

## PENERAPAN AGGLOMERATIVE HIERARCHICAL CLUSTERING DALAM PENGELOMPOKAN TINGKAT KERUSAKAN BANGUNAN PASCAGEMPA

Bagus Haryo Wibowo Kusumo, Agung Teguh Wibowo Almais,  
Yunifa Miftachul Arif, Muhammad Ainul Yaqin, Ririen Kusumawati  
Magister Informatika, Universitas Islam Negeri Maulana Malik Ibrahim Malang  
Jalan Gajayana No.50 Malang, Indonesia  
Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) Pasuruan, Indonesia  
240605220010@student.uin-malang.ac.id

### ABSTRAK

Indonesia merupakan wilayah dengan aktivitas seismik tinggi yang menyebabkan variasi tingkat kerusakan bangunan pascagempa. Permasalahan dalam penelitian ini adalah belum adanya pendekatan berbasis data yang secara sistematis mampu mengelompokkan tingkat kerusakan bangunan berdasarkan parameter numerik gempa bumi. Penelitian ini bertujuan menerapkan metode Agglomerative Hierarchical Clustering untuk mengidentifikasi pola pengelompokan tingkat kerusakan bangunan pascagempa berdasarkan variabel Magnitude, Depth (km), dan Epicenter (km). Data yang digunakan sebanyak 23 kejadian gempa periode April–Mei 2021 dari Pusat Gempa Regional VII BMKG. Tahapan penelitian meliputi preprocessing, normalisasi menggunakan StandardScaler, perhitungan Euclidean Distance, serta proses penggabungan kluster menggunakan Ward Linkage. Evaluasi dilakukan menggunakan Silhouette Score dan didukung validasi eksternal menggunakan data kerusakan bangunan aktual. Hasil penelitian menunjukkan konfigurasi optimal pada dua kluster dengan nilai Silhouette Score sebesar 0,63 yang mengindikasikan struktur kluster kuat dan terpisah dengan baik. Kluster yang terbentuk merepresentasikan dua karakteristik aktivitas seismik berbeda yang berkorelasi dengan variasi tingkat kerusakan bangunan. Dengan demikian, metode Agglomerative Clustering efektif digunakan sebagai pendekatan analitik dalam mendukung mitigasi dan pengambilan keputusan kebencanaan berbasis data.

**Kata kunci :** *Agglomerative Clustering, Gempa Bumi, Kerusakan Bangunan, Silhouette Score, Ward Linkage.*

### 1. PENDAHULUAN

Indonesia merupakan salah satu negara dengan tingkat aktivitas seismik tertinggi di dunia karena berada pada pertemuan Lempeng Indo-Australia, Eurasia, Pasifik, dan Laut Filipina [1]. Interaksi tektonik tersebut menyebabkan frekuensi gempa bumi relatif tinggi dengan intensitas yang bervariasi [2]. Dampak gempa bumi tidak hanya ditentukan oleh magnitudo, tetapi juga dipengaruhi oleh kedalaman, jarak terhadap sumber gempa, serta kondisi geologi lokal yang menyebabkan variasi tingkat kerusakan bangunan meskipun berada dalam wilayah yang berdekatan [3], [4].

Gempa bumi 10 April 2021 dengan magnitudo 6,1 di selatan Kabupaten Malang menjadi salah satu contoh kejadian yang menimbulkan dampak kerusakan signifikan di berbagai wilayah Jawa Timur [3]. Laporan resmi menunjukkan ribuan bangunan mengalami kerusakan dengan kategori ringan, sedang, hingga berat [5]. Namun demikian, distribusi tingkat kerusakan tersebut tidak bersifat linier terhadap parameter magnitudo semata. Variasi spasial dan karakteristik sumber gempa menyebabkan pola kerusakan yang kompleks dan sulit dipetakan secara konvensional [6].

Pendekatan analisis berbasis data menjadi penting dalam konteks mitigasi bencana [7]. Beberapa penelitian sebelumnya telah menerapkan metode clustering untuk mengelompokkan kejadian gempa berdasarkan karakteristik seismik [8], [9]. Metode Agglomerative Hierarchical Clustering terbukti efektif

dalam memetakan zona gempa dan struktur spasial aktivitas seismik [3], [10]. Selain itu, pendekatan clustering juga digunakan dalam analisis kerusakan bangunan pascabencana melalui integrasi metode numerik dan citra [11], [12], [13]. Meskipun demikian, sebagian besar penelitian masih berfokus pada pengelompokan kejadian gempa, bukan secara spesifik pada pengelompokan tingkat kerusakan bangunan berbasis parameter seismik. Dengan demikian, terdapat kebutuhan akan pendekatan analitik yang mampu mengelompokkan tingkat kerusakan bangunan secara sistematis berdasarkan data numerik gempa bumi, sehingga dapat mendukung perencanaan mitigasi dan pengambilan keputusan kebencanaan secara lebih objektif.

Selain itu, penggunaan Ward linkage terbukti mampu menghasilkan kluster yang lebih homogen karena meminimalkan variasi di dalam kluster [14]. Namun demikian, sebagian besar penelitian masih berfokus pada pengelompokan kejadian gempa, sementara pemanfaatannya untuk mengelompokkan tingkat kerusakan bangunan pascagempa masih relatif terbatas.

