

Pemanfaatan Algoritma K-Means dalam Klasterisasi Gempa Sulawesi

Wawan Gunawan ^{1,3}, Arief Wibowo ²

^{1,2}Magister of Computer Science, Universitas Budi Luhur, Indonesia

³Department of Informatics, Universitas Budi Luhur, Indonesia

Article Info

Article history:

Received 21 Apr 2024

Revised 25 Jul 2024

Accepted 13 Aug 2024

Keywords:

gempa bumi
sulawesi
elbow
klasterisasi
k-means

ABSTRACT

Indonesia is a region that frequently experiences earthquakes, especially in the Sulawesi area which has significant active faults. Sulawesi is an area that has relatively high seismic intensity, and there are several active faults, which are earthquake source zones. This study uses M-Means with a total of 9,710 records starting from 2019-2023, and the attributes consist of event_id, date_time, latitude, longitude, magnitude, mag_type, depth_km, phase_count, azimuth_gap, location, agencydengan. This data processing compares magnitude and depth consisting of 3 clusters, namely 51-132 Km depth with a total of 1,311, 3-50 Km depth with a total of 7,527, 133-300 Km depth with a total of 872, while the process with magnitude, depth and azimuth gap attributes consists of 4 clusters with each cluster respectively 3,957, 1,546, 1,458, and 2,749. By using a different set of input features, this research identifies that the results from 3 clusters or 4 clusters indicate that the province of South Sumatra shows a high level of earthquake proneness and frequent frequency in all clusters with the epicenter of the earthquake being in the Maluku Sea, between South Sulawesi with Southeast Sulawesi, as well as the province of Gorontalo. Based on the results obtained, there is a need for early prevention related to disasters, especially earthquakes that occurred on the island of Sumatra based on earth faults that run through the island..

Copyright © 2024 Universitas Indraprasta PGRI.
All rights reserved.

Corresponding Author:

Wawan Gunawan,
Magister of Computer Science,
Universitas Budi Luhur,
Jl. Nangka No. 58 C, Tanjung Barat, Ciledug, Tangerang.
Email: 2211601527@student.budiluhur.ac.id

1. PENDAHULUAN

Indonesia memiliki banyak potensi gempa bumi, karena secara geografis Indonesia dikelilingi oleh patahan dan sesar bumi yang aktif. Salah satu daerah yang cukup rawan akan gempa bumi yaitu daerah Sulawesi [1]. Sulawesi adalah salah satu wilayah di Indonesia yang sering terjadi gempa bumi dengan bencana gempa bumi yang pernah terjadi dengan kekuatan 7,4 skala richter pada tanggal 28 september 2018 di Sulawesi tengah, hingga saat ini Sulawesi masih terpantau sering terjadi gempa bumi.

Pulau Sulawesi adalah salah satu pulau yang terletak di Indonesia dan merupakan bagian dari Cincin Api Pasifik [2]. Pulau ini memiliki kompleksitas geologi yang tinggi dan dianggap sebagai salah satu daerah yang memiliki aktivitas sesar tektonik yang signifikan di wilayah Indonesia. Penelitian tentang sesar tektonik aktif di Pulau Sulawesi penting untuk pemahaman dan mitigasi risiko bencana terkait gempa bumi dan potensi tsunami. Sesar-sesar aktif yang merupakan zone sumber gempa tersebut antara lain adalah Sesar Palu-Koro

dan Sesar Walanae di Sulawesi bagian barat, Sesar Matano dan Sesar Lawanopo di Sulawesi bagian timur, serta Sesar Gorontalo di Sulawesi bagian utara [3]–[5].

Sulawesi merupakan daerah yang memiliki intensitas kegempaan yang cukup tinggi, dan terdapat beberapa sesar-sesar aktif yang merupakan zona sumber gempa. Tingginya Aktivitas gempa bumi di wilayah Sulawesi menunjukkan urgensi untuk meningkatkan kesiapsiagaan bencana. Pemahaman yang lebih baik tentang histori gempa dapat membantu meningkatkan rencana kesiapsiagaan.

Perlu adanya pencegahan dini terkait dengan kebencanaan, khususnya gempa. Penggunaan CNN dapat membantu untuk antisipasi awal terkait dengan penggunaan sensor untuk mengetahui seismik aktif sistem pemantauan dan sistem peringatan dini gempa. Penggunaan CNN menyebutkan bahwa dapat dengan fleksibel untuk mempelajari representasi fitur untuk berbagai jenis sensor [6]. Selain itu CNN pun mampu memberikan tingkat akurasi yang tinggi untuk magnitude, waktu, kedalaman, dan klasifikasi lokasi dengan nilai 93,67%, 89,55%, 92,54%, dan 89,5% [7].

Penelitian selanjutnya membandingkan algoritma SVM, RF, DT, dan ANN terkait dengan akselerasi, kecepatan, dan perpindahan menunjukkan korelasi yang baik dengan hasil yang membuktikan bahwa kombinasi akselerasi, kecepatan, dan perpindahan dapat digunakan untuk meningkatkan performa model dengan meningkatkan akurasi model RF dan ANN menjadi 0.974589 dan 0.965400 secara berturut-turut [8].

K-means merupakan algoritma klusterisasi yang sering digunakan dalam data science dan merupakan metode partisi yang bekerja dalam ruang Euclidean d-dimensi. Tujuan utama algoritma ini untuk membagi n titik data ke dalam c kluster sehingga titik-titik data dalam kluster yang sama memiliki kesamaan yang lebih besar dibandingkan dengan titik-titik data dari kluster yang berbeda. Setiap kluster akan diwakili oleh sebuah pusat kluster atau centroid. Dengan persamaan seperti nampak pada persamaan 1.

$$J(z, A) = \sum_{i=1}^n \sum_{k=1}^c z_{ik} \|x_i - a_k\|^2 \quad (1)$$

Di mana:

x_i adalah titik data ke i

a_k adalah centroid untuk kluster ke-k

z_{ik} adalah variabel biner yang menunjukkan apakah titik data ke-i termasuk dalam kluster ke-k atau tidak

$\|\cdot\|$ adalah norma euclidean yang mengukur jarak antara titik data dan centroid

