

# Penerapan Metode K-Medoids untuk Mengelompokkan Kecamatan di Kabupaten Cianjur Berdasarkan Dampak Kejadian Tanah Longsor Tahun 2024

Pingki Ratna Anjani<sup>1\*</sup>, Purnama Akbar<sup>2</sup>

<sup>1,2</sup> Program Studi Statistika, Universitas Islam Indonesia, Jl. Kaliurang KM 14,5, Kabupaten Sleman Daerah Istimewa Yogyakarta, 55584, Indonesia

\*Corresponding author: [22611021@students.uii.ac.id](mailto:22611021@students.uii.ac.id)



P-ISSN: 2986-4178  
E-ISSN: 2988-4004

## Riwayat Artikel

Dikirim: 14 Maret 2026  
Direvisi: 14 April 2026  
Diterima: 14 April 2026

## ABSTRAK

Tanah longsor merupakan bencana yang memiliki tingkat dampak berbeda antar wilayah. Kabupaten Cianjur termasuk daerah rawan dengan variasi dampak yang cukup tinggi pada tahun 2024. Namun, pemetaan berbasis data untuk mengelompokkan tingkat dampak masih terbatas. Penelitian ini bertujuan mengelompokkan kecamatan berdasarkan dampak tanah longsor menggunakan metode K-Medoids. Data diperoleh dari Badan Penanggulangan Bencana Daerah (BPBD) Provinsi Jawa Barat. Variabel yang digunakan meliputi jumlah jiwa terdampak, kerusakan rumah, dan kerusakan fasilitas umum. Analisis dilakukan melalui standarisasi data, reduksi dimensi menggunakan PCA, serta penentuan jumlah kluster optimal dengan metode Silhouette. Proses klusterisasi menggunakan K-Medoids berbasis jarak untuk menghasilkan kelompok yang stabil terhadap outlier.

Hasil penelitian menunjukkan terbentuknya tiga kluster. Kluster pertama memiliki dampak berat. Kluster kedua menunjukkan dampak ringan. Kluster ketiga memiliki dampak sangat berat. Setiap kluster memiliki karakteristik yang berbeda berdasarkan tingkat korban dan kerusakan. Penelitian ini berkontribusi dalam pemetaan tingkat dampak bencana secara berbasis data. Hasilnya dapat digunakan untuk mendukung penentuan prioritas penanganan dan mitigasi bencana.

**Kata Kunci:** dampak bencana, Kabupaten Cianjur, kluster, K-Medoids, tanah longsor

## ABSTRACT

Landslides are natural disasters with varying levels of impact across regions. Cianjur Regency is a landslide-prone area with significant variation in impacts in 2024. However, data-driven mapping to classify impact levels remains limited. This study aims to cluster sub-districts based on landslide impacts using the K-Medoids method. The data were obtained from the Regional Disaster Management Agency (BPBD) of West Java Province. The variables include the number of affected people, housing damage, and damage to public facilities. The analysis involves data

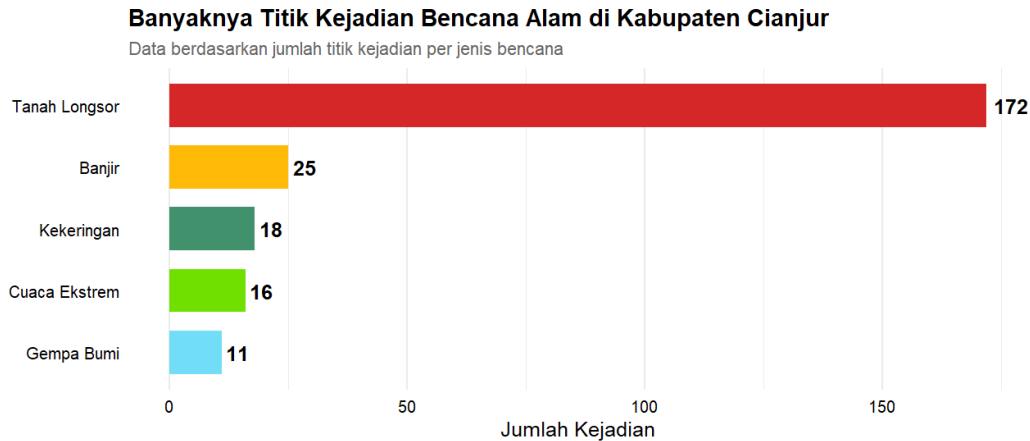
standardization, dimensionality reduction using PCA, and determination of the optimal number of clusters using the Silhouette method. The clustering process applies distance-based K-Medoids to produce groups that are robust to outliers. The results show three clusters. The first cluster represents high impact areas. The second cluster represents low impact areas. The third cluster represents very high impact areas. Each cluster has distinct characteristics based on the level of casualties and damage. This study contributes to data-driven mapping of disaster impact levels. The results can support decision-making in prioritizing disaster management and mitigation strategies.

**Keywords: disaster impact, Cianjur Regency, clustering, K-Medoids, landslide**

## 1. Pendahuluan

Indonesia merupakan salah satu negara dengan tingkat kerawanan bencana alam yang tinggi. Kondisi ini dipengaruhi oleh letak geografis Indonesia yang berada pada kawasan cincin api Pasifik (*Ring of Fire*), sehingga rentan terhadap berbagai aktivitas tektonik seperti gempa bumi, letusan gunung berapi, dan tsunami. Selain faktor geologis, kondisi klimatologis Indonesia yang beriklim tropis dengan dua musim utama, yaitu musim hujan dan musim kemarau, turut berkontribusi terhadap meningkatnya potensi bencana hidrometeorologi. Variabilitas curah hujan yang tinggi, dikombinasikan dengan kondisi topografi dan karakteristik tanah, menjadikan wilayah Indonesia rentan terhadap berbagai jenis bencana seperti banjir, tanah longsor, kekeringan, dan kebakaran hutan [1].

Secara konseptual, bencana alam didefinisikan sebagai peristiwa yang mengganggu kehidupan dan penghidupan masyarakat yang disebabkan oleh faktor alam, non-alam, maupun aktivitas manusia, yang berimplikasi pada timbulnya korban jiwa, kerusakan lingkungan, kerugian harta benda, serta dampak psikologis [2]. Salah satu bencana yang memiliki frekuensi kejadian tinggi di Indonesia adalah tanah longsor. Bencana ini umumnya terjadi pada wilayah dengan kemiringan lereng yang curam dan intensitas curah hujan yang tinggi. Dalam perspektif mekanika tanah, tanah longsor terjadi akibat ketidakseimbangan antara gaya penahan dan gaya penggerak yang bekerja pada lereng, sehingga menyebabkan pergerakan massa tanah atau batuan [3][4]. Provinsi Jawa Barat merupakan salah satu wilayah yang memiliki tingkat kerentanan tinggi terhadap bencana tanah longsor. Kondisi geografis yang didominasi oleh wilayah pegunungan serta tingginya curah hujan menjadikan provinsi ini rentan terhadap kejadian longsor. Kabupaten Cianjur, sebagai bagian dari Provinsi Jawa Barat, tercatat sebagai salah satu daerah dengan frekuensi kejadian tanah longsor yang signifikan..



