

# Pengelompokan Titik Kejadian Gempabumi di Wilayah Asia Tenggara Menggunakan Agglomerative Hierarchical Clustering

## *Clustering Earthquake Event Points in the Southeast Asia Region using Agglomerative Hierarchical Clustering*

Adi Arifin\*

Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta  
Jalan Padjajaran, Kabupaten Sleman, Indonesia  
\*e-mail: <mailto:adidongs@students.amikom.ac.id>

(received: 19 Juli 2023, revised: 23 Juli 2023, accepted: 29 Juli 2023)

### Abstrak

Gempabumi adalah suatu bencana alam yang tidak dapat dihindari dan paling berbahaya karena terjadi seketika dan tidak terkendali. Gempabumi juga sering terjadi di wilayah Asia Tenggara yang menyebabkan kerusakan infrastruktur, kehilangan nyawa, dan gangguan ekonomi. Oleh karena itu, perlu dilakukan upaya mitigasi risiko gempabumi. Salah satunya dengan cara mengelompokkan data gempabumi di Wilayah Asia Tenggara. Data yang digunakan dalam penelitian ini adalah data gempabumi yang diperoleh dari *The Seismological Facility for the Advancement of Geoscience (SAGE)* adalah fasilitas yang didanai NSF yang dioperasikan oleh *Incorporated Research Institutions for Seismology (IRIS)*. Penelitian ini menggunakan *Agglomerative Hierarchical Clustering (AHC)* dalam mengelompokkan data ke dalam beberapa *cluster*. Metode yang digunakan untuk menguji validitas *cluster* yang terbentuk dengan menggunakan *Silhouette Score Index (SSI)*, *Davies Doublin Index (DBI)* dan *Calinski Harabasz Index (CHI)*. Pada penelitian ini terdapat 2 proses pengelompokan yang menghasilkan 3 *cluster* untuk pengelompokan pertama yang bertujuan untuk membuat atribut baru yaitu, Area dan 3 *cluster* untuk pengelompokan kedua yang bertujuan untuk mengidentifikasi jenis gempa yang ada di wilayah Asia Tenggara. 3 *cluster* yang terbentuk ini memiliki nilai SSI, DBI, dan CHI sebesar 0,434353, 0,887791 dan 3769.030146 yang artinya AHC mampu mengelompokkan data gempabumi dengan cukup baik. Hasil penelitian ini menjadi acuan dalam mengembangkan penelitian lain seperti prediksi gempabumi yang akan datang dan dapat menjadi acuan untuk menyusun mitigasi bencana yang akan datang agar lebih waspada ke depannya

**Kata kunci:** Gempabumi, Asia Tenggara, *Agglomerative Hierarchical Clustering*, pengelompokan

### Abstract

*Earthquake is an inevitable and highly dangerous natural disaster due to its sudden and uncontrollable occurrence. Earthquakes frequently happen in the Southeast Asia region, resulting in infrastructure damage, loss of life, and economic disruption. Therefore, efforts should be made to mitigate earthquake risks, including the classification of earthquake data in the Southeast Asia region. The data used in this study were obtained from The Seismological Facility for the Advancement of Geoscience (SAGE), a facility funded by NSF and operated by the Incorporated Research Institutions for Seismology (IRIS). The study employed Agglomerative Hierarchical Clustering (AHC) to group the data into multiple clusters. The validity of the formed clusters was assessed using Silhouette Score Index (SSI), Davies Doublin Index (DBI), and Calinski Harabasz Index (CHI). The study involved two clustering processes, resulting in three clusters for the first clustering process aimed at creating new attributes, namely Area, and three clusters for the second clustering process aimed at identifying the types of earthquakes in the Southeast Asia region. These three formed clusters had SSI, DBI, and CHI values of 0.434353, 0.887791, and 3769.030146, respectively, indicating that AHC successfully classified the earthquake data. The findings of this research serve as a reference for further studies such as earthquake prediction and contribute to disaster mitigation strategies to enhance preparedness for future events.*

<http://sistemasi.ftik.unisi.ac.id>

**Keywords:** *Earthquake, Southeast Asia, Agglomerative Hierarchical Clustering, clustering.*

## 1 Pendahuluan

Gempabumi merupakan salah satu fenomena bencana alam yang tidak dapat di cegah maupun di hindari[1], Gempabumi juga merupakan bencana alam yang paling rusak karena menyebabkan banyak korban dan merupakan bencana alam yang paling berbahaya yang terjadi secara tiba-tiba dan tidak terkendali[2] Gempabumi menciptakan gelombang seismik, yaitu gelombang yang bergerak melalui bumi akibat deformasi struktural, tekanan atau tegangan sifat elastis kerak bumi [3] Tidak terkecuali Wilayah Asia Tenggara, wilayah Asia Tenggara terletak diantara beberapa lapisan lempeng bumi diantaranya India, Australia, dan Filipina yang merupakan wilayah “*Ring of Fire*” dimana sering terjadinya gempa, erupsi vulkanik, dan tsunami[4].

Menurut data yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), kejadian gempa terjadi setiap hari di wilayah Indonesia, tetapi tidak semua orang bisa merasakan gempa dikarenakan berkekuatan kecil dan titik pusat gempa terletak berjauhan dengan kediaman penduduk. Aktivitas seismik di wilayah Indonesia sering sekali terjadi, dengan rata-rata 400 gempa per bulan[5].

Di Indonesia pada hari Minggu, tanggal 5 Agustus 2018, pukul 18:45:35 WIB, terjadi gempabumi dengan magnitudo 7,0 dan kedalaman 15 km di koordinat 8,37° LS dan 116,48° BT. Gempa ini merupakan gempa susulan dari gempa sebelumnya yang terjadi pada tanggal 29 Juli 2018 dengan kekuatan M6,4 dan kedalaman 10 km. Dampak dari gempa tersebut sangat serius, menyebabkan kerusakan berat terutama di Kabupaten Lombok Utara dan Lombok Timur. Badan Nasional Penanggulangan Bencana (BNPB) melaporkan bahwa jumlah korban meninggal dunia akibat gempa ini mencapai 105 jiwa. Informasi ini penting untuk memahami dan mengkaji dampak gempabumi pada wilayah tersebut, serta memberikan wawasan untuk perencanaan mitigasi bencana di masa depan[6]. Pada hari Jumat, 28 September 2018, terjadi gempabumi tektonik di Kabupaten Donggala, Sulawesi Tengah, dengan magnitudo 7.7. Gempa tersebut terjadi pada pukul 17.02.44 WIB dan berlokasi di 0.18 LS dan 119.85 BT, sekitar 26 km di sebelah Utara Donggala, Sulawesi Tengah. Gempa ini memiliki kedalaman 10 km. Berdasarkan data sementara dari BPBD Kabupaten Donggala mencatat dampak dari gempa ini, yaitu 1 orang meninggal dunia, 10 orang luka-luka, dan puluhan rumah mengalami kerusakan. Para korban terluka dan rumah-rumah rusak akibat tertimpa bangunan yang roboh.[7]. Di Laos, Pada hari Kamis, 21 November 2019, telah terjadi serangkaian gempabumi di Laos berkekuatan 5.7 hingga 6.1 SR yang dapat dirasakan hingga ke wilayah perbatasan Laos dengan Thailand, yaitu di wilayah Propinsi Sainyabuli, sebelah barat laut Laos[8]. Pada Senin, 22 April 2019, terjadi gempabumi di Filipina dengan kekuatan 6,1 Skala Richter. Gempa ini berpusat di 81 km Barat Laut Filipina (Pampanga-Zambales) dengan kedalaman 71,2 km. Guncangan gempa dirasakan hingga Wilayah Manila dan sekitarnya. Akibat gempa ini, 8 orang meninggal dunia dan 50 orang lainnya masih terjebak di reruntuhan supermarket dan gedung bertingkat di Wilayah Pampanga. Gempabumi ini menyebabkan kerusakan serius di Bandara Internasional Clark-Filipina, yang sebelumnya pernah digunakan sebagai tempat mendarat pemimpin dunia saat menghadiri KTT ASEAN pada November 2017.[9].

Salah satu instansi pemerintah yang bertanggung jawab mengelola data dan informasi cuaca, iklim, dan bencana di Indonesia adalah Badan Meteorologi, Klimatologi, dan Geofisika (BMKG)[10]. Adapun Kemampuan komunitas geosains untuk merespons dengan cepat setelah peristiwa berbahaya atau tanda-tanda prekursor peristiwa ini, memberikan data penting untuk memahami proses fisik yang bertanggung jawab atas peristiwa destruktif ini. *The Seismological Facility for the Advancement of Geoscience (SAGE)* adalah fasilitas yang didanai NSF yang dioperasikan oleh *Incorporated Research Institutions for Seismology (IRIS)*[11]. SAGE telah membuat data gempabumi secara terbuka dan bisa dikelola ke dalam bentuk *datasheet*.

