JURNAL ELTEK
Vol. 24, No. 2, Oktober 2026, hal. 206~214
p-ISSN: 1693 – 4024 | e-ISSN: 2355-0740
DOI: 10.33795/eltek.v24i2.9955

Pelabelan tingkat guncangan gempa bumi menggunakan *clustering* berbasis *principal component analysis*

Agung Teguh Wibowo Almais[1], Renata Amalia Putri[2], Ahmad Zarkoni[3]
e-mail: agung.twa@ti.uin-malang.ac.id, renataamaliap04@gmail.com, ahmad.zarkoni@bmkg.go.id
[1,2]Jurusan Teknik Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia
[3]Badan Meteorologi, Klimatologi, dan Geofisika Malang, Indonesia

Informasi Artikel

Riwayat Artikel

Diterima 25 Mei 2026
Direvisi 21 Oktober 2026
Diterbitkan 30 Oktober 2026

Kata kunci:

Clustering
Gempa Bumi
Principal Component Analysis

ABSTRAK

Indonesia memiliki tingkat kerentanan tinggi terhadap bencana alam, khususnya gempa bumi. Perbedaan tingkat guncangan gempa memerlukan upaya mitigasi berbasis teknologi untuk mendukung analisis data gempa. Penelitian ini bertujuan mengelompokkan data gempa menggunakan metode *Principal Component Analysis* berbasis *clustering*. Data yang digunakan berupa data gempa harian tahun 2024 dari Badan Meteorologi, Klimatologi, dan Geofisika Stasiun Geofisika Kelas III Malang. Metode ini digunakan untuk mengekstraksi komponen utama yang merepresentasikan variasi dominan data. Hasil analisis menunjukkan bahwa komponen utama pertama mampu menjelaskan 34,1% variasi data dan didominasi oleh variabel *longitude* serta magnitudo. Nilai komponen utama pertama digunakan sebagai dasar proses *clustering* untuk mengidentifikasi pola gempa berdasarkan kemiripan karakteristik. Pengujian *Calinski-Harabasz Index* menghasilkan nilai sebesar 5,626,0269. Berdasarkan hasil tersebut, data gempa dikelompokkan menjadi kategori dirasakan dan tidak dirasakan mengacu pada Skala Intensitas Gempabumi Badan Meteorologi, Klimatologi, dan Geofisika. Hasil penelitian menunjukkan bahwa metode tersebut mampu mengurangi kompleksitas data dan membantu pengelompokan tingkat guncangan gempa secara efektif dan efisien.

Keywords:

Clustering
Earthquake
Principal Component Analysis

ABSTRACT

Indonesia is highly vulnerable to natural disasters, particularly earthquakes. This study aims to classify earthquake data using a clustering-based Principal Component Analysis method. The dataset consists of daily earthquake records in 2024 from the Badan Meteorologi, Klimatologi, dan Geofisika Geophysics Station Class III Malang. Principal component analysis was used to extract the main components representing dominant data variations. The results showed that the first principal component explained 34,1% of the variance and was dominated by longitude and magnitude variables. The first principal component scores were then used for clustering based on earthquake characteristics similarity. The Calinski-Harabasz Index produced a value of 5,626.0269. Based on the clustering results, earthquakes were categorized into felt and unfelt groups according to the Earthquake Intensity Scale of Badan Meteorologi, Klimatologi, dan Geofisika. The findings demonstrate that the method effectively reduces data complexity and supports efficient earthquake shaking classification.

Penulis Korespondensi:
Agung Teguh Wibowo Almais,
Jurusan Teknik Informatika,
Universitas Islam Negeri Maulana Malik Ibrahim Malang,
Jl. Gajayana No. 50, Malang, Jawa Timur, Indonesia.
Email: agung.twa@ti.uin-malang.ac.id

1. PENDAHULUAN

Indonesia merupakan negara dengan tingkat kerentanan yang tinggi terhadap bencana alam [1]. Secara geografis, Indonesia terletak pada pertemuan empat lempeng tektonik, yaitu lempeng Benua Asia, lempeng Benua Australia, lempeng Samudera Hindia, dan lempeng Samudera Pasifik [2]. Letak geografis tersebut menjadikan Indonesia sangat rawan mengalami berbagai bencana alam, seperti gempa bumi, tsunami, letusan gunung berapi, serta tanah longsor [3][4]. Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB) pada sepanjang tahun 2024 tercatat sebanyak 3.472 bencana alam yang didominasi oleh bencana hidrometeorologi seperti banjir, tanah longsor, kekeringan, dan puting beliung [5]. Di antara berbagai bencana tersebut, gempa bumi menjadi salah satu yang paling sering terjadi di Indonesia [6].

