IT-EXPLORE
Jurnal Penerapan Teknologi Informasi dan Komunikasi
ISSN 2828-7940, e-ISSN 2829-1727, DOI: 10.24246/itexplore.v4i1.2025.pp116-127
Volume 04 Nomor 01 Tahun 2025
Journal homepage: https://ejournal.uksw.edu/itexplore

Clustering zonasi daerah rawan bencana alam Provinsi Jawa Tengah menggunakan algoritma k-means dan *library* geopandas

Muhammad Faiq Adhitya Faqih[1], **Evangs Mailoa**[2]
Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana Salatiga

Informasi Artikel

Riwayat Artikel:

Dikirim:
14 Januari 2025
Direvisi:
28 Februari 2025
Diterima:
25 Februari 2025

Kata Kunci:

Banjir,
Tanah Longsor,
Jawa Tengah,
K-Means,
Mitigasi Bencana

ABSTRAK

Berdasarkan kajian risiko bencana Jawa Tengah tahun 2016–2020, banjir dan tanah longsor merupakan bencana paling sering terjadi, dengan 818 kasus banjir yang mencakup 31,33% dari total bencana dan tanah longsor sebesar 29,57%. Penelitian ini menggunakan algoritma K-Means dan library GeoPandas untuk mengelompokkan daerah rawan bencana di Jawa Tengah. Data kejadian banjir dan tanah longsor periode 2019–2023 diperoleh dari Badan Nasional Penanggulangan Bencana dan dianalisis melalui beberapa tahapan, yaitu pengumpulan data, standarisasi data menggunakan Standard Scaler, pengelompokan menggunakan metode K-Means, serta visualisasi hasil menggunakan GeoPandas. Hasil analisis menunjukkan bahwa Jawa Tengah terbagi menjadi 4 cluster: cluster 0 untuk daerah rawan bencana (3 wilayah), cluster 1 untuk daerah tidak rawan bencana (22 wilayah), cluster 2 untuk daerah rawan banjir (7 wilayah), dan cluster 3 untuk daerah rawan tanah longsor (3 wilayah). Hasil ini memberikan gambaran yang lebih akurat terkait sebaran daerah rawan bencana dan dapat digunakan sebagai dasar dalam pengambilan keputusan mitigasi bencana yang lebih tepat.

Keywords:

Flood,
Landslide,
Central Java,
K-Means,
Disaster Mitigation

ABSTRACT

Based on the 2016-2020 Central Java Disaster Risk Assessment, floods and landslides are the most frequent disasters, with 818 flood cases accounting for 31.33% of the total disasters and landslides accounting for 29.57%. This study aims to cluster disaster-prone areas in Central Java using the K-Means algorithm and the GeoPandas library. Data on disaster events for the period 2019-2023 was obtained from the National Disaster Management Agency, while administrative map data of Central Java was downloaded from the Geoportal of Central Java Province. The research stages include data collection, data cleaning, standardization using the Standard Scaler method, application of the K-Means algorithm for regional clustering, and visualization of results using GeoPandas. The results showed that Central Java was divided into four clusters, namely: cluster 0 (disaster-prone areas) includes 3 regions, cluster 1 (non-disaster-prone areas) has 22 regions, cluster 2 (flood-prone areas) consists of 7 regions, and cluster 3 (landslide-prone areas) has 3 regions. The results of this research provide spatial data-based information that can be used as a basis in decision-making for disaster mitigation in Central Java.

This is an open access article under the CC BY-SA license

Corresponding Author:

Muhammad Faiq Adhitya Faqih
Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
Jl. Dr. O. Notohamidjodjo No. 1-10, Blotongan, Salatiga, Jawa Tengah 50715
Email : 672020017@student.uksw.edu

1. PENDAHULUAN

Keragaman topografi di Jawa Tengah, yang terdiri dari dataran tinggi, pegunungan, dan sungai, menyuguhkan keindahan alam yang menakjubkan[1]. Namun, di balik keindahan tersebut, terdapat bahaya bencana yang mengancam[2]. Tanah longsor dan banjir merupakan contoh bencana yang mungkin terjadi[3],[4],[5]. Bencana-bencana ini bisa disebabkan oleh faktor alam maupun oleh tindakan manusia yang tidak bertanggung jawab[6].

Tanah longsor merupakan salah satu bencana yang sering terjadi di Jawa Tengah, yang dapat terlihat dari banyaknya dataran tinggi yang ada di daerah tersebut [7]. Beberapa penyebab terjadinya tanah longsor meliputi munculnya retakan di lereng tebing, intensitas hujan yang tinggi, dan rapuhnya struktur lereng[8]. Selain faktor alam, faktor lain yang mempengaruhi tanah longsor adalah aktivitas manusia seperti penebangan liar dan penambangan ilegal. Selain tanah longsor, banjir juga sering terjadi di Jawa Tengah. Banjir disebabkan oleh endapan sedimen dalam sungai, kurangnya daerah resapan air hujan, dan tingginya intensitas hujan[8].

Berdasarkan data yang terdapat di *Geoportal* Data Bencana Indonesia, tanah longsor dan banjir merupakan bencana yang sering terjadi di Jawa Tengah [9]. Selain itu, menurut Kajian Risiko Bencana (KRB) Jawa Tengah tahun 2016–2020, tanah longsor dan banjir menempati peringkat pertama dan kedua sebagai bencana yang paling sering terjadi di wilayah tersebut [3]. Banjir dilaporkan memiliki persentase 31,33% (818 kasus) dari total bencana yang terjadi di Jawa Tengah, sementara tanah longsor memiliki persentase 29,57% (772 kasus) [3] [10].

