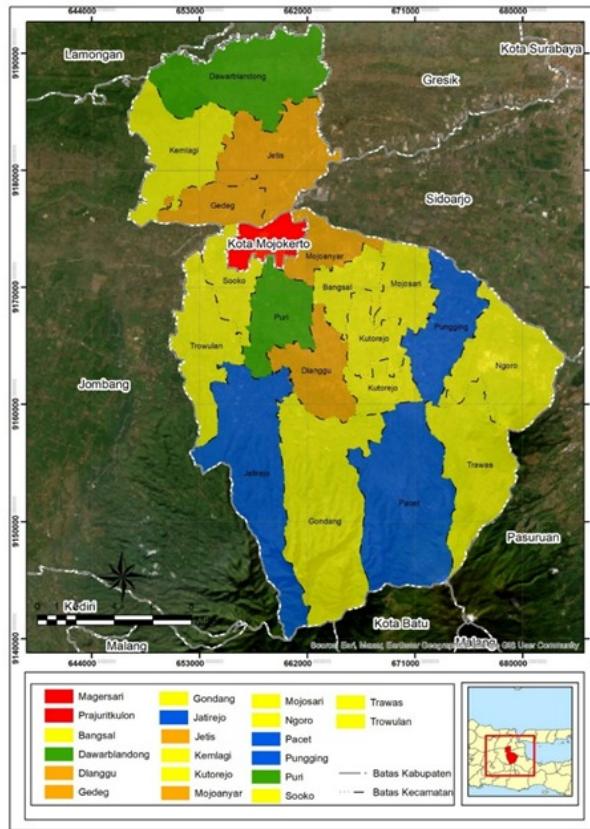
Penamaan dan deskripsi masing-masing kelompok hasil pengelompokan tersebut disajikan secara rinci dalam [Tabel 8](#).

**Tabel 8:** Hasil Pengelompokan Wilayah dengan Metode *K-Means Clustering*

Kelompok	Penamaan Khusus	Kecamatan di Kabupaten/Kota Mojokerto	Jumlah Anggota
1	Kawasan Perkotaan	Kecamatan Kranggan, Kecamatan Magersari, Kecamatan Prajurit Kulon.	3
2	Kawasan Perhutanan	Kecamatan Dawarblandong dan Kecamatan Puri.	2
3	Kawasan Bersejarah	Kecamatan Jetis, Kecamatan Gedeg, Kecamatan Trowulan, Kecamatan Dlanggu, dan Kecamatan Mojoanyar.	5
4	Kawasan Industri	Kecamatan Ngoro, Kecamatan Kemlagi, Kecamatan Bangsal, Kecamatan Sooko, Kecamatan Gondang, Kecamatan Mojosari, Kecamatan Trawas, Kecamatan Kutorejo.	8
5	Kawasan Pegunungan	Kecamatan Pungging, Kecamatan Jatirejo, dan Kecamatan Pacet.	3

Setelah proses pengelompokan dilakukan menggunakan metode *K-Means Clustering* hasil pengelompokan tersebut disajikan dalam bentuk visualisasi wilayah untuk memberikan gambaran

spasial mengenai distribusi setiap kelompok yang terbentuk berdasarkan karakteristik data. Dengan demikian, pola persebaran antar kelompok dapat dianalisis secara geografis dan lebih mudah diinterpretasikan. Hasil visualisasi pengelompokan wilayah tersebut disajikan pada Gambar 3.



Gambar 3: Visualisasi Hasil Pengelompokan *K-Means Clustering*

Berdasarkan Gambar 3, kelompok 1 dengan wilayah yang berwarna merah, diberi penamaan khusus sebagai Kawasan Perkotaan. Hal ini dikarenakan kelompok 1 terdiri dari tiga kecamatan dalam Kota Mojokerto. Ketiganya menunjukkan kesamaan nilai data pada seluruh indikator, dengan ciri khas wilayah seperti kepadatan tinggi, dominasi sektor jasa, serta fasilitas publik yang memadai dan lengkap. Kelompok 2 dengan wilayah yang berwarna hijau, diberi penamaan khusus sebagai Kawasan Perhutanan yang mencakup Kecamatan Dawarbandong dan Puri. Kedua kecamatan ini unggul pada indikator pekerja sosial profesional, taruna siaga bencana, dan wanita pemimpin kesejahteraan sosial. Secara geografis, wilayah ini sebagian besar merupakan kawasan hutan industri.

