

PENGELOMPOKAN DAERAH RAWAN BENCANA DI JAWA TENGAH MENGGUNAKAN ALGORITMA *K-MEANS CLUSTERING*

Robiyah Syah Hidayat, Muhammad Rafi Muttaqin, Dede Irmayanti

Teknik Informatika, STT Wastukencana Purwakarta

Jalan Cikopak No.53, Mulyamekar, Kec. Babakancikao, Kabupaten Purwakarta, Jawa Barat 41151

robiansyah27@wastukencana.ac.id

ABSTRAK

Provinsi Jawa Tengah, yang berlokasi di tengah Pulau Jawa, menghadapi beberapa peristiwa bencana alam seperti banjir, tanah longsor, dan gempa bumi. Berdasarkan penilaian ketahanan secara menyeluruh, Provinsi Jawa Tengah memiliki Indeks Ketahanan Daerah sebesar 0,76, yang menunjukkan tingkat kapasitas daerah yang sedang. Oleh karena itu untuk mengurangi efek negatif dari bencana, otoritas daerah Provinsi Jawa Tengah harus mengoptimalkan komitmen, kebijakan, dan kegiatan penanggulangan bencana alam. Selama 10 tahun terakhir, Provinsi Jawa Tengah menghadapi ancaman serius khususnya bencana alam yang salah satunya berupa bencana tanah longsor yang paling sering terjadi, tercatat sebanyak 3.022 kasus, banjir 1.366 kasus, serta 262 kasus gempa bumi yang tercatat. Untuk menghadapi tantangan ini, penting memahami daerah rawan bencana dengan pendekatan *data mining*. Penelitian ini menggunakan algoritma *k-means clustering* untuk pengelompokan data. Untuk menentukan jumlah *cluster* optimal, digunakan *Silhouette Coefficient* sebagai metode evaluasi. Data bersumber dari Badan Pusat Statistik Provinsi Jawa Tengah dan data diolah dengan melibatkan *Google Colaboratory*. Pengujian dilakukan dengan membentuk *cluster* sebanyak 4 untuk mencari evaluasi pengujian *Silhouette Coefficient* terbesar. Diketahui bahwa *n_cluster* 2 mendapatkan nilai 0.6373590606168947, dengan hasil *cluster* 0 terdiri dari 3 kabupaten yang memiliki indeks daerah bencana tinggi dan *cluster* 1 terdiri dari 32 kabupaten/kota yang memiliki indeks daerah bencana rendah.

Kata Kunci: Jawa Tengah, Bencana, *K-Means*, *Cluster*, *Silhouette Coefficient*

1. PENDAHULUAN

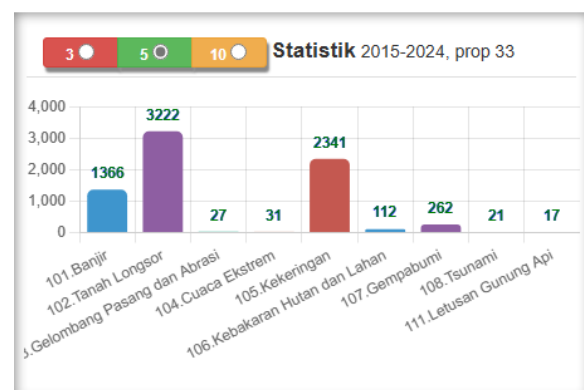
Indonesia, sebagai negara yang terletak di zona Cincin Api Pasifik, adalah salah satu wilayah yang sangat rawan terhadap bencana alam [1].

Provinsi Jawa Tengah, yang terletak di tengah Pulau Jawa, menghadapi berbagai ancaman bencana alam. Berdasarkan evaluasi keseluruhan, Provinsi Jawa Tengah memiliki Indeks Ketahanan Daerah sebesar 0,76, yang menunjukkan kapasitas sedang. Menurut Indeks Ketahanan Daerah, otoritas daerah Provinsi Jawa Tengah harus terus mengoptimalkan komitmen dan kebijakan untuk mengurangi dampak bencana alam dan memperbaiki jumlah dan kualitas upaya penanggulangan bencana untuk mengurangi dampak negatif yang ditimbulkan oleh bencana. [2].

Menurut halaman web dibi.bnpb.go.id [3] bahwa 10 tahun terakhir, Provinsi Jawa Tengah menghadapi ancaman serius dari bencana alam. Tanah longsor adalah salah satu bencana alam yang kerap terjadi dengan 3.022 kasus tercatat. Frekuensi tanah longsor yang tinggi ini mencerminkan kerentanan wilayah terhadap hujan deras dan kondisi geografis yang tidak stabil. Dampak dari tanah longsor seringkali meliputi kerusakan infrastruktur, kehilangan tempat tinggal, dan korban jiwa. Situasi ini memerlukan perhatian serius dari pemerintah dan masyarakat untuk menerapkan langkah-langkah pencegahan, seperti reboisasi, pembangunan terasering, dan sistem peringatan dini.

