

# Pengelompokan Data Gempa Bumi di Indonesia dengan Algoritma K-Means dan DBSCAN

Alifia Putri Widya Hadi<sup>1</sup>, Hasih Pratiwi<sup>2</sup>, Isnandar Slamet<sup>3</sup>

<sup>123</sup>Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Sebelas Maret

**Abstract.** Gempa bumi merupakan salah satu bencana yang tidak dapat dicegah dan terjadi secara tiba-tiba. Selain itu, lokasi pusat, waktu kejadian, dan kekuatan gempa bumi tidak dapat diprediksi secara tepat dan akurat. Indonesia dilewati oleh tiga titik lempeng tektonik besar dunia, sehingga sering mengalami gempa bumi. Perlu dilakukan upaya pencegahan atau mitigasi agar dapat mengurangi dampak gempa bumi. Mitigasi gempa bumi dapat dilakukan dengan cara pengolahan data dengan *clustering* sehingga diketahui karakteristik dari data gempa tersebut. Algoritma K-Means dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) merupakan metode *clustering* yang dapat digunakan untuk mengelompokkan data gempa bumi di Indonesia. Dari hasil analisis didapatkan bahwa algoritma K-Means menghasilkan 8 *cluster* optimal menggunakan metode *silhouette*. DBSCAN menghasilkan 14 *cluster* dan 102 dikategorikan sebagai *noise* dengan menggunakan nilai *MinPts* dan *Eps* optimal sebesar 12 dan 0,13. Nilai *silhouette coefficient* yang didapatkan K-Means sebesar 0,70 sedangkan DBSCAN sebesar 0,73. Dalam analisis ini K-Means mampu melakukan *clustering* lebih baik dibandingkan dengan DBSCAN karena memiliki *cluster error* yang lebih sedikit.

**Keyword.** Gempa Bumi, *Clustering*, K-Means, DBSCAN

## 1. Pendahuluan

Gempa bumi merupakan bencana yang tidak dapat dicegah, terjadi secara tiba-tiba, serta tidak dapat diperkirakan secara akurat lokasi pusatnya, waktu terjadinya, dan kekuatannya secara tepat dan akurat [1]. Indonesia merupakan negara yang dilewati tiga titik lempeng tektonik [2], sehingga sering mengalami gempa bumi. Perlu dilakukan upaya pencegahan atau mitigasi agar dapat mengurangi dampak gempa bumi, seperti berjatuhnya korban jiwa, kerugian materi, dan kerusakan lingkungan. Terdapat beberapa cara yang dapat dilakukan untuk mencegah terjadinya bencana gempa bumi, seperti memperkuat sistem dan sumber daya Badan Penanggulangan Bencana Daerah dan menentukan titik-titik evakuasi berdasarkan aktivitas gempa bumi sebelumnya [3]. Mitigasi gempa bumi dapat dilakukan dengan cara pengolahan data dengan *clustering* sehingga diketahui karakteristik dari data gempa tersebut. *Clustering* merupakan salah satu fungsi *data mining*.

Data mining merupakan sebuah proses penemuan informasi baru yang bermanfaat dari dalam suatu *database* yang sangat besar [4]. Salah satu kelompok *data mining* berdasarkan tugasnya melakukan *clustering* [5]. Algoritma K-Means merupakan salah satu algoritma *clustering*, dimana metode K-Means dapat melakukan pengelompokan atau *clustering* data dengan jumlah besar dengan waktu yang cepat dan efisien [6]. K-Means merupakan sebuah metode berbasis jarak yang membagi atau mengelompokkan data menjadi beberapa *cluster* berdasarkan kemiripan data [7].

Algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) pertama kali diperkenalkan oleh Ester, *et al.* [8]. DBSCAN merupakan sebuah algoritma *clustering* yang membangun area berdasarkan kepadatan terkoneksi (*density-connected*). Setiap objek dari sebuah *cluster* harus mengandung setidaknya sejumlah minimum data yang disimbolkan dengan *Eps* (jarak maksimal) dan *MinPts* (titik minimum).