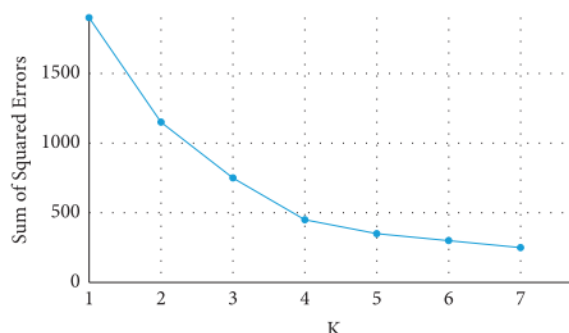
Permasalahan yang sering dihadapi dalam k-means adalah perlunya menentukan jumlah kluster sebelumnya (a priori), yang seringkali tidak diketahui dalam aplikasi nyata, serta k-means yang selalu dipengaruhi oleh inisialisasi. Untuk mengatasi masalah ini, indeks validitas kluster mendapatkan perhatian lebih karena dapat membantu dalam menentukan jumlah kluster yang optimal [9]. Algoritma k-means sudah terbukti secara efektif saat digunakan untuk bencana angin sehingga dapat mengidentifikasi faktor-faktor utama yang mempengaruhi berbagai jenis bencana getaran angin [10][11]. Penelitian selanjutnya terkait peningkatan dari algoritma ini pun telah dilakukan penelitian dengan hasil yang lebih stabil jika dibandingkan dengan k-means [12][13].

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

Algoritma k-means dapat digunakan untuk mengklasifikasikan data berdasarkan kategori daerah rawan hotspot dan juga kawasan rehabilitasi [14]. Pada penelitian tersebut menggunakan data tahun 2019 Januari-Juni terdapat 284 titik yang masuk wilayah rawan, 215 titik masuk wilayah tidak rentan, dan 129 titik masuk wilayah sangat rentan [15].

Penggunaan metode Elbow terkait dengan prediksi penyakit stroke telah berhasil meningkatkan nilai akurasi dan precision, dengan nilai akurasi meningkat sekitar 6% dan precision sekitar 0,12 [16], nilai SSE terendah dan sudut Elbow paling tajam pada grafik menentukan nilai cluster terbaik [17][18] seperti yang ditampilkan pada gambar 1, dan juga berdasarkan nilai rata-rata atau yang sering muncul [19]. Rumus SSE dapat digambarkan sebagai rumus pada persamaan 3.

$$SSE = \sum (y_i - \gamma_i) \quad (3)$$



Gambar. 1 Grafik penentuan nilai K dengan elbow

Dalam algoritma k-means, teknik siku memplot variasi yang dijelaskan melalui jumlah cluster dan memilih kurva siku yang paling tajam untuk mendapatkan jumlah cluster [20]. Hal ini bergantung pada penghitungan jumlah kesalahan kuadrat dalam sebuah cluster [21]. Setelah menggunakan elbow, selanjutnya pengujian akan dilakukan dengan menggunakan teknik silhouette untuk melihat hasil perbandingan dengan metode elbow [22].

Sebelumnya BMKG menggunakan Skala Richter (SR) sebagai ukuran intensitas gempa, namun saat ini beralih ke skala magnitudo karena dianggap lebih presisi. Selain itu, Skala Intensitas Mercalli Modifikasi (MMI) sering digunakan untuk mengukur getaran yang dirasakan oleh penduduk saat terjadi gempa [23]. Skala Mercalli adalah satuan untuk mengukur kekuatan gempa bumi. Satuan ini diciptakan oleh seorang vulkanologis dari Italia yang bernama Giuseppe Mercalli pada tahun 1902. Skala Mercalli terbagi menjadi 12 pecahan berdasarkan informasi dari orang-orang yang selamat dari gempa tersebut dan juga dengan melihat serta membandingkan tingkat kerusakan akibat gempa bumi tersebut [23].

Skala Magnitudo merupakan ukuran yang menggambarkan jumlah energi seismik yang dilepaskan oleh sumber gempa, yang diukur berdasarkan pengamatan seismograf. Semakin kuat gempa, maka nilai magnitudo yang dihasilkan juga akan semakin tinggi. Nilai yang tercatat berdasarkan Skala Magnitudo diekspresikan dalam angka bulat dan desimal. Menurut informasi dari Michigan Technological University, kisaran magnitudo lain yang tercatat seperti pada tabel 1.

TABEL I
KEKUATAN GEMPA

Kekuatan (Magnitudo)	Kerusakan yang diakibatkan
2,5-5,4	kerusakan yang tidak begitu parah
5,5-6,0	kerusakan ringan pada bangunan
6,1-6,9	kerusakan besar di kawasan dengan kepadatan penduduk yang tinggi
7,0-7,9	gempa besar yang dapat menimbulkan kerusakan parah
>8,0	gempa besar yang mampu meruntuhkan kawasan pusat gempa

BMKG saat ini menggunakan skala Mercalli yang masih sering digunakan terutama apabila tidak terdapat peralatan seismometer yang dapat mengukur kekuatan gempa bumi di tempat kejadian. Berdasarkan skala intensitas gempa yang digunakan [23], BMKG membaginya menjadi lima skala bagian, yaitu:

1. Skala MMI I-II
Tidak dirasakan atau dirasakan hanya oleh beberapa orang tetapi terekam oleh alat dengan percepatan tanah puncak < 2,9 gal.
2. Skala MMI III-V
Dirasakan oleh orang banyak tetapi tidak menimbulkan kerusakan. Benda-benda ringan yang digantung bergoyang dan jendela kaca bergetar dengan percepatan tanah puncak 2,9-88 gal.
3. Skala MMI VI
Bagian non struktur bangunan mengalami kerusakan ringan, seperti retak rambut pada dinding, genteng bergeser ke bawah dan sebagian berjatuh dengan percepatan tanah puncak 89-167 gal.
4. Skala MMI VII-VIII

Banyak Retakan terjadi pada dinding bangunan sederhana, sebagian roboh, kaca pecah. Sebagian plester dinding lepas. Hampir sebagian besar genteng bergeser ke bawah atau jatuh. Struktur bangunan mengalami kerusakan ringan sampai sedang dengan percepatan tanah puncak 168-564 gal.

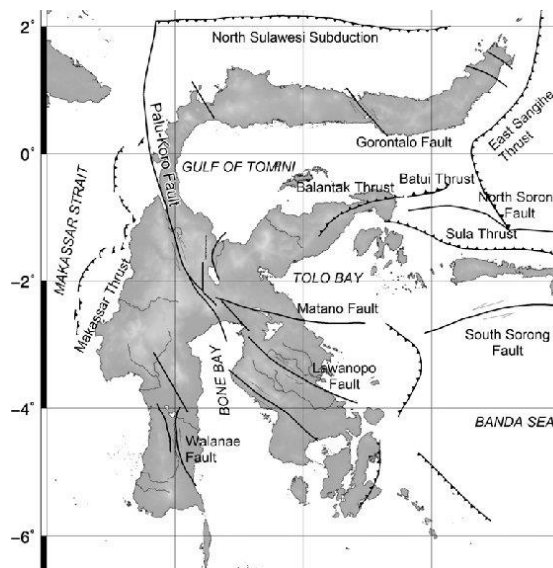
5. Skala MMI IX-XII

Sebagian besar dinding bangunan permanen roboh. Struktur bangunan mengalami kerusakan berat. Rel kereta api melengkung dengan percepatan tanah puncak >564 gal.