Wilayah yang berwarna orange merepresentasikan Kelompok 3 yang diberi penamaan khusus sebagai Kawasan Bersejarah. Kawasan ini terdiri dari lima kecamatan, yaitu Jetis, Gedeg, Trowulan, Dlanggu, dan Mojoanyar. Kesamaan wilayah tersebut terletak pada indikator lembaga kesejahteraan sosial dan karang taruna, serta nilai historis kawasan sebagai bagian dari pusat peradaban Kerajaan Majapahit. Kelompok 4 yang direpresentasikan dengan wilayah yang berwarna kuning, yang diberi penamaan khusus sebagai Kawasan Industri mencakup delapan kecamatan: Ngoro, Kemlagi, Bangsal, Sooko, Gondang, Mojosari, Trawas, dan Kutorejo. Wilayah ini memiliki nilai rendah pada indikator pekerja sosial masyarakat, namun didominasi oleh aktivitas industri dan perdagangan.

Wilayah yang direpresentasikan dengan warna biru merupakan Kelompok 5 yang diberi penamaan khusus sebagai Kawasan Pegunungan yang terdiri dari Kecamatan Pungging, Jatirejo, dan Pacet. Ketiganya memiliki nilai tinggi pada indikator wahana kesejahteraan sosial berbasis masyarakat, serta terletak di wilayah dataran tinggi yang berkembang sebagai kawasan wisata

alam dan agrowisata.

### 3.7 Penerapan Metode Evaluasi *V-Measure*

Hasil pengelompokan wilayah Mojokerto yang telah diperoleh dengan metode *K-Means Clustering* disajikan dalam [Tabel 7](#), kemudian dievaluasi menggunakan metode *V-Measure*. Evaluasi *V-Measure* dilakukan dengan membandingkan hasil pengelompokan *K-Means Clustering* dan *Agglomerative Hierarchical Clustering* dengan pendekatan *Centroid Linkage*. Pemilihan metode *Centroid Linkage* didasarkan pada kemiripan konsep dengan *K-Means Clustering*, yaitu sama-sama menggunakan titik pusat atau *centroid* sebagai dasar pengelompokan. Hasil pengelompokan dari kedua metode tersebut disajikan berturut-turut dalam [Tabel 9](#).

**Tabel 9:** Hasil Pengelompokan Data

No	Kecamatan	<i>K-Means</i> (C)	<i>Centroid Linkage</i> (Y)
1	Kranggan	1	1
2	Magersari	1	1
3	Prajurit Kulon	1	1
4	Dawarblandong	2	2
5	Gedeg	3	2
6	Jetis	3	2
7	Pungging	5	3
8	Jatirejo	5	4
9	Trowulan	3	2
10	Ngoro	4	3
11	Kemlagi	4	3
12	Pacet	5	3
13	Dlanggu	3	2
14	Bangsal	4	3
15	Sooko	4	3
16	Mojoanyar	3	5
17	Puri	2	2
18	Gondang	4	3
19	Mojosari	4	3
20	Trawas	4	3
21	Kutorejo	4	3

Berdasarkan [Tabel 9](#) yang telah disajikan, kemudian dilakukan perhitungan nilai *V-Measure* yang sebelumnya dilakukan perhitungan nilai homogenitas ( $H$ ) dengan [Pers. 6](#) dan nilai kelengkapan ( $K$ ) menggunakan [Pers. 7](#) pada nilai dalam [Tabel 10](#)

**Tabel 10:** Nilai Entropi  $H(C)$ ,  $H(C|Y)$ ,  $H(Y)$ , dan  $H(Y|C)$

$H(C)$	$H(C Y)$	$H(Y)$	$H(Y C)$
-0.401	0	-0.401	0
-0.323	0	-0.517	-0.918
-0.493	-0.723	-0.504	-0.721
-0.531	0	-0.210	0
-0.401	-0.919	-0.210	0

Berdasarkan [Tabel 10](#) perhitungan nilai homogenitas ( $H$ ) yang diperoleh dengan menggunakan nilai entropi bersama  $H(C)$  dan nilai entropi bersyarat hasil pengelompokan metode *K-Means Clustering* diberikan hasil pengelompokan pembanding, menggunakan [Pers. 6](#) sebagai berikut:

$$H = 1 - \frac{\frac{H(C | Y)}{n}}{H(C)} = 1 - \frac{\frac{6.372}{21}}{\frac{2.149}{21}} = 1 - \frac{0.303}{2.149} = 1 - 0.141 \\ = 0.859$$

Perhitungan nilai kelengkapan ( $K$ ) yang diperoleh dengan menggunakan nilai entropi bersama  $H(Y)$  dan nilai entropi bersyarat hasil pengelompokan pembanding diberikan hasil pengelompokan metode *K-Means Clustering*, menggunakan Pers. 7 sebagai berikut:

$$K = 1 - \frac{\frac{H(Y | C)}{n}}{H(Y)} = 1 - \frac{\frac{12.718}{21}}{\frac{1.847}{21}} = 1 - \frac{0.606}{1.847} = 1 - 0.328 \\ = 0.672$$