**Gambar 1.** Visualisasi Titik Kejadian Bencana Alam di Kabupaten Cianjur Tahun 2024

Berdasarkan data kejadian bencana tahun 2024 yang disajikan pada Gambar 1, tanah longsor merupakan jenis bencana dengan frekuensi tertinggi, yaitu sebanyak 172 titik kejadian. Angka tersebut jauh melampaui kejadian bencana lainnya, seperti banjir (25 kejadian), kekeringan (18 kejadian), cuaca ekstrem (16 kejadian), dan gempa bumi (11 kejadian). Temuan ini mengindikasikan bahwa tanah longsor merupakan bencana dominan di Kabupaten Cianjur dan berpotensi menimbulkan dampak yang luas terhadap masyarakat dan infrastruktur.

Tingginya frekuensi kejadian tersebut menunjukkan urgensi untuk mengidentifikasi wilayah dengan tingkat kerentanan yang berbeda secara lebih rinci. Pendekatan berbasis data melalui analisis pengelompokan wilayah dapat digunakan untuk mengklasifikasikan kecamatan berdasarkan karakteristik dampak bencana, seperti tingkat kerusakan dan jumlah korban terdampak. Informasi ini menjadi penting dalam mendukung perumusan kebijakan mitigasi yang lebih terarah dan berbasis bukti (*evidence-based policy*).

Meskipun demikian, studi-studi sebelumnya terkait pengelompokan wilayah rawan bencana masih menunjukkan sejumlah keterbatasan. Penelitian "*Klastering Bencana Alam Menggunakan K-Means pada Wilayah Jawa Barat*" cenderung berfokus pada intensitas kejadian tanpa mengintegrasikan variabel dampak bencana secara komprehensif, serta belum mempertimbangkan keberadaan *outlier* dalam data [17]. Di sisi lain, penelitian "*Hybrid Hierarchical Klastering dalam Pengelompokan Daerah Rawan Bencana Tanah Longsor di Sulawesi Selatan*" telah mengadopsi pendekatan klastering yang lebih kompleks, namun belum mengakomodasi karakteristik data seperti *outlier* maupun potensi redundansi informasi antar variabel [18]. Berdasarkan kajian tersebut, terdapat kesenjangan penelitian yang signifikan, khususnya dalam: (1) pengelompokan wilayah pada skala yang lebih rinci hingga tingkat kecamatan, (2) penggunaan variabel yang secara langsung merepresentasikan dampak bencana, serta (3) pengolahan data yang mempertimbangkan karakteristik seperti *outlier* dan redundansi informasi. Kesenjangan ini menunjukkan perlunya pendekatan analisis yang lebih robust dan representatif dalam menghasilkan struktur pengelompokan wilayah rawan bencana.

Sejalan dengan hal tersebut, penelitian ini mengusulkan pendekatan pengelompokan menggunakan metode K-Medoids yang memiliki ketahanan terhadap *outlier*, serta didukung dengan penerapan Principal Component Analysis (PCA) untuk mereduksi dimensi data dan menyederhanakan struktur variabel sebelum proses klasterisasi

[5][6][7]. Pendekatan ini diharapkan mampu menghasilkan pengelompokan yang lebih stabil dan informatif dalam merepresentasikan karakteristik wilayah terdampak bencana.

Tujuan penelitian ini adalah untuk mengelompokkan kecamatan di Kabupaten Cianjur berdasarkan dampak kejadian tanah longsor tahun 2024 guna mengidentifikasi wilayah dengan tingkat kerentanan tinggi sebagai dasar penyusunan strategi mitigasi bencana yang lebih efektif dan berbasis data.

## 2. Metodologi Penelitian

### 2.1. Jenis dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari Rekapitulasi Sebaran Kejadian Bencana Tahun 2024 yang diterbitkan oleh Badan Penanggulangan Bencana Daerah (BPBD) Provinsi Jawa Barat. Fokus penelitian ini adalah kejadian bencana tanah longsor di Kabupaten Cianjur dengan unit analisis berupa kecamatan. Variabel yang digunakan dalam penelitian ini meliputi Jiwa Menderita (X1), Rumah Rusak Berat (X2), Rumah Rusak Sedang (X3), Rumah Rusak Ringan (X4), Bangunan (X5), dan Fasilitas Umum (X6). Definisi operasional masing-masing variabel disajikan pada Tabel 1.

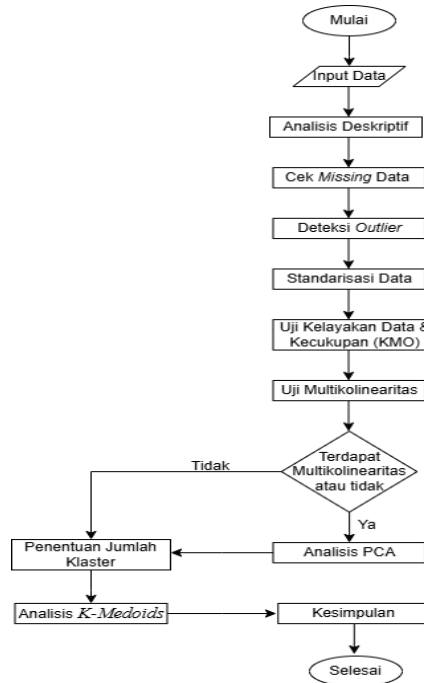
**Tabel 1** Variabel dan Definisi Variabel

Variabel	Kode	Keterangan	Satuan
Jiwa Menderita	X1	Individu ataupun kelompok yang terdampak bencana, mengalami kesulitan fisik, psikologis, atau ekonomi atau serta kehilangan nyawa akibat bencana.	Jiwa
Rumah Rusak Berat	X2	Bangunan rumah dengan kondisi roboh yang mana senagain besar dari komponen struktural rusak dengan kondisi bagian yang mengalami kerusakan lebih dari 45% dari keseluruhan bangunan.	Unit
Rumah Rusak Sedang	X3	Rumah dengan kondisi masih berdiri tetapi kondisi sebagian komponen struktural patah dan non struktural rusak, secara fisik kerusakan rumah diantara 31% - 45%.	Unit
Rumah Rusak Ringan	X4	Rumah dengan kondisi masih tegak/berdiri, terdiri dari beberapa komponen non struktural serta arsitektual rusak yang jika di presentasikan secara fisik tingkat keparahan kerusakan kurang dari 30%.	Unit
Bangunan	X5	Struktur fisik bangunan yang digunakan untuk berbagai keperluan, baik sebagai hunian, tempat usaha, dan lainnya.	Unit
Fasilitas Umum	X6	Infrastruktur yang disediakan untuk kepentingan masyarakat luas dan tidak bersifat pribadi.	Unit

Sumber: Pratinjau Data Rekapitulasi BPBD Provinsi Jawa Barat Tahun 2024

## 2.2. Tahapan Analisis

Tahapan analisis yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 2. yang menunjukkan alur proses secara sistematis, termasuk penerapan metode K-Medoids dalam proses pengelompokan wilayah per kecamatan di kabupaten Cianjur.