Penelitian ini menekankan pentingnya mitigasi bencana dan manajemen bencana alam, terutama tanah longsor dan banjir di Jawa Tengah menggunakan algoritma K-means untuk masalah publik di mana pengelompokan dilakukan pada variabel yang telah ditentukan seperti frekuensi, topografi, dan kondisi lingkungan. Hasil pengelompokan akan berguna bagi pemerintah dalam menetapkan prioritas untuk daerah-daerah yang memerlukan perhatian lebih dan sumber daya yang dialokasikan untuk mitigasi dan pencegahan bencana semacam itu[11]. Selain itu, pemahaman geospasial jauh lebih baik dibandingkan dengan menggunakan tabel, sehingga mengelompokkannya menggunakan Geopandas ke dalam peta membuat pemahaman hasil menjadi lebih mudah untuk dipahami. Peta memberikan pemahaman yang lebih jelas dan hubungan antara area geografis dan tingkat risiko bencana seperti bencana, memungkinkan pemerintah dan publik untuk lebih proaktif dan responsif dalam keadaan darurat[12].

Dari latar belakang yang sudah dijelaskan sebelumnya, penelitian ini bertujuan untuk membantu pemerintah Provinsi Jawa Tengah dalam memetakan wilayah rawan banjir dan tanah longsor menggunakan algoritma clustering K-means dan memvisualisasikan hasilnya dalam bentuk peta menggunakan Geopandas. Dengan data yang lebih akurat mengenai kejadian bencana, pemerintah dapat lebih efektif dalam menentukan skala prioritas untuk pencegahan dan penanggulangan bencana di daerah-daerah yang terdampak. Penelitian ini dapat memberikan panduan yang lebih jelas bagi pengambilan keputusan strategis dalam menghadapi tantangan bencana alam yang sering terjadi di wilayah ini..

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian mengenai zonasi daerah rawan bencana di Provinsi Sumatera Barat telah dilakukan dengan memanfaatkan data gempa bumi dan banjir dari tahun 2018-2021 di berbagai kota dan kabupaten di wilayah tersebut. Algoritma K-Means digunakan dalam proses *clustering*, sementara *library* geopandas diterapkan untuk memvisualisasikan data. Hasil dari clustering ini membantu pemerintah dalam menentukan wilayah yang rentan terhadap bencana alam dan merumuskan langkah preventif guna mengurangi risiko yang mungkin terjadi[6]. Perbedaan dengan penelitian ini terletak pada wilayah kajian dan jenis bencana yang dianalisis. Penelitian ini fokus pada Provinsi Jawa Tengah dengan dua bencana utama, yaitu banjir dan tanah longsor, yang menjadi perhatian utama dalam mitigasi risiko.

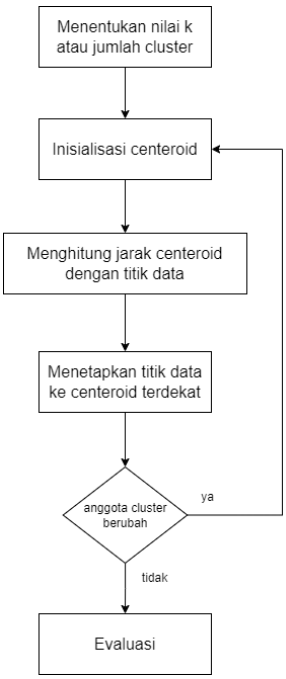
Pada penelitian lain yang menggunakan data bencana alam berdasarkan wilayah, waktu, jumlah korban, dan kerusakan fasilitas, algoritma K-Means diterapkan untuk melakukan pengelompokan. Hasil penelitian

ini menunjukkan adanya tiga cluster berdasarkan kriteria tersebut. Setiap provinsi dikelompokkan ke dalam cluster tertentu, kecuali dalam pengelompokan jumlah kejadian bencana yang dilakukan berdasarkan bulan. Pendekatan ini membantu dalam memahami pola dan distribusi bencana alam yang lebih terperinci berdasarkan faktor-faktor yang telah ditentukan[13]. Penelitian terdahulu telah mengelompokkan wilayah berdasarkan tingkat dampak bencana alam. Penelitian ini difokuskan pada identifikasi wilayah rawan bencana alam serta visualisasi hasil klasterisasi untuk mendukung upaya mitigasi bencana.

Terdapat penelitian *clustering* dan visualisasi bentuk kota di jawa timur. Aspek-aspek yang ditinjau dalam penelitian adalah compacness, sprawl, dan entropy. Untuk mengidentifikasi compacness menggunakan entropy index. *Nearest Neighbour* dan Z Score digunakan untuk menidentifikasi *compacness*. *Rose Diagram* dan QGIS digunakan untuk mengidentifikasi orientation. Ketiga hal tersebut digunakan untuk melakukan *clustering* bentuk kota menggunakan Geoda. t-SNE yang digunakan untuk clusturing menghasilkan 4 cluster dengan nilai korelasi 0,745[14]. Penelitian terdahulu lebih menitikberatkan pada karakteristik morfologi daerah kajian dan bukan pada analisis risiko bencana. Berbeda dengan penelitian ini yang fokus pada identifikasi daerah rawan bencana alam menggunakan data kejadian nyata dan analisis spasial untuk mendukung mitigasi bencana.

2.2. Landasan Teori

K-means merupakan salah satu algoritma dalam machine learning yang digunakan untuk mengelompokkan data menjadi beberapa cluster atau kelompok berdasarkan karakteristik yang serupa [11]. Pengelompokan k-means dilakukan berdasarkan centroid sampel yang disebarkan secara acak [15]. Data banjir dan tanah longsor yang berada dekat dengan centroid tersebut akan dikelompokkan ke dalam *cluster* yang bersesuaian. Dibandingkan dengan algoritma lainnya, k-means mudah diterapkan, mudah diadaptasi, dan sering digunakan dalam data mining [11],[16]. Berikut ini adalah tahapan-tahapan penerapan algoritma k-means hingga didapatkan *cluster* yang optimal.