Patahan

Patahan bumi terjadi ketika ada pergeseran antara dua bagian dari kerak bumi, patahan bumi seringkali terbentuk oleh gaya tektonik yang mengakibatkan batuan di sekitar patahan terlipat, dipatahkan, dan bergeser [24]. Selain itu aktivitas kegempaan yang seringkali ditimbulkan oleh sesar-sesar aktif yang ada di Sulawesi [5]. Sesar aktif di wilayah Sulawesi serta lokasinya dapat dikenali sebagai salah satu cara untuk melakukan mitigasi, diantaranya adalah sesar Gorontalo, sesar Palu Koro, sesar Matano, sesar Saddang seperti yang ditunjukkan pada gambar 2.

Patahan Palu-Koro melintang dengan arah relatif utara-selatan mulai dari Tanjung Mangkaliat di Kalimantan Utara hingga perbatasan Provinsi Sulawesi Tengah dan Sulawesi Tenggara dengan panjang +- 330 kilometer. Patahan Palu - Koro merupakan jenis patahan mendatar mengiri tipe bercabang (bifurcation) seperti patahan Alpina [1]. Selanjutnya Kota Gorontalo terdapat Sesar Geser Gorontalo dan struktur tanah khususnya di Kota Gorontalo disusun oleh endapan-endapan danau yang terdiri dari tanah liat dan pasir yang gampang berubah akibat getaran yang diakibatkan oleh gempa bumi [3].

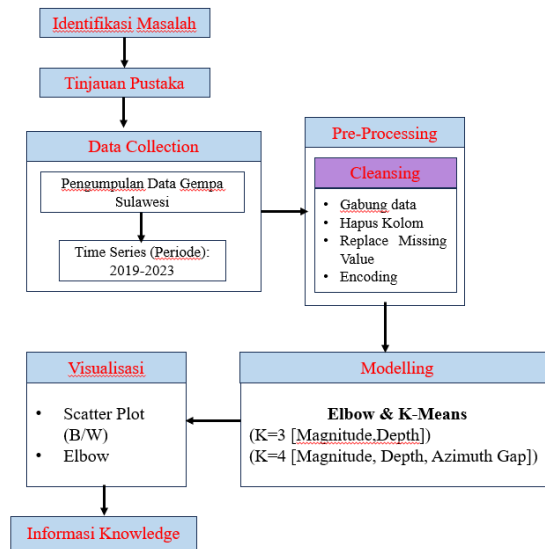


Gambar. 2 Patahan dan sesar pulau Sulawesi

Patahan Matano diketahui sebagai sesar aktif di Pulau Sulawesi dan masih merupakan keberlanjutan dari sesar Palu-Koro. Gempa terakhir dari patahan permukaan ini diperkirakan terjadi antara 1432 dan 1817 dengan kekuatan mencapai $M 7,4 \pm 0,3$ dan panjang retakan sekitar 110 km. Interval perulangan rata-rata dari pecahnya permukaan di jalur patahan gempa bumi ini dihitung pada 200-470 tahun dan slip rate diperkirakan sebesar 21 ± 9 mm per tahun [4].

2. METODE

Tahapan pada penelitian ini diawali dengan melakukan identifikasi terhadap masalah yang ada terkait dengan kebencanaan, dalam kasus ini adalah terkait dengan gempa bumi yang ada di Sulawesi, lalu dilanjutkan dengan literature review terkait dengan topik yang diambil. Bagian alur metode penelitian ini selanjutnya dapat dilihat pada gambar 2.



Gambar. 3 Flowchart tahapan penelitian

3. HASIL DAN PEMBAHASAN

Tahapan ini akan menjabarkan terkait proses yang dilaksanakan untuk mengetahui terkait dengan klasterisasi gempa yang ada di Sulawesi.

3.1. Pengumpulan Data

Dataset yang didapatkan berasal dari repositori milik BMKG yang dapat diakses secara publik (<https://re-pogempa.bmkg.go.id/>) dengan jumlah record sebanyak 9.710 dimulai dari tahun 2019-2023 dan atribut yang dimiliki terdiri atas event_id, date_time, latitude, longitude, magnitude, mag_type, depth_km, phase_count, azimuth_gap, location, agency seperti yang ditampilkan pada tabel 1.

TABEL II
DATASET GEMPA BUMI SULAWESI

No	Event ID	Date time	Latitude	Longitude	Magnitude	Mag Type	Depth (km)	Phase Count	Azimuth Gap	Location	Agency
1	bmg2019ccpo	2019-01-30T21:13:11,615709Z	0,252627194	122,3435059	3,021604766	MLv	131	7	139,9782639	Minahassa Peninsula, Sulawesi	BMKG
2	bmg2019ccfe	2019-01-30T15:57:07,126492Z	-2,94203496	119,4213715	2,439421611	MLv	12	7	149,2787399	Sulawesi, Indonesia	BMKG
3	bmg2019ccdi	2019-01-30T15:01:21,032063Z	-1,820792913	119,384491	2,882417838	MLv	10	10	189,9473495	Sulawesi, Indonesia	BMKG
4	bmg2019cbwt	2019-01-30T11:42:20,668034Z	1,805694461	119,502739	3,722964817	MLv	10	12	181,2581348	Sulawesi, Indonesia	BMKG
5	bmg2019cbwm	2019-01-30T11:35:38,805188Z	1,835213184	119,4530029	4,520827759	MLv	10	20	184,5670452	Sulawesi, Indonesia	BMKG
...
9706	bmg2023rdun	2023-09-01T23:53:15,099291Z	0,016925327	121,9864502	2,915738217	MLv	250	19	78,54156494	Minahassa Peninsula, Sulawesi	BMKG
9707	bmg2023rdpi	2023-09-01T21:15:24,115117Z	1,627227545	120,2110672	2,516981369	MLv	10	24	65,21083069	Sulawesi, Indonesia	BMKG
9708	bmg2023rdmq	2023-09-01T19:54:01,415669Z	2,445503712	120,8326416	2,842521675	MLv	10	33	48,9574585	Sulawesi, Indonesia	BMKG
9709	bmg2023rdfq	2023-09-01T16:22:21,315413Z	0,233703986	122,9118118	3,116213441	MLv	35	26	94,89404678	Minahassa Peninsula, Sulawesi	BMKG
9710	bmg2023rchv	2023-09-01T04:18:27,861924Z	2,997438431	118,5758286	2,776080996	M	28	13	144,5652351	Sulawesi, Indonesia	BMKG