Penelitian tentang Pengelompokan Titik Kejadian Gempabumi di Wilayah Asia Tenggara menjadi penting untuk memperoleh informasi yang lebih akurat tentang sebaran titik kejadian gempabumi di kawasan tersebut. Salah satu metode yang dapat digunakan untuk pengelompokan tersebut adalah Algoritma *Agglomerative Hierarchical Clustering*. Pengelompokan jangkauan diimplementasikan menggunakan kelompok dengan kelas rendah, sedang dan tinggi. AHC dengan menggunakan *bottom-up*, dimulai dari semua data sebagai sebuah kelompok dan kemudian secara rekursif mencari kelompok terdekat sebagai pasangan yang kemudian akan digabungkan menjadi kelompok yang lebih luas. Proses diulang sampai jumlah kelompok 1, sehingga akan tampak bergerak ke atas membentuk hirarki.

Keunggulan dari algoritma AHC adalah tidak perlu menentukan jumlah kluster yang diinginkan [12]. Metode ini memiliki tahapan dimulai dari pengumpulan data, pembagian kelompok kelas, perhitungan jarak euclidean, pembentukan dendrogram, dan evaluasi yang dilakukan menggunakan metode *silhouette score* (SS), *davies bouldin index* (DBI), dan *calinski harabasz index* (CHI). Data sampel yang digunakan pada penelitian ini adalah *Year, Quarter, Lat, Lon, Depth, Mag, Periode\_hari, Kategori gempa, Musim, dan Kedalaman gempa* untuk menentukan pengelompokan wilayah gempabumi di Asia Tenggara dan juga merupakan hasil dari pengolahan sebelum dilakukan proses *clustering*.

## 2 Tinjauan Literatur

Penelitian terhadap pengelompokan gempabumi telah banyak dilakukan oleh peneliti-peneliti terdahulu. Pada bagian ini akan disajikan beberapa literatur yang akan di jadikan sebagai acuan dan juga ilmu tambahan dalam penelitian ini.

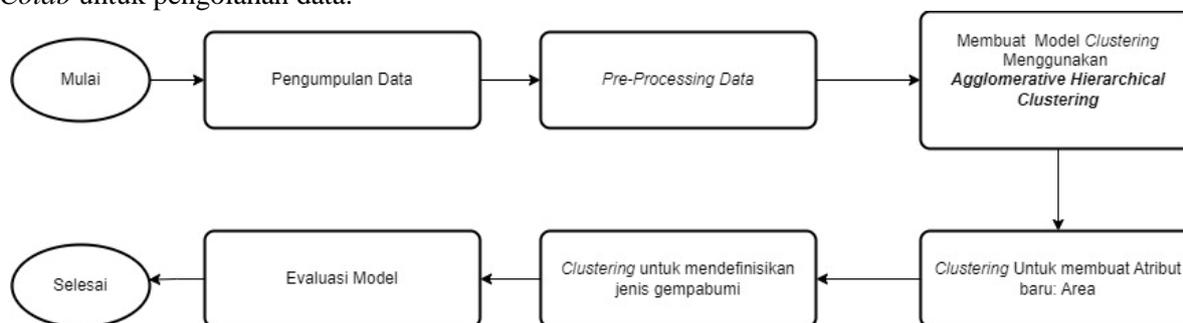
Penelitian terkait dengan pengelompokan dengan batasan mengelompokkan gempabumi ini telah dikaji oleh peneliti-peneliti sebelumnya. Salah satunya adalah penelitian yang dilakukan oleh [13] yang menggunakan Metode *Fuzzy C-Means* untuk mengelompokkan kejadian gempa berdasarkan dua parameter yaitu kekuatan gempa dan dampak akibat gempa. Dalam penelitian disimpulkan bahwa pengelompokan berdasarkan parameter kekuatan dengan parameter dampak gempabumi kurang memiliki keterikatan.

Halik M. dan Septiana L [14], juga menyajikan penelitian untuk prediksi daerah rawan bencana alam di Jawa Barat menggunakan algoritma *K-Means Clustering*. Penelitian ini menggunakan dataset yang diperoleh dari dokumen dan informasi yang di hasilkan oleh Dinas Pemberdayaan Masyarakat dan Desa dan juga Dinas Lingkungan Hidup Provinsi Jawa Barat, yang disusun dalam Data Terbuka Jawa Barat (<https://opendata.jabarprov.go.id>) sebagai portal data terbuka dan menyediakan data yang akurat dari semua perangkat daerah di seluruh Jawa Barat. Data yang diambil oleh [14], data bencana alam berdasarkan kabupaten/kota di Jawa Barat pada rentang tahun 2012 samai 2021. Dari hasil penelitian [14], tingkat kerawanan bencana alam di tiap daerah memiliki persentase di atas 50%. Bencana alam Tanah Longsor tercatat terjadi di daerah Bogor, Sukabumi, Kuningan, dan Kota Bogor. Sedangkan bencana alam Gempa Bumi terjadi di daerah Sukabumi, Garut, Tasikmalaya, dan Kota Sukabumi. Selanjutnya, bencana alam Banjir tercatat terjadi di daerah Kabupaten Bogor dan Bandung. Di sisi lain, bencana alam Angin Puting Beliung terjadi di daerah Kabupaten Bogor. Dan terakhir, bencana alam Kebakaran terjadi di daerah Sukabumi, Garut, Ciamis, Sumedang, dan Kota Bandung. Selain itu, hasil penelitian juga menghasilkan pembagian menjadi 3 cluster yaitu tingkat rendah (C0), tingkat sedang (C1), dan tinggi/rawan (C2) terhadap tingkat kerawanan bencana alam di wilayah-wilayah tersebut. Pengelompokan digunakan untuk mengetahui kelompok titik kejadian gempabumi juga disajikan oleh [15], yang melakukan penelitian tentang pemetaan kelompok gempa bumi di Mentawai dengan menggunakan *k-medoids clustering*. Hasil dari penelitian [15], menghasilkan 3 *cluster* dan juga akurasi pengelompokan nya, untuk *cluster* rendah sebesar 96,09%, *cluster* sedang sebesar 92,96% dan *cluster* tinggi sebesar 93.75%. sedangkan, persentase untuk keseluruhan sistem dalam mengelompokkan ketiga *cluster* adalah 94,867%. Penelitian lainnya yaitu oleh [16], menggunakan pendekatan penelitian terapan dengan menggunakan data sekunder dari *United States Geological Survey* (USGS). Data yang digunakan dalam penelitian ini berasal dari sekitar Provinsi Sumatera Barat dan mencakup rentang periode dari Desember 1970 hingga September 2021. Dalam penelitian ini, hanya tiga atribut data yang digunakan, yaitu *latitude, longitude, dan magnitude*. Hasil dari penelitian [16] menunjukkan bahwa kelompok paling optimal yang ditemukan adalah dengan menggunakan parameter  $K=2$ , dengan nilai *Silhouette Coefficient* (SS) tertinggi sebesar 0,448072. Kelompok ini menghasilkan titik pusat gempa di 99,29°E dan -1,05°S, dengan magnitude gempa sebesar 5,38 SR. Sebagian besar aktivitas gempa terjadi di wilayah Sumatera Barat, termasuk Kepulauan Mentawai. Selain itu, kejadian gempa juga terlihat di wilayah Mandailing Natal serta Kepulauan Nias dalam kelompok pertama. Gempa dengan magnitudo tertinggi yang pernah terjadi dalam kelompok pertama adalah gempa pada tanggal 30 September 2009 di lepas pantai Pariaman, dengan episentrum gempa berada pada 99,87°BT dan -0,72°LS, dengan magnitudo 7,6 SR. Sementara itu, pada kelompok kedua, pusat gempa terletak di 101,15°BT dan -3,44°S, dengan magnitudo 5,36 SR. Sebagian besar aktivitas seismik dalam kelompok kedua terjadi di wilayah Bengkulu Utara, Muko Muko, dan Pulau Pagai Selatan. Selain itu, juga terdapat aktivitas seismik di wilayah Jambi dan Pantai Selatan. Gempa dengan magnitudo seismik tertinggi

dalam kelompok kedua adalah gempa Bengkulu pada tanggal 12 September 2007, dengan episentrum gempa berada pada 100,84°E dan -2,63°S, dengan magnitudo 7,9 SR. Dalam penelitian ini, terdapat 372 aktivitas seismik yang terjadi dalam kelompok pertama dan 344 aktivitas seismik yang terjadi dalam kelompok kedua.