Menurut data *World Risk Report* (2017) menegaskan bahwa, kondisi kerentanan Indonesia memiliki persentase paparan bencana sebesar 19,36% dengan nilai kerentanan mencapai 54,19% [7]. Artinya, meskipun tingkat keterpaparan bencana tinggi, namun kesiapan Indonesia dalam menghadapi bencana masih tergolong rendah sehingga dampak bencana bisa menjadi lebih parah [8].

Gempa bumi dapat terjadi secara mendadak dengan intensitas yang bervariasi, mulai dari getaran ringan hingga guncangan besar yang menyebabkan kerusakan parah [9]. Setiap peristiwa gempa bumi direkam oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dalam bentuk data seismik dengan parameter numerik seperti magnitudo, kedalaman, koordinat geografis, dan waktu kejadian [10]. Kondisi ini menegaskan pentingnya upaya mitigasi untuk meminimalisasi dampak bencana. Oleh karena itu, teknologi berbasis *Principal Component Analysis* (PCA) diperlukan untuk pengelompokan terhadap tingkat guncangan yang disebabkan oleh gempa bumi [11].

Pendekatan menggunakan metode Principal Component Analysis telah dilakukan oleh Almais dkk. yang berhasil memanfaatkan PCA untuk proses *clustering* sehingga menghasilkan suatu teknik *clustering* baru yaitu PCA-*Clustering* dengan menggunakan data kerusakan sektor pasca-bencana [12]. PCA diterapkan dengan dua komponen utama, yakni PC1 dengan kontribusi variansi sebesar 85,17% dan PC2 dengan kontribusi variansi 9,36%. Hasil transformasi data kemudian dikelompokkan ke dalam tiga *cluster* berdasarkan nilai koordinat PC1, yaitu *cluster* 1 ($n < 0$) mewakili kerusakan ringan, *cluster* 2 ($0 \leq n < 2$) mewakili kerusakan sedang, dan *cluster* 3 ($n \geq 2$) mewakili kerusakan berat.

Principal Component Analysis (PCA) adalah istilah umum untuk sebuah teknik yang menggunakan prinsip-prinsip matematis yang kompleks untuk mentransformasikan sejumlah variabel yang mungkin saling berkorelasi menjadi sejumlah variabel yang lebih sedikit yang disebut komponen utama (*principal components*) [13]. PCA berbasis *clustering* digunakan untuk membentuk komponen utama dari variabel data gempa setelah melalui proses normalisasi data. Hasil transformasi PCA berupa nilai numerik kemudian dikelompokkan untuk menentukan label tingkat guncangan gempa bumi berdasarkan kategori yang telah ditetapkan yaitu kategori *dirasakan* dan *tidak dirasakan* menggunakan acuan Skala Intensitas Gempabumi (SIG-BMKG) yang disusun oleh BMKG [14].

Teknik *clustering* digunakan untuk menentukan label berdasarkan hasil nilai numerik dari teknik *Principal Component Analysis* (PCA), yang memanfaatkan 8539 data gempa yang telah melalui proses analisis dan memiliki nilai pada setiap kriteria. Selanjutnya, PCA memproses 8539 data tersebut untuk menghasilkan nilai numerik berdasarkan hasil visualisasi grafik PCA. Nilai-nilai numerik tersebut kemudian dikelompokkan (*clustered*) untuk menentukan jenis label tingkat guncangan gempa bumi.

Penelitian ini bertujuan untuk mengembangkan metode yang lebih akurat dan efisien dalam menentukan tingkat guncangan gempa berdasarkan data gempa bumi, sehingga dapat memberikan kontribusi yang berarti dalam mendukung upaya mitigasi bencana di Indonesia. Dengan demikian, penelitian ini diharapkan mampu memperluas penerapan metode PCA dalam analisis tingkat guncangan gempa serta menjadi rujukan bagi pengembangan teknik pengelompokan atau klasterisasi data pada penelitian kebencanaan di masa mendatang.

2. METODE PENELITIAN

2.1 Data dan Variabel Penelitian

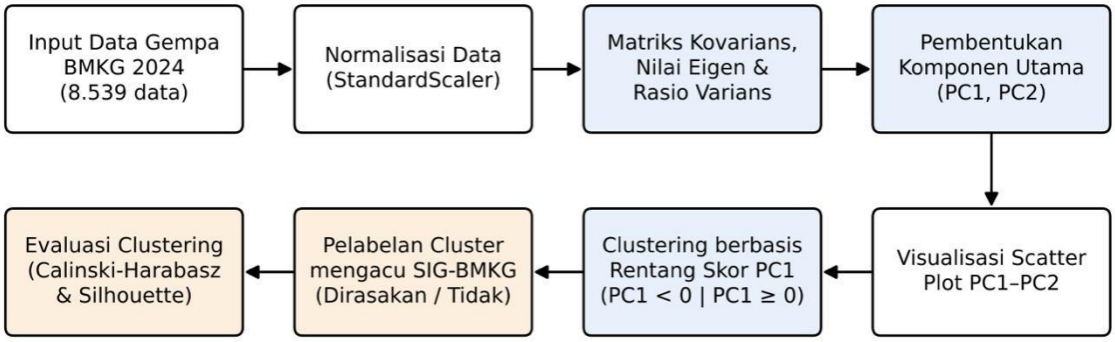
Penelitian ini menggunakan data sekunder yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang. Data yang digunakan merupakan data gempa bumi harian tahun 2024 yang mencakup beberapa variabel utama, yaitu longitude, latitude, magnitudo dan kedalaman (*depth*). Data gempa BMKG juga dapat diakses melalui laman Repositori Gempa BMKG di https://repogempa.bmkg.go.id. Jumlah data yang digunakan pada penelitian ini adalah 8.539 data, dan contoh lima data pertama dapat dilihat pada Tabel 1.