Berdasarkan latar belakang tersebut, permasalahan dalam penelitian ini adalah belum adanya model berbasis Agglomerative Hierarchical Clustering yang secara eksplisit digunakan untuk mengelompokkan tingkat kerusakan bangunan pascagempa berdasarkan parameter seismik numerik, serta divalidasi menggunakan metrik evaluasi internal dan data kerusakan aktual.

Ketiadaan pendekatan tersebut menyebabkan analisis dampak gempa masih bersifat deskriptif dan belum sepenuhnya memanfaatkan pola numerik yang terkandung dalam data seismik.

Penelitian ini bertujuan untuk menerapkan metode Agglomerative Hierarchical Clustering dalam mengelompokkan tingkat kerusakan bangunan pascagempa berdasarkan parameter Magnitude, Depth (km), dan Epicenter distance (km), serta mengevaluasi kualitas kluster menggunakan metrik validasi internal dan eksternal.

Penyelesaian permasalahan dilakukan melalui beberapa tahapan sistematis. Pertama, dilakukan preprocessing data gempa bumi yang meliputi data cleaning dan normalisasi menggunakan StandardScaler untuk menghindari bias skala antar variabel [15]. Kedua, pengukuran kemiripan antar data dilakukan menggunakan Euclidean Distance sebagai metrik jarak multidimensi [4]. Ketiga, proses penggabungan kluster dilakukan menggunakan Ward Linkage yang meminimalkan variansi intra-kluster sehingga menghasilkan struktur kluster yang homogen dan stabil [10].

Selanjutnya, kualitas kluster dievaluasi menggunakan Silhouette Score sebagai metrik validasi internal [16]. Untuk memastikan relevansi empiris, hasil kluster dibandingkan dengan data kerusakan bangunan aktual sebagai validasi eksternal [11], [12]. Dengan pendekatan ini, penelitian tidak hanya menghasilkan struktur

## 2. TINJAUAN PUSTAKA

Pendekatan clustering telah banyak digunakan dalam analisis data gempa bumi untuk mengidentifikasi pola spasial dan karakteristik aktivitas seismik. Penelitian di Sulawesi Utara menunjukkan bahwa metode Agglomerative Clustering mampu mengelompokkan kejadian gempa berdasarkan kemiripan magnitudo dan kedalaman sehingga menghasilkan kluster dengan karakteristik berbeda [1]. Hasil tersebut mendukung pemetaan zona risiko gempa secara lebih terstruktur.

Analisis serupa pada Pulau Sumatra menunjukkan bahwa metode hierarchical clustering efektif dalam membedakan wilayah berisiko tinggi dan rendah berdasarkan distribusi kedalaman dan magnitudo gempa [2]. Implementasi Agglomerative Hierarchical Clustering pada wilayah Sulawesi juga membuktikan bahwa pendekatan ini mampu memetakan area gempa dengan struktur spasial yang konsisten [3]. Temuan-temuan tersebut menegaskan bahwa teknik clustering dapat mengungkap struktur tersembunyi dalam data seismik yang sulit diidentifikasi melalui analisis deskriptif konvensional.

Agglomerative Hierarchical Clustering merupakan metode unsupervised learning dengan pendekatan bottom-up, di mana setiap data awalnya dianggap sebagai satu kluster kemudian digabungkan secara bertahap berdasarkan tingkat kemiripan hingga membentuk struktur hierarki lengkap [4]. Metode ini

banyak digunakan karena hasilnya dapat divisualisasikan dalam bentuk dendrogram sehingga memudahkan interpretasi struktur data.

Secara teoritis, clustering berbasis hierarki memiliki keunggulan dalam merepresentasikan hubungan antar data tanpa perlu menentukan jumlah kluster sejak awal [12]. Selain itu, penerapan hierarchical clustering pada data gempa di kawasan Asia Tenggara menunjukkan bahwa metode ini efektif dalam mengelompokkan titik kejadian gempa berdasarkan kedekatan spasial dan karakteristik numerik [13].

Algoritma hierarchical modern juga telah dioptimalkan untuk meningkatkan efisiensi komputasi dan stabilitas hasil kluster [14]. Hal ini menjadikan metode agglomerative relevan untuk analisis data seismik yang memiliki variasi numerik dan spasial cukup tinggi.

Dalam proses agglomerative clustering, metode linkage menentukan cara penggabungan antar kluster. Ward linkage merupakan pendekatan yang menggabungkan dua kluster dengan kenaikan variansi intra-kluster terkecil, sehingga menghasilkan kluster yang lebih homogen dan seimbang [4].

Penggunaan Ward linkage dalam konteks analisis risiko gempa terbukti menghasilkan struktur kluster yang stabil dan representatif terhadap zona seismik aktual [3]. Kombinasi metode ini dengan metrik Euclidean Distance memungkinkan pengukuran kemiripan antar titik gempa dalam ruang multidimensi secara objektif [13].

Secara konseptual, Euclidean Distance menghitung jarak geometris antar titik berdasarkan atribut numerik, sehingga cocok digunakan pada variabel seperti magnitudo, kedalaman, dan jarak episenter [12]. Optimalisasi pengelompokan zona bahaya pascagempa juga menunjukkan bahwa pemilihan metrik jarak yang tepat berpengaruh signifikan terhadap kualitas kluster [10]. Studi komparatif terbaru pada data gempa Indonesia menegaskan pentingnya normalisasi dan pemilihan metode clustering untuk menghindari bias akibat perbedaan skala variabel [11].