Gambar 1. Alur Pengelomokan Data

Langkah pertama dalam pengolahan data menggunakan k-means adalah menentukan nilai k atau jumlah cluster yang akan digunakan dalam pengelompokan data banjir dan tanah longsor di Jawa Tengah. Penentuan jumlah cluster dapat dilakukan menggunakan *elbow method* [17]. *Elbow method* melakukan percobaan pada setiap nilai k dari rentang yang ditentukan. Hasil dari *elbow method* berupa grafik, di mana titik varians mulai menurun secara signifikan dan membentuk sudut menyerupai siku (*elbow*) menunjukkan cluster optimal dalam data banjir dan tanah longsor di Jawa Tengah. Perhitungan *elbow method*

menggunakan rumus sebagai berikut [18]:

$$SSE = \sum_{k=1}^{k} \sum_{x_i} |x_i - x_k|^2 \quad (1)$$

Keterangan :
K = cluster ke-c
X_i = jarak data objek ke-i
Ck = pusat cluster ke-i

Data diubah menjadi standar menggunakan *Standard Scaler* untuk menyamakan skala data jumlah banjir dan tanah longsor yang bervariasi antar wilayah, sehingga meningkatkan akurasi dan efektivitas dalam pengelompokan dan perhitungan data[19]. Proses standarisasi ini melibatkan dua perhitungan utama, yaitu median dan standar deviasi [20],[21]. Nilai standar setiap data dihitung berdasarkan rumus berikut:

$$\mu = \frac{1}{N} \sum_{i=1}^{n} X_i \quad (2)$$

Keterangan :
μ = median
X_i = objek data ke i
N = jumlah data

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2} \quad (3)$$

Keterangan :
σ = standar deviasi
X_i= objek data ke i
N = jumlah data

$$Z = \frac{X_i - \mu}{\sigma} \quad (4)$$

Keterangan :
Z = nilai standar
X_i = objek data ke i
μ = median
σ = standar deviasi

Centroid iterasi pertama diinisialisasi secara acak. Rumus untuk menentukan centroid pada iterasi kedua dan seterusnya ditampilkan sebagai berikut:

$$v = \frac{\sum_{i=1}^{n} xi}{n};\ i - 1,2,3, \dots, n \quad (5)$$

Keterangan :
V = titik pusat kelompok
xi = objek data ke i dalam cluster
n = jumlah objek yang menjadi anggota cluster

Jarak antara setiap titik data dengan centroid yang telah diinisiasi dihitung menggunakan Euclidean Distance. Perhitungan jarak titik data dengan centeroid seperti berikut ini:

$$D(x_i, y_i) = \sqrt{(x_i - x_c)^2 - (y_i - y_c)^2} \quad (6)$$

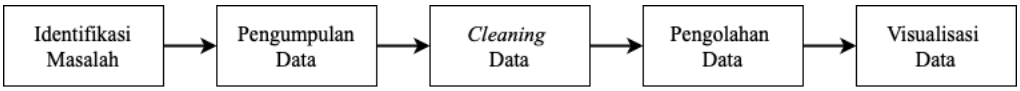
Keterangan :
D = jarak ecludian
P = jumlah variabel cluster
x_i:= data x ke-i
y_i: = data y ke-i
x_c = nilai x
y_c = nilai y
Perhitungan jarak antara titik data dengan centroid telah dilakukan, dan titik data dimasukkan ke centroid

terdekat berdasarkan perhitungan Euclidean Distance. Proses inisialisasi diulangi hingga pengelompokan titik data selesai. Hasil pengelompokan iterasi pertama dibandingkan dengan hasil pengelompokan iterasi kedua. Jika ada perbedaan anggota dalam cluster, proses perulangan akan terus dilakukan hingga perbandingan cluster terbaru dengan cluster sebelumnya tidak menunjukkan perubahan anggota [22].

GeoPandas merupakan library Python yang mengintegrasikan fungsionalitas Pandas untuk memanipulasi dan menganalisis data geospasial. [23]. Library ini mendukung berbagai format file geospasial, seperti shapefile dan GeoJSON, serta memungkinkan visualisasi data menggunakan Matplotlib atau library serupa. GeoPandas menyediakan beragam operasi geometri, termasuk penggabungan, pemotongan, dan pembuatan buffer, serta analisis spasial seperti perhitungan jarak dan area.[24]. Dalam penelitian ini Geopandas digunakan untuk memetakan hasil clustering banjir dan tanah longsor ke dalam peta Jawa Tengah.

3. METODE PENELITIAN

Metode penilitian yang digunakan dalam penelitian ini digambarkan pada Gambar 2, yang menunjukkan alur penelitian secara keseluruhan. Penelitian ini melibatkan lima tahapan, yaitu identifikasi masalah, pengumpulan data, *cleaning* data, pengolahan data dana visualisasi data.



Gambar 2 : Alur Penelitian

Tahap pertama, yaitu identifikasi masalah, dilakukan untuk menentukan jenis bencana yang paling sering terjadi di Provinsi Jawa Tengah. Berdasarkan Kajian Risiko Bencana (KRB) Jawa Tengah tahun 2016–2020, banjir dan tanah longsor diidentifikasi sebagai bencana yang paling sering terjadi di wilayah tersebut [3]. Dua jenis data dikumpulkan dalam penelitian ini, yaitu data peta Provinsi Jawa Tengah yang diunduh dari Geoportal Provinsi Jawa Tengah ((http://geoportal.jatengprov.go.id/) dan data kejadian banjir serta tanah longsor periode 2019–2023 yang diperoleh dari portal resmi Badan Nasional Penanggulangan Bencana (https://gis.bnpb.go.id/).