Tabel 1. Contoh Data Gempa Tahun 2024 (5 dari 8.539 data)

No	Latitude	Longitude	Depth	Magnitudo
1.	-8,9	112,93	39	2,8
2.	-9,57	113,86	15	3,3
3.	-7,8	111,33	11	2,5
4.	-7,27	108,68	10	2,2
5.	-7,94	110,46	20	1,3

2.2 Tahapan Penelitian

Tahapan penelitian ditunjukkan pada Gambar 1. Tahapan penelitian meliputi beberapa tahapan utama, yaitu pengumpulan data gempa, proses pra-pengolahan (*pre-processing*), reduksi dimensi menggunakan *Principal Component Analysis* (PCA), serta pengelompokan data berdasarkan hasil PCA. Tahapan dimulai dengan pengumpulan data yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang, yang mencakup data gempa bumi dengan berbagai atribut relevan. Selanjutnya, normalisasi data untuk menyamakan skala antarvariabel. Selanjutnya dilakukan PCA untuk mengidentifikasi komponen utama yang mampu menjelaskan sebagian besar variasi data. Hasil reduksi dimensi kemudian digunakan sebagai input dalam proses *clustering* untuk mengelompokkan data berdasarkan kemiripan karakteristik gempa. Setelah itu, data dikelompokkan berdasarkan nilai PC1, diberi label sesuai acuan SIG-BMKG, lalu dievaluasi dengan *Calinski-Harabasz Index* dan *Silhouette Coefficient*.



Gambar 1. Alur Penelitian PCA – Clustering

2.3 *Input* Data

Dataset yang digunakan pada penelitian ini adalah dataset Gempa Bumi tahun 2024 yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang. Dataset ini terdiri dari 4 fitur dan 8.539 baris data meliputi informasi rekapitulasi data gempa selama tahun 2024.

Pre-Processing

Pada tahap pre-processing, seluruh data dinormalisasi menggunakan metode *StandardScaler* agar setiap variabel berada pada skala yang sebanding [12]. Proses ini bertujuan untuk mencegah dominasi variabel tertentu akibat perbedaan satuan dan rentang nilai, sehingga analisis PCA dan clustering dapat berjalan secara optimal. Rumus normalisasi yang digunakan ditunjukkan pada persamaan (1).

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

Keterangan: z_{ij} merupakan hasil normalisasi data ke-*i* pada variabel ke-*j*, x_{ij} merupakan nilai data asli, $\bar{x}_j$ merupakan rata-rata, dan s_j merupakan standar deviasi dari variabel ke-*j*.

Proses

Tahap utama penelitian dimulai dengan penerapan PCA terhadap data yang telah dinormalisasi. PCA digunakan untuk mereduksi dimensi data serta menentukan komponen utama berdasarkan nilai eigen dan rasio varians. Komponen utama dengan kontribusi variansi terbesar dipilih sebagai representasi data. Dua komponen utama pertama (PC1 dan PC2) digunakan untuk visualisasi data dalam bentuk grafik dua dimensi guna mengamati pola persebaran dan kecenderungan pembentukan kluster.

Langkah PCA pada penelitian ini mengacu pada [13], [15]. Pertama, data hasil normalisasi disusun menjadi matriks Z berukuran 8.539 baris dan 4 kolom. Dari matriks tersebut dihitung matriks kovarians C seperti pada persamaan (2).

$$\mathbf{C} = \frac{1}{n-1}\mathbf{Z}^T\mathbf{Z} \quad (2)$$

Kedua, matriks kovarians diuraikan untuk mendapatkan nilai eigen (λ_k) dan vektor eigen (v_k) menggunakan persamaan (3). Nilai eigen kemudian diurutkan mulai dari yang paling besar.

$$\mathbf{C}\mathbf{v}_k = \lambda_k \mathbf{v}_k \quad (3)$$

Ketiga, rasio varians setiap komponen dihitung dengan persamaan (4). Komponen yang dipakai adalah komponen dengan nilai eigen lebih dari 1, sesuai kriteria Kaiser.