3.2. Pre-Processing

Pada tahap ini data mentah diolah untuk menghasilkan data yang bersih termasuk dengan melakukan pengisian atau penghapusan nilai kosong, penyamarataan format pada setiap atribut, dan mengubah isian yang tidak sesuai dengan format seharusnya. Berikut tahapan yang dilakukan dalam *pre-processing*:

1) *Data Transformasi*: Melakukan penghapusan pada record yang sama, pada atribut "Location" ditemukan 11 label yang memiliki makna yang serupa seperti "Minahassa Peninsula, Indonesia" dengan "Sulawesi, Indonesia". Berdasarkan temuan tersebut maka diputuskan untuk menggabungkan label-label tersebut mengingat

bahwa semenanjung Minahassa adalah bagian dari wilayah Sulawesi. Hal yang sama pun dilakukan untuk wilayah "Kepulauan Talaud" yang merupakan bagian dari provinsi Sulawesi Utara, selanjutnya dilakukan pelabelan menjadi "Sulawesi" untuk menyederhanakan klusterisasi.

2) *Encode atribut bertipe category*: melakukan encoding pada atribut Mag Type dan Location dengan menggunakan One Hot Encoding untuk mengubah informasi kategori menjadi format yang dapat dimengerti oleh mesin dengan menandai keberadaan atau ketiadaan suatu kategori.

3) *Mengkonfirmasi kebenaran outlier*: ditemukan adanya outlier pada atribut "Depth (km)" yang dinilai tidak wajar yaitu berupa nilai kedalaman di bawah 10Km, sedangkan mayoritas data kedalaman berada pada angka ratusan hingga ribuan kilo meter. Setelah dilakukan pengecekan terkait data tersebut dinyatakan data tersebut benar adanya dan tidak perlu dilakukan penghapusan data karena bukan masuk dalam data salah.

4) *Mengecek kebenaran data tiap atribut*: mencari tahu makna setiap atribut dan membandingkannya pada fakta yang ada, contohnya pada atribut Latitude dan Longitude, dilakukan pencocokan pada kondisi geografis Indonesia yang berada pada 6° LU – 11° LS dan 95° BT – 141° BT, hasilnya seperti ditunjukkan pada gambar 3.

Selected attribute		Type: Numeric	
Name: Latitude	Distinct: 9709	Unique: 9708 (100%)	
Missing: 0 (0%)			
Statistic		Value	
Minimum		-7.533	
Maximum		2.907	
Mean		-1.116	
StdDev		1.577	

Selected attribute		Type: Numeric	
Name: Longitude	Distinct: 9635	Unique: 9560 (98%)	
Missing: 0 (0%)			
Statistic		Value	
Minimum		117.954	
Maximum		125.058	
Mean		121.844	
StdDev		1.667	

Gambar. 4 Pencocokan kebenaran data atribut

Setelah data dilakukan pre-processing maka terlihat seluruh data yang dimiliki sudah menjadi numerik, selain itu terjadinya penambahan atribut setelah dilakukan proses encoding menggunakan one hot encoding dengan hasil seperti nampak pada tabel 2 dengan penjelasan masing-masing atribut seperti pada tabel 3.

TABEL III
DATASET SETELAH DILAKUKAN PRE-PROCESSING

No	Latitude	Longitude	Magnitude	Mag Type = Mlv	Mag Type = M	Mag Type = mb	Mag Type = Mw	Mag Type = Mw(mB)	Mag Type = ML	Depth (km)	...	Location = Ball Sea
1	0.7	0.6	0.4	1.0	0.0	0.0	0.0	0.0	0.0	0.4	...	0.0
2	0.4	0.2	0.3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
3	0.5	0.2	0.3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
4	0.5	0.2	0.5	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
5	0.5	0.2	0.6	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
6	0.5	0.2	0.4	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
7	0.5	0.2	0.3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
8	0.7	0.6	0.4	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
9	0.4	0.2	0.2	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
10	0.9	0.6	0.3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
...
9701	0.5	0.4	0.4	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
9702	0.6	0.3	0.3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
9703	0.7	0.7	0.5	1.0	0.0	0.0	0.0	0.0	0.0	0.5	...	0.0
9704	0.6	0.3	0.3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
9705	0.7	0.3	0.4	1.0	0.0	0.0	0.0	0.0	0.0	0.2	...	0.0
9706	0.7	0.5	0.4	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
9707	0.6	0.3	0.3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
9708	0.5	0.4	0.3	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
9709	0.7	0.6	0.4	1.0	0.0	0.0	0.0	0.0	0.0	0.1	...	0.0
9710	0.4	0.1	0.3	0.0	1.0	0.0	0.0	0.0	0.0	0.1	...	0.0

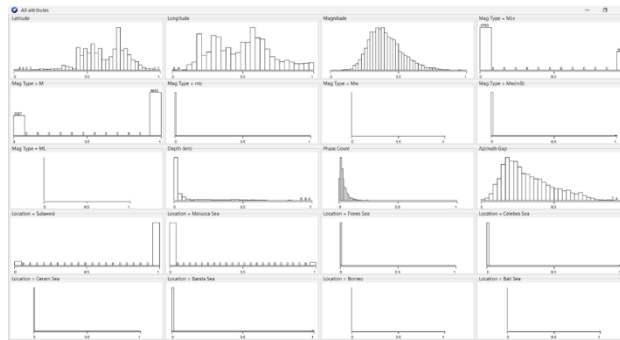
TABEL IV
ATRIBUT GEMPA

Atribut	Keterangan
Latitude	: Atribut yang berisikan informasi titik garis lintang dari pusat gempa yang terjadi.
Longitude	: Atribut yang berisikan informasi titik garis bujur dari pusat gempa yang terjadi.
Mag Type	: Atribut yang berisikan informasi tentang Tipe magnitudo gempa bumi seperti Mlv, Mw, MwP, dll.
Depth (KM)	: Atribut yang berisikan informasi kedalaman setiap gempabumi yang tercatat.
Phase Count	: Atribut yang Jumlah kedatangan fase stasiun dari gempa bumi.
Azimut Gap	: Atribut yang berisi informasi sudut maksimum yang memisahkan dua stasiun seismik yang berdekatan, keduanya diukur dari pusat episenter gempa bumi. Jika azimutal gap lebih dari 180°, akurasi lokasi menurun secara signifikan. Kemampuan deteksi suatu jaringan dapat dievaluasi berdasarkan hubungan antara magnitudo dan jarak pengamatan.