### 3 Metode Penelitian

Metode penelitian yang disajikan berupa alur tahapan yang akan menjadi acuan untuk menentukan *cluster* titik kejadian gempabumi di Asia Tenggara. Alur tahapan yang digunakan dalam penelitian ini disajikan pada Gambar 1. Penelitian menggunakan pemrograman *Python 3.10.11* dengan alat *Google Colab* untuk pengolahan data.



Gambar 1. Alur Tahapan Metode Penelitian

#### 3.1 Pengumpulan Data

Pada tahap pengumpulan data yang di gunakan adalah data sekunder. Data didapatkan dari *website* The Seismological Facility for the Advancement of Geoscience (SAGE) adalah fasilitas yang didanai NSF yang dioperasikan oleh Incorporated Research Institutions for Seismology (IRIS). Berupa data gempa bumi pada Asia Tenggara dan sekitarnya dengan periode tahun 2021 sampai 2023 sebanyak 5000 baris data.

#### 3.2 Pre-Processing Data

Pada tahap *pre-processing* data di lakukan pembentukan atribut data, pembuatan atribut data baru, seleksi atribut data, pengkodean label dan Standarisasi data. Tahap pemilihan data digunakan untuk memilih data yang akan dipergunakan dalam proses data *mining*[17]. Dalam penelitian ini tahapan pemilihan data dilakukan dengan pemilihan atribut, dari jumlah atribut 10 atribut di pilih menjadi 9 atribut (dengan simbol Y) awal yang digunakan untuk keperluan pengolahan data berikut yaitu pembagian data. Pemilihan awal atribut disajikan dalam Tabel 1. Data yang siap untuk proses pengolahan data memiliki ukuran dimensi 5000 baris  $\times$  9 kolom.

Tabel 1. Pemilihan Awal Atribut

Atribut	Penggunaan
<i>Year</i>	Y
<i>Day</i>	Y
<i>Timer</i>	Y
<i>Lat</i>	Y
<i>Lon</i>	Y
<i>Depth</i>	Y
<i>Mag</i>	Y
<i>Region</i>	T
<i>Timestamp</i>	

Tahapan berikutnya yaitu, tahapan pembentukan atribut baru dari atribut *Month* menjadi atribut *Quarter* (Kuartal). Kuartal menurut Kamus Besar Bahasa Indonesia (KBBI) adalah suku tahun (waktu seperempat tahun, yaitu tiga bulan) yang artinya dalam setahun di bagi 4. Kuartal 1 Januari-Maret, Kuartal 2 April-Juni, Kuartal 3 Juli-September dan Kuartal 4 Oktober-Desember. Pada atribut *Time* dipecah menjadi 3 atribut yaitu *Jam*, *Menit* dan *Periode\_hari* (Subuh, Pagi, Siang, Sore, Malam). Pada atribut *Mag*, dibuat atribut baru yang mengacu pada situs Britannica (<https://www.britannica.com/science/earthquake-geology#ref59550>) dimana, besarnya gempa akan dikategorikan berdasarkan besarnya gempa. Atribut baru ini di beri nama *Kategori\_Gempa* (*Micro*,

*Manor, Light, Moderate, Strong, Major, Great*). Kemudian membuat atribut baru *Musim*. Mengacu pada *website* Wonderful Indonesia (<https://www.indonesia.travel/gb/en/general-information/climate>) dan juga detikedu (<https://www.detik.com/edu/detikpedia/d-5737381/negara-negara-di-asean-secara-umum-beriklim-apa>), berdasarkan letak wilayah dan letak astronomisnya Indonesia dan Asia Tenggara memiliki musim yang sama dan tidak jauh berbeda, sehingga atribut *Musim* hanya memiliki 2 value (Kemarau dan Hujan). Mengacu pada *website* USGS (<https://www.usgs.gov/programs/earthquake-hazards/determining-depth-earthquake#:~:text=Shallow%20earthquakes%20are%20between%200,earthquakes%20deeper%20than%2070%20km.>) kedalaman gempa bumi dibagi menjadi 3 yaitu dangkal, menengah dan dalam. Hal ini lah yang menjadi dasar untuk membuat atribut baru berikutnya yaitu *Kedalaman\_gempa* berdasarkan atribut *Depth*.

Tahapan seleksi atribut data, pada tahap dilakukan seleksi atribut data yang sesuai dengan kebutuhan penelitian yang bertujuan untuk mengurangi redundansi data. Atribut yang telah memenuhi seleksi atribut data antara lain *Year, Quarter, Lat, Lon, Depth, Mag, Periode\_hari, Kategori\_gempa, Musim, Kedalaman\_gempa*. Data yang diseleksi ini memiliki rentang waktu periode tahun 2020-2023.

Setelah tahapan seleksi atribut data sesuai dengan kebutuhan, dilakukan standarisasi data dengan menggunakan teknik *StanderScaler*. Teknik ini menstandarisasi atribut dengan mengurangi rata-rata kemudian menskalakan ke varians unit [18]. *StanderScaler* menghasilkan distribusi yang mana standar deviasi adalah 1. Varians sama dengan 1 juga, karena varians = standar deviasi kuadrat. Dan  $1 \text{ kuadrat} = 1$ . *StandardScaler* membuat rata-rata distribusi mendekati 0 [18]. Untuk *clustering* yang mendefinisikan jenis gempa bumi di perlukan pengkodean label karena, menggunakan beberapa atribut bertipe data kategori. Menurut [19], Pengkodean label atau *Label Encoding* adalah proses konversi atribut bertipe data kategori ke dalam bentuk numerik/ proses pengkodean label ini mempermudah mesin untuk mengolah data.

### 3.3 Model Clustering Menggunakan Agglomerative Hierarchical Clustering (AHC)

AHC adalah metode pengelompokan hierarkis *bottom-up* yang menggabungkan  $n$  cluster menjadi satu cluster [20]. Alur dari proses AHC meliputi tahapan berikut [20], [21]:

- 1) Dimulai dengan  $n$  cluster, setiap cluster terdapat satu entitas dan matriks simetris  $N \times N$  jarak atau kemiripan
- 2) Menemukan matriks jarak tiap pasangan cluster terdekat. Misal: jarak antara cluster A dan B dilambangkan  $d_{ab}$
- 3) Cluster A dan B digabungkan dan dibuat cluster baru yaitu AB. Setelah itu, entri dalam matriks jarak diperbarui dengan cara berikut:
  - a. Menghapus baris dan kolom yang sesuai dengan cluster A dan B
  - b. Menambahkan baris dan kolom yang memberikan jarak antara cluster AB dengan cluster yang tersisa.
- 4) Mengulangi tahap 2 dan 3 sebanyak  $N-1$  kali dan kemudian semua objek akan berada di dalam satu cluster tunggal setelah algoritma berakhir.
- 5) Setelah itu, identitas cluster yang di gabung dan jarak kemiripannya diidentifikasi berdasarkan gabungan tersebut di tempatkan.

Algoritma pada AHC memiliki keragaman di antaranya sebagai berikut [21]:

#### 1) Single Linkage

Algoritma ini mengelompokkan dengan cara menggunakan jarak terdekat antara objek dengan objek. Dimana rumus untuk menentukan jarak antara cluster AB ke cluster C seperti berikut.

$$d_{(AB)C} = \min(d_{AC}, d_{BC}) \quad (1)$$

#### 2) Complete Linkage

Algoritma ini mengelompokkan dengan cara menggunakan jarak terjauh antara objek dengan objek. Dimana rumus untuk menentukan jarak antara cluster AB ke cluster C seperti berikut.

$$d_{(AB)C} = \max(d_{AC}, d_{BC}) \quad (2)$$

#### 3) Average Linkage

Algoritma ini mengelompokkan dengan cara menggunakan rata-rata jarak antara objek dengan objek. Dimana rumus untuk menentukan jarak antara cluster AB ke cluster C seperti berikut.

$$d_{(AB)C} = \frac{d_{AC}+d_{BC}}{n_{AC}n_{BC}} \quad (3)$$

4) *Metode Ward Linkage*

Algoritma ini adalah algoritma yang meminimalkan “kehilangan identitas” karena bergabung dengan dua kelompok. Dimana rumus untuk menentukan jarak antara cluster AB ke cluster C seperti berikut.

$$d_{(AB)C} = \frac{[(n_C+n_A)d_{AC}+(n_C+n_B)d_{BC}]-n_Cd_{AB}}{n_C+n_{AB}} \quad (4)$$