$$r_k = \frac{\lambda_k}{\sum_{j=1}^{p} \lambda_j} \quad (4)$$

Keempat, skor komponen utama diperoleh dengan mengalikan data hasil normalisasi dengan vektor eigen, seperti pada persamaan (5).

$$\mathbf{PC}_k = \mathbf{Z}\mathbf{v}_k \quad (5)$$

Terakhir, nilai faktor *loading* dihitung menggunakan persamaan (6). Nilai ini menunjukkan seberapa besar pengaruh setiap variabel terhadap komponen utama. Semakin besar nilai *loading*-nya, semakin dominan variabel tersebut.

$$l_{jk} = v_{jk}\sqrt{\lambda_k} \quad (6)$$

Output

Output penelitian berupa kelompok (*cluster*) data gempa yang memiliki karakteristik serupa. Metode *clustering* yang dipakai adalah PCA-*Clustering* yang mengacu pada penelitian Almais dkk. [12]. Pada metode ini, data tidak dikelompokkan menggunakan K-Means, tetapi berdasarkan rentang nilai PC1. PC1 dipilih karena memiliki rasio varians paling besar dan banyak dipengaruhi oleh variabel magnitudo. Jumlah *cluster* yang digunakan adalah dua, yaitu *cluster* 1 untuk data dengan PC1 < 0 dan *cluster* 2 untuk data dengan PC1 ≥ 0. Angka 0 dipakai sebagai batas karena setelah normalisasi nilai rata-rata PC1 adalah nol. Pada grafik PC1–PC2 juga terlihat bahwa data mulai terpisah di sekitar nilai tersebut.

Proses *clustering* tidak menentukan label, tetapi hanya membagi data menjadi dua kelompok. Label diberikan oleh peneliti setelah *cluster* terbentuk, dengan melihat karakteristik masing-masing *cluster* dan mengacu pada Skala Intensitas Gempabumi (SIG-BMKG) [14]. Dari hasil faktor *loading* (Tabel 3), magnitudo memiliki nilai positif yang besar pada PC1 (0,801), sedangkan kedalaman bernilai negatif (−0,222). Artinya, data dengan PC1 positif cenderung memiliki magnitudo lebih besar dan kedalaman lebih dangkal. Karena itu, *cluster* dengan PC1 ≥ 0 diberi label "Dirasakan" (SIG-BMKG-skala II ke atas) dan *cluster* dengan PC1 < 0 diberi label "Tidak Dirasakan" (SIG-BMKG-skala I).

2.4 Skenario Pengujian

Setelah tahap implementasi, dilakukan skenario pengujian untuk mengevaluasi kinerja hasil pengelompokan berbasis PCA yang telah dibangun. Evaluasi dilakukan berdasarkan validasi internal dan eksternal.

Validasi Internal

Validasi internal menggunakan skala *Calinski-Harabasz Index* (CHI) untuk menilai kualitas *clustering* berdasarkan tingkat kedekatan data dalam satu *cluster* dan jarak pemisah antar *cluster*. Semakin besar nilai Indeks CH menunjukkan jumlah cluster yang semakin baik [16]. Nilai indeks yang tinggi menunjukkan bahwa pemisahan antar cluster cukup jelas dan data dalam masing-masing cluster memiliki tingkat homogenitas yang baik. Rumus CHI ditunjukkan pada persamaan (7) [17].

$$CH = \frac{\mathrm{tr}(\mathbf{B}_K)/(K-1)}{\mathrm{tr}(\mathbf{W}_K)/(n-K)} \quad (7)$$

Keterangan: tr (B_K) adalah nilai *trace* matriks sebaran antar-*cluster*, tr (W_K) adalah nilai *trace* matriks sebaran di dalam *cluster*, *K* adalah banyaknya *cluster*, dan *n* adalah banyaknya data.

Pengujian kedua menggunakan *Silhouette Coefficient* [18]. Metode ini dipakai untuk melihat apakah setiap data sudah masuk ke *cluster* yang tepat. Nilai *silhouette* dihitung dengan persamaan (8).

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \quad (8)$$

Keterangan: *a*(*i*) adalah rata-rata jarak data ke-*i* dengan data lain di dalam *cluster* yang sama, sedangkan *b*(*i*) adalah rata-rata jarak data ke-*i* dengan data pada *cluster* terdekat. Nilai *silhouette* berada di antara −1 sampai 1. Jika nilainya mendekati 1, data sudah berada pada *cluster* yang tepat. Jika mendekati 0, data berada di perbatasan dua *cluster*, dan jika negatif, data kemungkinan masuk ke *cluster* yang salah. Nilai akhir yang digunakan adalah rata-rata *silhouette* dari seluruh data.