Selain diterapkan pada pengelompokan kejadian gempa, metode clustering juga digunakan dalam analisis tingkat kerusakan bangunan pascabencana. Integrasi Principal Component Analysis (PCA) dan clustering berhasil digunakan untuk melabeli tingkat kerusakan sektor bangunan secara objektif [5].

Pendekatan berbasis tekstur citra menggunakan GLCM dan PCA juga mampu mengklasifikasikan tingkat kerusakan bangunan dengan akurasi tinggi dibandingkan metode manual [6]. Di sisi lain, pendekatan deep learning berbasis Convolutional Neural Network (CNN) memungkinkan identifikasi kerusakan bangunan secara otomatis dari citra satelit sebelum dan sesudah gempa [7].

Metode clustering juga digunakan untuk mensimulasikan kerusakan bangunan pada skala kota besar dengan efisiensi komputasi yang lebih baik

dibanding simulasi individual per bangunan [8]. Selain itu, integrasi machine learning dalam analisis risiko gempa menunjukkan bahwa kombinasi model numerik dan spasial dapat meningkatkan akurasi pemetaan risiko [9].

Meskipun berbagai penelitian telah mengkaji pengelompokan gempa maupun kerusakan bangunan, kajian yang secara langsung menghubungkan parameter seismik numerik dengan pengelompokan tingkat kerusakan bangunan masih terbatas.

Evaluasi kualitas kluster menjadi aspek penting dalam penelitian berbasis unsupervised learning. Salah satu metrik yang umum digunakan adalah Silhouette Score, yang mengukur keseimbangan antara kohesi intra-kluster dan separasi antar-kluster [16]. Nilai Silhouette yang mendekati 1 menunjukkan struktur kluster yang kuat dan terpisah dengan baik.

Dalam konteks kerusakan bangunan pascabencana, validasi eksternal menggunakan data lapangan menjadi langkah penting untuk memastikan bahwa hasil clustering memiliki relevansi empiris. Studi terbaru menunjukkan bahwa pengelompokan tingkat kerusakan bangunan perlu diverifikasi menggunakan data aktual agar model yang dihasilkan tidak hanya valid secara matematis tetapi juga representatif secara praktis [15].

Dengan demikian, kombinasi validasi internal menggunakan Silhouette Score dan validasi eksternal berbasis data kerusakan aktual memberikan dasar evaluasi yang komprehensif dalam penelitian ini.

### 3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksploratif dengan metode unsupervised learning, yaitu Agglomerative Hierarchical Clustering, untuk mengelompokkan tingkat kerusakan bangunan pascagempa bumi berdasarkan parameter seismik. Pendekatan ini dipilih karena mampu mengidentifikasi pola dan struktur hierarkis alami pada data gempa tanpa memerlukan label awal, sehingga sesuai untuk analisis fenomena kebencanaan yang bersifat kompleks dan tidak terstruktur.

#### 3.1. Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder gempa bumi yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) melalui repositori resmi repogempa.bmkg.go.id. Data mencakup kejadian gempa bumi di Wilayah Pusat Gempa Regional VII (Jawa Timur) pada periode April 2021, dengan total 23 kejadian gempa.

Variabel awal terdiri dari latitude dan longitude yang kemudian ditransformasikan menjadi Epicenter distance (km) menggunakan perhitungan jarak Haversine terhadap titik referensi pusat gempa. Dengan demikian, variabel akhir yang digunakan dalam proses clustering adalah Magnitude, Depth (km), dan Epicenter distance (km). Keempat variabel tersebut dipilih karena secara langsung

merepresentasikan karakteristik spasial dan kekuatan gempa yang berpotensi memengaruhi tingkat kerusakan bangunan. Selain itu, data kerusakan bangunan pascagempa yang dilaporkan oleh BPBD Jawa Timur digunakan sebagai data pembanding (validasi eksternal) untuk mengevaluasi relevansi hasil kluster dengan kondisi nyata di lapangan.

#### 3.2. Preprocessing Data

Tahap preprocessing bertujuan untuk memastikan data berada dalam kondisi siap dianalisis dan bebas dari bias akibat kesalahan pencatatan maupun perbedaan skala antar variabel. Preprocessing dilakukan melalui dua tahap utama, yaitu data cleaning dan normalisasi data.

Data cleaning dilakukan dengan menghilangkan data duplikat, memastikan tidak terdapat nilai kosong (missing values) pada variabel utama, serta menyeragamkan format data agar konsisten. Tahap ini penting karena kualitas data yang rendah dapat menghasilkan struktur kluster yang tidak stabil dan sulit diinterpretasikan secara ilmiah.

Selanjutnya, dilakukan normalisasi data menggunakan metode StandardScaler. Normalisasi diperlukan karena setiap variabel memiliki rentang nilai yang berbeda, misalnya kedalaman gempa yang bernilai puluhan hingga ratusan kilometer, sedangkan magnitudo berada pada rentang yang jauh lebih kecil [10]. Proses normalisasi dilakukan dengan mengubah setiap variabel agar memiliki nilai rata-rata nol dan simpangan baku satu, sebagaimana ditunjukkan pada Persamaan (1):

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (1)$$

Normalisasi ini bertujuan untuk mencegah dominasi variabel tertentu dalam proses perhitungan jarak, sehingga setiap variabel memberikan kontribusi yang seimbang dalam pembentukan kluster.