### 3.4 Evaluasi Model

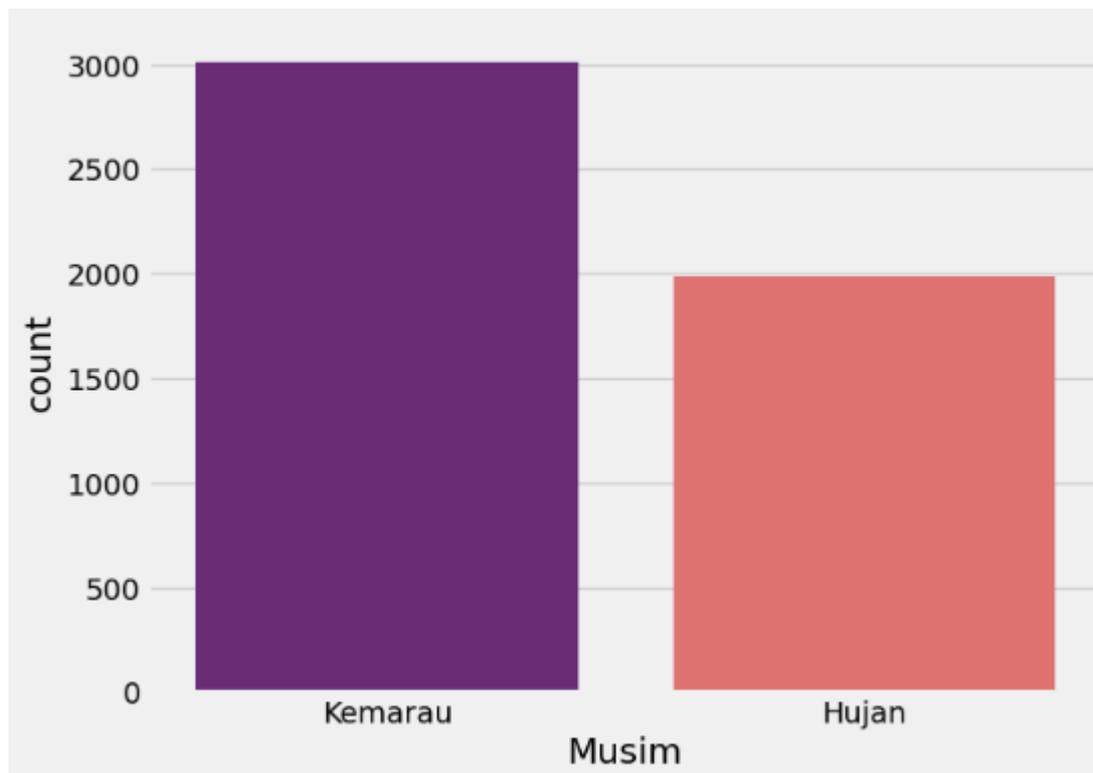
Evaluasi model bertujuan untuk menentukan jumlah *cluster* (nilai k) yang paling optimal[17]. Pada penelitian ini, evaluasi model *cluster* menggunakan Silhouette Score Index (SSI), Davies Doublin Index (DBI) dan Calinski Harabasz Index (CHI). SSI adalah metode evaluasi untuk menguji akurasi suatu *cluster* yang telah terbentuk dari proses *clustering*. Nilai SSI memiliki rentang nilai -1 sampai dengan 1. Nilai SSI yang maximum menunjukkan nilai k yang optimal[17], [20]. Menurut [22], DBI adalah salah satu teknik pengukuran untuk mengevaluasi hasil *clustering*, dengan memaksimalkan jarak antar *cluster* dan sekaligus akan berupaya meminimalkan jarak antar titik dalam *cluster*, dimana jumlah *cluster* yang optimal adalah jumlah *cluster* dengan nilai DBI yang minimum[17]. CHI adalah suatu teknik mengevaluasi validitas *cluster* berdasarkan perhitungan *Sum of Square* antar *cluster* (BSS) dan *Sum of Square* Dalam-Kluster (WSS). CHI mengukur rasio pemisahan berdasarkan jarak maksimum antar *centroid* dan mengukur kekompakan berdasarkan besarnya jarak antara setiap data dengan *centroid*[23].

## 4 Hasil dan Pembahasan

Dataset gempabumi di Asia Tenggara yang berhasil dikumpulkan sebanyak 5000 data. Selanjutnya, dilakukan *preprocessing data* dengan memeriksa data *null* dalam setiap atribut dan pemilihan data atau dengan menghapus atribut yang tidak perlukan. Hasil dari pemeriksaan ini tidak terdapat data *null*/kosong. Pada tahap pemilihan data dilakukan pemilihan beberapa atribut yang diperlukan untuk dilakukan pengolahan data lebih lanjut. Dataset gempabumi yang telah dilakukan pemilihan atribut berdimensi 5.000 baris x 9 kolom. Selanjutnya, Tahap pembentukan atribut baru atau ekstraksi fitur yang bertujuan untuk memperjelas dan melabelkan data. Hasil dari tahap pembentukan atribut baru membuat dimensi dari dataset gempabumi di Asia Tenggara bertambah menjadi 5.000 baris x 16 kolom. Selanjutnya, dilakukan seleksi atribut untuk menghindari redudansi data. Hasil dari seleksi data termasuk juga hasil akhir tahapan *preprocessing data* yang disajikan pada Tabel 2 dan diketahui sebanyak 3.000 titik gempabumi yang terjadi di musim kemarau dan paling banyak terjadi di rentang tahun 2021-2022 yang didukung oleh grafik pada Gambar 1.

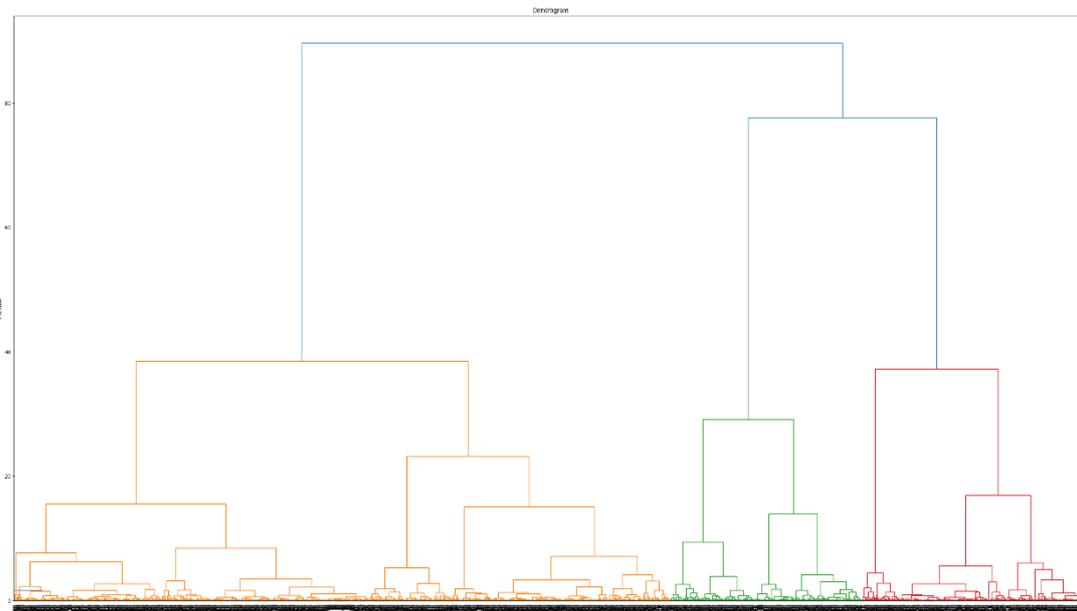
Tabel 2. Hasil Seleksi Data

Year	Quarter	Lat	Lon	Depth	Magnitude	Period	Kategori_gempa	Musim	Kedalaman_gempa
2023	2	75,848	119,277	281,0	4,3	Malam	Light	Kemarau	Menengah
2023	2	57,205	124,4945	63,5	4,8	Malam	Light	Kemarau	Dangkal
2023	2	57,954	124,5511	76,1	4,5	Sore	Light	Kemarau	Menengah
2023	2	57,618	124,5428	73,4	4,9	Sore	Light	Kemarau	Menengah
2023	2	31,135	126,7521	45,9	5,1	Sore	Moderate	Kemarau	Dangkal

