



Klasterisasi Dampak Bencana Gempa Bumi Menggunakan Algoritma K-Means di Pulau Jawa

Aan Wahyu^{#1}, Rushendra^{#2}

[#]Fakultas Ilmu Komputer, Universitas Mercu Buana
Jl. Meruya Selatan No. 1 Kembangan Jakarta Barat

¹41517120054@student.mercubuana.ac.id

²Rushendra@mercubuana.ac.id

Abstrak—Pulau Jawa merupakan salah satu daerah rawan gempa bumi dan memiliki populasi penduduk yang padat. Dalam hal ini pemerintah diharuskan untuk memberikan perhatian lebih pada penanggulangan bencana khususnya di pulau jawa. Maka dari itu penelitian ini hadir untuk mengetahui dampak gempa bumi berdasarkan tingkat keparahannya dengan cara melakukan klasterisasi data persebaran dampak bencana gempa bumi (2012 - 2021) dari Badan Nasional Penanggulangan Bencana (BNPB) menggunakan algoritma *K-Means Clustering*. Dari penelitian ini ditemukan bahwa dampak bencana dapat dibagi menjadi 4 klaster. Klaster 1 memiliki dampak bencana paling banyak meliputi meninggal, luka, menderita, mengungsi, kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, kantor, jembatan, dan kios. Klaster ini memiliki Mean Absolute Error (MAE) senilai 0,017, Mean Square Error (MSE) senilai 0,002, standar deviasi senilai 0,255 dan variance senilai 0,065. Klaster 2 meliputi meninggal, luka, menderita, mengungsi, rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, pada korban menderita, mengungsi, kerusakan rumah dan fasilitas ibadah. Klaster ini memperoleh MAE senilai 0,053, MSE senilai 0,011, standar deviasi senilai 0,249 dan variance senilai 0,062. Klaster 3 meliputi korban luka, menderita, mengungsi, kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah dan kantor. klaster ini memperoleh MAE senilai 0,102, MSE senilai 0,039, standar deviasi senilai 0,212 dan variance senilai 0,045. Klaster 0 memiliki dampak bencana paling sedikit meliputi kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, kantor, dan kios. klaster ini memperoleh MAE sebesar 0,021, MSE senilai 0,005, standar deviasi senilai 0,251 dan variance senilai 0,063.

Kata kunci— Bencana, Klaster, Dampak, K-Means, Gempa Bumi

I. PENDAHULUAN

Indonesia merupakan salah satu Negara yang tidak asing dengan bencana Gempa Bumi. BNPB mencatat sebanyak 8.624 kali pada 2020, 11.515 kali gempa terjadi pada 2019, 11.920 kali pada 2018. Beberapa faktor yang mendukung yaitu faktor geologi meliputi jenis tanah dan batuan, struktur dan tekstur tanah, topografi, tektonik, dan pola pengaliran sungai [1]. Lokasi geografis Indonesia terletak di lempeng tektonik yang dikenal dengan sebutan wilayah

ring of fire (cincin api). Dimana wilayah *ring of fire* terkenal dengan banyaknya gempa bumi yang terjadi disana. Dengan pergerakan lempeng tektonik ini dapat menyebabkan terbentuknya rangkaian gunung api aktif dan jalur gempa bumi. [2].

Berdasarkan kajian seismogenetik, pulau jawa merupakan daerah rawan gempa dengan sistem satuan seismotektonik busur sangat aktif yang meliputi Jawa Barat dan sumatra, serta busur aktif bagian Jawa Barat, Jawa Tengah, Jawa Timur [3]. Kekuatan gempa, jarak sumber, kepadatan penduduk, kondisi geologi setempat, dan infrastruktur merupakan hal-hal penentu indeks bencana dan resiko di pulau Jawa.[3]. Berdasarkan pemetaan daerah rawan bencana, terdapat 4 provinsi yang memiliki tingkat kerawanan yang tinggi yaitu Aceh, Jawa Tengah, Jawa Timur, dan Jawa Barat [4]. Jika dilihat dari tingkatan tersebut, pulau jawa merupakan wilayah yang mendominasi rawan. Seperti diketahui bahwa pulau jawa memiliki populasi penduduk yang padat. Dikarenakan alasan tersebut, pemerintah diharuskan untuk memberikan perhatian lebih pada penanggulangan bencana gempa bumi khususnya di pulau Jawa. Dalam hal ini, perlunya mengetahui bagaimana dampak gempa bumi yang terjadi di pulau Jawa. Maka dari itu dibutuhkan sebuah algoritma untuk mempermudah proses klasterisasi dampak gempa bumi.