**Gambar 2.** Tahapan Penelitian

Pada **Gambar 2** proses analisis dimulai dengan penginputan data dan eksplorasi awal menggunakan analisis deskriptif untuk memperoleh gambaran umum karakteristik data, seperti nilai minimum, maksimum, rata-rata, dan sebaran data [8]. Selanjutnya dilakukan pemeriksaan terhadap *missing data* untuk memastikan kelengkapan dataset. Apabila ditemukan data yang tidak lengkap, dilakukan penanganan melalui penghapusan atau imputasi sederhana sesuai kebutuhan analisis [9].

Tahap berikutnya adalah deteksi *outlier* menggunakan visualisasi *boxplot* untuk mengidentifikasi nilai ekstrem yang berpotensi memengaruhi hasil analisis. Setelah itu, seluruh variabel distandarisasi menggunakan metode *Z-score* untuk menyamakan skala antar variabel [10]. Evaluasi kelayakan data dilakukan menggunakan Kaiser-Meyer-Olkin (KMO) dan Measure of Sampling Adequacy (MSA). Data dinyatakan layak apabila nilai  $KMO > 0,5$  [13][14]. Apabila terdapat korelasi tinggi antar variabel yang menyebabkan redundansi informasi, maka dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) [12].

Penentuan jumlah kluster optimal dilakukan menggunakan metode *Silhouette Coefficient*, dengan memilih jumlah kluster yang menghasilkan nilai rata-rata *silhouette* tertinggi. Selanjutnya, proses klasterisasi dilakukan menggunakan metode *K-Medoids* dengan jarak Euclidean. Proses ini dilakukan secara iteratif hingga diperoleh medoid yang stabil [15][16]. Hasil kluster kemudian diinterpretasikan berdasarkan karakteristik masing-masing kelompok kecamatan untuk mendukung pengambilan keputusan mitigasi bencana.

## 2.3. Statistika Deskriptif

Statistika deskriptif digunakan untuk memberikan gambaran umum mengenai karakteristik data, khususnya melalui ukuran rata-rata dan simpangan baku [8]. Berikut ini merupakan rumus rata-rata:

$$\bar{X} = \frac{\sum x_i}{n} \quad (1)$$

dengan  $\bar{X}$  digunakan untuk menyatakan nilai rata-rata dari suatu kumpulan data,  $\sum x_i$  menunjukkan jumlah seluruh nilai data, sedangkan  $x_i$  mewakili setiap nilai dalam data. Selanjutnya merupakan rumus simpangan baku atau *standard deviation*, sebagai berikut:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}} \quad (2)$$

Dibagian  $x_i$  merupakan tiap nilai data,  $\bar{x}$  adalah rata-rata, dan  $n$  merupakan jumlah data. Selisih antara tiap nilai dan rata-rata dikuadratkan, dijumlahkan, lalu dibagi  $n - 1$  untuk mengoreksi bias karena data hanya sampel. Hasil akhirnya diakar untuk mendapatkan simpangan baku.

#### 2.4. Pra-pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum analisis klusterisasi. Tahap ini meliputi identifikasi *missing data*, deteksi *outlier*, dan standarisasi data. *Missing data* ditangani melalui penghapusan data yang tidak lengkap atau imputasi sederhana sesuai kebutuhan analisis [9]. Deteksi *outlier* dilakukan menggunakan *boxplot*, sedangkan standarisasi data dilakukan menggunakan metode *Z-score* untuk menyamakan skala antar variabel [10].

#### 2.5. Standarisasi Data

Standardisasi adalah dasar atau pedoman yang digunakan sebagai acuan minimal agar data menjadi seragam [10]. Proses ini dilakukan dengan menyesuaikan data agar nilai rata-ratanya menjadi 0 dan simpangan bakunya menjadi 1. Langkah ini digunakan untuk menyamakan skala antar variabel, sehingga tidak ada variabel yang mendominasi analisis hanya karena memiliki satuan atau rentang nilai yang lebih besar.

$$X' = \frac{X - \mu}{\sigma} \quad (3)$$

$X$  merupakan nilai asli data,  $\mu$  adalah nilai dari rata-rata keseluruhan data, serta  $\sigma$  adalah sebuah nilai yang didapat dari proses standar deviasi data.

#### 2.6. Uji Kecukupan dan Kelayakan Data

Uji kelayakan data dilakukan untuk memastikan bahwa data memenuhi syarat untuk analisis multivariat. Salah satu ukuran yang digunakan adalah *Kaiser-Meyer-Olkin* (KMO) dengan rumus:

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p a_{ij}^2} \quad (4)$$

Nilai  $KMO > 0,5$  menunjukkan bahwa data memiliki kecukupan sampel yang memadai [13]. Selain itu, dilakukan pengujian *Measure of Sampling Adequacy* (MSA) untuk menilai kelayakan masing-masing variabel. Variabel dengan nilai  $MSA < 0,5$  akan dikeluarkan dari analisis [14].

## 2.7. Uji Multikolinearitas

Dalam analisis kluster, keberadaan korelasi yang tinggi antar variabel dapat menyebabkan redundansi informasi dan memengaruhi hasil pengelompokan. Oleh karena itu, perlu dilakukan evaluasi keterkaitan antar variabel sebelum proses klusterisasi. Apabila ditemukan korelasi yang tinggi antar variabel, maka dilakukan reduksi dimensi menggunakan *Principal Component Analysis* (PCA) untuk mengurangi redundansi informasi dan menghasilkan variabel baru yang saling tidak berkorelasi [12].

## 2.8. *Principal Component Analysis* (PCA)

Reduksi dimensi dilakukan menggunakan *Principal Component Analysis* (PCA) untuk menyederhanakan struktur data dengan mengurangi redundansi informasi akibat korelasi antar variabel [12]. Sebelum penerapan PCA, dilakukan evaluasi kelayakan data menggunakan nilai Kaiser-Meyer-Olkin (KMO) dan Measure of Sampling Adequacy (MSA), dengan kriteria nilai KMO > 0,5 [13][14]. Komponen utama yang dipertahankan ditentukan berdasarkan nilai *eigenvalue* > 1 serta proporsi kumulatif varians yang mampu menjelaskan sebagian besar informasi data.

## 2.9. *K-Medoids* Klustering

Metode *K-Medoids* digunakan untuk mengelompokkan kecamatan berdasarkan kemiripan karakteristik dampak bencana. Metode ini dipilih karena lebih robust terhadap *outlier* dibandingkan metode berbasis rata-rata [15][16].

Jarak antar objek dihitung menggunakan jarak Euclidean:

$$d(i, j) = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (5)$$

Dengan  $x_{ik}$  dan  $x_{jk}$  adalah nilai atribut ke- $k$  dari objek  $k$  dan objek  $i, n$  merupakan jumlah atribut atau dimensi.

Penentuan jumlah kluster dilakukan menggunakan metode *Silhouette Coefficient*, dengan memilih jumlah kluster yang menghasilkan nilai *silhouette* rata-rata tertinggi. Nilai *silhouette* mendekati 1 menunjukkan kualitas kluster yang baik, sedangkan nilai mendekati 0 menunjukkan adanya tumpang tindih antar kluster. Implementasi klusterisasi dilakukan menggunakan algoritma PAM (*Partitioning Around Medoids*) hingga diperoleh medoid yang stabil (konvergen) [15][16].