Berdasarkan hasil perhitungan diperoleh nilai nilai homogenitas ( $H$ ) dan nilai kelengkapan ( $K$ ): ke dalam Pers. 8 untuk perhitungan nilai *V-Measure*:

$$V = 2 \times \frac{0.859 \times 0.672}{0.859 + 0.672} = 0.756$$

Berdasarkan perhitungan menggunakan Pers. 8, diperoleh nilai *V-Measure* sebesar 0.756, yang menunjukkan bahwa hasil pengelompokan menggunakan metode *K-Means Clustering* merupakan hasil pengelompokan yang memiliki kategori akurat. Nilai homogenitas sebesar 0.859 menunjukkan bahwa anggota dalam satu kelompok cenderung memiliki kesamaan karakteristik. Sedangkan nilai kelengkapan sebesar 0.672 menunjukkan bahwa kelompok yang terbentuk mencakup sebagian besar anggota yang seharusnya berada dalam satu kelompok yang sama. Dengan demikian, metode *K-Means Clustering* terbukti memiliki akurasi pengelompokan yang baik dalam mengelompokkan data potensi sumber kesejahteraan sosial di wilayah Mojokerto.

## 4 Kesimpulan

Penelitian ini mengelompokkan wilayah Kabupaten dan Kota Mojokerto berdasarkan potensi sumber kesejahteraan sosial menggunakan metode *K-Means Clustering* serta mengevaluasi kualitas pengelompokan dengan *V-Measure*. Berdasarkan metode *Elbow*, diperoleh jumlah kelompok optimal sebanyak lima klaster yang merepresentasikan kesamaan karakteristik antar kecamatan dalam hal potensi sosial.

Evaluasi hasil pengelompokan menghasilkan nilai *V-Measure* sebesar 0.756, dengan homogenitas sebesar 0.859 (kategori sangat baik) dan kelengkapan sebesar 0.672 (kategori baik). Hal ini menunjukkan bahwa metode yang digunakan mampu membentuk kelompok yang cukup akurat dalam hal kesamaan anggota dalam setiap klaster. Penelitian ini berkontribusi dalam menunjukkan potensi penerapan *unsupervised learning* untuk mendukung pemetaan wilayah berbasis data sosial secara lebih terstruktur.

## Pernyataan Kontribusi Penulis (CReditT)

**Riza Mar'atus Sholihah:** Konseptualisasi, Metodologi, Penulisan–Draf Awal, Perangkat Lunak, Validasi, Visualisasi. **Fachrur Rozi:** Kurasi Data, Analisis Formal, Penulisan–Telaah dan Penyuntingan.

## Deklarasi Penggunaan AI atau Teknologi Berbasis AI

Model ChatGPT versi 4 digunakan untuk membantu penyusunan draf awal dan koreksi struktur kalimat.

## Deklarasi Konflik Kepentingan

Penulis menyatakan tidak ada konflik kepentingan.

## Ucapan Terima Kasih

Ucapan terima kasih yang tulus ditujukan kepada dosen pembimbing yang telah membimbing dan memberikan arahan selama proses penelitian. Peneliti juga menyampaikan terima kasih kepada Badan Pusat Statistik (BPS) serta Dinas Sosial Kabupaten dan Kota Mojokerto atas bantuan data dan informasi yang sangat berharga. Ucapan terima kasih juga disampaikan Program Studi Matematika Universitas Islam Negeri Maulana Malik Ibrahim Malang atas fasilitas dan dukungan akademik yang diberikan. Peneliti tidak lupa mengucapkan terima kasih kepada keluarga tercinta dan teman-teman yang telah memberikan semangat, bantuan, dan kerja sama selama pelaksanaan penelitian ini.

## Ketersediaan Data

Data pendukung dalam penelitian ini dapat diperoleh dari penulis korespondensi atas permintaan kepada Dinas Pemerintah Daerah Kabupaten dan Kota Mojokerto, dengan ketentuan bahwa akses terhadap data tersebut tunduk pada perjanjian kerahasiaan yang berlaku.