Beberapa penelitian yang terkait dengan bencana alam, data mining, dan data science pernah dilakukan. Hasil dari pengelompokan data menggunakan metode k-means dalam penelitian menunjukkan bahwa, provinsi Jawa Tengah dan Jawa Barat adalah dua provinsi yang paling rawan dilanda bencana gempa bumi di pulau Jawa [5]. Pengelompokan episentrum gempa di Provinsi Bengkulu menggunakan k-means clustering dengan euclidean distance. Variabelnya adalah garis lintang, garis bujur dan besaran. Jumlah cluster optimum ditentukan dengan menggunakan indeks Krzanowski dan Lai (KL) yaitu 7 [6]. Klasterisasi gempa di Provinsi Nusa Tenggara Barat menggunakan metode k-means yang bertujuan sebagai acuan terkait mitigasi bencana gempa bumi dan evaluasi wilayah dan tata ruang [7]. Algoritma Geo K-means dengan memasukan informasi kontekstual dalam proses pengelompokan. Pendekatan ini

untuk diagnosis pasca-gempa yang lebih akurat tentang perilaku geodinamika daerah yang terkena dampak [8]. Pengelompokan produk menggunakan metode k-means clustering untuk merancang ukuran kotak dan menemukan lapisan paling banyak yang dapat ditumpuk pada palet [9]. K-means sebagai sumber titik yang dapat digunakan untuk mempartisi katalog hiposenter gempa dan sebagai pengembangan sumber baris baru dari algoritma yang sesuai dalam seismologi [10].

Hal ini membuktikan klusterisasi k-means mampu memaksimalkan persamaan karakteristik dalam kluster dan memaksimalkan perbedaan antar kluster untuk mendapatkan kelompok data yang diinginkan. Algoritma K-means ini memiliki keunggulan yaitu sederhana dan mudah dipahami, cepat pemrosesannya, mudah penerapannya, dan tersedia di berbagai tools.

Dengan demikian penelitian ini hadir dengan memanfaatkan algoritma klusterisasi k-means untuk melihat dampak gempa bumi di pulau Jawa secara

menyeluruh dengan cara mengelompokan tingkat keparahan dampak gempa bumi. Hal ini diharapkan dapat mempermudah proses penanggulangan bencana dan dapat membantu pihak – pihak terkait dalam pengambilan keputusan.

II. METODOLOGI

A. Pengumpulan Dataset

Dataset yang digunakan adalah data persebaran dampak bencana gempa bumi (2012 – 2021) dari sumber gis.bnpb.go.id dan dbi.bnpb.go.id [11]. Gambar 1 menunjukkan dataset persebaran dampak gempa bumi memiliki 15 parameter: meninggal, hilang, luka, menderit, mengungsi, rumah, fasilitas pendidikan (fasdik), fasilitas kesehatan (faskes), fasilitas ibadah (fasib), fasilitas umum (fasum), kantor, jembatan, pabrik, dan kios. Dataset yang belum dinormalisasi dengan total seluruh data 147 data atau baris. Tabel 1 merupakan contoh dataset yang digunakan.