#### 3.3. Agglomerative Hierarchical Clustering

Metode Agglomerative Hierarchical Clustering digunakan sebagai inti dari proses pengelompokan data [17]. Metode ini bekerja dengan pendekatan bottom-up, di mana setiap data awalnya dianggap sebagai satu kluster, kemudian secara bertahap digabungkan dengan kluster lain yang paling mirip hingga membentuk struktur hierarki [18].

Pengukuran tingkat kemiripan antar data dilakukan menggunakan Euclidean Distance, yang menghitung jarak lurus antar dua titik data dalam ruang multidimensi berdasarkan variabel latitude, longitude, depth, dan magnitude. Jarak Euclidean dirumuskan pada Persamaan (2):

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2)$$

Semakin kecil nilai jarak yang dihasilkan, semakin tinggi tingkat kemiripan antar kejadian

gempa, sehingga kedua data tersebut berpotensi berada dalam kluster yang sama.

### 3.4. Metode Penggabungan Kluster (Ward Linkage)

Dalam proses penggabungan kluster, penelitian ini menggunakan Ward Linkage, yaitu metode yang menggabungkan dua kluster dengan kenaikan variansi intra-kluster terkecil [19]. Pendekatan ini bertujuan untuk menghasilkan kluster yang lebih homogen dan stabil, khususnya pada data seismik yang memiliki variasi spasial dan numerik yang cukup tinggi.

Secara matematis, kriteria penggabungan Ward Linkage dinyatakan pada Persamaan (3):

$$\Delta(C_a, C_b) = \frac{n_a n_b}{n_a + n_b} \|\mu_a - \mu_b\|^2 \quad (3)$$

Dua kluster dengan nilai  $\Delta$  terkecil akan digabungkan terlebih dahulu. Dengan cara ini, proses pengelompokan tidak hanya mempertimbangkan kedekatan antar data, tetapi juga menjaga konsistensi dan keseragaman karakteristik dalam setiap kluster.

### 3.5. Visualisasi Dendrogram

Hasil proses hierarchical clustering divisualisasikan dalam bentuk dendrogram, yang menggambarkan tahapan penggabungan kluster secara hierarkis. Dendrogram digunakan untuk mengamati struktur alami data dan menentukan jumlah kluster yang paling representatif dengan memotong dendrogram pada tingkat jarak tertentu. Visualisasi ini membantu peneliti dalam menafsirkan hubungan antar kluster secara lebih intuitif.

### 3.6. Validasi Hasil Kluster

Evaluasi kualitas kluster dilakukan melalui validasi internal dan validasi eksternal.

Validasi internal menggunakan Silhouette Score, yang mengukur keseimbangan antara kohesi intra-kluster dan separasi antar-kluster. Nilai Silhouette untuk setiap data dihitung menggunakan Persamaan (4):

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (4)$$

Nilai Silhouette berada pada rentang -1 hingga 1, di mana nilai mendekati 1 menunjukkan struktur kluster yang semakin baik.

Selanjutnya, validasi eksternal dilakukan dengan membandingkan hasil kluster dengan data kerusakan bangunan pascagempa. Kesesuaian pola kluster dengan tingkat kerusakan bangunan digunakan sebagai indikator bahwa hasil pengelompokan memiliki relevansi empiris terhadap kondisi nyata di lapangan dan dapat digunakan sebagai dasar analisis mitigasi bencana.

### 3.7. Perangkat dan Lingkungan Analisis

Seluruh proses pengolahan dan analisis data dilakukan menggunakan bahasa pemrograman Python dengan dukungan pustaka scikit-learn untuk

normalisasi dan clustering, SciPy untuk pembentukan dendrogram, serta Matplotlib untuk visualisasi hasil.

## 4. HASIL DAN PEMBAHASAN

### 4.1. Deskripsi Data dan Seleksi Fitur

Penelitian ini menggunakan 23 data kejadian gempa bumi di wilayah selatan Malang periode April–Mei 2021. Variabel awal terdiri dari Latitude, Longitude, Depth (km), dan Magnitude sebagaimana disajikan pada Tabel 1.

Tabel 1. Variabel Input

| Latitude | Longitude | Depth (km) | Magnitude |
|----------|-----------|------------|-----------|
| -8,83    | 112,50    | 80         | 6,1       |
| -8,82    | 112,48    | 46         | 3,2       |
| -8,80    | 112,50    | 41         | 3,6       |
| -8,78    | 112,50    | 43         | 3,3       |
| ...      | ...       | ...        | ...       |
| -8,94    | 112,32    | 32         | 2,9       |
| -9,01    | 112,31    | 88         | 2,8       |
| -8,80    | 112,36    | 86         | 2,7       |
| -8,48    | 112,50    | 10         | 3,0       |

Berdasarkan kebutuhan analisis clustering, dilakukan seleksi fitur untuk menghindari redundansi spasial antara latitude dan longitude. Oleh karena itu, kedua koordinat tersebut ditransformasikan menjadi variabel Epicenter distance (km), sehingga fitur akhir yang digunakan adalah Depth, Magnitude, dan Epicenter.