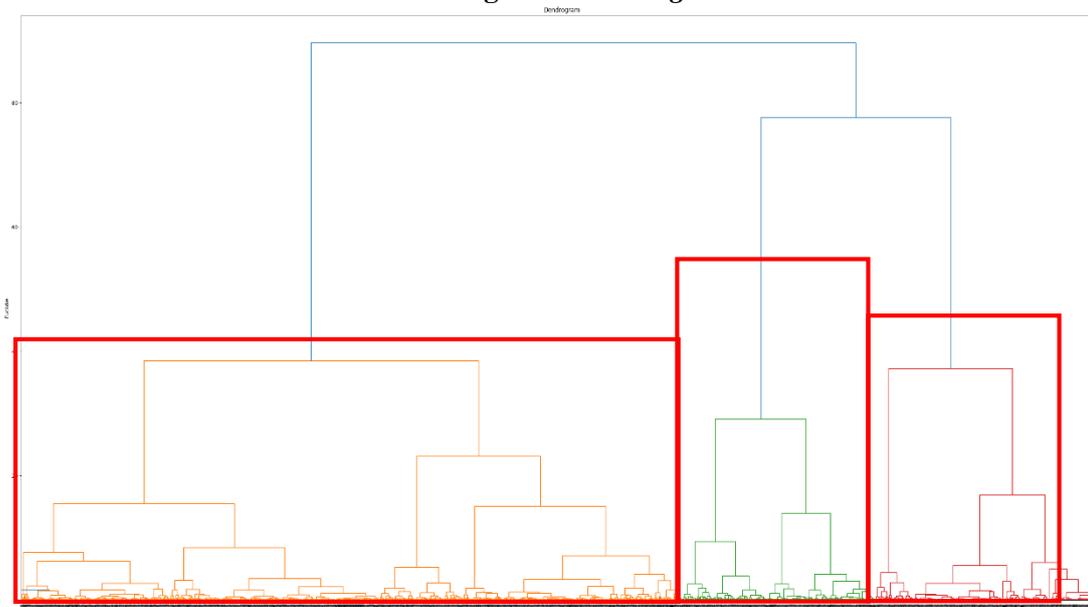


**Gambar 1. Grafik Distribusi Jumlah Kejadian Gempa pada Atribut Musim**

Pada pembuatan model *clustering* menggunakan AHC akan diterapkan pada hasil akhir dari *preprocessing data*. Pada penelitian ini dilakukan 2 proses *clustering*. Pertama, *clustering* untuk membuat fitur baru (Area) berdasarkan atribut *Lat* dan *Lon*. Kedua, *clustering* untuk mendefinisikan jenis gempa bumi yang dengan menggunakan atribut baru hasil dari tahap pembentukan atribut baru atau ekstraksi fitur. Pada *clustering* pertama untuk membuat atribut *Area* digunakan 2 atribut (*Lat*, *Lon*) yang kemudian data kedua atribut itu distandardisasi dengan teknik *StandardScaler*. Hasil dari standarisasi ini kemudian diolah ke tahap berikutnya yaitu mencari jumlah *cluster* optimal dengan menggunakan dendrogram pendekatan *hierarchical* yang diukur menggunakan *Euclidean distance* dengan metode *Ward*. Hasil dendrogram *clustering* pertama disajikan pada Gambar 2 dan diketahui bahwa data hasil standarisasi memiliki 4 *cluster* dan jumlah *cluster* optimal menurut dendrogram adalah 3 yang di ambil dari  $n \text{ cluster} - 1$ . Jumlah *cluster* tersebut disajikan pada Gambar 3. Hasil dari mencari *cluster* optimal tersebut kemudian digunakan untuk ke tahap berikutnya proses *clustering* dan pada Tabel 3 disajikan hasil dari *clustering* untuk membuat atribut *Area* dan Gambar 4 menyajikan distribusinya.



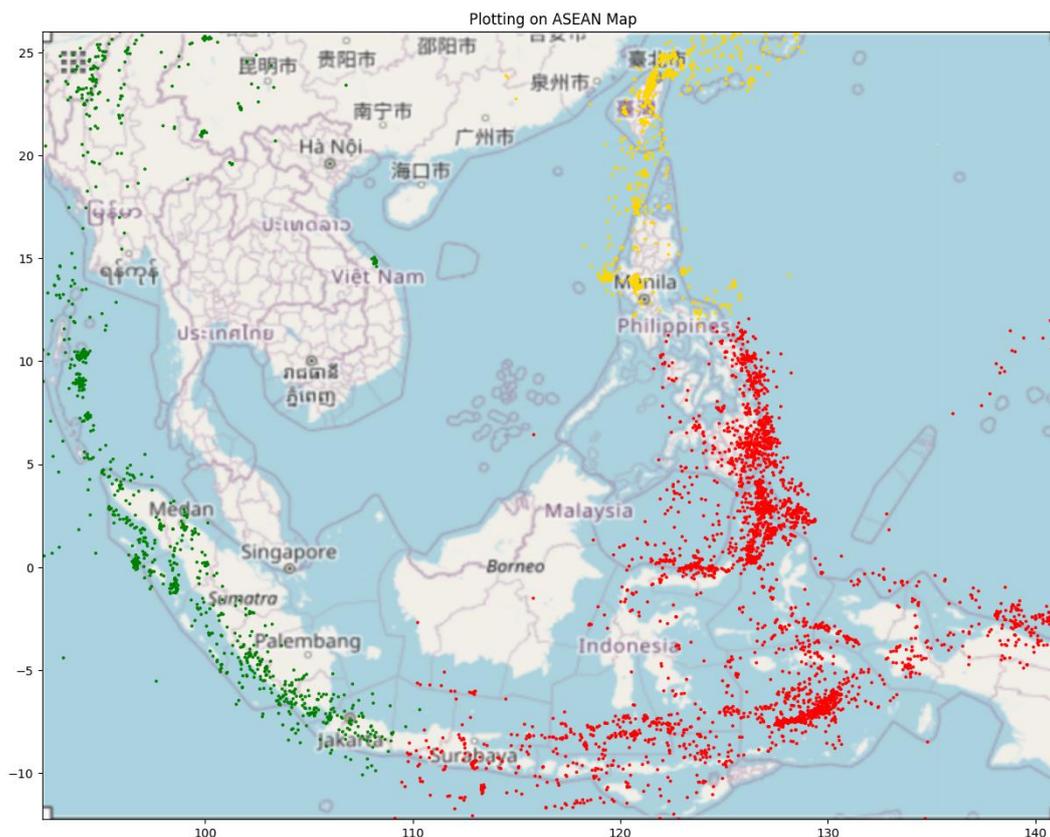
Gambar 2. Dendrogram Clustering Pertama



Gambar 3. Jumlah Cluster Optimal Berdasarkan Dendrogram

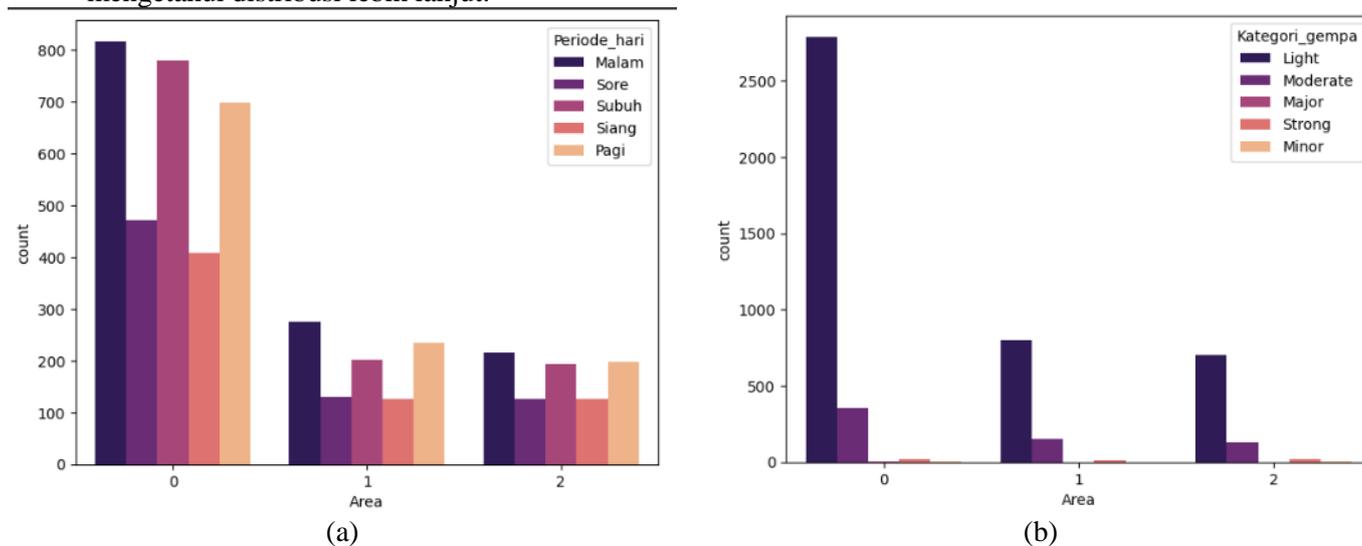
Tabel 3. Hasil dari Clustering Untuk Membuat Atribut Area