TABEL I
CONTOH DATASET

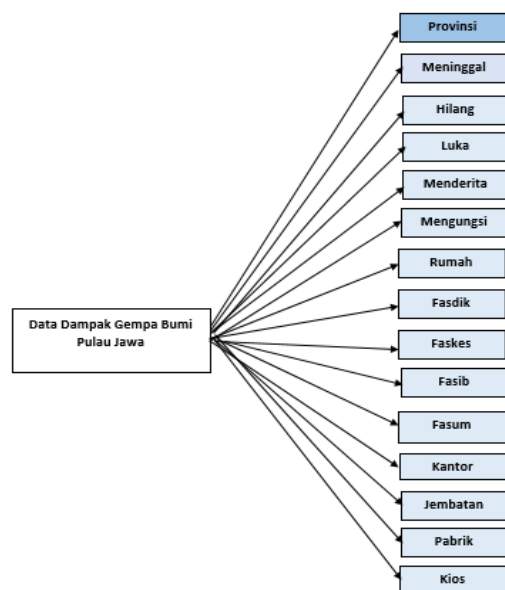
prov	meni nggal	hilang	luka	mende rita	meng ungsi	rumah	fasdik	faskes	fasib	fasum	kantor	jem bata	pabrik	kios
DKI JAKARTA	0	0	1	0	0	0	0	0	0	0	3	0	0	0
JAWA BARAT	1	0	0	0	285	1,847	15	0	25	0	2	0	0	0
BANTEN	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JAWA TIMUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JAWA TENGAH	0	0	1	0	0	0	0	0	0	0	0	0	0	0
JAWA TIMUR	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DI YOGYAKARTA	1	0	0	0	0	0	0	0	0	0	0	0	0	0
DI YOGYAKARTA	0	0	10	0	0	1	0	0	0	0	0	0	0	0
JAWA BARAT	0	0	0	0	0	0	0	0	0	0	0	0	0	0
JAWA TIMUR	0	0	1	0	0	77	1	135	2	0	145	0	0	0

Dari informasi dataset didapatkan data masih sangat acak, tidak urut berdasarkan tanggal, masih terdapat data float, dan masih terdapat beberpa parameter yang tidak digunakan dan tidak diperlukan dalam proses eksperimen dan pembuatan model namun tidak ada data yang bernilai null atau hilang.

Dataset ini merupakan gambaran dari beberapa provinsi di pulau Jawa, Dengan jumlah terbanyak berasal dari Jawa Barat sebanyak 58 dan terbanyak kedua yaitu berasal dari Jawa Timur. Tabel 2 menunjukkan jumlah tiap provinsi yang memiliki catatan gempa bumi dari tahun 2012 hingga tahun 2021.

TABEL II
DATA JUMLAH TIAP PROVINSI

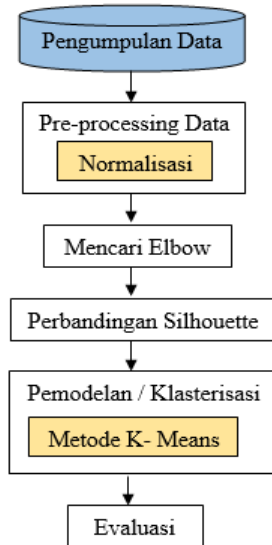
Provinsi	Jumlah
Jawa Barat	58
Jawa Timur	48
Jawa Tengah	22
Banten	11
DI Yogyakarta	7
DKI Jakarta	1



Gambar. 1 Dataset dampak gempa

B. Metode Penelitian

Gambar 2 menunjukkan cara kerja sistem atau alur eksperimen penelitian, terlihat bahwa terdapat beberapa tahapan system sebagai berikut:



Gambar. 2 Metode penelitian

Langkah pertama adalah mengumpulkan Data Persebaran Dampak Bencana Gempa Bumi (2012 – 2021) yang didapat dari BNPB (Badan Nasional Penanggulangan Bencana) Republik Indonesia. Namun data ini masih mentah, sehingga perlu dilakukan preprocessing berupa uji data hilang dan normalisasi untuk mempersiapkan data sebelum digunakan untuk clustering. Sebelum ke tahap pemodelan, dilakukan pencarian kluster terbaik menggunakan metode elbow, serta melakukan perbandingan silhouette menggunakan PCA (*Principal Component Analysis*) dan tanpa PCA. Tahap selanjutnya adalah tahap pemodelan klasterisasi menggunakan k-means untuk mencari pola sebaran dampak bencana dan kemudian di evaluasi menggunakan MSE, MAE, Standar Deviasi dan Analysis of Variance. Berikut penjelasan tiap tahap:

1) *Pre-Processing*: Pada tahap pre-processing dilakukan pengujian missing value dan normalisasi. Pengujian missing value berguna untuk menghindari error dengan cara menghilangkan data yang kosong atau null [12]. Pada tahap pre-processing data yang digunakan sebelum pre-processing sebanyak 147 data atau baris, kemudian setelah dilakukan pre-processing data masih sama sebanyak 147 data atau rows, hal ini menunjukkan bahwa data tidak ada yang kosong atau bernilai null. Setelah itu melakukan tahap normalisasi yang bertujuan untuk mengubah data agar memiliki distribusi yang serupa. Pada kasus ini menerapkan metode min – max scaler yang bekerja dengan menyesuaikan data dalam rentang tertentu yaitu mulai dari 0 hingga 1. Rumus yang digunakan untuk normalisasi adalah sebagai berikut:

$$x' = \frac{(x - \min(x))}{(\max(x) - \min(x))} \quad (3)$$

x' = Nilai dari masing – masing fitur
 $\min(x)$ = Nilai terkecil pada fitur
 $\max(x)$ = Nilai terbesar pada fitur

Sebagai contoh data nilai 15 dengan nilai fitur 15, 20, 25, 30 yang akan dinormalisasi dengan contoh sebagai berikut :

$$x' = \frac{15 - 15}{30 - 15}$$

$$x' = \frac{0}{15}$$

$$x' = 0$$

Dari hasil tersebut dapat disimpulkan nilai sebelum dinormalisasi 15 dan telah dinormalisasi menjadi 0.

2) *Mencari Elbow*: Pada tahap ini akan menentukan jumlah kluster terbaik menggunakan metode Elbow. Metode ini bekerja paling baik dalam menentukan nilai k yang optimal [13]. Dalam metode ini, sebenarnya akan memvariasikan jumlah kluster (k) dari 1-10. Kemudian menghitung jumlah sebuah kluster atau Within- Cluster Sum of Square (WCSS) untuk setiap nilai K. Grafik akan bergerak cepat pada suatu titik sehingga menciptakan bentuk siku. Kemudian grafik akan bergerak mendekati atau sejajar sumbu x. jumlah kluster atau nilai k yang optimal akan menyesuaikan titik ini.

3) *Silhouette Coefficient*: Pada tahap ini akan dilakukan perbandingan metode Silhouette Score menggunakan Principal Component Analysis (PCA) dan tanpa PCA. Silhouette Coefficient digunakan untuk mengevaluasi kluster dan melihat kualitas penempatan data dalam sebuah kluster. Silhouette Coefficient adalah contoh evaluasi yang dilakukan jika kebenaran dasar tidak diketahui, di mana model kluster yang lebih jelas ditentukan oleh skor silhouette coefficient yang lebih tinggi. Silhouette Coefficient terbagi menjadi dua skor sebagai berikut [14]:

- a) Jarak rata – rata antara sampel dan semua titik lain di cluster yang sama.
- b) Jarak rata – rata antara sampel dan semua titik lain di cluster terdekat berikutnya.

Silhouette Coefficient s untuk sampel tunggal diberikan sebagai berikut [14] :

$$s = \frac{b - a}{\max(a, b)} \quad (2)$$

4) *Pemodelan / Klasterisasi*: Pada penelitian ini akan dilakukan pengukuran jarak Euclidean untuk algoritma k-means. Jenis pengukuran jarak ini biasa digunakan dalam data mining. Dalam jarak euclidean, jarak antara dua titik yang didefinisikan sebagai garis lurus. Jarak Euclidean dihitung dengan [15]:

$$DEuclidean(x_i, x_j) = \sqrt{\sum_{m=1}^n (x_{im} - x_{jm})^2} \quad (1)$$

DEuclidean(*x_i*, *x_j*) : Euclidean Distance

x_i: Data -i

x_j: Data -j

x_{im}: Data -i attribute -m

x_{jm}: Data -j attribute -m

Hasil dari pengujian euclidean distance berupa kluster yang sudah built-in ke dalam algoritma k-means dan manifestasi dalam bentuk data centroid. Algoritma *K-means* bertujuan untuk menjadikan nilai rata-rata dari suatu kluster sebagai centroid dari k tersebut [12]. Grup metode pengelompokan K-Means data berdasarkan kedekatan mereka satu sama lain sesuai dengan jarak Euclidean.