Selain tanah longsor, banjir juga menjadi ancaman signifikan di Jawa Tengah dengan 1.366 kasus dalam dekade terakhir. Penyebab utama banjir

biasanya adalah curah hujan tinggi, drainase yang buruk, dan perubahan penggunaan lahan yang tidak terkendali. Banjir menyebabkan gangguan besar pada kehidupan sehari-hari, termasuk kerusakan properti, gangguan transportasi, dan risiko kesehatan. Di samping itu, Provinsi Jawa Tengah juga mencatat 262 kasus gempa bumi, meskipun jumlahnya lebih sedikit dibandingkan tanah longsor dan banjir, dampaknya tetap signifikan. Gempa bumi berpotensi menyebabkan kerusakan struktural yang parah dan memicu bencana sekunder seperti tanah longsor. Kombinasi ketiga jenis bencana ini menuntut pendekatan holistik dalam manajemen bencana untuk meminimalkan kerugian dan melindungi kehidupan serta harta benda warga Jawa Tengah.



Gambar 1. Statistik bencana alam di Jawa Tengah
Sumber : dibi.bnpb.go.id

Frekuensi kejadian bencana pada gambar 1 ini membantu dalam memahami dampak relatif dari setiap jenis bencana dan memberikan dasar untuk prioritas dalam upaya mitigasi dan perencanaan. Tujuan dari penelitian ini adalah untuk mengelompokkan daerah rawan bencana menggunakan metode algoritma *K-Means Clustering*. Pengelompokan daerah berdasarkan tingkat kerawanan bencana dapat membantu pemerintah dan pemangku kepentingan yang ada di Jawa Tengah dalam mengalokasikan sumber daya secara efektif, mengarahkan upaya mitigasi, dan merencanakan respons darurat.

2. TINJAUAN PUSTAKA

2.1. Pengelompokan

Proses mengatur kumpulan objek tertentu ke dalam kelas-kelas yang sebanding dikenal sebagai pengelompokan. Sebuah *cluster* terdiri dari sekumpulan objek yang memiliki perbedaan yang jelas dan masing-masing memiliki banyak kesamaan. Salah satu fitur *data mining*, pengelompokan juga dikenal sebagai analisis *cluster*, dapat digunakan untuk mengelompokkan data dari daerah yang mungkin rawan bencana [4].

2.2. Daerah

Daerah adalah kawasan menurut batasan ruang lingkup administrasi. Sementara itu, ruang didefinisikan sebagai area baik dalam konteks geografis sebagai lokasi aktivitas manusia atau fenomena alam, maupun dalam konteks ekonomi yang diwakili oleh interaksi antara elemen-elemen ekonomi [5].

2.3. Rawan Bencana

Kondisi atau karakteristik geografis, biologis, hidrologis, klimatologis, geografis, politik, ekonomi, budaya, sosial, dan teknologi yang mengurangi kemampuan untuk mencegah, mengurangi dampak, mempersiapkan, dan merespons bahaya tertentu disebut sebagai rawan bencana. [2].

2.4. Algoritma *K-means*

Data dikelompokkan ke dalam dua kelompok atau lebih menggunakan metode partisi non-hierarkis yang dikenal sebagai *K-Means* berdasarkan kesamaan atribut. Metode ini dirancang untuk mengurangi perbedaan di dalam setiap kelompok dan meningkatkan perbedaan antar kelompok dengan cara meminimalkan fungsi objektif yang telah ditentukan. [6]. Untuk melakukan *clustering* dengan memanfaatkan algoritma *K-Means*, dapat mengikuti langkah-langkah berikut [7]:

- Tentukan berapa banyak kelompok (k) yang diinginkan.
- Inisialisasi *centroid*, atau pusat kelompok.
- Tempatkan semua data dan objek ke *cluster* yang paling dekat. Jarak dua objek menentukan seberapa dekat mereka. Teori jarak *Euclidean* dapat untuk mengukur seberapa jauh setiap data dari setiap titik

pusat *cluster*. Teori ini dirumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2} \quad (1)$$

Keterangan :

$D(i,j)$ = Jarak antara data pada titik i dan j

X_{ki} = Data ke i atribut data ke k

X_{kj} = Titik data *centroid* ke k pada atribut data ke j

- Anggota kelompok saat ini dimasukkan dalam perhitungan ulang untuk menentukan pusat kelompok. Pusat kelompok dapat dihitung dengan menggunakan persamaan berikut, yang menghitung rata-rata dari semua data atau objek yang berada dalam kelompok tersebut. Sebagai berikut, teori ini dirumuskan.

$$R_k = \frac{1}{N_k} (X_{1k} + X_{2k} + \dots + X_{nk}) \quad (2)$$

Keterangan:

R_k = Rata rata k

N_k = Jumlah data pada *cluster* k

X_{nk} = Pola pada urutan ke n yang termasuk dalam *cluster* k

- Jika pusat kelompok mengalami perubahan, ulangi langkah nomor 3.

2.5. *Clustering*

Clustering adalah teknik untuk mengklasifikasikan data berdasarkan karakteristik yang mirip. Berbeda dengan klasifikasi yang menggunakan variabel target, *clustering* tidak membutuhkan variabel tersebut dan digunakan sebagai langkah awal dalam analisis data. Ada berbagai macam algoritma untuk *clustering*, seperti *K-Means*, *Fuzzy C-Means*, *Clarans*, dan *K-Medoids* yang pada setiapnya memiliki kelebihan dan kekurangannya. Meskipun berbeda, prinsip dasar dari setiap metode *data mining* ini adalah mengelompokkan data berdasarkan jarak kesamaan dan menemukan karakteristik yang cocok dalam satu kelompok [8].