Location : Atribut Nama lokasi episenter gempa bumi merujuk pada Wilayah Flinn-Engdahl (F-E). Wilayah F-E terdiri dari serangkaian zona seismik yang saling berdekatan yang meliputi permukaan bumi. Dalam seismologi, wilayah-wilayah ini menjadi standar untuk lokalitas gempa bumi. Skema ini diusulkan pada tahun 1965 oleh Edward A.

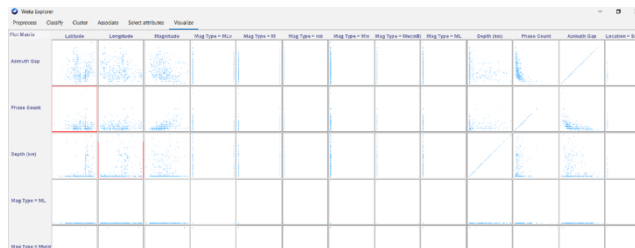
3.3. Visualisasi Data

Pada gambar 4 diinformasikan terkait dengan perilaku data terhadap kelas yang digunakan dan ditunjukkan dengan diagram bar untuk 20 atribut data yang dimiliki.



Gambar. 5 Visualisasi rentang data atribut

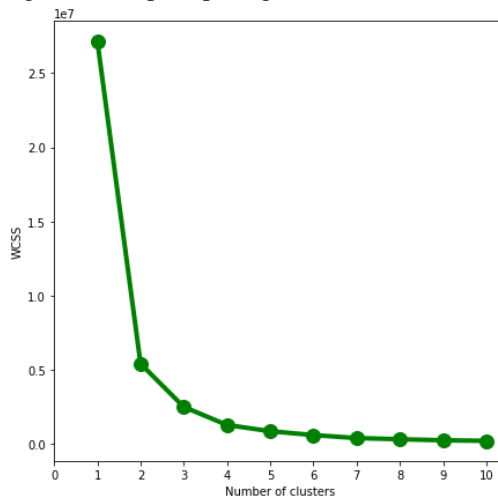
Selanjutnya dilakukan pengamatan terkait dengan korelasi antar atribut dengan masing-masing menunjukkan karakteristik yang unik. Dalam grafik yang ditunjukkan pada gambar 5 teramati beberapa pola penyebaran data yang tercampur maupun data yang terpisah.



Gambar. 6 Visualisasi scatter plot

1) K-Means

Pada proses ini akan menggunakan atribut Magnitude dan Depth (km), untuk mengetahui nilai K terdekat menggunakan metode elbow dengan hasil seperti pada gambar 6.



Gambar. 7 Visualisasi Elbow

Berdasarkan grafik pada gambar 6 yang ditampilkan, maka akan digunakan nilai $k=3$, sehingga data hasil proses k-means dapat ditampilkan seperti pada tabel 5.

TABEL V
HASIL KLASTERISASI NILAI $K=3$ BERDASARKAN MAGNITUDE DAN DEPTH

No	Latitude	Longitude	Magnitude	Depth (km)	Phase Count	Cluster
1	0,252627194	122,3435059	3,021604766	131	7	0
2	-2,94203496	119,4213715	2,439421611	12	7	1
3	-1,82079291	119,384491	2,882417838	10	10	1
4	-1,80569446	119,502739	3,722964817	10	12	1
5	-1,83521318	119,4530029	4,520827759	10	20	1
...
9706	0,016925327	121,9864502	2,915738217	250	19	2
9707	-1,62722754	120,2110672	2,516981369	10	24	1
9708	-2,44550371	120,8326416	2,842521675	10	33	1
9709	-0,23370398	122,9118118	3,116213441	35	26	1
9710	-2,99743843	118,5758286	2,776080996	28	13	1

Gambaran sejauh mana data dalam masing-masing cluster tersebar dari centroidnya, di mana nilai yang lebih rendah mengindikasikan kekompakan dan homogenitas cluster. Analisis centroid pada setiap atribut memberikan wawasan tentang karakteristik khusus dari setiap kelompok. Pada tabel 6 ditampilkan terkait jumlah data terkait gempa pada masing-masing cluster, sedangkan pada tabel 7 ditampilkan terkait banyaknya frekuensi gempa yang terjadi pada masing-masing cluster.

TABEL VI
TOTAL GEMPA BUMI BERDASARKAN CLUSTER

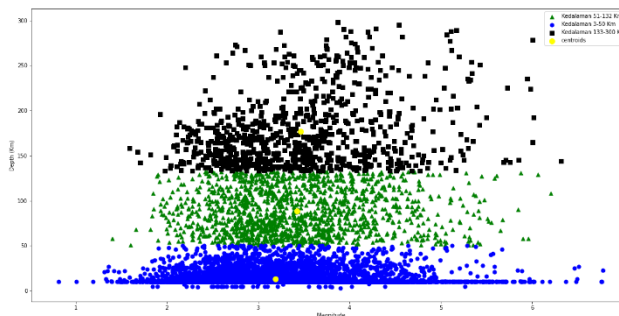
Cluster	Total Gempa	Prosentase	Simbol
Kedalaman 51-132 Km	1.311	13,5%	▲
Kedalaman 3-50 Km	7.527	77,52%	●
Kedalaman 133-300 Km	872	9,98%	■
Total	9.710		

TABEL VII
FREKUENSI GEMPA BUMI BERDASARKAN CLUSTER

Cluster	Frekuensi Gempa	Prosentase	Simbol
Kedalaman 51-132 Km	42.223	17,93%	▲
Kedalaman 3-50 Km	167.979	71,34%	●
Kedalaman 133-300 Km	25.273	10,73%	■
Total	235.475		