K-means analisis kluster adalah proses iteratif karena merupakan algoritma partisi keras [6]. Pertama, data awalnya dipartisi. Setiap grup dihitung meannya dan kemudian data dipartisi lagi dengan mengalokasikan setiap data ke mean terdekatnya posisi kluster. Dalam bentuknya yang paling sederhana, algoritma ini terdiri dari langkah-langkah berikut [12] :

1. Tentukan jumlah k-cluster (secara acak)
2. Menghasilkan nilai acak untuk cluster centroid sebanyak k-cluster.
3. Hitung menggunakan Euclidean Distance untuk mengetahui jarak data input ke setiap centroid.
4. Klasifikasi data yang memiliki jarak terdekat terhadap centroid.
5. Perbarui nilai pusat atau centroid dari rata – rata cluster.
6. Melakukan iterasi atau pengulangan sampai tidak ada yang berubah.

Setelah itu gunakan nilai akhir rata- rata pusat cluster tersebut sebagai parameter untuk menentukan klasifikasi data.

5) *Evaluasi*: Setelah data diklasifikasikan kemudian di evaluasi. Pada tahap evaluasi, model akan diukur performanya. Pemodelan akan diukur menggunakan MSE (Main Squared Error), MAE (Mean Absolute Error), Standar Deviasi, Analysis Variance. Receiver Operating Characteristic Curve.

Nilai tengah galat absolut (mean absolute error)

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (4)$$

Nilai tengah galat kuadrat (mean squared error)

$$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (5)$$

Deviasi standar galat (standard deviation of error)

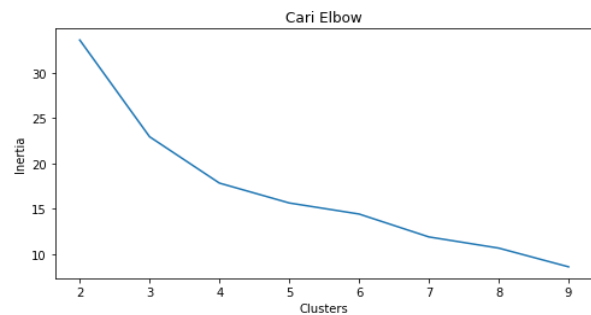
$$SDE = \sqrt{\sum_{i=1}^n e_i^2 / (n-1)} \quad (6)$$

Analysis Variance

$$SDE = \sqrt{\sum_{i=1}^n e_i^2 / (n-1)} \quad (7)$$

III. HASIL DAN PEMBAHASAN

Pada pengujian kluster terbaik menggunakan silhouette membandingkan metode menggunakan PCA (Principal Component Analysis) dan tanpa PCA yang sudah didekomposisi, dengan menggunakan PCA memperoleh hasil 67 %, sedangkan tanpa PCA memperoleh 70%. Hal ini membuktikan parameter terbaik yaitu tanpa menggunakan PCA. Sedangkan dalam mencari jumlah cluster terbaik menggunakan metode elbow. Gambar 2 merupakan grafik elbow yang menunjukkan hasil cluster terbaik berjumlah 4.



Gambar. 1 Grafik Elbow

Berdasarkan hasil klusterisasi menggunakan K-Means, Klusterisasi terbagi menjadi 4 kluster yaitu cluster 0, 1, 2, dan 3. Pada Tabel 3 menampilkan jumlah kerusakan dari tiap – tiap kluster.