Year	Quarter	Lat	Lon	Depth	Magnitude	...	Musim	Kedalaman_gempa	Area
2023	2	75,848	119,277	281,0	4,3	...	Kemaru	Menengah	0
2023	2	57,205	124,4945	63,5	4,8	...	Kemaru	Dangkal	0
2023	2	57,954	124,5511	76,1	4,5	...	Kemaru	Menengah	0
2023	2	57,618	124,5428	73,4	4,9	...	Kemaru	Menengah	0
2023	2	31,135	126,7521	45,9	5,1	...	Kemaru	Dangkal	0



**Gambar 4 Distribusi dari Atribut Area**

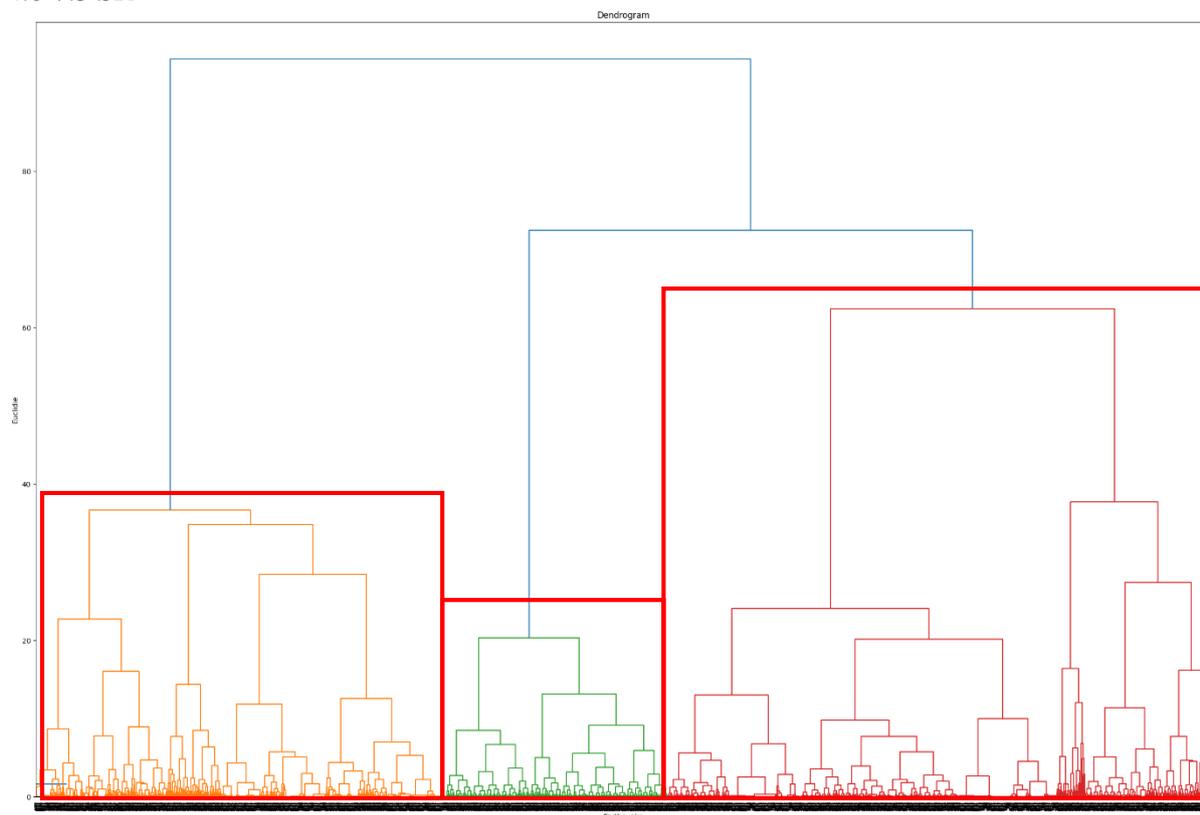
Hasil dari *clustering* pertama ini diketahui bahwa Area 0 (titik merah pada Gambar 4) merupakan Area yang paling banyak titik kejadian gempa bumi, pada Area 0 kejadian titik gempa bumi juga lebih banyak terjadi pada periode hari malam hingga subuh, Gempa yang terjadi pada seluruh Area paling banyak adalah gempa dengan kategori ringan atau *light*. Perhatikan grafik pada Gambar 5 untuk mengetahui distribusi lebih lanjut.



**Gambar 5 (a) Grafik antara Atribur Area dengan Periode\_hari (b) Grafik antara Atribut Area dengan Kategori\_gempa**

Selanjutnya, *clustering* kedua yang bertujuan untuk mendefinisikan Jenis Gempa Bumi di Asia Tenggara yang menggunakan atribut *Mag* dan *Depth*. Prosesnya sama seperti pada *clustering* pertama dan hasil dari *clustering* kedua ini mendapati bahwa nilai optimal jumlah *cluster* adalah 3 *cluster*. Pada

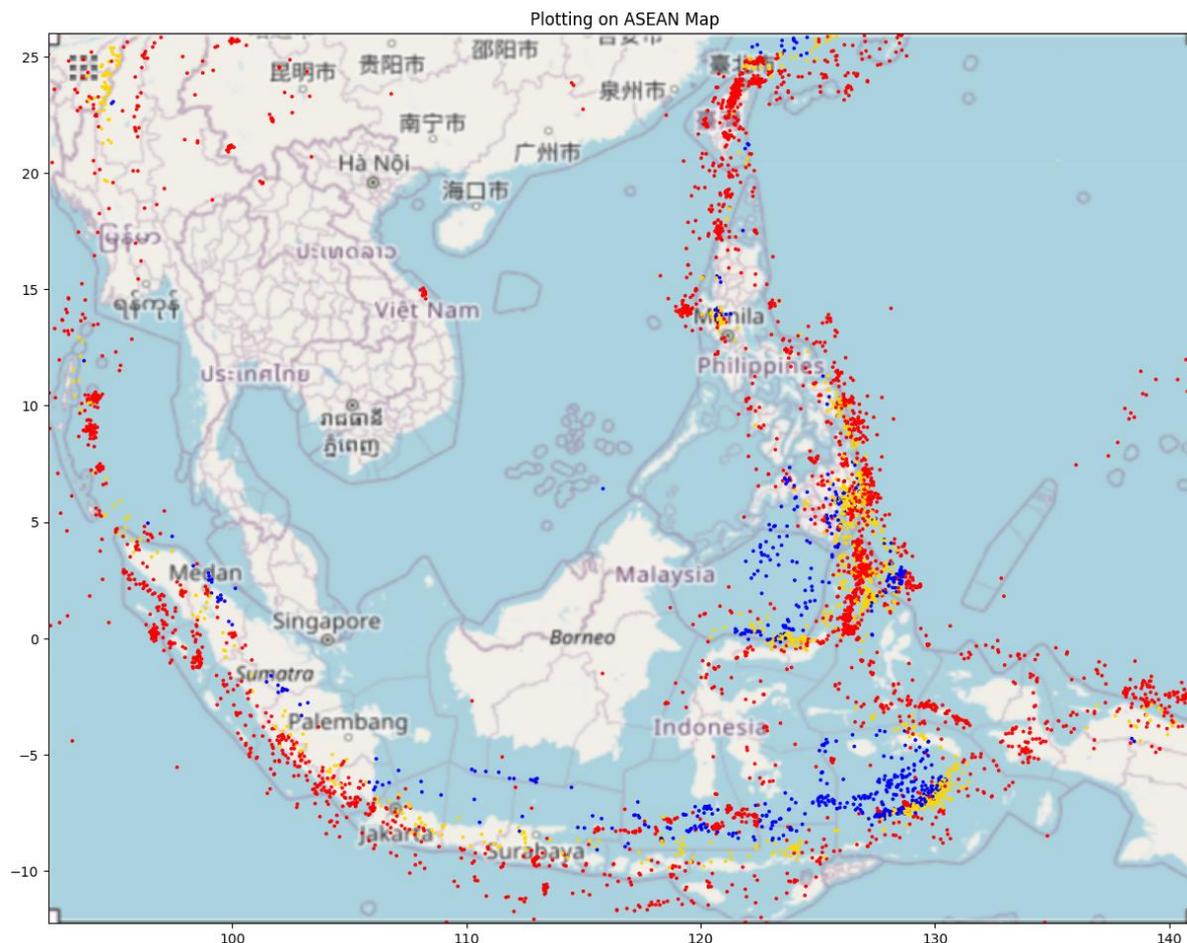
Gambar 6 disajikan hasil dendrogram beserta *cluster* yang diambil. Dari hasil *clustering* kedua dapat dilihat pada Tabel 4 dan Distribusi sebaran hasil *cluster* pada Gambar 7. Hasil tahapan *clustering* kedua, diketahui jenis gempa dibagi menjadi 3 *cluster* disajikan pada Tabel 5 pembagian jenis *cluster*. *Cluster* 0 dengan karakteristik kedalaman gempa 0.0 – 62.8 dan Magnitudo 3.0 – 7.3 SR. *Cluster* 1, kedalaman gempa 62.6-138.6 dan Magnitudo 4.0-7.6 SR. *Cluster* 2, kedalaman gempa 138.8-632.7 dan Magnitudo 4.0-7.3 SR