Tabel 2. Pemilihan Variabel Fitur

| Depth (km) | Magnitude | Hypocenter (km) |
|------------|-----------|-----------------|
| 80         | 6,1       | 77,26           |
| 46         | 3,2       | 76,89           |
| 41         | 3,6       | 74,08           |
| 43         | 3,3       | 71,97           |
| ...        | ...       | ...             |
| 32         | 2,9       | 96,05           |
| 88         | 2,8       | 103,54          |
| 86         | 2,7       | 80,21           |
| 10         | 3,0       | 41,76           |

Statistik deskriptif terhadap ketiga variabel tersebut disajikan pada Tabel 3.

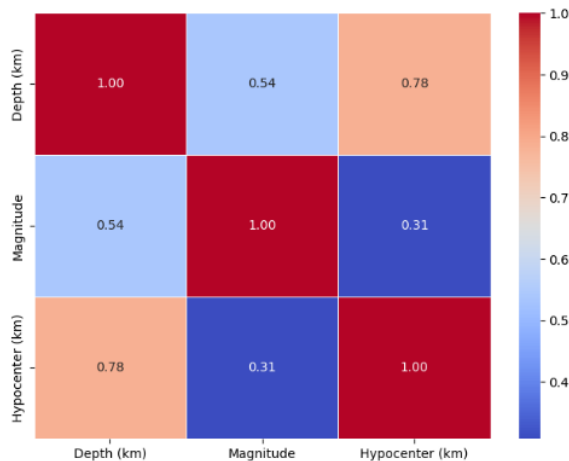
Tabel 3. Deskripsi Statistik

| Deskripsi | Magnitude | Depth (km) | Hypocenter (km) |
|-----------|-----------|------------|-----------------|
| count     | 23.00     | 23.00      | 23.00           |
| mean      | 3.50      | 52.57      | 98.94           |
| std       | 0.97      | 26.80      | 19.36           |
| min       | 2.7       | 10         | 43              |
| 25%       | 2.9       | 40         | 89              |
| 50%       | 3.2       | 45         | 96              |
| 75%       | 3.5       | 65         | 111             |
| max       | 6.1       | 110        | 136             |

Dari Tabel 3 dapat dilihat bahwa magnitudo memiliki rata-rata 3,50 dengan nilai maksimum 6,1, sedangkan kedalaman memiliki rata-rata 52,57 km dengan rentang 10–110 km. Variabel epicenter

memiliki rata-rata 98,94 km dengan deviasi standar 19,36 km. Variasi nilai yang cukup besar pada kedalaman dan epicenter menunjukkan adanya heterogenitas spasial yang berpotensi membentuk kluster alami.

Hubungan antar variabel dianalisis menggunakan matriks korelasi sebagaimana ditampilkan pada Gambar 1.



Gambar 1. Korelasi matriks fitur gempa bumi

Dari Gambar 1 terlihat bahwa tidak terdapat korelasi sangat tinggi ( $>0,8$ ) antar variabel, sehingga tidak terjadi multikolinearitas yang dapat mendistorsi perhitungan jarak dalam clustering. Dengan demikian, ketiga fitur dinyatakan layak digunakan dalam proses Agglomerative Clustering.

#### 4.2. Evaluasi Jumlah Kluster

Proses clustering dilakukan menggunakan metode Agglomerative dengan variasi jumlah kluster 2–6 serta tiga pendekatan linkage (ward, complete, dan average). Evaluasi dilakukan menggunakan Silhouette Score, Davies–Bouldin Index (DBI), dan Calinski–Harabasz Index (CHI). Hasil lengkap ditunjukkan pada Tabel 4.

Tabel 4. Hasil Tuning Data dan Perbandingan Indeks Validasi

| N_clusters | Linkage  | Silhouette Score | DBI Score | CHI Score |
|------------|----------|------------------|-----------|-----------|
| 2          | ward     | 0,63             | 0,53      | 39,98     |
| 2          | complete | 0,63             | 0,53      | 39,98     |
| 2          | average  | 0,63             | 0,53      | 39,98     |
| 3          | ward     | 0,62             | 0,41      | 35,75     |
| 3          | complete | 0,62             | 0,41      | 35,75     |
| 3          | average  | 0,62             | 0,41      | 35,75     |
| 4          | ward     | 0,49             | 0,49      | 40,72     |
| 4          | complete | 0,49             | 0,49      | 40,72     |
| 4          | average  | 0,49             | 0,37      | 30,57     |
| 5          | ward     | 0,44             | 0,46      | 42,88     |
| 5          | complete | 0,44             | 0,46      | 42,88     |
| 5          | average  | 0,41             | 0,34      | 29,59     |
| 6          | ward     | 0,44             | 0,49      | 46,48     |
| 6          | complete | 0,43             | 0,43      | 44,93     |
| 6          | average  | 0,41             | 0,38      | 28,93     |

Berdasarkan Tabel 4 diperoleh temuan sebagai berikut:

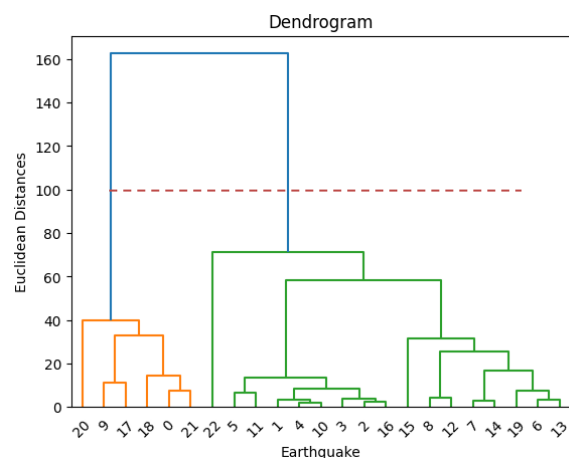
- Silhouette Score tertinggi diperoleh pada jumlah kluster 2 dengan nilai 0,63.
- DBI terendah diperoleh pada jumlah kluster 3 (0,41).
- CHI tertinggi diperoleh pada jumlah kluster 6 (46,48).

Namun demikian, pemilihan jumlah kluster tidak dapat hanya mengacu pada satu metrik. Silhouette Score merupakan indikator paling seimbang karena mempertimbangkan kohesi internal dan separasi antar kluster secara simultan. Nilai 0,63 menunjukkan struktur kluster kuat (strong structure).

Walaupun DBI minimum terjadi pada tiga kluster, nilai DBI pada dua kluster (0,53) masih berada pada kategori baik ( $<0,6$ ). Sementara itu, peningkatan CHI pada kluster lebih besar cenderung disebabkan oleh fragmentasi data (over-partitioning), yang justru menurunkan nilai Silhouette secara signifikan. Dengan mempertimbangkan keseimbangan ketiga indeks, jumlah kluster optimal dalam penelitian ini adalah dua kluster dengan Ward linkage.

#### 4.3. Analisis Struktur Herarki

Struktur hierarki hasil Agglomerative Clustering divisualisasikan menggunakan dendrogram pada Gambar 2.



Gambar 2. Hasil dendrogram Agglomerative Clustering

Dari Gambar 2 terlihat adanya pemisahan alami menjadi dua kelompok utama sebelum penggabungan akhir. Tinggi percabangan yang signifikan menunjukkan bahwa pemotongan pada dua kluster merupakan struktur natural dalam data. Hal ini mengindikasikan adanya dua pola aktivitas seismik yang berbeda secara karakteristik.

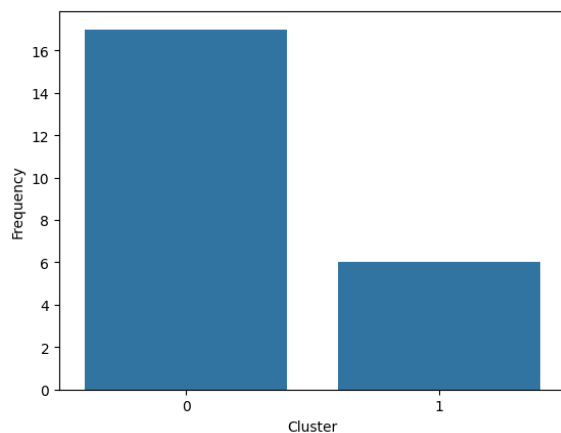
#### 4.4. Karakteristik Kluster

Hasil pelabelan data terhadap dua kluster disajikan pada Tabel 5.

Tabel 5. Hasil Label Cluster

| Depth (km) | Magnitude | Hypocenter (km) | Cluster |
|------------|-----------|-----------------|---------|
| 80         | 6,1       | 77,26           | 1       |
| 46         | 3,2       | 76,89           | 0       |
| 41         | 3,6       | 74,08           | 0       |
| 43         | 3,3       | 71,97           | 0       |
| ...        | ...       | ...             | ...     |
| 32         | 2,9       | 96,05           | 0       |
| 88         | 2,8       | 103,54          | 1       |
| 86         | 2,7       | 80,21           | 1       |
| 10         | 3,0       | 41,76           | 0       |

Distribusi frekuensi masing-masing klaster ditampilkan pada Gambar 3.

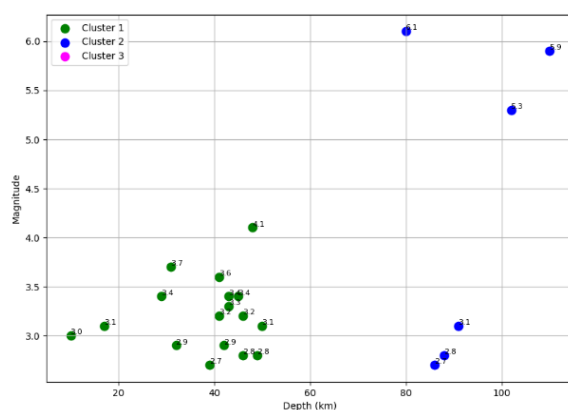


Gambar 3. Distribusi frekuensi data pada setiap klaster

Dari hasil tersebut diperoleh karakteristik sebagai berikut:

- Klaster 0: Magnitudo relatif rendah hingga menengah (2,7–3,6), Kedalaman dangkal hingga menengah, Epicenter relatif lebih dekat, Jumlah anggota lebih dominan.
- Klaster 1: Magnitudo lebih tinggi (hingga 6,1), Kedalaman lebih dalam ( $\geq 80$  km), Epicenter relatif lebih jauh, Mewakili kejadian ekstrem.