2.6. *Silhouette Coefficient*

Koefisien *silhouette* adalah ukuran yang digunakan untuk mengevaluasi seberapa baik hasil *clustering* dalam analisis data. Ukuran ini menilai kualitas setiap objek data dengan membandingkan kesesuaian objek tersebut dalam kelompoknya sendiri dengan kelompok lainnya. Secara sederhana, metrik ini menggabungkan aspek kohesi dan pemisahan untuk menilai efektivitas *clustering*. Nilai koefisien *silhouette* berkisar antara -1 hingga 1, dan *clustering* dianggap baik jika nilai mendekati 1 [7].

Untuk menentukan validasi, teknik *clustering* saat ini menggunakan *Silhouette Coefficient*. Nilai koefisien memiliki rentang nilai antara -1 dan 1. Semakin mendekati nol nilai tersebut, semakin baik jumlah *cluster* yang dipilih, karena menunjukkan bahwa pengelompokan lebih optimal. Sebaliknya, nilai yang lebih dekat dengan nol menunjukkan jumlah *cluster* yang lebih buruk [9].

Tabel 1. Kategori evaluasi nilai koefisien silhouette

Rentang	Kategori
$0.71 < s \leq 1.00$	Kuat
$0.51 < s \leq 0.70$	Baik
$0.26 < s \leq 0.50$	Lemah
≤ 0.25	Buruk

Tabel 1 menunjukkan kategori kuat/buruknya nilai koefisien.

2.7. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) adalah metode yang sering dimanfaatkan untuk mengurangi dimensi data dengan mengubah data ke dalam bentuk baru yang memiliki nilai berbeda di bidang yang baru. [10].

Teknik PCA sangat efektif pada saat data memiliki banyak variabel yang saling berkorelasi. PCA memungkinkan pengurangan jumlah variabel dari n membentuk k variabel baru yang disebut *principal component*, dengan k yang jumlahnya lebih sedikit dibandingkan n . Meskipun hanya melibatkan k *principal component*, hasilnya akan setara dengan menggunakan n variabel. Variabel yang terpilih ini dikenal sebagai *principal component*. [11].

2.8. Python

Python adalah bahasa pemrograman tingkat tinggi yang mendukung eksekusi instruksi secara langsung dan menggunakan pendekatan berorientasi objek. Python menggunakan sistem semantik dinamis yang meningkatkan keterbacaan sintaksisnya. Selain itu, Python juga mudah dipelajari karena memiliki fitur manajemen memori otomatis [12].

Python adalah bahasa pemrograman yang mengutamakan keterbacaan kode, menggabungkan berbagai kemampuan dan fitur dengan kode yang jelas serta dilengkapi dengan banyak *library* yang komprehensif. Python juga mendukung berbagai paradigma pemrograman, termasuk pemrograman fungsional, berbasis objek, dan imperatif [13].

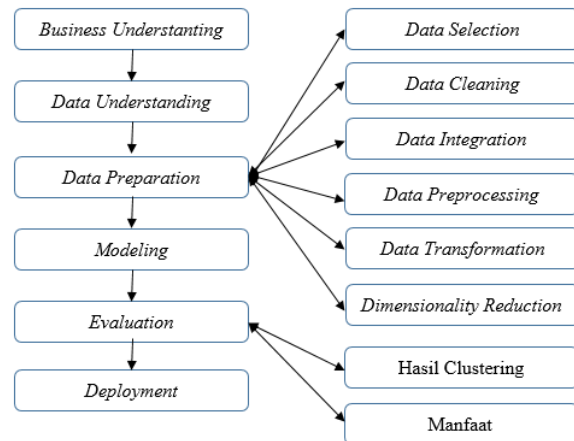
2.9. Google Colab

Google Colab, atau Google Colaboratory, adalah dokumen yang dapat dieksekusi yang memungkinkan penyimpanan, penulisan, dan berbagi kode program melalui Google Drive. Keuntungan menggunakan Google Colab termasuk kemampuan untuk berkolaborasi secara online dengan pengguna lain, akses gratis ke GPU dengan spesifikasi tinggi (GPU Tesla, RAM 12GB, DISK 300GB), dan fleksibilitas dalam menjalankan program *deep learning*. Google Colab tidak memerlukan konfigurasi tambahan karena menggunakan teknologi *cloud computing*, menyediakan akses gratis ke mesin berkecepatan tinggi (GPU), dan mudah terhubung dengan Google Drive dan GitHub. Dalam dunia *Data Science*, Google Colab dapat digunakan untuk menganalisis dan memvisualisasikan data menggunakan *library Python*. Untuk *machine learning*, Google Colab

memungkinkan berbagai kegiatan seperti mengimpor *dataset* gambar, melatih klasifikasi gambar, dan mengevaluasi model, semuanya dapat dilakukan dengan beberapa baris kode [14].