## 3. Hasil dan Pembahasan

Berikut merupakan data yang digunakan dalam penelitian, data yang di tampilkan pada Tabel 2 merupakan data Pratinjau Rekapitulasi BPBD Provinsi Jawa Barat Tahun 2024, pada penelitian ini data yang digunakan merupakan data sebaran bencana alam tanah longsor yang terjadi di daerah kabupaten Cianjur. Data yang digunakan mencakup data kecamatan yang ada di kabupaten Cianjur, jumlah jiwa menderita, rumah dengan rusak berat, sedang, ringan, serta bangunan dan fasilitas umum yang terdampak akibat bencana tanah longsor.

**Tabel 2** Data Dampak Kejadian Bencana Tanah Longsor

Kecamatan	X1	X2	X3	X4	X5	X6
Argabinta	696	73	54	19	1	26
Bojongcipung	204	4	6	26	1	7
Campaka	253	8	33	1	0	10
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
Takokak	1479	32	77	142	1	47

Kecamatan	X1	X2	X3	X4	X5	X6
Tanggeung	2520	128	123	407	0	52
Warungkondang	0	0	0	0	0	0

**Tabel 2** menampilkan data serta variabel yang akan digunakan dalam analisis *K-Medoids* dalam penelitian ini, data digunakan sebagai dasar untuk mengidentifikasi pola kerentanan wilayah melalui pendekatan klusterisasi berbasis karakteristik dampak bencana..

### 3.1. Analisis Deskriptif

Analisis deskriptif pada tahap ini bertujuan untuk memberikan gambaran awal mengenai karakteristik dan dampak bencana tanah longsor di Kabupaten Cianjur pada tahun 2024. Tabel berikut menyajikan ringkasan statistik dari sejumlah variabel yang berkaitan dengan dampak kejadian, seperti jumlah jiwa menderita, rumah rusak (berat, sedang, dan ringan), bangunan terdampak, serta fasilitas umum yang rusak. Penyajian nilai minimum, maksimum, median, dan rata-rata pada tiap variabel bertujuan untuk mengidentifikasi pola umum serta potensi kejadian ekstrem yang memengaruhi tingkat kerusakan akibat tanah longsor. Berikut merupakan ringkasan data dari setiap variable yang ada, terdapat pada **Tabel 3**.

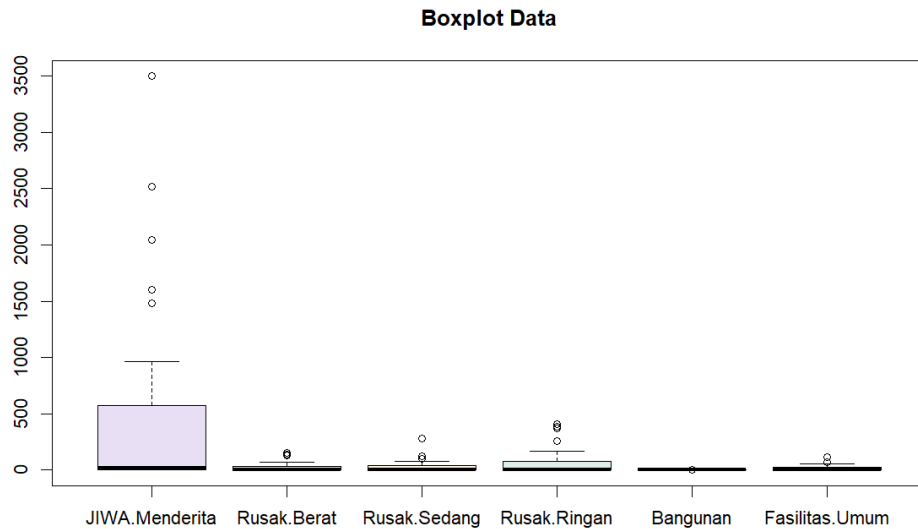
**Tabel 3** Ringkasan Data

Variabel	Minimum	Maksimum	Median	Mean
X1	0	3497	20	487
X2	0	153	1	24.56
X3	0	284	1	28.78
X4	0	407	0	65.81
X5	0	1	0	0.22
X6	0	114	1.5	15.94

Pada **Tabel 3**, hasil analisis deskriptif menunjukkan bahwa seluruh variabel memiliki distribusi yang tidak merata, yang ditandai dengan nilai rata-rata yang jauh lebih besar dibandingkan median. Kondisi ini mengindikasikan adanya ketimpangan dampak antar kecamatan, di mana sebagian besar wilayah mengalami dampak yang relatif kecil, namun terdapat beberapa kecamatan dengan dampak yang sangat besar. Secara substantif, hal ini menunjukkan bahwa kejadian tanah longsor di Kabupaten Cianjur bersifat tidak merata secara spasial, sehingga pendekatan penanganan bencana tidak dapat dilakukan secara seragam. Perbedaan yang signifikan antara median dan rata-rata, khususnya pada variabel jumlah jiwa menderita, mengindikasikan adanya konsentrasi dampak pada wilayah tertentu yang dapat dikategorikan sebagai hotspot bencana. Pola serupa juga terlihat pada variabel kerusakan rumah dan fasilitas umum, yang menunjukkan adanya kejadian ekstrem yang berkontribusi besar terhadap nilai rata-rata.

### 3.2. Uji Outlier

Uji *outlier* dilakukan untuk mendeteksi nilai-nilai ekstrem yang menyimpang jauh dari pola umum data. *Outlier* adalah observasi yang berada di luar rentang sebaran mayoritas data, dan keberadaannya dapat memengaruhi hasil analisis sehingga menjadi kurang akurat dan tidak representatif terhadap populasi. Untuk mengidentifikasi outlier, digunakan visualisasi *boxplot* yang memperlihatkan sebaran data dan posisi nilai-nilai yang menyimpang.



**Gambar 3.** Visualisasi *Boxplot Data Outlier*

Berdasarkan hasil pengecekan *outlier* menggunakan *boxplot* pada **Gambar 3**, menunjukkan bahwa hampir seluruh variabel memiliki nilai ekstrem yang berada di luar rentang sebaran mayoritas data. Dalam konteks penelitian ini, keberadaan outlier tidak hanya dipandang sebagai penyimpangan statistik, tetapi juga mencerminkan adanya kejadian bencana dengan dampak yang sangat besar pada wilayah tertentu. Oleh karena itu, outlier justru memiliki makna substantif sebagai indikator wilayah prioritas penanganan. Kondisi ini menjadi dasar pemilihan metode *K-Medoids*, yang lebih robust terhadap outlier dibandingkan metode seperti *K-Means*. Dengan menggunakan *K-Medoids*, pusat kluster ditentukan berdasarkan data aktual (*medoid*), sehingga hasil pengelompokan lebih representatif terhadap kondisi nyata di lapangan.

### 3.3. Standarisasi Data

Standarisasi data dilakukan karena terdapat perbedaan skala antar variabel yang cukup signifikan, seperti jumlah jiwa terdampak yang dapat mencapai ribuan, sementara variabel bangunan hanya memiliki nilai biner. Tanpa proses standarisasi, variabel dengan skala besar akan mendominasi perhitungan jarak dalam proses klusterisasi, sehingga dapat menghasilkan pengelompokan yang bias. Oleh karena itu, standarisasi bertujuan untuk menyetarakan skala seluruh variabel agar setiap variabel memiliki kontribusi yang seimbang dalam analisis. Dengan demikian, hasil klusterisasi yang diperoleh mencerminkan kombinasi seluruh indikator dampak secara proporsional..