Gambar 6. Dendrogram dan *Cluster* dari *Clustering* Kedua

Tabel 4. Hasil dari *Clustering* Kedua

Year	Quarter	Lat	Lon	Depth	Mag	...	Kedalaman_gempa	Area	Cluster
2023	2	-75,848	119,277	281,0	4,3	...	Menengah	0	2
2023	2	57,205	124,494 5	63,5	4,8	...	Dangkal	0	1
2023	2	57,954	124,551 1	76,1	4,5	...	Menengah	0	1
2023	2	57,618	124,542 8	73,4	4,9	...	Menengah	0	1
2023	2	31,135	126,752 1	45,9	5,1	...	Dangkal	0	0



**Gambar 7. Distribusi sebaran hasil *cluster* kedua**  
**Tabel 5. Pembagian Jenis *Cluster***

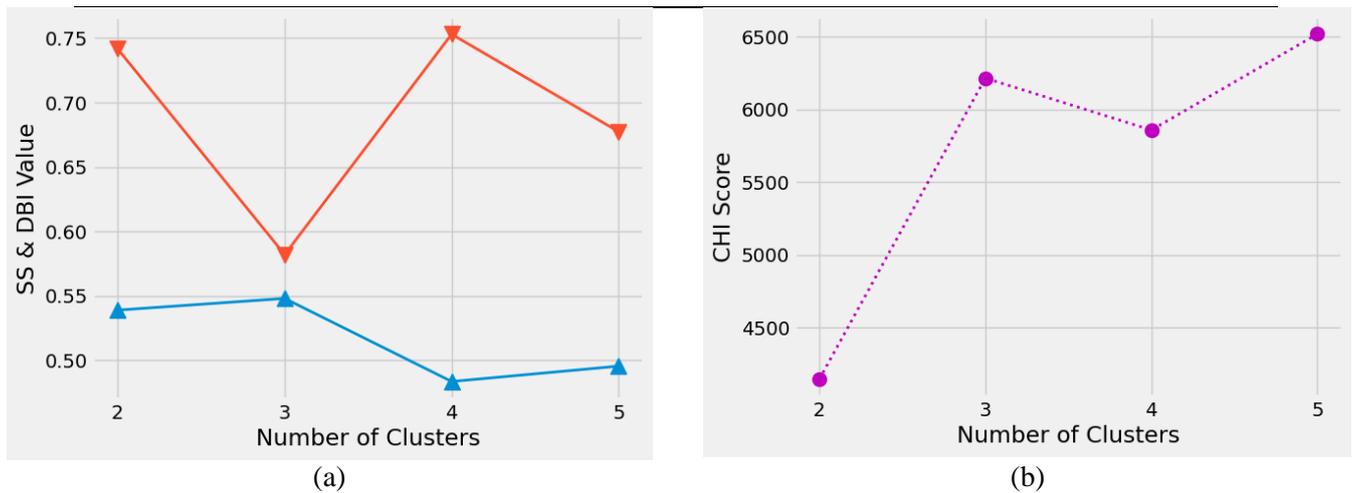
JenisGempa	JumlahData	Persentase	Depth	Mag
0	3241	64,82	(0.0) - (62.8)	(3.0) - (7.3)
1	994	19,88	(62.6) - (138.6)	(4.0) - (7.6)
2	765	15,30	(138.8) - (632.7)	(4.0) - (7.3)

Tahapan Terakhir adalah Uji Validitas dengan mengevaluasi menggunakan 3 indeks (SSI, DBI, CHI). Hasil dari evaluasi *clustering* pertama disajikan pada Tabel 6 dan di visualisasikan dalam grafik yang disajikan pada Gambar 8. Sesuai dengan penjelasan pada metode penelitian didapatkan evaluasi untuk penentuan *cluster* optimal pada *clustering* pertama dalam penelitian adalah 3 *cluster* dengan SSI bernilai 0.434353, DBI bernilai 0.887791 dan CHI 3.769.30. Sehingga, 3 *cluster* untuk *clustering* pertama sudah optimal.

**Tabel 6. Evaluasi *Clustering* Pertama**

<i>Cluster</i>	<i>Silhouette Score Index</i>	<i>Davies Bouldin Index</i>	<i>Calinski Harabasz Index</i>
2	0,538880	0,741559	4146,563325
3	0,547961	0,582004	6214,774108

4	0,483486	0,752997	5857,962792
5	0,495413	0,677370	6518,150922

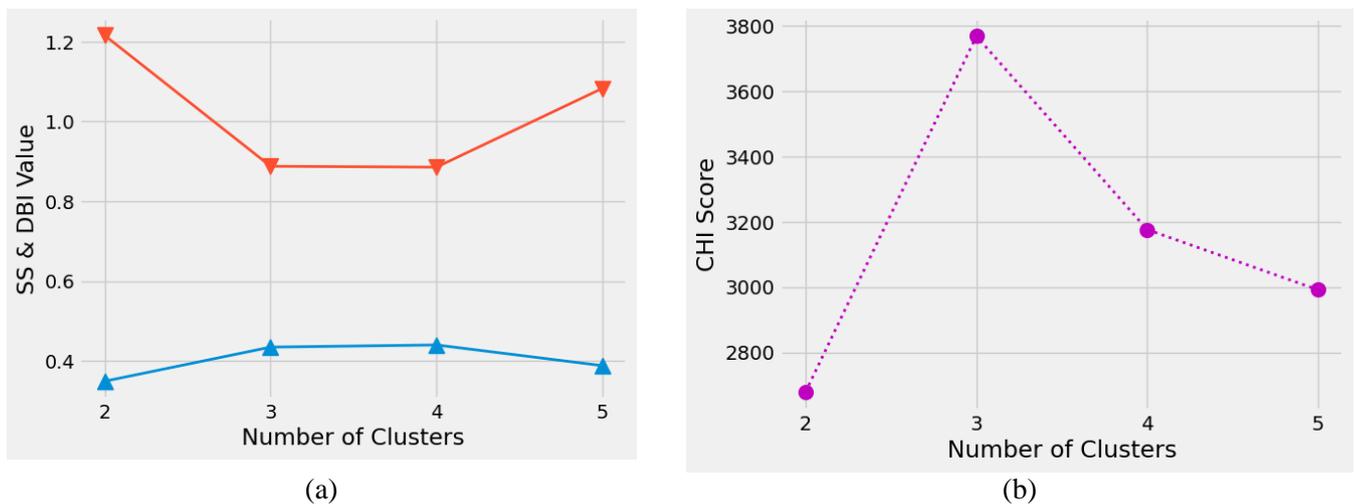


**Gambar 8. Grafik Untuk Clustering Pertama (a) SSI (biru) dan DBI (Orange) (b) CHI**

Selanjutnya, evaluasi *Clustering* kedua yang disajikan pada Tabel 7 dan di visualisasikan dalam grafik yang disajikan dalam Gambar 9. Hasil dari evaluasi ini mendapati bahwa jumlah *cluster* yang optimal adalah 3 dan 4 *cluster*, dengan selisih sangat kecil. Dari Tabel 7 terlihat bahwa sesuai dengan penjelasan pada metode penelitian jumlah *cluster* paling optimal adalah 4 *cluster* bila mengacu pada SSI dan DBI, tetapi, dilihat pada CHI bahwa jumlah *cluster* sama dengan 3 lebih optimal. Sehingga penelitian ini menggunakan 3 *cluster* untuk *clustering* kedua yang bertujuan untuk mendefinisikan jenis gempabumi.