3. METODE PENELITIAN

Pada studi ini, metode CRISP-DM yang terstruktur dimanfaatkan, mencakup enam tahapan: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*



Gambar 2. Metode *crisp-dm*

Metode yang digunakan yaitu *crisp-dm* dapat dilihat pada gambar 2 diatas

3.1. Business Understanding

Pada langkah pemahaman bisnis, berbagai langkah dilakukan untuk memahami tujuan dan persyaratan dari penelitian ini. Langkah-langkah tersebut meliputi menetapkan tujuan bisnis, mengevaluasi situasi, menetapkan tujuan *data mining*, dan merancang rencana proyek.

3.2. Data Understanding

Pada langkah pemahaman data dimulai dengan pengumpulan data yang diperlukan untuk penelitian ini, kemudian mencakup deskripsi data, pemeriksaan kualitas data, dan eksplorasi data.

3.3. Data Preparation

Tahap persiapan data dimulai dengan langkah-langkah seperti pemilihan atribut yang relevan untuk pemodelan, penanganan nilai yang hilang, perubahan nama atribut, dan pembuatan atribut baru berdasarkan atribut yang ada (*feature engineering*).

3.4. Modeling

Pada langkah ini, pemodelan dilakukan dengan menerapkan teknik *clustering* menggunakan algoritma *k-means*. Sebelum memulai pemodelan, langkah pertama adalah menentukan jumlah *cluster* yang optimal.

3.5. Evaluation

Setelah proses pemodelan, langkah berikutnya adalah peninjauan kembali untuk menentukan model mana yang optimal untuk digunakan.

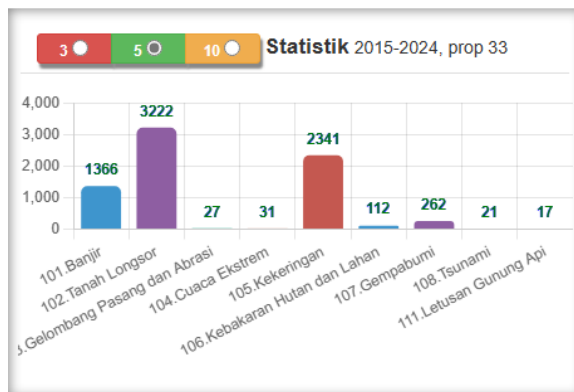
3.6. Deployment

Setelah evaluasi selesai, langkah berikutnya adalah menilai hasil model secara mendetail.

4. HASIL DAN PEMBAHASAN

4.1. Business Understanding

Penelitian ini bertujuan untuk memahami kerentanan daerah rawan bencana di Jawa Tengah. Menurut dibi.bnppb.go.id, Jawa Tengah menghadapi bahaya bencana alam dalam sepuluh tahun terakhir.



Gambar 3. Statistik bencana alam di Jawa Tengah
Sumber : dibi.bnppb.go.id

Gambar 3 diatas menunjukkan grafik bencana alam di Jawa Tengah. Dengan 3.022 kasus tercatat, Tanah longsor yaitu salah satu jenis bencana alam yang paling sering terjadi di beberapa daerah. Ini menunjukkan bahwa wilayah ini rentan terhadap hujan deras dan kondisi geografis yang tidak stabil. Dampaknya meliputi kerusakan infrastruktur, kehilangan tempat tinggal, dan korban jiwa. Oleh karena itu, langkah-langkah pencegahan seperti sistem peringatan dini dan reboisasi diperlukan dan 1.366 kasus dalam sepuluh tahun terakhir, banjir juga menjadi ancaman signifikan selain tanah longsor, Curah hujan tinggi, drainase yang buruk, dan perubahan penggunaan lahan yang tidak terkendali adalah penyebab utama banjir, yang menyebabkan kerusakan properti dan gangguan transportasi. Selain itu, Jawa Tengah memiliki 262 kasus gempa bumi, meskipun angka ini lebih sedikit.

4.2. Data Understanding

Dataset yang digunakan dalam penelitian ini meliputi data bencana alam di Provinsi Jawa Tengah dari tahun 2018 sampai 2022. Sumber data berasal dari situs jateng.bps.go.id. Dataset ini meliputi tiga jenis bencana utama: banjir, gempa bumi, dan tanah longsor, dengan total 175 data catatan yang berisi informasi mengenai nama kabupaten/kota serta jumlah kejadian banjir, gempa bumi, dan tanah longsor.

Untuk tahun 2022, dataset ini diperluas dengan tambahan jenis bencana, yaitu letusan gunung api, tsunami, banjir bandang, kekeringan, kebakaran hutan dan lahan, angin puting beliung, dan gelombang pasang, sehingga total terdapat 11 atribut dalam dataset tersebut.

4.3. Data Preparation

Tahap ini melibatkan seleksi, pembersihan, integrasi, pemrosesan awal data dan transformasi data untuk analisis lebih lanjut. Data dipilih berdasarkan fitur-fitur relevan dan dinormalisasi menggunakan StandardScaler untuk memastikan setiap fitur memiliki skala yang sama. Selanjutnya, PCA digunakan untuk mereduksi dimensi data, memfasilitasi visualisasi dan mengurangi kompleksitas.