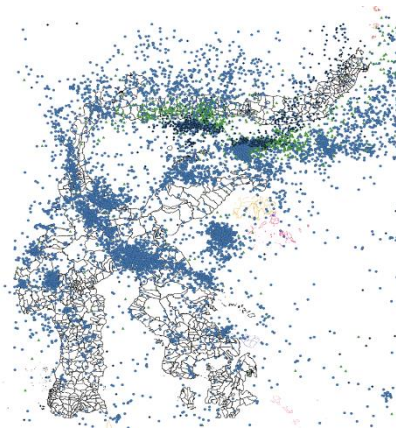
Hasil evaluasi pada data latih menunjukkan distribusi clustered instances di dalam kelompok-kelompok hasil analisis k-Means. Terdapat tiga cluster yang terbentuk, dan distribusi instance dalam masing-masing cluster memberikan gambaran proporsi relatif dari setiap kelompok. Cluster 1 mencapai proporsi tertinggi dengan 77,52% dari total data, sementara Cluster 2 memiliki kontribusi paling kecil hanya 9,98%, sedangkan Cluster 0 memberikan kontribusi sebesar 13,5%. Penerapan hasil ini dapat memberikan pemahaman cepat tentang proporsi dan ukuran relatif dari masing-masing kelompok dalam dataset gempa bumi. Sebagai contoh, Cluster 1, sebagai kelompok terbesar, menunjukkan pola atau karakteristik yang dominan dalam data. Selain itu jika berdasarkan data frekuensi gempa yang terjadi menunjukkan bahwa Cluster 0 sebanyak 17,93%, cluster 1 sebanyak 71,34%, dan cluster 2 sebesar 10,73%.

Visualisasi terkait dengan klasterisasi kondisi gempa dapat dilihat pada gambar 7.



Gambar. 8 Visualisasi Klastering gempa bumi

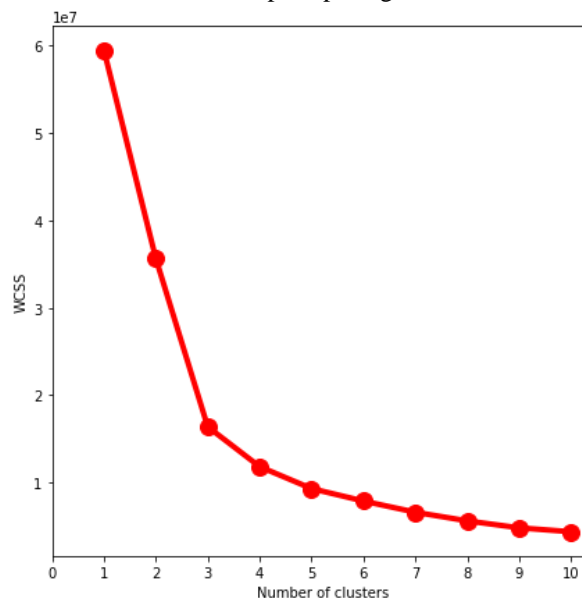
Dapat dijelaskan bahwa hasil dari pemodelan menggunakan k-means didapatkan tingkat kedalaman yang dangkal dengan magnitude yang tinggi berada pada warna biru, sehingga dapat menjadi perhatian pemangku kebijakan terkait dengan kondisi di Sulawesi yang sering terkena gempa, dengan kedalaman seperti itu perlu pengawasan terkait dengan diiringi bencana lain seperti gelombang tsunami. Untuk lebih jelas terkait dengan lokasi gempa yang terjadi pada pulau Sulawesi tersebut, dapat dilihat pada gambar 8 terkait dengan pemetaan wilayah gempa.



Gambar. 9 Visualisasi mapping klastering gempa bumi

Berdasarkan pada hasil visualisasi ditampilkan bahwa daerah pada provinsi Sumatera Selatan menunjukkan tingkat rawan gempa bumi yang tinggi dan frekuensi yang sering pada seluruh cluster dengan pusat titik gempa berada pada Lautan Maluku, antara Sulawesi Selatan dengan Sulawesi Tenggara, serta pada provinsi Gorontalo. Berdasarkan data yang nampak pada gambar 8 terlihat bahwa terjadi aktivitas gempa yang sangat dominan pada provinsi tersebut. Selanjutnya dapat dilihat juga pada pesisir pantai di sebelah Teluk Tomoni berdekatan dengan provinsi Gorontalo.

Selanjutnya dengan menggunakan atribut Magnitude, Depth (km) dan Azimuth Gap juga menggunakan metode elbow untuk mengetahui nilai K terdekat seperti pada gambar 9.



Gambar. 10 Visualisasi Elbow

Berdasarkan grafik pada gambar 9 yang ditampilkan, maka akan digunakan nilai $k=4$, sehingga data hasil proses k-means dapat ditampilkan seperti pada tabel 8.

TABEL VIII
HASIL KLASTERISASI NILAI K=4 BERDASARKAN MAGNITUDE DAN DEPTH

Latitude	Longitude	Magnitude	Depth (km)	Azimuth Gap	Phase Count	Cluster
0,252627194	122,3435059	3,021604766	131	139,9782639	7	2
-2,94203496	119,4213715	2,439421611	12	149,2787399	7	3
-1,82079291	119,384491	2,882417838	10	189,9473495	10	1
-1,80569446	119,502739	3,722964817	10	181,2581348	12	1
-1,83521318	119,4530029	4,520827759	10	184,5670452	20	1
...
0,016925327	121,9864502	2,915738217	250	78,54156494	19	2
-1,62722754	120,2110672	2,516981369	10	65,21083069	24	0
-2,44550371	120,8326416	2,842521675	10	48,9574585	33	0
-0,23370398	122,9118118	3,116213441	35	94,89404678	26	0
-2,99743843	118,5758286	2,776080996	28	144,5652351	13	3

Gambaran sejauh mana data dalam masing-masing cluster tersebar dari centroidnya, di mana nilai yang lebih rendah mengindikasikan kekompakan dan homogenitas cluster. Analisis centroid pada setiap atribut memberikan wawasan tentang karakteristik khusus dari setiap kelompok. Pada tabel 9 ditampilkan terkait jumlah data terkait gempa pada masing-masing cluster, sedangkan pada tabel 10 ditampilkan terkait banyaknya frekuensi gempa yang terjadi pada masing-masing cluster.