Validasi Eksternal

Validasi eksternal ini dilakukan dengan cara membandingkan hasil clustering berbasis PCA dengan data target asli yang diperoleh dari BMKG seperti, magnitudo dan kedalaman gempa. Tujuannya adalah untuk menilai sejauh mana akurasi hasil pengelompokan sesuai dengan label atau target yang tersedia dari BMKG.

3. HASIL DAN PEMBAHASAN

Hasil dan pembahasan menyajikan hasil analisis serta pembahasan yang diperoleh dari penerapan metode *Principal Component Analysis* (PCA) dan *clustering* terhadap data gempa harian. Penyajian hasil difokuskan pada tahapan utama penelitian, mulai dari proses *pre-processing* data, reduksi dimensi menggunakan PCA, hingga pembentukan dan evaluasi *cluster*. Setiap hasil yang diperoleh dianalisis secara sistematis untuk mengungkap pola utama yang terkandung dalam data serta keterkaitannya dengan karakteristik gempa.

3.1. Hasil *Pre-Processing Data*

Pada tahap *pre-processing data* dilakukan untuk mendapatkan data gempa yang berkualitas dan siap untuk diproses ke dalam model menggunakan metode *Principal Component Analysis* (PCA). Tahap *pre-processing* data ini terdapat dua tahap yaitu teknik pengumpulan data dan normalisasi data. Data yang digunakan untuk tahap *pre-processing* berjumlah 8.539 baris yang telah diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Stasiun Geofisika Kelas III Malang. Pada data ini berisi rekapan data gempa pada tahun 2024.

Selanjutnya, seluruh variabel dinormalisasi menggunakan metode *StandardScaler*. Hasil normalisasi menunjukkan bahwa setiap variabel memiliki nilai rata-rata mendekati nol dan standar deviasi sebesar satu. Hal ini memastikan bahwa tidak ada variabel yang mendominasi analisis akibat perbedaan satuan atau rentang nilai, sehingga PCA dapat menangkap variasi data secara proporsional.

3.2. Hasil *Principal Component Analysis*

Hasil PCA menunjukkan bahwa struktur variasi data gempa bumi dapat direpresentasikan secara efektif oleh sejumlah kecil komponen utama. Berdasarkan nilai eigen dan *variance ratio*, dua komponen utama pertama (PC1 dan PC2) mampu menjelaskan sekitar 60,7% total variasi data. Komponen utama pertama (PC1) memiliki nilai eigen sebesar 1,364353 dengan *variance ratio* 0,341048, yang berarti komponen ini mampu menjelaskan sekitar 34,1% dari total variasi data. Komponen utama kedua (PC2) memiliki nilai eigen sebesar 1,064056 dengan *variance ratio* 0,265983, atau sekitar 26,6% dari total variasi, sebagaimana ditunjukkan pada Tabel 2.

Tabel 2. Hasil Nilai Eigen dan Rasio Varians

Komponen	Nilai Eigen	Rasio Varians
PC1	1,364353	0,341048
PC2	1,064056	0,265983
PC3	0,973796	0,243421
PC4	0,598263	0,149548

Komponen utama pertama (PC1) didominasi oleh variabel longitude dan magnitudo, yang menunjukkan PC1 sebagai komponen merepresentasikan lokasi wilayah dan faktor besaran getaran gempa. Komponen utama kedua (PC2) didominasi oleh variabel latitude dan longitude artinya komponen tersebut menggambarkan dinamika lokasi wilayah terjadi gempa. Nilai faktor loading setiap variabel ditunjukkan pada Tabel 3.

Tabel 3. Hasil Faktor Loading Variabel

Variabel	PC1	PC2	PC3	PC4
Latitude	0,055667	0,897257	0,351669	-0,261297
Longitude	0,818659	0,136841	0,21202	0,515981
Depth	-0,222158	-0,381343	0,897281	-0,015092
Magnitudo	0,80106	-0,307958	0,007727	-0,513344