TABEL III
HEAT MAP KLASTERISASI

	meninggal	hilang	luka	menderita	mengungsi	rumah	fasdik	faskes	fasib	fasum	kantor	jembatan	pabrik	kios
Cluster 0	1	0	2	0	0	9814	53	1	23	0	9	0	0	3
Cluster 1	12	0	136	1059348	361624	9927	36	14	128	0	35	3	0	4
Cluster 2	12	0	148	6915	2415093	107339	207	162	282	0	201	0	0	1
Cluster 3	5	0	3	2000	109	712	9	0	8	0	0	0	0	0

Berdasarkan Tabel 3 kluster tinggi atau cluster 1 merupakan data yang memiliki dampak bencana paling banyak meliputi meninggal, luka, menderita, mengungsi, kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, kantor, jembatan, dan kios. Berdasarkan data yang diperoleh untuk kluster ini menghasilkan korban menderita paling parah sebanyak 1,059,348 orang.

Klaster tengah atas atau cluster 2 merupakan data yang memiliki dampak bencana paling banyak kedua meliputi meninggal, luka, menderita, mengungsi, rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, pada korban menderita, mengungsi, kerusakan rumah dan fasilitas ibadah. Klaster tengah bawah atau cluster 3 merupakan data yang memiliki dampak bencana pada korban luka, menderita, mengungsi, kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah dan kantor dengan data. Klaster bawah atau cluster 0 merupakan data yang memiliki dampak bencana paling sedikit pada kerusakan rumah, fasilitas pendidikan, fasilitas kesehatan, fasilitas ibadah, kantor, dan kios. Tabel 4 menunjukkan centroid setiap cluster sebagai berikut:

TABEL IV
CENTROID

	Cluster 0	Cluster 1	Cluster 2	Cluster 3
meninggal	125661,336	531772,356	1131164,72	5526394,51
hilang	0.00000000e+00	0.00000000e+00	0.00000000e+00	0.00000000e+00
luka	219483,739	2476397,95	2262329,44	17507039,2
menderita	91911793,8	1933230,14	0.00000000e+00	69211,7282
mengungsi	5899031,08	1651931,14	6,93889E-10	19949658,5
rumah	28312992,9	97944647,5	19245914,7	7141463,46
fasdik	176443,515	1556005,28	88249317,8	6573765,26
faskes	-1,73472E-10	317272,139	282114,822	5379640,49
fasib	272679,813	2705308,94	11582940,3	16634798,6
fasum	0.00000000e+00	0.00000000e+00	0.00000000e+00	0.00000000e+00
kantor	-6,93889E-10	837315,923	4388925,74	16653927,6
jembatan	0.00000000e+00	4881,39048	0.00000000e+00	0.00000000e+00
pabrik	0.00000000e+00	0.00000000e+00	0.00000000e+00	0.00000000e+00
kios	1,73472E-10	245979,903	7407310,78	5930,17354

Tabel 4 merupakan evaluasi pada tiap cluster, berdasarkan hasil evaluasi dapat diperoleh bahwa cluster 1 memiliki nilai MAE dan MSE terendah yaitu 0,017 dan 0,002, hal ini menunjukkan bahwa semakin rendah nilai error yang terjadi maka semakin baik cluster tersebut begitu pula juga sebaliknya. Sedangkan cluster yang memiliki standar deviasi dan variance terbaik yaitu cluster 3 memperoleh standar deviasi 0,212 dan variance 0,045.

Cluster yang memiliki MAE dan MSE yang paling tinggi diperoleh cluster 3 yaitu MAE 0,102 dan MSE 0,039. Sedangkan Standar deviasi dan variance tertinggi diperoleh oleh cluster 1 dengan standar deviasi 0,255 dan variance 0,065.

TABEL V
EVALUASI KLASSTER

	MAE	MSE	Stdv	Variance
Cluster 0	0,021	0,005	0,251	0,063
Cluster 1	0,017	0,002	0,255	0,065
Cluster 2	0,053	0,011	0,249	0,062
Cluster 3	0,102	0,039	0,212	0,045

IV. KESIMPULAN

Setelah dilakukan pengujian dan analisis pada data persebaran dampak bencana gempa bumi (2012 – 2021)

menggunakan algoritma k-means diperoleh hasil bahwa metode k-means bekerja dengan baik tanpa menggunakan (Principal Component Analysis). Klasterisasi menggunakan elbow menghasilkan 4 cluster yaitu cluster tinggi, cluster menengah atas, cluster menengah bawah dan cluster bawah. Setiap cluster memiliki evaluasi yang berbeda, cluster atas atau cluster 1 memiliki nilai MAE dan MSE terbaik yaitu 0,017 dan 0,02 sedangkan cluster yang memiliki standar deviasi dan variance terbaik yaitu cluster menengah bawah atau cluster 3 memperoleh standar deviasi 0,212 dan variance 0,45.