Data yang diperoleh dari Badan Nasional Penanggulangan Bencana dirangkum untuk menyederhanakan informasi yang awalnya mencatat setiap kejadian banjir dan tanah longsor secara rinci. Perangkuman data dilakukan ke dalam tiga kolom, yaitu nama kota/kabupaten, jumlah kejadian tanah longsor, dan jumlah kejadian banjir. Data ini tidak memerlukan proses pembersihan lebih lanjut karena telah disesuaikan selama proses perangkuman. Untuk data peta dari Geoportal Provinsi Jawa Tengah, kolom yang tidak relevan dihapus, sehingga hanya menyisakan kolom nama kota dan geometry. Selanjutnya, kolom ID ditambahkan, dan penomoran ID kota/kabupaten disesuaikan dengan ID pada data jumlah kejadian banjir dan tanah longsor.

Pada tahap pengolahan data sebelum pengelompokan menggunakan algoritma K-Means, jumlah cluster ditentukan dengan menggunakan metode elbow. Data jumlah kejadian banjir dan tanah longsor digabungkan dengan data peta Provinsi Jawa Tengah, kemudian dilakukan proses standarisasi menggunakan Standard Scaler. Setelah itu, centroid diinisialisasi secara acak dan jarak antara titik data kota/kabupaten dengan setiap centroid dihitung. Setiap titik data dimasukkan ke dalam cluster dengan jarak terdekat ke centroid. Setelah pengelompokan awal, posisi centroid diperbarui berdasarkan rata-rata posisi titik data dalam masing-masing cluster. Proses ini diulang dengan menghitung kembali jarak Euclidean antara titik data dan centroid baru. Iterasi dilakukan hingga tidak ada perubahan anggota cluster antara dua iterasi berturut-turut. Setelah proses pengelompokan selesai, kolom baru ditambahkan untuk menunjukkan cluster masing-masing kota/kabupaten.

Hasil pengelompokan divisualisasikan menggunakan GeoPandas. Setiap kota/kabupaten dalam cluster yang sama diberi warna yang seragam, sedangkan warna yang berbeda digunakan untuk membedakan antarcluster. Visualisasi bersifat interaktif, sehingga saat salah satu wilayah diklik, informasi seperti nama kota/kabupaten, jumlah kejadian banjir, jumlah kejadian tanah longsor, dan nomor cluster akan ditampilkan.

4. HASIL DAN PEMBAHASAN

Data untuk penelitian ini diambil dari website Geoportal Data Bencana Indonesia dari tahun 2019 hingga 2023 (https://gis.bnpb.go.id/). Data yang diperoleh dari website tersebut telah melalui proses pembersihan dan penambahan ID agar dapat digabungkan dengan file Excel, sehingga dihasilkan data seperti yang ditunjukkan pada Tabel 1.

Tabel 1 Kategori Hujan Berdasarkan Intensitasnya

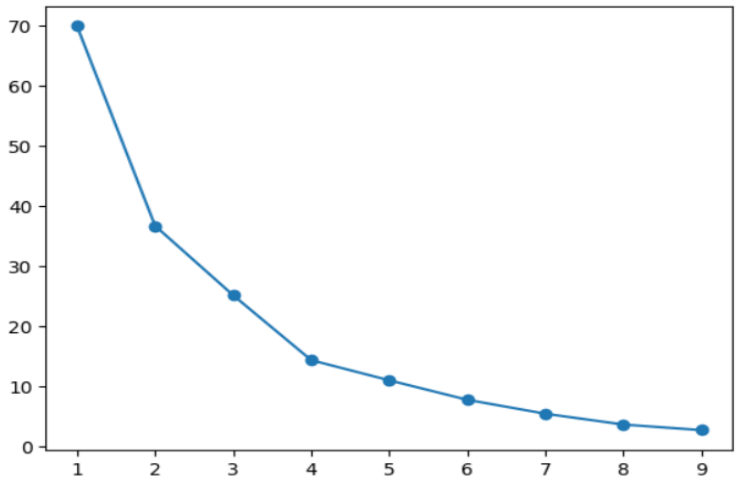
No	Daerah	Tanah Longsor	Banjir
1	Kabupaten Banjarnegara	79	4
2	Kabupaten Banyumas	130	42
3	Kabupaten Batang	13	14
4	Kabupaten Blora	10	30
…	…	…	…
34	Kota Surakarta	4	14
35	Kota Tegal	0	16

Peta Jawa Tengah diperoleh dari Geoportal Provinsi Jawa Tengah (http://geoportal.jatengprov.go.id/). Peta yang diunduh berbentuk shapefile (SHP) dan perlu diubah menjadi GeoJSON agar memudahkan penggabungan dengan file data bencana. Data yang terdapat dalam file peta ditunjukkan pada Tabel 2.

Tabel 1 Data GeoJSON Jawa Tengah

fid_batas_	a2name	a1name	dakk	dak_nkab	…
192	Banjarnegara	JAWA TENGAH	3304	Banjarnegara	…
190	Banyumas	JAWA TENGAH	3302	Banyumas	…
213	Batang	JAWA TENGAH	3325	Batang	…
204	Blora	JAWA TENGAH	3316	Blora	…
…	…	…	…	…	…
223	Tegal	JAWA TENGAH	3376	Kota Tegal	…

Data dalam file GeoJSON Jawa Tengah hanya memuat nama kota, sementara data lain dihapus. Satu kolom ID ditambahkan, dan nomor ID disesuaikan dengan data Excel. Penyamaan ID dilakukan agar penggabungan file dapat dilakukan.



Gambar 3 : Hasil Elbow Method

Jumlah cluster optimal ditentukan menggunakan *elbow method*. Berdasarkan grafik di atas, jumlah cluster yang paling optimal adalah 4. Hal ini disebabkan oleh grafik yang mulai melandai pada cluster ke-4, menandakan bahwa penambahan jumlah cluster setelahnya tidak memberikan dampak signifikan terhadap pengelompokan.