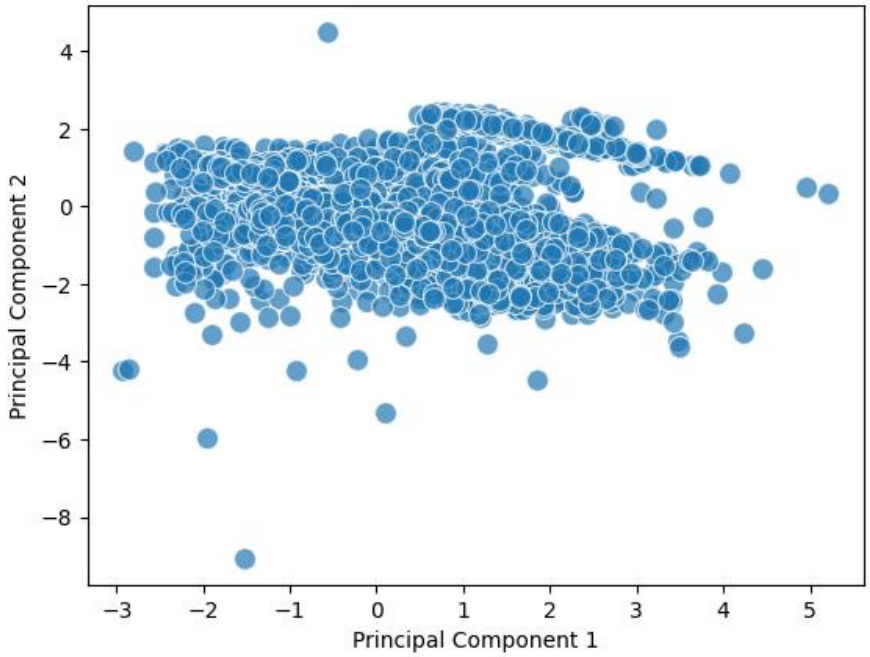
Dominasi PC1 dan PC2 menunjukkan bahwa reduksi dimensi menjadi dua komponen utama sudah memadai untuk visualisasi dan analisis lanjutan, karena sebagian besar informasi penting data tetap dipertahankan.

3.3. Visualisasi dan Pola Persebaran

Visualisasi *scatter plot* menggunakan komponen utama pertama (PC1) dan komponen utama kedua (PC2) dilakukan untuk menggambarkan struktur penyebaran data gempa setelah proses reduksi dimensi dengan PCA. Pada Gambar 2. menunjukkan bahwa sebaran titik pada grafik memperlihatkan bagaimana data gempa tersebar setelah direduksi dimensinya dari variabel asli ke dua komponen utama yang memiliki kontribusi

informasi terbesar. Terlihat bahwa titik-titik data menumpuk atau membentuk satu kelompok bukan menyebar ke beberapa area dengan pola distribusi yang berbeda.

Grafik dua dimensi ini menampilkan proyeksi setiap data gempa pada bidang PC1–PC2, di mana sumbu horizontal merepresentasikan PC1 sebagai komponen dengan kontribusi variansi terbesar, sedangkan sumbu vertikal merepresentasikan PC2 sebagai komponen dengan kontribusi variansi kedua terbesar. Penyebaran titik pada grafik menunjukkan bagaimana data terdistribusi setelah direduksi dari ruang variabel asli yang berdimensi tinggi ke dalam dua komponen utama.



Gambar 2. Grafik *Scatter Plot* Hasil PCA

3.4. Hasil *Clustering* dan Pelabelan

Berdasarkan pola persebaran datanya, grafik ini menampilkan hasil visualisasi PCA yang telah diberi label berdasarkan hasil prediksi *clustering*, yaitu kategori “Dirasakan” dan “Tidak Dirasakan”. Setiap titik pada grafik merepresentasikan satu data observasi yang telah direduksi ke dalam dua dimensi utama, yaitu *Principal Component* 1 (PC1) dan *Principal Component* 2 (PC2). Warna merah menunjukkan data dengan label “Dirasakan”, sedangkan warna oranye menunjukkan data dengan label “Tidak Dirasakan”.



Gambar 3. Grafik *Cluster* berdasarkan Label

Berdasarkan sebarannya, terlihat bahwa pemisahan antara kedua label cukup jelas terjadi pada sumbu *Principal Component* 1. Data dengan label "Tidak Dirasakan" cenderung terkonsentrasi pada nilai PC1 yang negatif (sebelah kiri grafik), sedangkan data dengan label "Dirasakan" lebih banyak berada pada nilai PC1 yang positif (sebelah kanan grafik). Hal ini menunjukkan bahwa komponen utama pertama memiliki peran yang sangat dominan dalam membedakan kedua kelompok tersebut. Dengan kata lain, PC1 menjadi faktor pembeda utama antara data yang termasuk kategori "Dirasakan" dan "Tidak Dirasakan". Batas kedua kelompok berada di sekitar nilai PC1 = 0, sesuai dengan aturan yang digunakan pada tahap metode.

Secara keseluruhan, visualisasi ini menunjukkan bahwa proses reduksi dimensi menggunakan PCA berhasil mempertahankan struktur pemisahan data berdasarkan label prediksi. Komponen utama pertama terbukti efektif dalam merepresentasikan perbedaan karakteristik antara kategori "Dirasakan" dan "Tidak Dirasakan", sehingga mendukung kualitas hasil *clustering* atau pengelompokan yang diperoleh dalam ruang dua dimensi PCA.