TABEL IX
TOTAL GEMPA BUMI BERDASARKAN CLUSTER

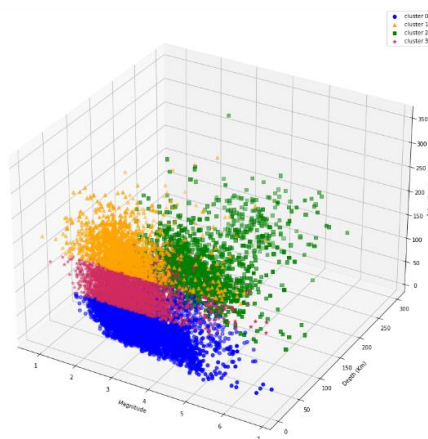
Cluster	Total Gempa	Prosentase	Simbol
Cluster 0	3.957	40,75%	●
Cluster 1	1.546	15,92%	▲
Cluster 2	1.458	15,02%	■
Cluster 3	2.749	28,31%	★
Total	9.710		

TABEL X
FREKUENSI GEMPA BUMI BERDASARKAN CLUSTER

Cluster	Frekuensi Gempa	Prosentase	Simbol
Cluster 0	131.054	55,68%	●
Cluster 1	16.672	7,08%	▲
Cluster 2	44.353	18,84%	■
Cluster 3	43.396	18,44%	★
Total	235.475		

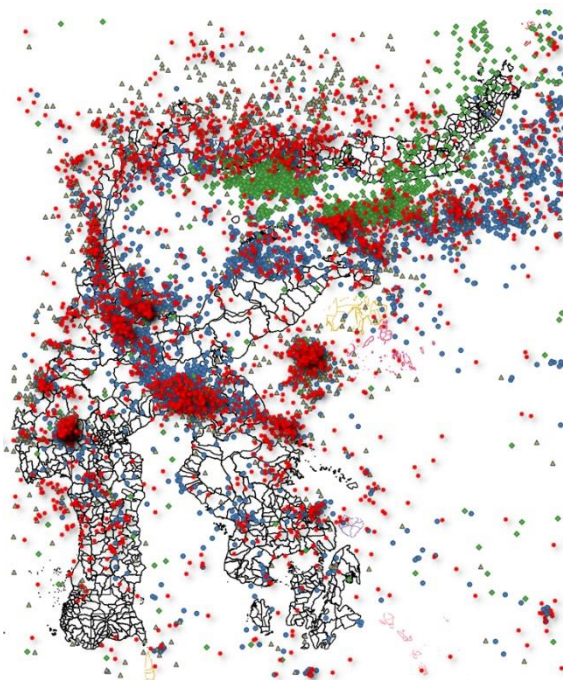
Hasil evaluasi pada data latihan menunjukkan distribusi clustered instances di dalam kelompok-kelompok hasil analisis k-Means. Terdapat tiga cluster yang terbentuk, dan distribusi instance dalam masing-masing cluster memberikan gambaran proporsi relatif dari setiap kelompok. Cluster 0 mencapai proporsi tertinggi dengan 40,75% dari total data, sementara Cluster 2 memiliki kontribusi paling kecil hanya 15,02%, sedangkan Cluster 1 dan 3 masing-masing memberikan kontribusi sebesar 15,92% dan 28,31%. Penerapan hasil ini dapat memberikan pemahaman cepat tentang proporsi dan ukuran relatif dari masing-masing kelompok dalam dataset gempa bumi. Sebagai contoh, Cluster 0, sebagai kelompok terbesar, menunjukkan pola atau karakteristik yang dominan dalam data. Selain itu jika berdasarkan data frekuensi gempa yang terjadi menunjukkan bahwa Cluster 0 sebanyak 55,68%, cluster 1 sebanyak 7,08%, cluster 2 sebesar 18,84%, dan cluster 3 sebesar 18,44%.

Visualisasi terkait dengan klasterisasi kondisi gempa dapat dilihat pada gambar 10.



Gambar. 11 Visualisasi Klastering gempa bumi

Dapat dijelaskan bahwa hasil dari pemodelan menggunakan k-means didapatkan tingkat kedalaman 40-90 Km dengan magnitudo yang tinggi berada pada warna biru, sehingga dapat menjadi perhatian pemangku kebijakan terkait dengan kondisi di Sulawesi yang sering terkena gempa, dengan kedalaman seperti itu perlu pengawasan terkait dengan diiringi bencana lain seperti gelombang tsunami. Untuk lebih jelas terkait dengan lokasi gempa yang terjadi pada pulau Sulawesi tersebut, dapat dilihat pada gambar 11 terkait dengan pemetaan wilayah gempa.



Gambar. 12 Visualisasi mapping klastering gempa bumi

Berdasarkan pada hasil visualisasi ditampilkan bahwa daerah pada provinsi Sumatera Selatan menunjukkan tingkat rawan gempa bumi yang tinggi dan frekuensi yang sering pada seluruh cluster dengan pusat titik gempa berada pada Lautan Maluku, antara Sulawesi Selatan dengan Sulawesi Tenggara, serta pada provinsi Gorontalo. Berdasarkan data yang nampak pada gambar 11 terlihat bahwa terjadi aktivitas gempa yang sangat dominan pada provinsi tersebut. Selanjutnya dapat dilihat juga pada pesisir pantai di sebelah Teluk Tomoni berdekatan dengan provinsi Gorontalo. Hasil tersebut yang ditampilkan menghasilkan lokasi yang sama dengan yang diinformasikan pada gambar 8.

Dalam konteks gambar, hal ini dapat membantu para peneliti atau pengambil kebijakan untuk mengidentifikasi pola seismik dan wilayah yang memiliki risiko gempa serupa, yang mana bisa sangat bermanfaat dalam manajemen risiko bencana dan perencanaan respons darurat.

4. PENUTUP

Hasil penelitian ini memiliki dampak signifikan dalam memahami pola gempa bumi dan dapat dijadikan landasan untuk pengembangan strategi mitigasi risiko bencana gempa. Melalui pemrosesan data yang cermat dan penggunaan algoritma K-Means telah berhasil mengelompokkan pola gempa ke dalam klaster yang memiliki karakteristik berbeda.

Pemahaman ini dapat membantu pihak berwenang dalam mengidentifikasi pola gempa secara lebih efisien, memahami variasi geografis dan magnitudo, serta mengidentifikasi wilayah yang lebih rentan terhadap gempa bumi. Provinsi Sumatera Selatan menunjukkan tingkat rawan gempa bumi yang tinggi dan frekuensi yang sering pada seluruh cluster dengan pusat titik gempa berada pada Lautan Maluku, antara Sulawesi Selatan dengan Sulawesi Tenggara, serta pada provinsi Gorontalo. Selanjutnya, data hasil pengelompokan ini dapat digunakan untuk mengembangkan model prediktif, memperkuat peringatan dini, dan merancang rencana mitigasi bencana yang lebih spesifik. Dengan menyatukan informasi ini, penelitian ini berpotensi memberikan kontribusi positif dalam upaya melindungi masyarakat dan infrastruktur dari risiko gempa bumi yang serius.