Kode program 1 Kode Program untuk Visualisasi Data Clustering

Clustering Data

```
cluster=KMeans(n_clusters=4).fit_predict(scaler.fit_transform(jateng2[["TANAH
LONGSOR", "BANJIR"]]))
jateng2["Cluster"]=cluster
```

Data dikelompokkan ke dalam empat klaster menggunakan kode pada baris pertama. Pada kode tersebut juga diterapkan proses standarisasi nilai untuk data tanah longsor dan banjir guna meningkatkan akurasi algoritma k-means. Tahapan perhitungan untuk memperoleh nilai standar dari setiap data dijelaskan sebagai berikut:

$$\mu \text{ Banjir} = \frac{1}{35} \quad 804$$
$$= 29{,}97$$

$$\mu \text{ Tanah Longsor} = \frac{1}{35} \quad 1288$$
$$= 36{,}8$$

Perhitungan standar deviasi dari tanah longsor dan banjir

$$\sigma \text{ Banjir} = \sqrt{\frac{1}{35} \quad 10298{,}97}$$
$$= \ 17.15$$

$$\sigma \text{ Tanah longsor} = \sqrt{\frac{1}{35} \quad 50587{,}6}$$
$$= \ 38.02$$

Contoh perhitungan nilai standar data (Tanah Longsor Kota Blora)

$$Z \ \ Tanah \ Longsor \ Blora = \frac{10 - \ 29{,}97}{17.15}$$
$$= \ -1.164$$

Tabel 3 Data setelah Standar Deviasi

No	Daerah	Tanah Longsor	Banjir
1	Kabupaten Banjarnegara	2,859	-0,863
2	Kabupaten Banyumas	5,833	0,137
3	Kabupaten Batang	-0,990	-0,600
4	Kabupaten Blora	-1,164	-0,179
5	Kabupaten Boyolali	-1,164	-0,731
…	…	…	…
33	Kota Semarang	5,833	0,295
34	Kota Surakarta	-1,514	-0,600
35	Kota Tegal	-1,748	-0,547

Nilai awal centroid diinisialisasi dengan memilih secara acak dari data yang telah distandarisasi. Dalam penelitian ini, parameter tanah longsor digunakan sebagai variabel x, sedangkan parameter banjir digunakan sebagai variabel y. Sebagai contoh, empat nilai acak yang dipilih dari data untuk centroid awal adalah (2,859; -0,863), (-1,748; -0,547), (0,468; -0,626), dan (-1,398; -0,6).

Contoh perhitungan *ecludian distance* Kota Magelang dengan centeroid (2.859,-0.863)

$$D(-1{,}748, -0{,}942) = \sqrt{(-1{,}748 - 2.859)^2 - (-0{,}942 - (-0.863))^2}$$
$$= 4.606$$

Tabel 4 Hasil Perhitungan *Ecludian Distance*

Daerah	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster
Kabupaten Banjarnegara	0,000	4,596	2,379	4,249	1
Kabupaten Banyumas	2,801	7,550	5,310	7,193	1
Kabupaten Batang	3,839	0,757	1,457	0,408	4
Kabupaten Blora	3,965	0,453	1,570	0,350	4
Kabupaten Boyolali	4,021	0,554	1,629	0,193	4
…	…	…	…	…	…
Kota Surakarta	4,365	0,228	1,982	0,116	4
Kota Tegal	4,596	0,000	2,214	0,345	2

Jarak antara data dan titik yang diinisialisasi dihitung untuk menentukan kedekatan relatif setiap data terhadap titik tersebut. Data kemudian dikelompokkan ke dalam cluster berdasarkan hasil perhitungan tersebut. Pada iterasi selanjutnya, titik centroid diperbarui berdasarkan hasil perhitungan dari iterasi sebelumnya untuk digunakan pada proses clustering berikutnya. Berikut adalah contoh perhitungan centeroid (cluster 2).

$$v \text{ Tanah Longsor} = \frac{-0{,}931 + (-1{,}281) + (-1{,}631) + (-1{,}748)}{4}$$
$$= -1.398$$

$$v \text{ Banjir} = \frac{0{,}558 + 0{,}373 + (-0{,}652) + (-0.547)}{4}$$
$$= -0.067$$

Centroid baru untuk cluster 2 dihitung dengan hasil (-1,398; -0,067), di mana sumbu x merepresentasikan kejadian tanah longsor dan sumbu y merepresentasikan kejadian banjir. Perhitungan serupa diterapkan pada cluster lainnya untuk menentukan centroid baru. Dari centroid yang diperbarui, dilakukan iterasi kedua dengan menghitung jarak setiap titik data terhadap centroid menggunakan metode *Euclidean*. Data dikelompokkan cluster ke dalam cluster berdasarkan kedekatannya dengan centroid baru.