Tabel 4 menampilkan hasil pelabelan data gempa berdasarkan nilai dua komponen utama hasil reduksi dimensi menggunakan *Principal Component Analysis* (PCA), yaitu *Principal Component* 1 (PC1) dan *Principal Component* 2 (PC2). Setiap baris pada tabel merepresentasikan satu data observasi yang telah ditransformasikan dari variabel asli ke dalam bentuk skor komponen utama, kemudian diberikan label sesuai hasil pengelompokan atau klasifikasi, yaitu "Dirasakan" dan "Tidak Dirasakan". Data pada Tabel 4 hanya lima contoh dari 8.539 data.

Tabel 4. Contoh Hasil Pelabelan (5 dari 8.539 data)

No	PC1	PC2	Label
1.	1,004969	-0,352439	Dirasakan
2.	2,183259	-0,773692	Dirasakan
3.	-0,050502	0,699144	Tidak Dirasakan
4.	-1,842468	0,969191	Tidak Dirasakan
5.	-1,95835	1,019267	Tidak Dirasakan

Berdasarkan tabel tersebut, terlihat bahwa data dengan nilai PC1 positif cenderung memperoleh label "Dirasakan", seperti pada data nomor 1 dan 2 yang memiliki nilai PC1 sebesar 1,004969 dan 2.183259. Sebaliknya, data dengan nilai PC1 negatif cenderung termasuk dalam kategori "Tidak Dirasakan", seperti pada data nomor 3, 4, dan 5 yang memiliki nilai PC1 kurang dari nol. Hal ini menunjukkan bahwa komponen utama pertama (PC1) memiliki peran dominan dalam membedakan karakteristik gempa yang "Dirasakan" dan "Tidak Dirasakan".

3.5. Evaluasi Model

Evaluasi kualitas *clustering* dilakukan dengan menggunakan *Calinski-Harabasz Index* (CHI). Hasil pengujian dengan menghasilkan dua *cluster* memiliki nilai CH *Index* sebesar 5.626,0269, yang menunjukkan bahwa struktur *cluster* yang terbentuk memiliki tingkat pemisahan yang tergolong baik. Nilai indeks CH yang lebih besar menunjukkan bahwa titik-titik data dalam setiap cluster memiliki kemiripan internal yang kuat dan jarak yang jelas terhadap *cluster* lain. Selain itu, jumlah 2 *cluster* juga sesuai dengan kategori label BMKG, yang hanya membagi kondisi gempa bumi ke dalam 2 kelas, yaitu "Tidak Dirasakan" dan "Dirasakan".

Pengujian juga dilakukan menggunakan *Silhouette Coefficient* pada data hasil PCA (PC1 dan PC2), sama seperti perhitungan CHI. Nilai rata-rata *silhouette* yang diperoleh adalah 0.4146. Nilai ini positif, sehingga sebagian besar data sudah berada pada *cluster* yang sesuai. Namun, nilainya belum mendekati 1 karena masih ada data di sekitar batas PC1 = 0 yang letaknya dekat dengan *cluster* lain. Dari 8.539 data, sebanyak 3.567 data masuk ke *cluster* "Dirasakan" dan 4.972 data masuk ke *cluster* "Tidak Dirasakan". Hasil pengujian dirangkum pada Tabel 5.

Tabel 5. Ringkasan Hasil Evaluasi *Clustering*

Metrik	Nilai	Keterangan
Calinski-Harabasz Index	5.626,0269	Semakin besar semakin baik
Silhouette Coefficient	0,4146	Rentang −1 s.d. 1; mendekati 1 semakin baik
Jumlah data per cluster	3.567 dan 4.972	Dirasakan dan Tidak Dirasakan

4. KESIMPULAN

Berdasarkan hasil penelitian, metode PCA berbasis *clustering* dapat digunakan untuk melabelkan tingkat guncangan pada 8.539 data gempa BMKG tahun 2024. Dari empat variabel yang digunakan, PCA menghasilkan dua komponen utama dengan nilai eigen di atas 1. PC1 memiliki nilai eigen 1,364 dan menjelaskan 34,1% variasi data, sedangkan PC2 memiliki nilai eigen 1,064 dan menjelaskan 26,6% variasi data. Jika digabungkan, kedua komponen ini menjelaskan 60,7% variasi data.

PC1 paling dipengaruhi oleh longitude (*loading* 0,819) dan magnitudo (*loading* 0,801), sehingga PC1 dipakai sebagai dasar pengelompokan. Dengan batas PC1 = 0, data terbagi menjadi dua *cluster*, yaitu 3.567 data "Dirasakan" (PC1 $\geq$ 0) dan 4.972 data "Tidak Dirasakan" (PC1 < 0) sesuai acuan SIG-BMKG. Hasil pengujian menunjukkan nilai *Calinski-Harabasz Index* sebesar 5.626,0269 dan *Silhouette Coefficient* sebesar 0.4146. Kedua nilai ini menunjukkan bahwa *cluster* yang terbentuk sudah cukup terpisah, meskipun masih ada data yang berada di sekitar batas kedua *cluster*.