DAFTAR PUSTAKA

- [1] ESDM, "Gempa Bumi Di Wilayah Donggala Mengikuti Pola Patahan Palu-Koro," 2018. <https://www.esdm.go.id/id/media-center/arsip-berita/gempa-bumi-di-wilayah-donggala-mengikuti-pola-patahan-palu-koro>.
- [2] "Indonesia di Lingkaran Api Pasifik," 2019. <https://indonesiabaik.id/infografis/indonesia-di-lingkaran-api-pasifik>.
- [3] L. D. Muda, "Gorontalo Dikepung Sesar Aktif Pemicu Gempa, Masyarakat Pesisir dan Lereng Butuh Mitigasi Bencana," 2021. <https://www.ung.ac.id/home/berita/gorontalo-dikepung-sesar-aktif-pemicu-gempa-masyarakat-pesisir-dan-lereng-butuh-mitigasi-bencana>.
- [4] A. Arif, "Sesar Matano Berpotensi Jadi Sumber Gempa Besar Berikutnya," *KOMPAS*, 2017. <https://www.kompas.id/baca/humaniora/2023/03/01/sesar-matano-berpotensi-jadi-sumber-gempa-besar-berikutnya>.
- [5] P. Setyaningrum, "4 Sesar Aktif di Sulawesi, Ada Sesar Palu Koro yang Pernah Memicu Gempa dan Tsunami," 2022. <https://makassar.kompas.com/read/2022/11/27/201914378/4-sesar-aktif-di-sulawesi-ada-sesar-palu-koro-yang-pernah-memicu-gempa-dan?page=all>.
- [6] B. Ku, G. Kim, J. K. Ahn, J. Lee, and H. Ko, "Attention-Based Convolutional Neural Network for Earthquake Event Classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 12, pp. 2057–2061, 2021, doi: 10.1109/LGRS.2020.3014418.
- [7] O. M. Saad, A. G. Hafez, and M. S. Soliman, "Deep Learning Approach for Earthquake Parameters Classification in Earthquake Early Warning System," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 7, pp. 1293–1297, 2021, doi: 10.1109/LGRS.2020.2998580.
- [8] M. A. Murti, R. Junior, A. N. Ahmed, and A. Elshafie, "Earthquake multi-classification detection based velocity and displacement data filtering using machine learning algorithms," *Sci. Rep.*, vol. 12, no. 1, pp. 1–12, 2022, doi: 10.1038/s41598-022-25098-1.
- [9] K. P. Sinaga and M. S. Yang, "Unsupervised K-means clustering algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
- [10] L. Peng, W. Jing-Chao, L. Bin, L. Dan-Yu, C. Yong-Feng, and Y. Jia-Lun, "Research on data mining of wind disaster of power transmission line based on clustering analysis," *Proc. - 2019 6th Int. Conf. Inf. Sci. Control Eng. ICISCE 2019*, pp. 459–463, 2019, doi: 10.1109/ICISCE48695.2019.00098.
- [11] P. Prihandoko and B. Bertalya, "a Data Analysis of the Impact of Natural Disaster Using K-Means Clustering Algorithm," *Kursor*, vol. 8, no. 4, p. 169, 2017, doi: 10.28961/kursor.v8i4.109.
- [12] W. Jianguo and X. Linyao, "Application of K-means Algorithm in Geological Disaster Monitoring System," *Int. J. Adv. Network, Monit. Control.*, vol. 3, no. 3, pp. 16–22, 2019, doi: 10.21307/ijanmc-2019-002.
- [13] D. Wu, Y. Zhang, F. Yang, S. Cheng, H. Sun, and T. Sun, "Improved k-means algorithm based on optimizing initial cluster centers," *ICIC Express Lett.*, vol. 7, no. 3 B, pp. 991–996, 2013.
- [14] A. Budiman, S. Defit, and Y. Yunus, "Comparison of Priority Areas and Rehabilitation Risk Areas for Post Disaster by K-Means Method," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 70–75, 2020, doi: 10.37034/jsisfotek.v3i2.112.

- [15] N. A. Khairani and E. Sutoyo, "Application of K-Means Clustering Algorithm for Determination of Fire-Prone Areas Utilizing Hotspots in West Kalimantan Province," *Int. J. Adv. Data Inf. Syst.*, vol. 1, no. 1, pp. 9–16, 2020, doi: 10.25008/ijadis.v1i1.13.
- [16] F. Sutomo *et al.*, "Optimization of the K-Nearest Neighbors Algorithm Using the Elbow Method on Stroke Prediction," *J. Tek. Inform.*, vol. 4, no. 1, pp. 125–130, 2023, doi: 10.52436/1.jutif.2023.4.1.839.
- [17] F. Marisa, A. R. Wardhani, W. Purnomowati, A. V. Vitianingsih, A. L. Maukar, and W. Puspitarini, "K-MEANS WITH ELBOW METHOD," vol. 7, no. 2, pp. 307–312, 2023.
- [18] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 336, no. 1, 2018, doi: 10.1088/1757-899X/336/1/012017.
- [19] E. Umargono, J. E. Suseno, and V. G. S. K., "K-Means Clustering Optimization using the Elbow Method and Early Centroid Determination Based-on Mean and Median," vol. 474, no. Isstec 2019, pp. 234–240, 2020, doi: 10.5220/0009908402340240.
- [20] G. Racolte *et al.*, "Spherical K-Means and Elbow Method Optimizations with Fisher Statistics for 3D Stochastic DFN from Virtual Outcrop Models," *IEEE Access*, vol. 10, pp. 63723–63735, 2022, doi: 10.1109/ACCESS.2022.3182332.
- [21] R. Sammouda and A. El-Zaart, "An Optimized Approach for Prostate Image Segmentation Using K-Means Clustering Algorithm with Elbow Method," *Comput. Intell. Neurosci.*, vol. 2021, 2021, doi: 10.1155/2021/4553832.
- [22] H. Humaira and R. Rasyidah, "Determining The Appropriate Cluster Number Using Elbow Method for K-Means Algorithm," 2020, doi: 10.4108/eai.24-1-2018.2292388.
- [23] BMKG, "Skala MMI (Modified Mercalli Intensity)," 2024. <https://www.bmkg.go.id/gempabumi/skala-mmi.bmkg>.
- [24] A. Situmorang, "Jenis-jenis patahan bumi, penyebab gempa dan tsunami," 2023. <https://sains.sindonews.com/read/1097019/766/jenis-jenis-patahan-bumi-penyebab-gempa-dan-tsunami-1683972315>.