**Tabel 4** Data Hasil standarisasi

X1	X2	X3	X4	X5	X6
0.254	1.137	0.442	-0.381	1.860	0.384
-0.321	-0.482	-0.400	-0.324	1.860	-0.341
-0.263	-0.389	0.074	-0.528	-0.520	-0.227
-0.331	-0.248	-0.242	-0.495	-0.520	-0.303
-0.559	-0.576	-0.505	-0.536	-0.520	-0.571
...	...	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
1.170	0.174	0.846	0.621	1.860	1.187
2.388	2.429	1.654	2.781	-0.520	1.379
-0.559	-0.576	-0.505	-0.536	-0.520	-0.609

### 3.4. Uji Kecukupan dan Kelayakan Data

Nilai KMO digunakan untuk melihat apakah data layak digunakan untuk analisis selanjutnya. Adapun kriteria uji KMO adalah ketika nilai KMO > 0.5 maka data dianggap memiliki tingkat kelayakan yang cukup untuk dilakukan analisis lebih lanjut, karena struktur korelasi antar variabel dinilai memadai. Berikut merupakan hasil perhitungan KMO. Sama halnya dengan *Measure of Sampling Adequacy* (MSA) sebagai uji kelayakan model, jika nilai MSA lebih dari 0.5 maka model dari setiap variabel layak digunakan. Berikut merupakan hasil perhitungan KMO dan MSA.

**Tabel 5** Nilai KMO dan MSA

KMO	0.79
Variabel	MSA
X1	0.70
X2	0.89
X3	0.73
X4	0.70
X5	0.90
X6	0.95

Dari hasil perhitungan pada **Tabel 5** nilai *Kaiser-Meyer-Olkin* (KMO) sebesar 0,79 menunjukkan bahwa data memiliki tingkat kecukupan yang baik untuk dilakukan analisis lebih lanjut, khususnya dalam konteks reduksi dimensi menggunakan PCA. Dalam penelitian ini, KMO tidak diinterpretasikan dalam kerangka analisis regresi, melainkan sebagai indikator bahwa struktur korelasi antar variabel cukup kuat dan layak untuk direduksi. Selain itu, nilai *Measure of Sampling Adequacy* (MSA) pada seluruh variabel yang berada di atas 0,5 menunjukkan bahwa setiap variabel memiliki kontribusi yang memadai dalam membentuk struktur data. Hal ini mengindikasikan bahwa data memiliki redundansi informasi yang cukup tinggi, sehingga penggunaan PCA menjadi relevan untuk mengurangi kompleksitas tanpa kehilangan informasi penting.

### 3.5. Evaluasi Kelayakan Kluster

Pada tahapan ini data perlu dilakukan Evaluasi kelayakan kluster, sebelum dilakukan analisis kluster, terlebih dahulu dilakukan pengujian asumsi untuk memastikan kelayakan data dalam analisis kluster. Salah satu pengujian yang dilakukan adalah menghitung nilai Variance Inflation Factor (VIF) guna mendeteksi adanya multikolinearitas antar variabel. Nilai VIF digunakan untuk menilai apakah variabel bebas memiliki korelasi tinggi satu sama lain, yang dapat memengaruhi validitas hasil klusterisasi.

#### 3.5.1. Uji Multikolinearitas

**Tabel 6** Nilai VIF Suatu Variabel Dengan Variabel Bebas Lainnya

Variabel	X1	X2	X3	X4	X5	X6
X1	-	8.463	11.080	4.316	1.249	14.965
X2	70.239	-	23.521	18.452	1.229	14.770
X3	39.680	10.148	-	9.993	1.286	13.723
X4	20.143	10.374	13.022	-	1.273	15.418
X5	87.409	10.361	25.133	19.092	-	15.508
X6	86.868	10.329	22.243	19.178	1.286	-

Pada **Tabel 6** didapatkan nilai VIF dari setiap variabel yang ada, yang mana jika nilai VIF < 10 maka terbebas dari multikolinearitas dan jika VIF > 10 maka akan terjadi multikolinearitas. Dari hasil perhitungan tersebut terdapat beberapa variabel memiliki nilai VIF >10, untuk itu perlu melakukan penanganan terhadap multikolinearitas tersebut menggunakan *Principal Component Analysis* (PCA).

### 3.6. Principal Component Analysis (PCA)

Analisis Komponen Utama *Principal Component Analysis* (PCA) merupakan suatu metode statistik yang digunakan untuk menangani permasalahan multikolinearitas antar variabel dalam suatu dataset. Metode ini bekerja dengan mentransformasikan sekumpulan variabel asal yang saling berkorelasi menjadi sejumlah variabel baru yang jumlahnya lebih sedikit, namun saling bebas (tidak berkorelasi). Transformasi ini dilakukan melalui proses reduksi dimensi data, dengan tetap mempertahankan sebanyak mungkin informasi yang terkandung dalam variabel-variabel awal. Untuk menentukan jumlah komponen utama yang relevan, digunakan pendekatan nilai *eigen* (*eigenvalue*), di mana komponen dengan nilai *eigen* lebih dari satu dianggap layak untuk dipertahankan dalam analisis. Berikut disajikan hasil perhitungan nilai eigen pada **Tabel 7**.

**Tabel 7 Eigen Value**

Komponen Utama	Nilai Eigen
1	4.803
2	0.813
3	0.238
4	0.090
5	0.045
6	0.008

Hasil perhitungan nilai eigen yang ditampilkan pada **Tabel 7** menunjukkan bahwa komponen utama pertama memiliki nilai eigen yang jauh lebih besar dibandingkan komponen lainnya dan mampu menjelaskan sebagian besar variasi data. Meskipun secara kriteria Kaiser hanya satu komponen yang memenuhi syarat ( $eigenvalue > 1$ ), penelitian ini menggunakan dua komponen utama dengan pertimbangan metodologis yang lebih komprehensif.

**Tabel 8 Ringkasan PCA**

Komponen	Standar Deviasi	Variansi (%)	Kumulatif (%)
1	1.480	80.05	80.05
2	0.949	13.56	93.61
3	0.698	3.983	97.59
4	0.549	1.514	99.11
5	0.461	0.756	99.87
6	0.300	0.135	100.00

Dua komponen tersebut mampu menjelaskan lebih dari 90% variasi data, sehingga hampir seluruh informasi dalam data tetap terjaga. Selain itu, penggunaan dua komponen memungkinkan representasi struktur data yang lebih kompleks dan meningkatkan kemampuan pemisahan kluster. Dengan demikian, keputusan penggunaan dua komponen tidak hanya didasarkan pada kriteria statistik semata, tetapi juga mempertimbangkan kebutuhan analisis kluster yang lebih optimal. Pada **Tabel 9** berdasarkan dua komponen utama tersebut untuk masing-masing kecamatan yang menjadi objek penelitian.

**Tabel 9 Komponen Utama PCA**

No	Kecamatan	PC1	PC2
1	Argabinta	-1.223	-1.594
2	Bojongpicung	0.404	-1.993
3	Campaka	0.691	0.332
...	...	...	...