**Tabel 7. Evaluasi Clustering Kedua**

<i>Cluster</i>	<i>Silhouette Score Index</i>	<i>Davies Bouldin Index</i>	<i>Calinski Harabasz Index</i>
2	0,348947	1.215641	2678.730316
3	0,434353	0.887791	3769.030146
4	0,439762	0.885528	3176.245248
5	0,387973	1.082878	2991.610802



Gambar 9. Grafik Untuk Clustering Kedua (a) SSI (biru) dan DBI (Orange) (b) CHI

## 5 Kesimpulan

Dalam penelitian ini dilakukan proses *data mining* terhadap data gempabumi di wilayah Asia Tenggara pada rentang tahun 2021-2023 untuk tujuan mengelompokkan titik kejadian gempabumi yang ada di wilayah Asia Tenggara. Dari hasil penelitian yang telah dilakukan, pengelompokkan dengan menerapkan AHC pada 2 proses *clustering* menghasilkan 3 *cluster* untuk *clustering* pertama yang bertujuan untuk membuat atribut baru yaitu, Area dan 3 *cluster* untuk *clustering* kedua yang bertujuan untuk mengidentifikasi jenis gempa yang ada di wilayah Asia Tenggara. Dari penelitian ini juga disimpulkan Area 0 (Titik Merah pada Gambar 4) merupakan wilayah yang sering terjadi gempa dari tahun 2021 hingga sekarang. Hasil pengelompokkan titik gempabumi di wilayah Asia Tenggara dapat menjadi acuan dalam mengembangkan penelitian lain seperti prediksi gempabumi yang akan datang. Penelitian ini juga dapat menjadi acuan untuk menyusun mitigasi bencana yang akan datang agar lebih waspada ke depannya.

## Referensi

- [1] R. R. A. Rahman dan A. W. Wijayanto, "Pengelompokan Data Gempa Bumi menggunakan Algoritma DBSCAN. Grouping Earthquakes Data Using DBSCAN Algorithm," 2021.
- [2] S. B. Mariani, "Kesiapsiagaan Masyarakat dalam Menghadapi Bencana Gempa Bumi," 2021, Diakses: Mar. 31, 2023. [Online]. Tersedia: <http://eprints.ums.ac.id/92135/>
- [3] D. Jales Manalu, R. Rahmawati, dan T. Widiharih, "Pengelompokan Titik Gempa di Pulau Sulawesi menggunakan Algoritma ST-DBSCAN (Spatio Temporal-Density Based Spatial Clustering Application with Noise)," vol. 10, hal. 554–561, 2021, [Online]. Tersedia: <https://ejournal3.undip.ac.id/index.php/gaussian/>
- [4] N. Amaliah Selvyana dan D. Rahmah Fitriani, "Hubungan Pengalaman dengan Kesiapsiagaan Remaja Dalam Menghadapi Banjir di Samarinda," 2021.
- [5] K. Pakuani dan R. Kurniawan, "Kajian Penentuan Nilai Epsilon Optimal Pada Algoritma DMDBSCAN dan Pemetaan Daerah Rawan Gempa Bumi Di Indonesia Tahun 2014-2020," 2021.
- [6] MAGMA Indonesia, "Gempa bumi M 7.0 SR di Lombok Utara, NTB." Agustus 06, 2018. Diakses: Mar. 31, 2023. [Online]. Tersedia: <https://magma.esdm.go.id/v1/press-release/162/gempa-bumi-m-70-sr-di-lombok-utara-ntb>
- [7] Badan Meteorologi Klimatologi dan Geofisika, "Gempabumi Tektonik M=7.7 Kabupaten Donggala, Sulawesi Tengah pada hari Jumat, 28 September 2018, Berpotensi Tsunami." Badan Meteorologi Klimatologi dan Geofisika, Jakarta, 2018. Diakses: Mar. 31, 2023. [Online]. Tersedia: <https://www.bmkg.go.id/press-release/?lang=ID&p=gempabumi-tektonik-m7-7-kabupaten->

donggala-sulawesi-tengah-pada-hari-jumat-28-september-2018-berpotensi-tsunami&tag=press-release

- [8] R. D. R. L. Kedutaan Besar Republik Indonesia di Vientiane, "Himbauan Kepada WNI di Laos." 2019. Diakses: Mar. 31, 2023. [Online]. Tersedia: <https://kemlu.go.id/vientiane/id/news/3328/himbauan-kepada-wni-di-laos>
- [9] R. M. Kedutaan Besar Republik Indonesia di Manila, "Gempa 6,1 Skala Richter guncang Manila dan sekitarnya." 2019. Diakses: Mar. 31, 2023. [Online]. Tersedia: <https://kemlu.go.id/manila/id/news/205/gempa-61-skala-richter-guncang-manila-dan-sekitarnya>
- [10] A. Prakasa dan F. D. Utami, "Sistem Informasi Radar Cuaca Terintegrasi BMKG," 2019.
- [11] J. Sweet, K. Anderson, dan R. Woodward, "NSF SAGE Facility Begins Procurement of Rapid Response Instrumentation," vol. 2021, Desember 2021, Diakses: Mar. 31, 2023. [Online]. Tersedia: <https://ui.adsabs.harvard.edu/abs/2021AGUFM.T35B0203S/abstract>
- [12] Kusumawati Rajnaparmitha, E. Bayunanda, A. M. Rifa'i, M. R. G. Asgar, F. I. Ilmawati, dan Kusrini, "Clustering Titik Panas Menggunakan Algoritma Agglomerative Hierarchical Clustering (AHC)," vol. 8, 2022.
- [13] R. R. Arista, R. A. Asmara, dan D. Puspitasari, "Pengelompokan Kejadian Gempa Bumi Menggunakan Fuzzy C-Means Clustering," 2017.
- [14] M. F. Al Halik dan L. Septiana, "Analisa Data Untuk Prediksi Daerah Rawan Bencana Alam Di Jawa Barat Menggunakan Algoritma K-Means Clustering," *Journal of Information System, Applied, Management, Accounting and Research*, vol. 6, no. 4, hal. 856–870, 2022, doi: 10.52362/jisamar.v6i4.939.
- [15] Y. Setiawan, S. Aziz Suprianto, A. Wijanarko, D. Setyo Rini, dan M. Yusa, "Pemetaan Kelompok Sebaran Titik Gempa Bumi Mentawai dengan Metode K-Medoids Clustering," 2022.
- [16] J. Jefrianto dan D. Prima Sari, "Analisis K-Medoids Clustering pada Episentrum Gempa Bumi di Provinsi Sumatera Barat dan Sekitarnya," *Journal Of Mathematics UNP*, vol. 7, no. 2, hal. 22–27, 2022.
- [17] A. Ayu, D. Sulistyawati, dan M. Sadikin, "Penerapan Algoritma K-Medoids untuk Menentukan Segmentasi Pelanggan." [Online]. Tersedia: <http://sistemasi.ftik.unisi.ac.id>
- [18] T. Zulhaq Jasman, E. Hasmin, Sunardi, C. Susanto, dan W. Musu, "Perbandingan Logistic Regression, Random Forest, dan Perceptron pada Klasifikasi Pasien Gagal Jantung," Oktober, vol. 14, no. 3, hal. 271–286, 2022, doi: 10.22303/csrid.14.3.2022.271-286.
- [19] N. Amini, T. H. Saragih, M. R. Faisal, A. Farmadi, dan F. Abadi, "Implementasi Algoritma Genetika untuk Seleksi Fitur pada Klasifikasi Genre Musik menggunakan Metode Random Forest," vol. 9, 2022.
- [20] E. M. S. Rochman, dkk., "A Combination of Algorithm Agglomerative Hierarchical Cluster (AHC) and K-Means for Clustering Tourism in Madura-Indonesia," *Journal of Mathematical and Computational Science*, 2022, doi: 10.28919/jmcs/7086.
- [21] A. Azzahra dan A. W. Wijayanto, "Perbandingan Agglomerative Hierarchical dan K-Means dalam Pengelompokan Provinsi Berdasarkan Pelayanan Kesehatan Maternal," *Comparison of Agglomerative Hierarchical and K-Means in Grouping Provinces Based on Maternal Health Services*.
- [22] A. K. Singh, S. Mittal, P. Malhotra, dan Y. V. Srivastava, "Clustering Evaluation by Davies-Bouldin Index (DBI) in Cereal Data using K-Means," in *Proceedings of the 4th International Conference on Computing Methodologies and Communication, ICCMC 2020*, Institute of Electrical and Electronics Engineers Inc., Maret 2020, hal. 306–310. doi: 10.1109/ICCMC48092.2020.ICCMC-00057.

- [23] R. Gustriansyah, N. Suhandi, dan F. Antony, "Clustering Optimization in RFM Analysis based on K-Means," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 1, hal. 470–477, 2019, doi: 10.11591/ijeecs.v18.i1.pp470-477.