Visualisasi spasial hasil clustering ditampilkan pada Gambar 4.



Gambar 4. Visualisasi hasil clustering (Agglomerative)

Dari Gambar 4 terlihat bahwa Klaster 1 membentuk kelompok terpisah yang merepresentasikan gempa dengan intensitas dan kedalaman lebih besar, sedangkan Klaster 0 merepresentasikan gempa susulan dengan magnitudo lebih kecil.

#### 4.5. Interpretasi dan Implikasi

Hasil clustering menunjukkan bahwa kombinasi Magnitude, Depth, dan Epicenter distance mampu membentuk dua pola utama aktivitas gempa, yaitu:

- Kelompok gempa dangkal dengan magnitudo rendah (potensi dampak lokal)
- Kelompok gempa dalam dengan magnitudo tinggi (potensi dampak luas)

Secara statistik, model memenuhi kriteria validasi Silhouette Score tinggi (0,6287), DBI rendah (0,5275), CHI stabil (39,9820).

Hal ini menunjukkan bahwa pengelompokan yang dihasilkan signifikan dan memiliki struktur yang jelas. Dengan demikian, hipotesis penelitian yang menyatakan bahwa Agglomerative Clustering mampu menghasilkan pengelompokan tingkat kerusakan bangunan yang bermakna dapat diterima.

Secara praktis, hasil ini dapat digunakan sebagai dasar pengelompokan wilayah prioritas mitigasi pascagempa, di mana klaster dengan magnitudo dan kedalaman tinggi perlu mendapat perhatian lebih dalam perencanaan tanggap darurat dan rehabilitasi infrastruktur.

## 5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk menerapkan teknik Agglomerative Clustering dalam mengelompokkan tingkat kerusakan bangunan pascagempa berdasarkan parameter seismik. Variabel yang digunakan meliputi Magnitude, Depth (km), dan Epicenter (km) yang telah melalui tahap preprocessing dan standardisasi sebelum dilakukan proses clustering menggunakan metode Ward Linkage dan Euclidean Distance.

Berdasarkan hasil evaluasi model, konfigurasi terbaik diperoleh pada jumlah klaster  $k = 2$  dengan nilai Silhouette Score sebesar 0,6287. Nilai tersebut menunjukkan kualitas pengelompokan yang baik, dengan kohesi internal klaster yang kuat dan separasi antar klaster yang jelas. Hasil ini mengindikasikan bahwa struktur data memiliki pola alami yang dapat diidentifikasi secara hierarkis menggunakan metode agglomerative.

Dua klaster yang terbentuk secara umum merepresentasikan perbedaan karakteristik kejadian gempa yang berpotensi memengaruhi tingkat kerusakan bangunan, yaitu kelompok gempa dengan parameter seismik lebih signifikan dan kelompok dengan parameter relatif lebih rendah. Validasi eksternal menggunakan data kerusakan bangunan menunjukkan konsistensi antara hasil klaster dan kondisi riil di lapangan.

Dengan demikian, teknik Agglomerative Clustering terbukti efektif dalam mengelompokkan tingkat kerusakan bangunan berbasis data gempa bumi, serta mampu mendukung analisis mitigasi dan pengambilan keputusan kebencanaan secara lebih sistematis. Hasil penelitian ini mendukung metode yang digunakan untuk menghasilkan struktur klaster yang signifikan dan representatif.

Untuk pengembangan penelitian selanjutnya, disarankan agar:

- a. Menambahkan variabel geoteknik dan karakteristik struktur bangunan guna meningkatkan sensitivitas model terhadap faktor lokal.
- b. Membandingkan metode Agglomerative dengan algoritma clustering lain seperti DBSCAN atau Gaussian Mixture Model untuk menguji stabilitas hasil.
- c. Mengembangkan model ke dalam sistem pendukung keputusan berbasis spasial (GIS) guna membantu BMKG dan BPBD dalam perencanaan mitigasi dan tanggap darurat.