Hasil ini diharapkan dapat membantu analisis kebencanaan dan upaya mitigasi gempa bumi. Pada penelitian berikutnya, hasil pelabelan dapat dibandingkan dengan data intensitas guncangan dari BMKG, serta dapat dicoba menggunakan metode *clustering* lain seperti K-Means atau DBSCAN.

5. DAFTAR PUSTAKA

[1] A. Harijoko, A. Anggraini, A. Setianto, A. Hendratno, and ..., *Geologi Gempa Bumi Indonesia*. books.google.com, 2024. [Online]. Available: https://books.google.com/books?hl=en&lr=&id=ojPuEAAAQBAJ&oi=fnd&pg=PA1&dq=gempa&ots=qf95JtN03j&sig=39GC7PdnKn0EHZrtQO9bkw411Gg

[2] J. Inayah, A. Fanani, and W. D. Utami, "Klasterisasi Data Kejadian Gempa Bumi di Indonesia Menggunakan Metode K-Medoids," *JUSTIN (Jurnal Sistem dan …*, 2024.

[3] T. Dartanto, "Natural disasters, mitigation and household welfare in Indonesia: Evidence from a large-scale longitudinal survey," *Cogent Economics and Finance*, vol. 10, no. 1, 2022, doi: 10.1080/23322039.2022.2037250.

[4] A. Wibowo, "Clustering Indonesian Provinces by Disaster Intensity using K-Means Algorithm: a Data Mining Approach," *Disaster Advances*, vol. 17, no. 12, pp. 1–8, 2024, doi: 10.25303/1712da0108.

[5] BNPB, "Indeks Risiko Bencana Indonesia Tahun 2024," *Badan Nasional Penanggulangan Bencana*, vol. 03, pp. 1–184, 2024.

[6] H. Hadi, S. Agustina, and A. Subhani, "Penguatan Kesiapsiagaan Stakeholder dalam Pengurangan Risiko Bencana Alam Gempabumi," *Geodika: Jurnal Kajian Ilmu dan Pendidikan Geografi*, vol. 3, no. 1, p. 30, 2019, doi: 10.29408/geodika.v3i1.1476.

[7] M. Fuady, R. Munadi, and M. A. K. Fuady, "Disaster mitigation in Indonesia: between plans and reality," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1087, no. 1, p. 012011, 2021, doi: 10.1088/1757-899x/1087/1/012011.

[8] T. Y. Susanto, "On the earthquake distribution modeling in sumatra by cauchy cluster process: Comparing log-linear and log-Additive intensity models," *Sains Malays.*, vol. 52, no. 2, pp. 655–667, 2023, doi: 10.17576/jsm-2023-5202-25.

[9] K. H. AlKhaldi, "96 - Earthquakes," G. B. T.-C. D. M. (Third E. Ciottone, Ed., New Delhi: Elsevier, 2024, pp. 601–604. doi: https://doi.org/10.1016/B978-0-323-80932-0.00096-3.

[10] P. Supendi, S. Rohadi, D. Karnawati, and A. D. Nugraha, "Analysis of the destructive earthquakes end of West Java , Indonesia," vol. 02003, pp. 1–6, 2020.

[11] R. Pourramezan, H. Karimi, and J. Mahseredjian, "Real-time disturbance detection and classification using principal component analysis of PMU data," *IEEE Power and Energy Society General Meeting*, vol. 2020-Augus, no. 1, 2020, doi: 10.1109/PESGM41954.2020.9281677.

[12] A. T. W. Almais *et al.*, "Principal Component Analysis-Based Data Clustering for Labeling of Level Damage Sector in Post-Natural Disasters," *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3275852.

[13] B. M. S. Hasan, "A Review of Principal Component Analysis Algorithm for Dimensionality Reduction," *Journal of Soft Computing and Data Mining*, vol. 2, no. 1, pp. 20–30, 2021, doi: 10.30880/jscdm.2021.02.01.003.

[14] M. Muzli, M. Masturyono, J. Murjaya, and M. Riyadi, "Studi Awal Penyusunan Skala Intensitas Gempabumi Badan Meteorologi Klimatologi Dan Geofisika," 2017. doi: 10.31172/jmg.v17i2.440.

[15] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.

[16] A. F. Khairati, A. A. Adlina, G. F. Hertono, and B. D. Handari, “Kajian Indeks Validitas pada Algoritma K-Means Enhanced dan K-Means MMCA,” vol. 2, pp. 161–170, 2019.
[17] T. Caliński and J. Harabasz, “A dendrite method for cluster analysis,” *Communications in Statistics*, vol. 3, no. 1, pp. 1–27, Jan. 1974, doi: 10.1080/03610927408827101.
[18] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.