No	Kecamatan	PC1	PC2
...	...	...	
30	Takokak	-2.167	-1.477
31	Tenggeung	-4.509	1.639
	Warungkondang	1.331	0.226
32			

### 3.7. Penentuan Jumlah Klaster

Penentuan jumlah klaster optimal dalam penelitian ini dilakukan melalui dua pendekatan, yaitu metode NbClust yang mencakup 30 indeks evaluasi, serta metode validasi alternatif menggunakan fungsi *cValid*. Penggunaan NbClust didasarkan pada keunggulannya dalam mengintegrasikan berbagai metode evaluasi *internal* dan *stability* index secara komprehensif, sehingga memberikan dasar statistik yang kuat untuk menentukan jumlah klaster yang paling sesuai dengan struktur data. Hasil dari NbClust menunjukkan bahwa dari 30 metode yang digunakan, sebanyak 8 metode merekomendasikan 10 klaster, 6 metode menyarankan 3 klaster, dan 5 metode menyarankan 2 klaster. Sisanya mengusulkan 4, 5, 6, dan 8 klaster seperti ditampilkan dalam **Tabel 10** berikut.

**Tabel 10** Jumlah Klaster Berdasarkan 30 Metode NbClust

Jumlah Klaster	Jumlah Metode yang Mendukung
2	5
3	6
4	1
5	2
6	1
8	1
10	8

Berdasarkan prinsip *majority rule*, NbClust menyimpulkan bahwa jumlah klaster optimal adalah 10. Namun demikian, jumlah klaster sebanyak 10 dianggap kurang relevan secara praktis, terutama dalam konteks penanganan bencana tanah longsor yang membutuhkan kategorisasi wilayah terdampak yang sederhana, terfokus, dan implementatif. Terlalu banyak klaster justru dapat mempersulit proses identifikasi prioritas dan penetapan strategi penanggulangan di lapangan. Sebagai respons terhadap hal tersebut, dilakukan validasi tambahan menggunakan metode *cValid* dengan algoritma Partitioning Around Medoids (PAM). Hasil evaluasi dari metode ini ditampilkan dalam **Tabel 11** berikut.

**Tabel 11** Hasil Validasi Jumlah Klaster

Indeks Validasi	Nilai	Jumlah Klaster Optimal
APN	0.119	2
AD	1.231	3
ADM	0.681	3
FOM	1.338	3
<i>Connectivity</i>	9.272	2
<i>Dunn</i>	0.287	3
<i>Silhouette</i>	0.696	3

Hasil validasi jumlah klaster menunjukkan bahwa sebagian besar indeks mendukung pemilihan tiga klaster sebagai jumlah yang paling optimal. Pemilihan ini

terutama didasarkan pada nilai indeks *Silhouette* sebesar 0.696, yang mengindikasikan bahwa struktur klaster yang terbentuk memiliki pemisahan yang cukup jelas dan konsisten. Indeks *Silhouette* mengukur tingkat keseragaman objek di dalam suatu klaster dibandingkan dengan objek di klaster lainnya, sehingga nilai yang mendekati satu mencerminkan bahwa klaster yang terbentuk memiliki homogenitas internal yang tinggi serta heterogenitas yang kuat antar klaster.

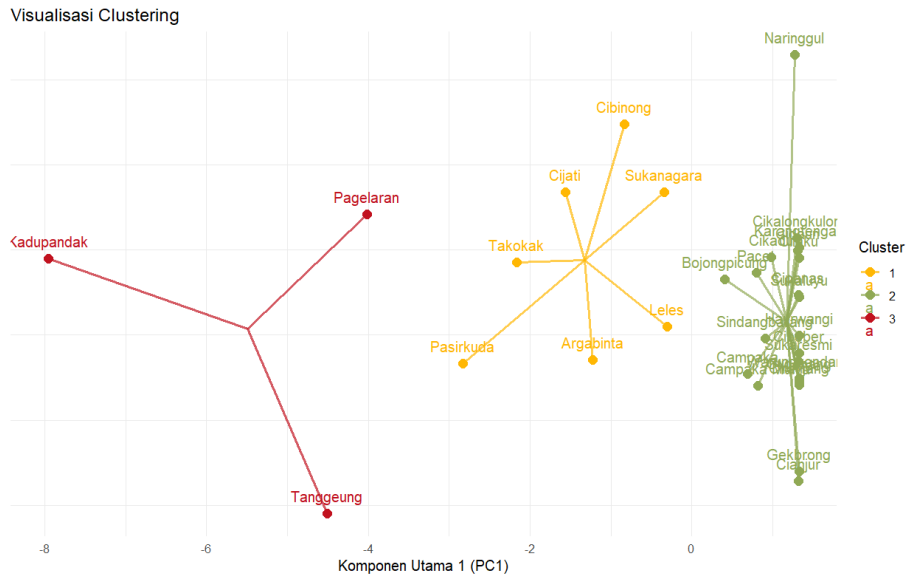
Selain itu, indeks validasi lain seperti *Average Distance* (AD), *Average Distance Between Means* (ADM), dan *Figure of Merit* (FOM) juga menunjukkan nilai yang relatif rendah pada konfigurasi tiga klaster, yang mengindikasikan bahwa klaster yang terbentuk memiliki kualitas internal yang baik dan tidak saling tumpang tindih. Meskipun indeks *Average Proportion of Non-overlap* (APN) dan *Connectivity* menunjukkan preferensi terhadap dua klaster, nilai-nilai tersebut tidak memberikan keunggulan yang substansial bila dibandingkan dengan konfigurasi tiga klaster. Dengan mempertimbangkan keseluruhan hasil evaluasi serta relevansinya terhadap tujuan analisis, maka ditetapkan bahwa jumlah klaster optimal dalam penelitian ini adalah sebanyak tiga klaster

### 3.8. *K-Medoids* Klustering

Setelah permasalahan multikolinearitas berhasil diatasi dengan menerapkan metode *Principal Component Analysis* (PCA), tahap selanjutnya adalah melakukan analisis pengelompokan (klustering) menggunakan algoritma *K-Medoids*. Pemilihan metode ini didasarkan pada karakteristik data yang mengandung *outlier*, yang terdeteksi pada beberapa variabel dan berpotensi memengaruhi hasil analisis jika digunakan metode klustering yang sensitif terhadap *outlier* seperti *K-Means*. Algoritma *K-Medoids* dipilih karena memiliki keunggulan dalam mengelompokkan data yang mengandung nilai pencilan, sehingga lebih *robust* terhadap pengaruh *outlier*. Proses klustering dilakukan terhadap data dari 32 kecamatan yang ada di kabupaten Cianjur, dan hasilnya menunjukkan objek-objek yang terpilih sebagai *medoids* atau titik pusat dari masing-masing klaster yang tersaji pada **Tabel 12** sebagai berikut.