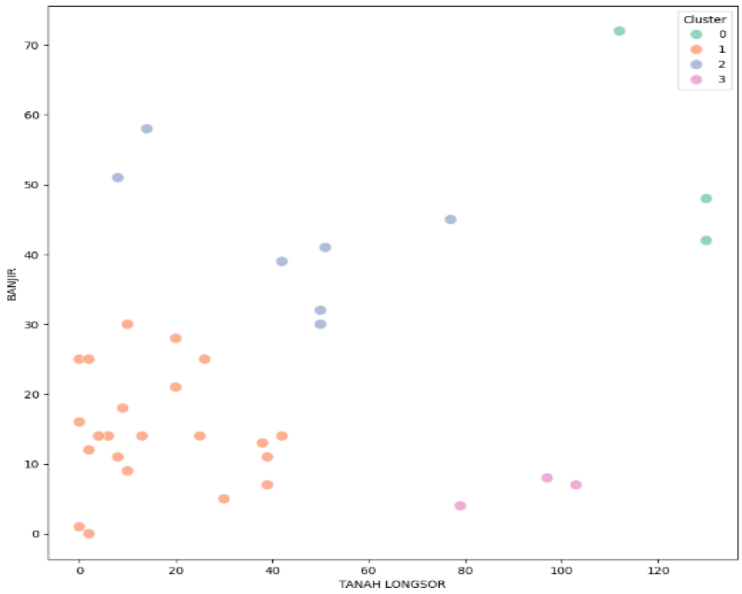
Perbandingan anggota cluste pada iterasi terbaru dengan iterasi sebelumnya digunakan untuk memastikan adanya perubahan anggota. Jika ditemukan perubahan dalam komposisi cluster, iterasi berikutnya dilakukan hingga tidak terdapat perubahan anggota antara dua iterasi terakhir. Pada baris pertama kode, penamaan cluster hanya bersifat acak yang dihasilkan oleh algoritma K-means dan tidak menentukan atribut spesifik cluster tersebut.

Kode program 2 Kode Program untuk Visualisasi Data Clustering

Visualisasi Data Clustering

```
plt.figure(figsize=[10,10])
sns.scatterplot(data = jateng2,x= "TANAH LONGSOR", y= "BANJIR", hue =
"Cluster" , palette="Set2", s = 100, alpha = 0.7)
```

Jumlah cluster yang optimal telah ditentukan, dan data tersebut telah dikelompokkan ke dalam 4 cluster. Visualisasi cluster yang terbentuk disajikan menggunakan kode program pada baris 1 dan 2.



Gambar 4 : Scatter Plot Hasil Clustering

Dari visualisasi data yang dikelompokkan menjadi 4 cluster, terlihat bahwa data pada cluster 0 ditandai oleh tingginya jumlah bencana banjir dan tanah longsor. Data pada cluster 1 memiliki karakteristik jumlah tanah longsor dan banjir yang relatif rendah. Cluster 3 dicirikan oleh tingginya jumlah bencana banjir dan rendahnya jumlah bencana tanah longsor. Sementara itu, cluster terakhir menunjukkan jumlah bencana banjir yang rendah dan jumlah bencana tanah longsor yang tinggi.

Tabel 5 Data setelah Clustering

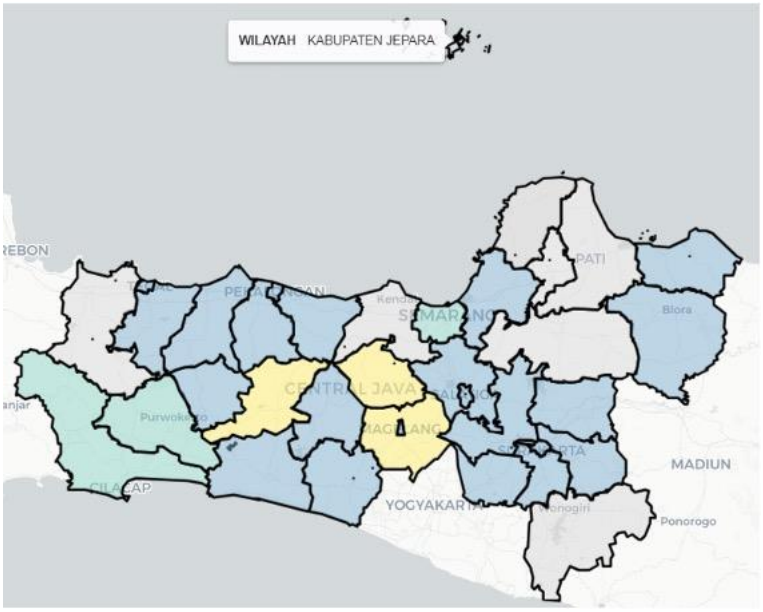
No	Daerah	Tanah Longsor	Banjir	Cluster
1	Kabupaten Banjarnegara	79	4	2
2	Kabupaten Banyumas	130	42	1
3	Kabupaten Batang	13	14	0
4	Kabupaten Blora	10	30	0
5	Kabupaten Boyolali	10	9	0
…	…	…	…	…
34	Kota Surakarta	4	14	0
35	Kota Tegal	0	16	0

Kode Program 3 Kode Program Visualisasi Clustering pada Peta

Visualisasi Clustering pada Peta

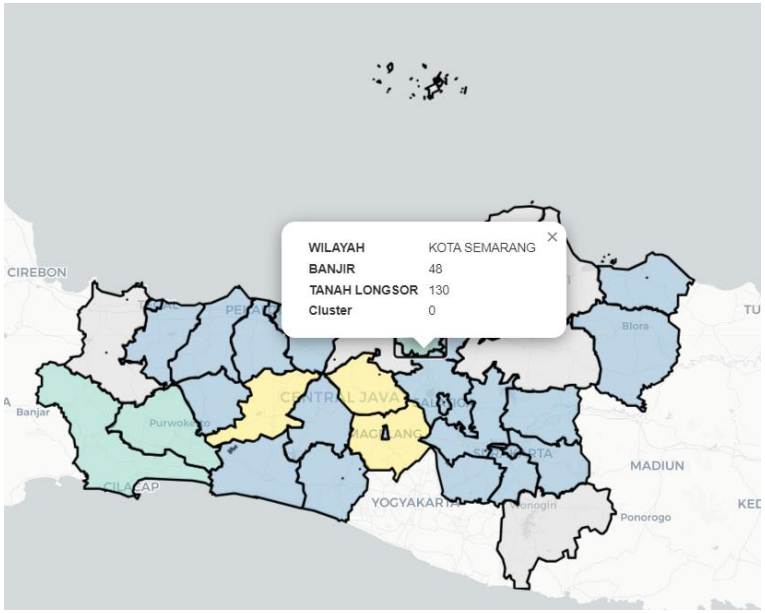
```
jateng2.explore(
    column="Cluster",
    tooltip="WILAYAH",
    popup=True,
    tiles="CartoDB positron",
    cmap="Set3",
    style_kwds=dict(color="black"),
)
```