REFERENSI

- [1] Yusra Agustin, "BENCANA GEOLOGI, (Seri Pengetahuan Bencana)," 2016. <https://sumbarprov.go.id/home/news/8753-bencana-geologi-seri-pengetahuan-bencana> (accessed Apr. 18, 2022).
- [2] M. R. Amri *et al.*, *RBI (Risiko Bencana Indonesia)*, vol. 9, no. 3. 2016.
- [3] A. S. Pusat, S. Geologi, B. Geologi, J. Diponegoro, N. 57, and B. Sari, "Seismotektonik dan Potensi Kegempaan Wilayah Jawa."
- [4] B. Supriyadi, A. P. Windarto, T. Soemartono, and Mungad, "Classification of natural disaster prone areas in Indonesia using K-means," *Int. J. Grid Distrib. Comput.*, vol. 11, no. 8, pp. 87–98, 2018, doi: 10.14257/IJGDC.2018.11.8.08.
- [5] A. M. Siregar, "PENERAPAN ALGORITMA K-MEANS UNTUK PENGELOMPOKAN DAERAH RAWAN BENCANA DI INDONESIA," *Intern. (Information Syst. Journal)*, vol. 1, no. 2, pp. 1–10, Feb. 2019, doi: 10.32627/INTERNAL.V1I2.42.
- [6] P. Novianti, D. Setyorini, and U. Rafflesia, "K-means cluster analysis in earthquake epicenter clustering," *Int. J. Adv. Intell. Informatics*, vol. 3, no. 2, pp. 81–89, Jul. 2017, doi: 10.26555/IJAIN.V3I2.100.
- [7] H. J. Wattimanela, "Grouping of Tectonic Earthquakes in the Province of Nusa Tenggara Barat Indonesia with K-Means Cluster Method Approach and Determination of Distribution Type," *Sci. Nat.*, vol. 2, no. 3, pp. 177–191, Sep. 2019, doi: 10.30598/SNVOL2ISS3PP177-191YEAR2019.
- [8] F. Mato and T. Toulkeridis, "An unsupervised K-means based clustering method for geophysical post-earthquake diagnosis," *2017 IEEE Symp. Ser. Comput. Intell. SSCI 2017 - Proc.*, vol. 2018-January, pp. 1–8, Feb. 2018, doi: 10.1109/SSCI.2017.8285216.
- [9] S. Plungsri and K. Puntusavase, "K-means Clustering for Grouping Product Size for Reducing Cost Packaging," *J. Ind. Technol.*, vol. 16, no. 2, pp. 30–44, Jul. 2020, doi: 10.14416/J.IND.TECH.2020.07.003.
- [10] G. Weatherill and P. W. Burton, "Delineation of shallow seismic source zones using K-means cluster analysis, with application to the Aegean region," *Geophys. J. Int.*, vol. 176, no. 2, pp. 565–588, Feb. 2009, doi: 10.1111/J.1365-246X.2008.03997.X/3/176-2-565-FIG016.JPEG.
- [11] "Geoportal Kebencanaan Indonesia." <https://gis.bnpp.go.id/#tabel> (accessed Feb. 01, 2022).
- [12] J. Han, M. Kamber, and J. Pei, "Data Mining. Concepts and Techniques, 3rd Edition (The Morgan Kaufmann Series in Data Management Systems)," 2011.
- [13] J. Han, M. Kamber, and J. Pei, "Data Mining : Concepts and Techniques : Concepts and Techniques (3rd Edition)," *Data Min.*, pp. 1–38, 2012, [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/B9780123814791000010>.
- [14] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, no. C, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.
- [15] T. Hertz, *Learning Distance Functions: Algorithms and Applications*. Citeseer, 2006.