4.3.1. Data Selection

Dataset bencana alam yang mencakup tahun 2018-2022 terdiri dari 175 data. Hanya data tahun 2022 saja yang diseleksi karna data 2018-2021 sudah sesuai kebutuhan. Data tahun 2022 ini diseleksi, dan dari total 11 atribut yang ada, hanya 5 atribut yang dipilih meliputi nama kabupaten/kota, banjir, gempa bumi, tanah longsor, dan tahun.

Tabel 2. Hasil data selection

Kabupaten / Kota	Banjir	Gempa Bumi	Tanah Longsor	Tahun
Kabupaten Cilacap	78	139	56	2018
Kabupaten Banyumas	34	67	61	2018
Kabupaten Purbalingga	9	3	29	2018
Kabupaten Banjarnegara	6	16	59	2018
Kabupaten Kebumen	92	52	83	2018
..
Kota Tegal	0	16	0	2022

Pada tabel 2 Atribut yang tidak diperlukan dalam dataset awal akan dihapus karena tidak sesuai dengan kebutuhan penelitian.

4.3.2. Data Cleaning

Pada tahap pembersihan data, salah satu tugasnya adalah menangani nilai yang hilang. Namun, dataset yang digunakan dalam penelitian ini lengkap, tanpa yang hilang atau data yang tidak konsisten. Oleh sebab itu, proses dapat dilanjutkan ke tahap selanjutnya, yaitu data integration.

4.3.3. Data Integration

Dataset yang sudah di seleksi akan di sesuaikan dengan kebutuhan yang ada yaitu dengan menambahkan atribut nomor, latitude, dan longitude untuk masing-masing wilayah. Selain itu, data insiden

bencana (banjir, gempa bumi, dan longsor) akan dijumlahkan per tahun sehingga jumlah data yang semula 175 menjadi 35. Data yang sudah di tambahkan

dan dijumlahkan yang ditunjukkan pada tabel 3 berikut.

Tabel 3. Data yang sudah disesuaikan

No	Kabupaten / Kota	Banjir	Gempa Bumi	Tanah Longsor	latitude	longitude
1	Kabupaten Cilacap	211	237	150	-7,733333	109
2	Kabupaten Banyumas	127	93	230	-7,4832133	109,140438
3	Kabupaten Purbalingga	59	9	129	-7,390747	109,3638
4	Kabupaten Banjarnegara	27	34	274	-7,402706	109,681396
5	Kabupaten Kebumen	341	152	286	-7,678682	109,656502
..
35	Kota Tegal	24	16	0	-6,8666667	109,1333333

Penyesuaian yang dilakukan pada Tabel 3 ini memungkinkan analisis yang lebih terfokus dan efisien dengan cara menyederhanakan data yang ada, sehingga memudahkan identifikasi pola dan tren yang relevan.

4.3.4. Data Preprocessing

Pada tahap *data preprocessing*, data mentah akan dipersiapkan agar siap digunakan dalam pemodelan. tiga kolom utama - Banjir, Gempa Bumi, dan Tanah Longsor - dipilih untuk pemodelan dengan algoritma *k-means* Seperti yang terlihat pada Gambar 4 berikut :

```
# Preprocessing
features = ['Banjir', 'Gempa Bumi', 'Tanah Longsor']
X = df[features]
```

Gambar 4. Source code data preprocessing

Pemilihan pada gambar 4 ini didasarkan pada pentingnya ketiga jenis bencana ini dalam menentukan tingkat kerentanan suatu wilayah.

4.3.5. Data Transformation

Tahap berikutnya adalah tahap transformasi data. Pada tahap ini, kita mengubah nilai-nilai numerik dalam *dataset* menjadi rentang 0-1. Pendekatan yang digunakan untuk transformasi data adalah *Standard Scaler*. Tujuannya adalah untuk menstandarkan fitur-fitur data sehingga memiliki rata-rata 0 dan variansi 1.

```
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

Gambar 5. Source code data transformation

Dengan melakukan seperti gambar 5, kita memastikan bahwa data tersebar secara merata dan tidak ada fitur yang terlalu menonjol dibandingkan yang lain dalam hal skala.

4.3.6. Dimensionality Reduction

PCA untuk mereduksi dimensi data agar memudahkan visualisasi dan mengurangi

kompleksitas model. Ini akan membantu dalam mengidentifikasi pola utama dalam data.

```
[24] # PCA untuk Reduksi Dimensi
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)
```

Gambar 6. Source code mereduksi dimensi oleh PCA

Tujuannya gambar 6 adalah untuk menyusun data dengan lebih sederhana dan menghapus faktor atau indikator yang kurang signifikan tanpa mengurangi inti dan maksud dari data aslinya.

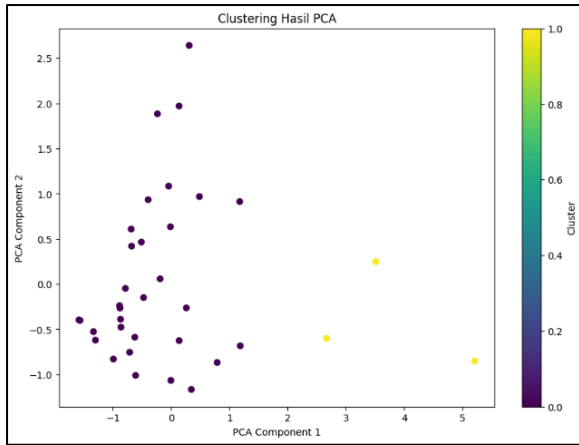
4.4. Modeling

Model dilakukan dengan visualisasi hasil *clustering* menggunakan dua komponen utama dari *PCA*. Dalam kasus ini, *PCA* digunakan untuk mengubah data yang terdiri dari tiga kolom bencana menjadi dua komponen utama. Hal ini mempermudah visualisasi dan analisis selanjutnya. Setiap titik data di *plot* dalam dua dimensi berdasarkan dua komponen *PCA* ini, yang menggambarkan variansi terbesar dalam data. Setelah reduksi dimensi dengan *PCA*, algoritma *K-Means* diterapkan pada data yang telah direduksi. Berikut proses *k-means clustering* yang ditunjukkan pada Gambar 7 dan Visualisasi pada Gambar 8 berikut :