Hasil clustering divisualisasikan dalam bentuk peta menggunakan GeoPandas melalui kode program 2. Pada kode program tersebut, kolom 'Cluster' digunakan untuk menentukan pewarnaan peta. Tooltip berfungsi untuk menampilkan informasi saat kursor meng-hover peta, seperti yang terlihat pada gambar 4. Tiles digunakan untuk menentukan tipe peta yang ditampilkan. Hasil visualisasi pengelompokan ditunjukkan pada Gambar 5. Saat kursor diarahkan ke suatu wilayah, akan ditampilkan nama kota atau kabupaten dari wilayah tersebut. Melalui visualisasi ini, distribusi klaster dapat terlihat dengan lebih jelas.



Gambar 5 : Visualisasi Clustering

Gambar 6 menunjukkan fitur interaktif dari peta hasil clustering. Saat pengguna mengarahkan kursor atau mengklik sebuah wilayah, informasi detail mengenai wilayah tersebut akan muncul, seperti nama daerah, jumlah kasus banjir dan tanah longsor, serta kategori cluster. Fitur ini memungkinkan analisis yang lebih mendalam tentang karakteristik masing-masing wilayah dan memudahkan dalam pengambilan keputusan terkait kebijakan penanggulangan bencana.



Gambar 6 : Visualisasi Keterangan Clustering

Penelitian ini memiliki beberapa keunggulan dibandingkan dengan artikel ilmiah lainnya, meskipun sama-sama menggunakan algoritma K-Means untuk clustering wilayah rawan bencana. Dalam artikel dari Analisis Bencana Tanah Longsor di Wilayah Jawa Tengah, analisis lebih berfokus pada identifikasi risiko bencana tanpa menerapkan pendekatan berbasis data spasial untuk mendukung hasil clustering[4]. Sementara itu, artikel Pemetaan Potensi Bencana di Jawa Tengah Menggunakan Google Maps API dan KML dengan Metode K-Means juga menggunakan K-Means, tetapi pendekatan visualisasi hasil kurang

menonjol. Pemetaan pada penelitian tersebut memvisualisasikan daerah rawan bencana yang paling banyak terjadi[25]. Dalam penelitian ini, pustaka GeoPandas dimanfaatkan untuk memvisualisasikan hasil clustering secara lebih jelas, sehingga pola distribusi risiko bencana di setiap wilayah dapat diinterpretasikan dengan lebih mudah. Dengan demikian, penelitian ini tidak hanya menghasilkan pembagian cluster yang lebih terstruktur, tetapi juga menyediakan representasi visual yang lebih intuitif bagi pengambil kebijakan dalam merancang strategi mitigasi bencana.

5. SIMPULAN DAN SARAN

Dari hasil penelitian, diidentifikasi 4 cluster daerah rawan bencana di Jawa Tengah. Cluster 0 ditandai dengan tingginya jumlah bencana banjir dan tanah longsor, sehingga wilayah dalam cluster ini memerlukan penanganan bencana yang lebih intensif. Cluster 1 dicirikan oleh jumlah tanah longsor dan banjir yang relatif rendah, menjadikannya wilayah yang aman atau tidak rawan bencana. Cluster 2 menunjukkan rendahnya jumlah bencana banjir namun tingginya jumlah bencana tanah longsor. Sementara itu, cluster 3 memiliki jumlah bencana banjir yang tinggi dan jumlah bencana tanah longsor yang rendah. Kabupaten Banyumas, Kabupaten Cilacap, dan Kota Semarang tercatat sebagai wilayah yang paling sering mengalami banjir dan tanah longsor pada periode 2019–2023. Cluster 1 terdiri dari 22 wilayah, cluster 2 memiliki 7 wilayah, dan cluster 3 mencakup 3 wilayah. Berdasarkan hasil tersebut, disarankan agar penelitian selanjutnya mempertimbangkan variabel tambahan seperti curah hujan, jenis tanah, dan perubahan penggunaan lahan untuk analisis yang lebih komprehensif.

DAFTAR PUSTAKA

[1] R. Wijianto and Y. Nurdeviyanti, “Aplikasi Dewisita Pesona Wisata Di Jawa Tengah Berbasis Android,” *J. Informatics Comput. Eng. J.*, vol. 3, no. 2, pp. 125–132, 2023.

[2] S. Sudaryanto, “Pengurangan Resiko Bencana Alam Melalui Pengembangan Kompetensi Aparatur Sipil Negara Badan Penangulangan Bencana Daerah di Jawa Tengah,” *J. Public Policy Appl.*, vol. 3, no. 1, pp. 60–82, 2021.

[3] BNPB, “Kajian Risiko Bencana Jawa Tengah 2016 - 2020,” *Badan Nas. Penanggulangan Bencana*, p. 68, 2020.

[4] R. Isnaini, “Analisis Bencana Tanah Longsor di Wilayah Provinsi Jawa Tengah,” *Islam. Manag. Empower. J.*, vol. 1, no. 2, pp. 143–160, 2019, doi: 10.18326/imej.v1i2.143-160.

[5] D. Prabowo and R. Ayu Permatasari Rahman, “Identifikasi Potensi Banjir Di Kawasan Sub Urban Kota Semarang (Studi Kasus Kab Kendal),” *J. Planol.*, vol. 20, no. 1, pp. 1829–9172, 2023.