#### DAFTAR PUSTAKA

- [1] H. Lbath, A. Petersen, W. Meiring, and S. Achard, "Clustering-based inter-regional correlation estimation," *Comput. Stat. Data Anal.*, vol. 191, p. 107876, Mar. 2024, doi: 10.1016/j.csda.2023.107876.
- [2] A. R. Rio and B. M. Siahaan, "Agglomerative Clustering of 2022 Earthquakes in North Sulawesi, Indonesia," *Buana Inf. Technol. Comput. Sci. (BIT CS)*, vol. 4, no. 2, pp. 76–84, Jul. 2023, doi: 10.36805/bit-cs.v4i2.5361.
- [3] H. Pratiwi, S. S. Handajani, I. Susanto, S. Sangadji, R. Meilawati, and I. S. Khairunnisa, "Hierarchical clustering algorithm for analyzing risk of earthquake on sumatra island," in *International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2021*, IEEE, Oct. 2021, pp. 1–6. doi: 10.1109/ICECCME52200.2021.9590890.
- [4] M. B. Prayogo, I. Slamet, and Respatiwan, "Implementation of agglomerative hierarchical clustering algorithm for earthquake areas in Sulawesi Island," 2022, p. 080006. doi: 10.1063/5.0075119.
- [5] W. Wiyono, "Study of Earthquake Damage Potential Based on B-Value Variation in the Special Region of Yogyakarta," *Sunan Kalijaga J. Phys.*, vol. 6, no. 2, pp. 54–63, Nov. 2024, doi: 10.14421/physics.v6i2.5156.
- [6] M. Prasetyo, H. Sulistiani, O. Y. Inonu, K. Magda, and B. Santosa, "Klasifikasi Tingkat Risiko Gempa di Indonesia Menggunakan Pola Spasial dan Temporal Berbasis Decision Tree," *Bull. Comput. Sci. Res.*, vol. 5, no. 5, pp. 1059–1066, Aug. 2025, doi: 10.47065/bulletincsr.v5i5.624.
- [7] R. Steinmann, L. Seydoux, É. Beaucé, and M. Campillo, "Hierarchical Exploration of Continuous Seismograms With Unsupervised Learning," *J. Geophys. Res. Solid Earth*, vol. 127, no. 1, Jan. 2022, doi: 10.1029/2021JB022455.
- [8] C. Swastikawati, M. A. Pamoengkas, A. C. Wahyudi, and K. Kusriani, "Penerapan Data Sains Untuk Klasifikasi Wilayah Yang Terdampak Kerusakan Gempa Bumi Dengan Metode C.45," *J. Econ. Manag. Account. Technol.*, vol. 7, no. 2, pp. 390–402, Aug. 2024, doi: 10.32500/jematech.v7i2.7679.
- [9] W. Ma, Y. Qin, Z. Xu, and G. Chen, "Prediction model for generation trend of axial deformation of saturated coral sand under cyclic loading," *Soil Dyn. Earthq. Eng.*, vol. 184, p. 108861, Sep. 2024, doi: 10.1016/j.soildyn.2024.108861.
- [10] E. K. Tokuda, C. H. Comin, and L. da F. Costa, "Revisiting agglomerative clustering," *Phys. A Stat. Mech. its Appl.*, vol. 585, p. 126433, Jan. 2022, doi: 10.1016/j.physa.2021.126433.
- [11] A. Priyono et al., "Seismic Attenuation Tomography From 2018 Lombok Earthquakes, Indonesia," *Front. Earth Sci.*, vol. 9, Mar. 2021, doi: 10.3389/feart.2021.639692.
- [12] A. T. W. Almaiz et al., "Characterization of Structural Building Damage in Post-Disaster Using GLCM-PCA Analysis Integration," *IEEE Access*, vol. 12, pp. 146190–146201, 2024, doi: 10.1109/ACCESS.2024.3469637.
- [13] K. Brudko, R. Okhrimchuk, and V. Demidov, "Automatic recognition and damage evaluation of building infrastructure in seismic active zones using machine learning," in *16th International Conference Monitoring of Geological Processes and Ecological Condition of the Environment, Monitoring 2022*, European Association of Geoscientists & Engineers, 2022, pp. 1–5. doi: 10.3997/2214-4609.2022580199.
- [14] X. Chen, L. Zhang, X. Lin, K. A. Skalomenos, and Z. Chen, "A clustering-based analysis method for simulating seismic damage of buildings in large cities," *Eng. Struct.*, vol. 307, p. 117860, May 2024, doi: 10.1016/j.engstruct.2024.117860.
- [15] Z. Fahrezi, R. Misliniyati, K. Amri, L. Z. Mase, and Hardiansyah, "Prediction of Megathrust Impact on Seismic Response in Selebar District of Bengkulu City," *J. PenSil*, vol. 14, no. 1, pp. 50–64, Jan. 2025, doi: 10.21009/jpensil.v14i1.50133.
- [16] Z. Fahrezi, R. Misliniyati, K. Amri, L. Z. Mase, and Hardiansyah, "Earthquake Hazard Evaluation Study on the Central Area of Selebar District, Bengkulu City, based on Seismic Response Analysis," *Ris. Geol. dan Pertamb.*, vol. 35, no. 1, pp. 11–24, Jul. 2025, doi: 10.55981/risetgeotam.2025.1331.
- [17] R. Abid and M. Qaiser, "Pollen morphology of

- Lactuca L. (s. lat.) (Cichorieae: Asteraceae) from Hindukush, Western Himalayan and Karakorum ranges and its taxonomic significance,” *Palynology*, vol. 47, no. 1, Jan. 2023, doi: 10.1080/01916122.2022.2089927.
- [18] A. Schaefer, T. P. Scolaro, and E. Ghisi, “Cluster analysis applied to obtaining reference models for building thermal performance studies,” *J. Build. Eng.*, vol. 89, p. 109273, Jul. 2024, doi: 10.1016/j.jobbe.2024.109273.
- [19] Doli *et al.*, “Diversity and Phylogenetic Relationships of *Cylindrospermum* Species from the Aquatic and Semi-Aquatic Habitats of Meerut, Western Uttar Pradesh, Bharat,” *Egypt. J. Aquat. Biol. Fish.*, vol. 29, no. 5, pp. 1575–1601, Sep. 2025, doi: 10.21608/ejabf.2025.362549.5424.