```
# Clustering dengan K-Means
kmeans = KMeans(n_clusters=2, random_state=42)
df['Cluster'] = kmeans.fit_predict(X_pca)
# Visualisasi Hasil Clustering
plt.figure(figsize=(10, 7))
plt.scatter(X_pca[:, 0], X_pca[:, 1], c=df['Cluster'], cmap='viridis')
plt.xlabel('PCA Component 1')
plt.ylabel('PCA Component 2')
plt.title('Clustering Hasil PCA')
plt.colorbar(label='Cluster')
```

Gambar 7. Source code proses k-means dan visualisasi hasil reduksi dari PCA

Pada Gambar 7 *K-Means* mengelompokkan data ke dalam beberapa kelompok berdasarkan seberapa dekat titik data satu sama lain.

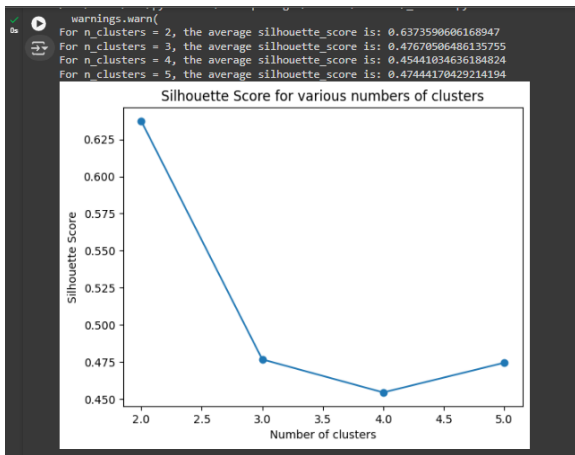


Gambar 8. Visualisasi pengelompokan cluster

Melalui visualisasi gambar 8 ini, kita dapat mengidentifikasi pola-pola yang mungkin tidak terlihat dari analisis tabel atau grafik konvensional. Visualisasi ini menunjukkan distribusi data dalam dua dimensi dan memberikan gambaran tentang seberapa baik model clustering telah mengelompokkan data.

4.5. Evaluation

Tahap ini melibatkan evaluasi untuk menentukan model yang paling sesuai. Algoritma K-Means dimanfaatkan untuk mengklasifikasikan data. Untuk menemukan jumlah cluster optimal, evaluasi dilakukan menggunakan Silhouette Score pada berbagai jumlah cluster. Seperti yang terlihat pada Gambar 9, peneliti menggunakan metode silhouette coefficient untuk menentukan jumlah cluster terbaik.



Gambar 9. Grafik silhouette coefficient

Hasil dari evaluasi menggunakan Silhouette Score menunjukkan seberapa baik model K-Means membagi data ke dalam cluster yang berbeda. Berdasarkan hasil grafik pada Gambar 9 menunjukkan bahwa setelah menggunakan teknik silhouette coefficient, didapatkan nilai berikut: untuk n_cluster = 2, silhouette_score adalah 0.6373590606168947; untuk n_cluster = 3, silhouette_score adalah 0.47670506486135755; untuk n_cluster = 4, silhouette_score adalah 0.45441034636184824; dan untuk n_cluster = 5,

silhouette_score adalah 0.47444170429214194. Karena n_cluster = 2 memiliki nilai rata-rata silhouette coefficient tertinggi, yaitu 0.6373590606168947, penelitian ini menggunakan cluster = 2 sebagai nilai k untuk pemodelan k-means.

4.5.1. Hasil Clustering

Dataset di bawah ini terdiri dari 35 entri yang mewakili berbagai kabupaten/kota di Jawa Tengah. Kolom 'Cluster' menunjukkan hasil clustering yang dilakukan menggunakan algoritma K-Means dengan jumlah klaster yang ditentukan sebanyak 2 cluster, seperti yang terlihat pada Tabel 4 di bawah ini.

Tabel 4. Hasil data setelah di clustering

No	Kabupaten / Kota	Cluster
1	Kabupaten Cilacap	1
2	Kabupaten Banyumas	0
3	Kabupaten Purbalingga	0
4	Kabupaten Banjarnegara	0
5	Kabupaten Kebumen	1
6	Kabupaten Purworejo	1
7	Kabupaten Wonosobo	0
8	Kabupaten Magelang	0
9	Kabupaten Boyolali	0
10	Kabupaten Klaten	0
11	Kabupaten Sukoharjo	0
12	Kabupaten Wonogiri	0
13	Kabupaten Karanganyar	0
14	Kabupaten Sragen	0
15	Kabupaten Grobogan	0
16	Kabupaten Blora	0
17	Kabupaten Rembang	0
18	Kabupaten Pati	0
19	Kabupaten Kudus	0
20	Kabupaten Jepara	0
21	Kabupaten Demak	0
22	Kabupaten Semarang	0
23	Kabupaten Temanggung	0
24	Kabupaten Kendal	0
25	Kabupaten Batang	0
26	Kabupaten Pekalongan	0
27	Kabupaten Pemalang	0
28	Kabupaten Tegal	0
29	Kabupaten Brebes	0
30	Kota Magelang	0
31	Kota Surakarta	0
32	Kota Salatiga	0
33	Kota Semarang	0
34	Kota Pekalongan	0
35	Kota Tegal	0

Keterangan Cluster:

- a. Cluster 0: Mewakili daerah-daerah yang mungkin memiliki karakteristik bencana yang rendah, berdasarkan jumlah kejadian banjir, gempa bumi, dan tanah longsor.
- b. Cluster 1: Mewakili daerah-daerah dengan karakteristik bencana yang lebih tinggi dibandingkan dengan Cluster 0.

4.5.2. Manfaat

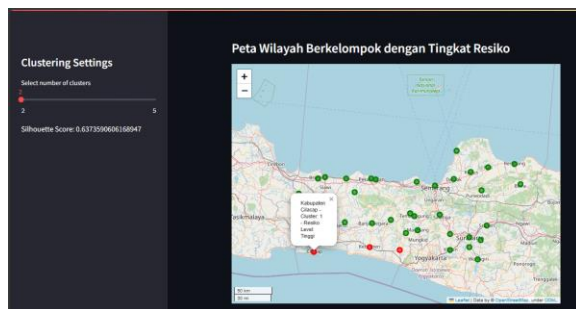
Hasil *clustering* ini dapat digunakan untuk:

- Mengidentifikasi daerah-daerah yang memerlukan perhatian khusus dalam hal mitigasi dan penanganan bencana.
- Mengalokasikan sumber daya dengan lebih efektif sesuai dengan kebutuhan masing-masing klaster.

Dengan menggunakan analisis ini, badan mitigasi bencana dapat lebih memahami pola kejadian bencana dan mengambil tahapan-tahapan yang tepat untuk meminimalkan dampak bencana di masa depan.

4.6. Deployment

Tahap akhir dari proses ini adalah penerapan hasil *clustering* ke dalam peta interaktif menggunakan *Streamlit* dan *Folium*, yang memvisualisasikan titik data di Provinsi Jawa Tengah dengan marker yang diwarnai berdasarkan klaster masing-masing. Yang ditunjukkan pada gambar 10 berikut :



Gambar 10. Tampilan hasil *clustering* dengan *map* menggunakan *streamlit*

Peta interaktif gambar 10 ini sangat membantu dalam pengambilan keputusan terkait penanggulangan bencana, karena dapat digunakan oleh pemangku kepentingan untuk mengidentifikasi wilayah-wilayah yang membutuhkan perhatian khusus.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa pengelompokan daerah rawan bencana di Jawa Tengah menggunakan algoritma *K-Means clustering* berhasil diterapkan. Untuk menentukan jumlah klaster yang optimal, digunakan metode *Silhouette Coefficient*. Dalam proses klasterisasi, teridentifikasi bahwa nilai *silhouette* tertinggi dicapai pada klasterisasi *K-Means* dengan jumlah klaster sebanyak 2. Pada tahap evaluasi, hasil yang diperoleh adalah 0,6373590606168947, yang termasuk dalam kategori baik. Dari klaster tersebut, 26 kabupaten dan 6 kota berada dalam klaster dengan indeks daerah bencana rendah, yaitu: Kabupaten Banyumas, Kabupaten Purbalingga, Kabupaten Banjarnegara, Kabupaten Temanggung, Kabupaten Kudus, Kabupaten Blora, Kabupaten Klaten, Kabupaten Pemalang, Kabupaten Batang, Kabupaten Grobogan, Kabupaten Jepara, Kabupaten Wonosobo, Kabupaten Karanganyar,

Kabupaten Sukoharjo, Kabupaten Boyolali, Kabupaten Pati, Kabupaten Brebes, Kabupaten Rembang, Kabupaten Semarang, Kabupaten Magelang, Kabupaten Tegal, Kabupaten Sragen, Kabupaten Demak, Kabupaten Pekalongan, Kabupaten Kendal, Kota Magelang, Kota Semarang, Kota Tegal, Kota Salatiga, Kota Surakarta, Kota Pekalongan. Sebanyak 3 kabupaten berada dalam klaster dengan indeks daerah bencana tinggi, yaitu Kabupaten Cilacap, Kabupaten Kebumen, dan Kabupaten Purworejo.

Penelitian ini diharapkan dapat menjadi sumber acuan yang berguna bagi penelitian berikutnya dalam bidang yang sama. Selain itu, diharapkan penelitian mendatang tidak hanya menggunakan algoritma, variabel, dan metode evaluasi yang sama, tetapi juga mencoba variasi yang lebih bervariasi untuk mendapatkan hasil yang lebih baik. Peneliti juga berharap penelitian mendatang dapat menghasilkan peta lokasi daerah rawan bencana yang lebih rinci hingga ke titik lokasi terjadinya bencana.

DAFTAR PUSTAKA

- W. A. Wibawana, "Apa itu Ring of Fire? Penyebab Indonesia Rawan Dilanda Gempa," 05 Desember 2022. Accessed: Apr. 24, 2024. [Online]. Available: <https://news.detik.com/berita/d-6444291/apa-itu-ring-of-fire-penyebab-indonesia-rawan-dilanda-gempa>