No	Kecamatan	ID	PC1
1	Argabinta	1	-1.223
2	Cianjur	5	1.314
3	Tenggeung	31	-4.509

**Tabel 12** menampilkan daftar kecamatan yang teridentifikasi sebagai titik pusat (*medoid*) dari hasil klusterisasi menggunakan metode *K-Medoids* menunjukkan bahwa data terbagi ke dalam tiga kelompok dengan karakteristik yang berbeda. *Medoid* yang terpilih pada masing-masing klaster merepresentasikan kondisi tipikal dari setiap kelompok. Kecamatan Argabinta, Cianjur, dan Tanggeung menjadi pusat klaster yang menggambarkan variasi tingkat dampak bencana, mulai dari rendah hingga sangat tinggi. Secara visual, hasil klusterisasi menunjukkan adanya pemisahan yang cukup jelas antar kelompok, yang menandakan bahwa metode yang digunakan mampu mengidentifikasi pola dalam data dengan baik. Klaster yang terbentuk juga menunjukkan variasi tingkat homogenitas, di mana beberapa klaster lebih terkonsentrasi dibandingkan yang lain. Berikut merupakan hasil pengelompokan kecamatan di kabupaten Cianjur tahun 2024 berdasarkan dampak kejadian tanah longsor.



**Gambar 4.** Plot Klaster

Pada **Gambar 4.** merupakan visualisasi hasil dari algoritma *K-Medoids Klustering* yang membagi data ke dalam tiga klaster berbeda. Setiap klaster dibedakan berdasarkan warna, yaitu kuning (klaster 1), hijau (klaster 2), dan merah (klaster 3). Titik-titik pada visualisasi merepresentasikan data poin, sementara garis-garis menunjukkan hubungan antara masing-masing titik dengan *medoid* klasternya, menggambarkan ruang cakupan dari setiap klaster.

Klaster hijau (2) tampak lebih terkonsentrasi di satu area, yang mengindikasikan bahwa data dalam klaster tersebut memiliki kemiripan yang tinggi atau bersifat lebih homogen. Sebaliknya, klaster kuning (1) dan merah (3) terlihat lebih menyebar, menunjukkan bahwa data di dalamnya memiliki variasi yang lebih besar. Pemisahan antara klaster merah dan hijau terlihat cukup jelas, sedangkan klaster kuning tampak terpisah dari keduanya, yang mungkin menunjukkan bahwa data dalam klaster tersebut memiliki karakteristik yang berbeda secara signifikan dari dua klaster lainnya. Berikut merupakan hasil dari klasterisasi data dampak kejadian tanah longsor 2024 berdasarkan kecamatan di kabupaten Cianjur.

**Tabel 13** Hasil Klasterisasi

Klaster	Jumlah	Anggota
1	7	Argabinta, Cibinong, Cijati, Leles, Pasirkuda, Sukanegara, Takokak.
2	22	Bojongpicung, Campaka, Campaka Mulya, Cianjur, Cibeber, Cidaun, Cikadu, Cikalongkulon, Cilaku, Cipanas, Ciranjang, Cugenang, Gekbrong, Haurwangi, Karangtengah, Mandeh, Naringgul, Pacet, Sidangbarang, Sukaluyu, Sukaresmi, Warungkondang.
3	3	Kadupandak, Pagelaran, Tanggeung.

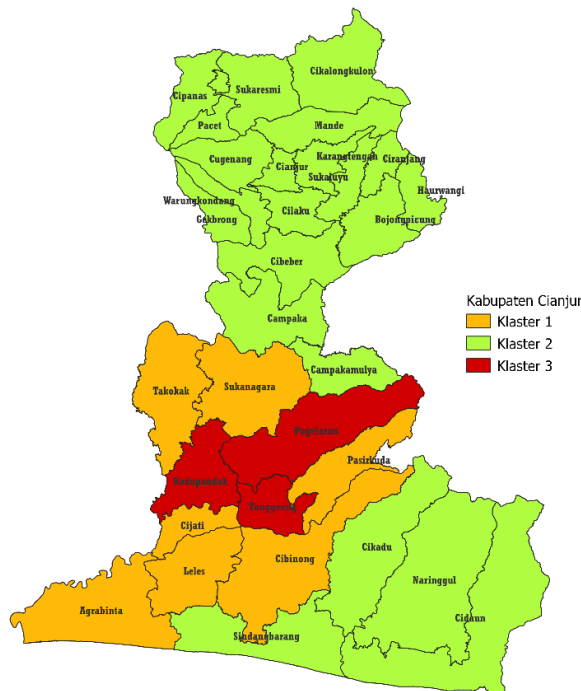
Pada **Tabel 13** pengelompokan berdasarkan jumlah dampak kejadian tanah longsor di kabupaten Cianjur tahun 2024 didapatkan kelompok 1 terdiri dari 7 anggota kecamatan, kelompok 2 berjumlah 22 anggota, dan kelompok 3 memiliki anggota sebanyak 3 kecamatan di kabupaten Cianjur.

**Tabel 14** Profilisasi Klaster *K-Medoids*

Klaster	X1	X2	X3	X4	X5	X6	Keterangan
1	907	44	47.3	146.3	0.6	33.6	Dampak Berat

Klaster	X1	X2	X3	X4	X5	X6	Keterangan
2	41.1	2.7	3.6	2	0.1	1.8	Dampak Ringan
3	2686.7	139.7	170	345.7	0.3	78.7	Dampak Sangat Berat

Hasil profilisasi menunjukkan bahwa klaster yang terbentuk memiliki karakteristik yang berbeda secara signifikan. Klaster pertama menunjukkan tingkat dampak yang tergolong berat, dengan nilai yang cukup tinggi pada variabel jumlah jiwa terdampak dan kerusakan rumah, meskipun belum mencapai kondisi ekstrem. Klaster kedua memiliki nilai yang relatif rendah pada seluruh variabel, sehingga dapat dikategorikan sebagai wilayah dengan dampak ringan dan kondisi yang relatif stabil. Sementara itu, klaster ketiga menunjukkan nilai yang sangat tinggi pada hampir seluruh variabel, yang mengindikasikan bahwa wilayah dalam klaster ini mengalami dampak yang sangat berat dan dapat dikategorikan sebagai zona prioritas penanganan. Perbedaan karakteristik ini menunjukkan bahwa setiap klaster memiliki pola dominan yang berbeda, sehingga memerlukan pendekatan penanganan yang berbeda pula.



**Gambar 5.** Pemetaan Klaster

**Gambar 5.** memperlihatkan hasil pemetaan klaster menggunakan perangkat lunak QGIS. Peta Kabupaten Cianjur ditampilkan dalam tiga warna berbeda untuk membedakan masing-masing klaster yang terbentuk. Warna merah merepresentasikan klaster ketiga, yaitu wilayah dengan Dampak Berat warna jingga menunjukkan klaster pertama, yaitu Dampak Sedang dan warna hijau merepresentasikan klaster kedua, yaitu wilayah dengan Dampak Ringan. Hasil klasterisasi yang diperoleh dalam penelitian ini secara langsung menjawab tujuan penelitian, yaitu mengelompokkan kecamatan berdasarkan dampak bencana tanah longsor. Melalui pengelompokan ini, dapat diidentifikasi wilayah dengan tingkat kerentanan yang berbeda, mulai dari rendah hingga sangat tinggi. Selain itu, hasil ini juga menunjukkan adanya ketimpangan dampak antar wilayah, yang menegaskan pentingnya pendekatan berbasis data dalam perencanaan penanggulangan bencana.