[6] D. A. Wicaksono and Y. A. Susetyo, “Clustering Zonasi Daerah Rawan Bencana Alam Di Provinsi Sumatera Barat Menggunakan Algoritma K-Means Dan Library Geopandas,” *J. Indones. Manaj. Inform. dan Komun.*, vol. 4, no. 2, pp. 426–438, 2023, doi: 10.35870/jimik.v4i2.225.

[7] S. U. Darmatazia, D. P. Setianingrum, and M. P. Masyudi, “Analisis Penyebab Tanah Longsor di Gunung Manggah, Jalan Otto Iskandardinata, Kelurahan Sungai Dama, Kecamatan Samarinda Ilir, Kota Samarinda,” *Dep. Ris. Dan Penelit.*, vol. 2, no. 1, pp. 1–9, 2022.

[8] F. A. Kurnianto, V. Elfiani, and A. F. Alfani, “Analisis Spasial Kerentanan Banjir dan Longsor di Kabupaten Banyuwangi,” *JPIG (Jurnal Pendidik. dan Ilmu Geogr.*, vol. 6, no. 1, pp. 49–60, 2021, doi: 10.21067/jpig.v6i1.5323.

[9] B. N. P. Bencana, “Geoportal Data Bencana Indonesia,” *Badan Nasional Penanggulangan Bencana*, 2023. https://gis.bnpb.go.id/ (accessed Nov. 26, 2023).

[10] O. sari Pinda rini and R. Khoirudin, “Analisis Sektor-Sektor Ekonomi Dengan Potensi Unggulan Di Kabupaten/Kota Se-Provinsi Jawa Tengah Tahun 2011-2016,” *J. Ilm. Ekon. Dan Bisnis*, vol. 17, no. 1, pp. 15–23, 2020, doi: 10.31849/jieb.v17i1.2790.

[11] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, “Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food,” *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574–583, 2022, [Online]. Available: https://doi.org/10.5281/zenodo.7052276.

[12] N. L. Erawati, I. K. N. Adi Jaya, and I. K. A. Asmarajaya, "Visualisasi Data Pemetaan Nasabah Kredit Pada LPD Desa Adat Kesiman Menggunakan Python," *RESI J. Ris. Sist. Inf.*, vol. 2, no. 1, pp. 129–138, 2023, doi: 10.32795/resi.v2i1.4266.

[13] M. Murdiaty, A. Angela, and C. Sylvia, "Pengelompokkan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means," *J. Media Inform. Budidarma*, vol. 4, no. 3, p. 744, 2020, doi: 10.30865/mib.v4i3.2213.

[14] F. Afrianto and A. D. Hariyanto, "Clustering Dan Visualisasi Bentuk Kota Di Jawa Timur," *J. Plano Buana*, vol. 2, no. 2, pp. 57–67, 2022, doi: 10.36456/jpb.v2i2.5303.

[15] E. Umargono, J. E. Suseno, and V. G. S. K., "K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median," vol. 474, no. Isstec 2019, pp. 234–240, 2020, doi: 10.5220/0009908402340240.

[16] H. Y. Heraldi, N. C. Aprilia, and H. Pratiwi, "Analisis Cluster Intensitas Kebencanaan di Indonesia Menggunakan Metode K-Means," *Indones. J. Appl. Stat.*, vol. 2, no. 2, p. 137, 2019, doi: 10.13057/ijas.v2i2.34911.

[17] V. A. Ekasetya and A. Jananto, "Klusterisasi Optimal Dengan Elbow Method Untuk Pengelompokan Data Kecelakaan Lalu Lintas Di Kota Semarang," *J. Din. Inform.*, vol. 12, no. 1, pp. 20–28, 2020, doi: 10.35315/informatika.v12i1.8159.

[18] H. Humaira and R. Rasyidah, "Determining The Appropiate Cluster Number Using Elbow Method for K-Means Algorithm," no. January 2020, 2020, doi: 10.4108/eai.24-1-2018.2292388.

[19] D. Cournapeau, "StandardScaler," *scikit-learn developers*, 2010. https://scikit-learn.org/dev/modules/generated/sklearn.preprocessing.StandardScaler.html (accessed Nov. 15, 2024).

[20] H. Hernandez, "Probability Distribution and Bias of the Sample Standard Deviation," *ForsChem Res. Reports*, vol. 8, pp. 1–26, 2023, doi: 10.13140/RG.2.2.22144.51205.

[21] E. Priambodo, T. Prahasto, and A. Widodo, "Deteksi Dini Kerusakan Bearing Menggunakan Machine Learning Pendekatan Support Vector Regression (Svr)," *J. Tek. Mesin S-1*, vol. 11, no. 3, pp. 101–112, 2023.

[22] E. Oktaviana, *Clustering bencana alam di Indonesia menggunakan algoritma K-Means*. Surabaya, 2022.

[23] B. Suharjo, "Application of K-Means Cluster and Spatial Statistics using Python to Analyze the Indicators of Indonesia Information Technology," *Digit. Zo. J. Teknol. Inf. dan Komun.*, vol. 12, no. 1, pp. 11–18, 2021, doi: 10.31849/digitalzone.v12i1.4310.

[24] M. R. N. Yusuf, Y. A. Pranoto, and F. X. Ariwibisono, "Rancang Bangun Sistem Informasi Geografis Pemetaan Upah Minimum Kota (Umk) Dan Biaya Kebutuhan Hidup Di Provinsi Jawa Timur Berbasis Web," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 4, no. 2, pp. 247–253, 2020, doi: 10.36040/jati.v4i2.2722.

[25] M. F. H. Adi and Y. A. Susetyo, "Pemetaan Potensi Bencana di Jawa Tengah Menggunakan Google Maps API dan KML dengan Metode," *Indones. J. Comput. Model.*, vol. 3, no. 1, pp. 43–51, 2018.