- BNPB, "Dokumen kajian risiko bencana nasional provinsi Jawa Tengah 2022 - 2026," *Kedeputian Bid. Sist. dan Strateg. Direktorat Pemetaan dan Eval. Risiko bencana*, p. 173, 2021.
- Badan Nasional Penanggulangan Bencana, "Statistik Bencana di Indonesia," Badan Nasional Penanggulangan Bencana (BNPB). Accessed: Jun. 20, 2024. [Online]. Available: <https://dibi.bnpb.go.id/>
- A. Rofifah, R. Goenjantoro, and D. Yuniarti, "Perbandingan Pengelompokan *K-Means* dan *K-Medoids* Pada Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Studi Kasus: Data Titik Panas Di Indonesia Pada 28 April 2018)," *J. EKSPONENSIAL*, vol. 10, no. 2, pp. 143–152, 2019.
- P. Hardati, "Hierarki Pusat Pelayanan di Kecamatan Ungaran Barat dan Ungaran Timur Kabupaten Semarang," *J. Geogr.*, vol. 13 No 2, no. 2, pp. 205–224, 2016.
- A. Baldah, A. V. Duarisah, and R. A. Maulana, "Clustering Daerah Rawan Bencana Alam Di Indonesia Berdasarkan Provinsi Dengan Metode *K-Means*," *J. Ilm. Inform. Glob.*, vol. 14, no. 2, pp. 31–36, 2023, doi: 10.36982/jiig.v14i2.3186.
- A. Atira and B. Nurina Sari, "Penerapan *Silhouette Coefficient*, Elbow Method dan Gap Statistics untuk Penentuan *Cluster Optimum* dalam Pengelompokan Provinsi di Indonesia

- Berdasarkan Indeks Kebahagiaan,” *J. Ilm. Wahana Pendidik.*, vol. 9, no. 17, pp. 76–86, 2023, [Online]. Available: <https://doi.org/10.5281/zenodo.8282638>
- [8] E. Tasia and M. Afdal, “Comparison Of *K-Means* And *K-Medoid* Algorithms For *Clustering* Of Flood-Prone Areas In Rokan Hilir District Perbandingan Algoritma *K-Means* Dan *K-Medoids* Untuk *Clustering* Daerah Rawan Banjir Di Kabupaten Rokan Hilir,” vol. 3, no. 1, pp. 65–73, 2023.
- [9] I. Nabilla Audy, T. Nur Padilah, and B. Nurina Sari, “Pengelompokan Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma Fuzzy *C-Means*,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2799–2803, 2024, doi: 10.36040/jati.v7i4.7205.
- [10] B. N. Azmi, A. Hermawan, and D. Avianto, “Analisis Pengaruh *PCA* Pada Klasifikasi Kualitas Air Menggunakan Algoritma *K-Nearest Neighbor* dan *Logistic Regression*,” *J. Sist. dan Teknol. Inf.*, vol. 7, no. 2, pp. 94–103, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/view/8190%0Ahttp://jurnal.unmuhjember.ac.id/index.php/JUSTINDO/article/download/8190/4143>
- [11] M. Z. Nasution, “PENERAPAN *PRINCIPAL COMPONENT ANALYSIS (PCA)* DALAM PENENTUAN FAKTOR DOMINAN YANG MEMPENGARUHI PRESTASI BELAJAR SISWA (Studi Kasus : SMK Raksana 2 Medan),” *J. Teknol. Inf.*, vol. 3, no. 1, p. 41, 2019, doi: 10.36294/jurti.v3i1.686.
- [12] F. N. Dhewayani, D. Amelia, D. N. Alifah, B. N. Sari, and M. Jajuli, “Implementasi *K-Means Clustering* untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model *CRISP-DM*,” *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, 2022, doi: 10.34010/jati.v12i1.6674.
- [13] M. Maesaroh, T. Nur Padilah, and J. Haerul Jaman, “Penerapan Algoritma *K-Means Clustering* Pada Pengelompokan Daerah Penyebaran Diare Di Provinsi Jawa Barat,” *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2783–2787, 2024, doi: 10.36040/jati.v7i4.7208.
- [14] R. O. Felani, “Analisis Prilaku Pengguna e-learning menggunakan Algoritma *K-Means Clustering*,” *Jusikom J. Sist. Komput. Musirawas*, vol. 7, no. 1, pp. 61–73, 2022, doi: 10.32767/jusikom.v7i1.1538.