Dengan demikian, penelitian ini tidak hanya menghasilkan klasifikasi wilayah, tetapi juga memberikan dasar untuk pengambilan keputusan yang lebih tepat sasaran.

#### 4. Kesimpulan

Berdasarkan hasil analisis dan pembahasan, penelitian ini menunjukkan bahwa metode K-Medoids mampu mengelompokkan kecamatan di Kabupaten Cianjur berdasarkan karakteristik dampak bencana tanah longsor secara efektif dan robust terhadap keberadaan outlier. Penentuan jumlah kluster optimal menghasilkan tiga kluster, yang didukung oleh nilai evaluasi Silhouette yang menunjukkan kualitas pemisahan kluster yang baik.

Hasil klusterisasi mengidentifikasi adanya perbedaan tingkat dampak yang jelas antar wilayah. Kluster pertama dikategorikan sebagai wilayah dengan dampak berat, kluster kedua sebagai wilayah dengan dampak ringan, dan kluster ketiga sebagai wilayah dengan dampak sangat berat. Kluster dampak sangat berat memiliki karakteristik nilai tertinggi pada hampir seluruh variabel, yang menunjukkan tingginya jumlah korban terdampak serta kerusakan infrastruktur dan fasilitas umum. Sementara itu, kluster dampak ringan menunjukkan kondisi sebaliknya, dengan tingkat kerusakan dan jumlah korban yang relatif rendah. Kluster dampak berat berada di antara keduanya dan mencerminkan wilayah yang memiliki potensi peningkatan risiko jika tidak ditangani secara tepat.

Secara keseluruhan, penelitian ini memberikan kontribusi dalam bentuk pemetaan wilayah terdampak bencana tanah longsor berbasis data, yang mampu mengidentifikasi pola ketimpangan dampak antar kecamatan. Hasil ini menunjukkan bahwa kejadian tanah longsor di Kabupaten Cianjur tidak terjadi secara merata, sehingga memerlukan pendekatan penanganan yang berbasis prioritas wilayah. Implikasi dari penelitian ini menunjukkan bahwa wilayah dengan kategori dampak sangat berat perlu menjadi prioritas utama dalam penanganan bencana, baik dalam bentuk respons darurat, rehabilitasi, maupun perencanaan mitigasi jangka panjang. Wilayah dengan dampak berat memerlukan penguatan sistem mitigasi dan infrastruktur, sedangkan wilayah dengan dampak ringan dapat difokuskan pada upaya pencegahan dan peningkatan kesiapsiagaan.

Namun demikian, penelitian ini memiliki keterbatasan, antara lain penggunaan variabel yang masih terbatas pada data dampak kejadian tanpa mempertimbangkan faktor penyebab seperti kondisi geografis, curah hujan, atau penggunaan lahan. Selain itu, penggunaan satu metode klusterisasi juga membatasi kemungkinan eksplorasi hasil yang lebih beragam.

#### 5. Daftar Pustaka

- [1] Raditya, J. (2021). *Indeks Risiko Bencana Indonesia Tahun 2021*. Badan Nasional Penanggulangan Bencana.
- [2] Ibrahim, I. A. (2022). Penanggulangan bencana daerah Kota Cirebon perspektif Undang-Undang Nomor 24 Tahun 2007. *FOCUS: Jurnal of*, 3(1). <https://doi.org/10.47685/focus.v3i1.312>.
- [3] Raditya, J. (2021). *Indeks Risiko Bencana Indonesia Tahun 2021*. Badan Nasional Aryanto, H. S., et al. (2019). Analisis penyebab kejadian dan evaluasi bencana tanah longsor di Desa Banaran, Kecamatan Pulung, Kabupaten Ponorogo, Provinsi Jawa Timur tanggal 1 April 2017. *Jurnal Ilmu Lingkungan*, 17(2), 272–282. <https://doi.org/10.14710/jil.17.2.272-282>.
- [4] Az'zahra, Y., et al. (2022). Menyiapkan generasi muda sigap menghadapi bencana tanah longsor. *Abdimas Universal*, 4(2). <https://doi.org/10.36277/abdimasuniversal.v4i2.249>
- [5] Herviany, M., et al. (2021). Perbandingan algoritma K-means dan K-medoids untuk pengelompokan daerah rawan tanah longsor di Provinsi Jawa Barat. *Indonesian*

- Journal of Machine Learning and Computer Science*, 1(1), 34–40.  
<https://journal.irpi.or.id/index.php/malcom/article/view/60>
- [6] Han, J., & Kamber, M. (2012). *Data mining: Concepts and techniques* (3rd ed.). Elsevier.
- [7] Sari, D. R. P. (2023). Metode principal component analysis (PCA) sebagai penanganan asumsi multikolinearitas (studi kasus: data produksi tapioka). *Parameter*, 2(2), 115–124. <https://ojs3.unpatti.ac.id/index.php/parameter>
- [8] Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, 22(1), 67–72. [https://doi.org/10.4103/aca.ACA\\_157\\_18](https://doi.org/10.4103/aca.ACA_157_18)
- [9] Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576. <https://doi.org/10.1146/annurev.psych.58.110405.085530>
- [10] Ramadan, Z. H. (2023). Standarisasi kualitas riset di sekolah dasar. *Journal on Education*, 5(2), 1932–1939. <http://jonedu.org/index.php/joe>
- [11] Ghozali, I. (2007). *Aplikasi analisis multivariat dengan program SPSS* (Edisi 4). Badan Penerbit Universitas Diponegoro.
- [12] Rasidia, F., Goejantoro, R., & Fathurahman, M. (2025). Analisis kluster menggunakan metode average linkage dengan validasi multiscale bootstrap (studi kasus: indikator pendidikan di Indonesia tahun 2021). *Jurnal Eksponensial*, 16(1). <https://doi.org/10.30872/eksponensial.v16i1.1392>
- [13] Muhajir, M. (2021). *Modul praktikum statistika multivariat terapan*. Universitas Islam Indonesia. [14] Raditya, J. (2021). *Indeks Risiko Bencana Indonesia Tahun 2021*. Badan Nasional Penanggulangan Bencana.
- [14] Santoso, S. (2006). *Menggunakan SPSS untuk statistik multivariat*. Elex Media Komputindo.
- [15] Arora, P., Deepali, & Varshney, S. (2016). Analysis of K-means and K-medoids algorithm for big data. *Procedia Computer Science*, 78, 507–512. <https://doi.org/10.1016/j.procs.2016.02.095>
- [16] Maharani, S., & Yotenka, R. (2024). Pengelompokan kecamatan di Daerah Istimewa Yogyakarta berdasarkan jumlah pengguna alat kontrasepsi tahun 2022 dengan K-medoids kluster. *Emerging Statistics and Data Science Journal*, 2(2)
- [17] Rohman, D., Annisa, R., Efendi, D. I., & Solahudin, D. (2024). Klustering bencana alam menggunakan K-means pada wilayah Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1).
- [18] Azzahrah, F., Annas, S., & Rais, Z. (2022). Hybrid hierarchical clustering dalam pengelompokan daerah rawan bencana tanah longsor di Sulawesi Selatan. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 4(3), 153–161. <https://doi.org/10.35580/variansiunm38>.