## Daftar Pustaka

- [1] E. N. R. Khakim, “Perbandingan Algoritma Klasifikasi Data Kesejahteraan Sosial Kabupaten Bantul,” *Jurnal Processor*, vol. 17, no. 2, pp. 91–100, 2022.
- [2] F. I. Putri, R. Damayanti, and Kismiantini, “Penerapan Algoritma K-Means Untuk Mengelompokan Kecamatan Di Kabupaten Gunungkidul Berdasarkan Program Keluarga Harapan,” *Prosiding Seminar Nasional Matematika, Statistika, dan Aplikasinya*, vol. 2, pp. 408–418, 2022.
- [3] D. T. Kusuma and N. Agani, “Prototipe Komparasi Model Clustering Menggunakan Metode K-Means Dan FCM Untuk Menentukan Strategi Promosi : Study Kasus Sekolah Tinggi Teknik-PLN Jakarta,” *Jurnal TICOM*, vol. 3, no. 3, p. 93 460, 2015. [Available online](#).
- [4] E. P. P. Bryan Orleans, “Clustering Algoritma (K-Means),” pp. 1–5, 2022. [Available online](#).
- [5] A. Rosenberg and J. Hirschberg, “V-Measure: A conditional entropy-based external cluster evaluation measure,” *EMNLP-CoNLL 2007 - Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, no. June, pp. 410–420, 2007.
- [6] Yuhenny, Dwi, “Profil Kemiskinan di Kabupaten Mojokerto Maret 2024,” no. 43, pp. 1–12, 2024.
- [7] E. Rahma Meilaniwati and M. Fauzan, “Klasifikasi Penduduk Miskin Penerima PKH Menggunakan Metode Naive Bayes dan KNN,” *Jurnal Kajian dan Terapan Matematika*, vol. 8, no. 2, pp. 75–84, 2022. [Available online](#).

- [8] W. T. Wijaya and R. Herawati, "the Analysis of Relationship Between Height of Body and Gain of Medals in Olympics," *Proxies : Jurnal Informatika*, vol. 4, no. 2, pp. 149–158, 2024. DOI: [10.24167/proxies.v4i2.12441](https://doi.org/10.24167/proxies.v4i2.12441)
- [9] U. Rahardja, Q. Aini, and M. Iqbal, "Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage dan Ward," *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 5, no. 1, pp. 40–43, 2020.
- [10] M. P. Lawuna, Y. Zega, and R. N. Mendorfa, "Analisis Cluster Mahasiswa Pendidikan Matematika Universitas Nias," *Jurnal Cendekia : Jurnal Pendidikan Matematika*, vol. 8, no. 2, pp. 946–961, 2024. DOI: [10.31004/cendekia.v8i2.2940](https://doi.org/10.31004/cendekia.v8i2.2940)
- [11] R. Silvi, "Analisis Cluster dengan Data Outlier Menggunakan Centroid Linkage dan K-Means Clustering untuk Pengelompokan Indikator HIV/AIDS di Indonesia," *Jurnal Matematika "MANTIK"*, vol. 4, no. 1, pp. 22–31, 2018. DOI: [10.15642/mantik.2018.4.1.22-31](https://doi.org/10.15642/mantik.2018.4.1.22-31)
- [12] I. Ghazali, "Ghozali\_imam\_2011\_aplikasi\_analisis\_mult.pdf," p. 129, 2011. Available online.
- [13] A. Zaki, I. Irwan, and I. A. Sembe, "Penerapan K-Means Clustering dalam Pengelompokan Data (Studi Kasus Profil Mahasiswa Matematika FMIPA UNM)," *Journal of Mathematics Computations and Statistics*, vol. 5, no. 2, p. 163, 2022. DOI: [10.35580/jmathcos.v5i2.38820](https://doi.org/10.35580/jmathcos.v5i2.38820)
- [14] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 14, no. 2, pp. 277–288, 2023. DOI: [10.24176/simet.v14i2.9630](https://doi.org/10.24176/simet.v14i2.9630)
- [15] V. A. Ekasetya and A. Jananto, "Klusterisasi Optimal Dengan Elbow Method Untuk Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Semarang," *Jurnal Dinamika Informatika*, vol. 12, no. 1, pp. 20–28, 2020. DOI: [10.35315/informatika.v12i1.8159](https://doi.org/10.35315/informatika.v12i1.8159)
- [16] W. K. Hardle and L. Simar, "Cluster Analysis BT - Applied Multivariate Statistical Analysis," in W. K. Hardle and L. Simar, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2015, pp. 385–405. DOI: [10.1007/978-3-662-45171-7\\_13](https://doi.org/10.1007/978-3-662-45171-7_13) Available online.
- [17] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis.: Pearson Prentice Hall*. 2007, p. 773.
- [18] "Penerapan Rekayasa Ulang Iteratif Pada Sistem Informasi Akademik FTIF Its Surabaya," *Juti*, vol. 5, no. 1, pp. 19–25, 2006.
- [19] C. Lancia et al., "Method to measure the mismatch between target and achieved received dose intensity of chemotherapy in cancer trials: A retrospective analysis of the MRC BO06 trial in osteosarcoma," *BMJ Open*, vol. 9, no. 5, pp. 1–8, 2019. DOI: [10.1136/bmjopen-2018-022980](https://doi.org/10.1136/bmjopen-2018-022980)