Algoritma K-Means dan DBSCAN dapat digunakan untuk *clustering* berbagai permasalahan, salah satunya adalah untuk *clustering* data gempa bumi yang terjadi di Indonesia. Data gempa bumi yang diperoleh dari *website* United States Geological Survey (USGS) dapat diolah dan dilakukan *clustering* guna mengetahui karakteristik gempa bumi di Indonesia.

## 2. Metode Penelitian

Pada penelitian ini akan digunakan proses pengambilan data, *pre-processing data*, dan metode *clustering* yang meliputi K-Means dan DBSCAN.

### 2.1 Pengambilan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari *United States Geological Survey* (USGS). Data yang digunakan merupakan data gempa bumi di Indonesia pada tahun 2012 – 2021. Data tersebut berisikan 1.922 data dan 4 variabel, yakni *depth* dan *magnitude* sebagai variabel komponen utama, serta variabel atribut yakni *latitude* dan *longitude*. Penentuan nilai minimal dan maksimal dari variabel penting dalam batasan pembahasan. Ditentukan minimal *magnitude* sebesar 4,5 Mb dengan batasan maksimal kedalaman (*depth*) 300 KM. Penentuan besar minimal dan maksimal variabel tersebut dipilih karena merupakan batasan dari kategori gempa yang berpengaruh atau menimbulkan potensi kerusakan.

### 2.2 Pre-Processing Data

Dalam analisis ini, terdapat dua tahapan *pre-processing data* yang digunakan. *Pre-processing data* tersebut meliputi *cleaning data* dan standarisasi data

- Cleaning data* dilakukan dengan melengkapi data dan menghapus data kosong ataupun data yang tidak digunakan.
- Standarisasi data dilakukan untuk menyamakan data yang tidak sesuai dan untuk menyamakan variabel agar memiliki distribusi yang serupa.

### 2.3 Clustering

**2.3.1 K-Means.** K-Means merupakan algoritma *clustering* yang mengarah pada pembagian  $n$  objek pengamatan ke dalam  $k$  kelompok (*cluster*) dan setiap objek pengamatan dimiliki oleh suatu kelompok dengan rata-rata (mean) terdekat [9]. Prinsip utama dari teknik ini adalah menyusun  $k$  buah mean dari sekumpulan data. Pada algoritma K-Means, dibutuhkan jarak setiap data input terhadap masing-masing *centroid* dengan menggunakan jarak Euclid. Jarak Euclid merupakan jarak yang umumnya digunakan pada *clustering*. Menurut Ediyanti, dkk. [10] jarak Euclid didefinisikan dengan

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1, 2, \dots, n \quad (1)$$

dimana

- $d(x, y)$  : jarak Euclid data  $x$  dan  $y$   
 $x_i$  : objek  $x$  ke  $i$   
 $y_i$  : data  $y$  ke  $i$   
 $n$  : banyaknya objek

Jarak terpendek antara *centroid* dengan objek menentukan titik *cluster* antar-objek. Adapun *centroid* didefinisikan dengan

$$v = \frac{\sum_{i=1}^n x_i}{n}; i = 1, 2, \dots, n \quad (2)$$

dimana

- $v$  : centroid data  
 $x_i$  : objek  $x$  ke  $i$   
 $n$  : banyaknya objek anggota cluster

2.3.2 *DBSCAN. Density-Based Spatial Clustering Algorithm with Noise* (DBSCAN) adalah algoritma untuk *clustering* data berdasarkan kepadatan. Konsep kepadatan yang dimaksud dalam DBSCAN adalah jumlah minimal data dalam radius *Eps*, data tersebut termasuk dalam kategori kepadatan yang diinginkan [11].

Terdapat dua parameter penting untuk *clustering* dengan menggunakan algoritma DBSCAN, yakni:

- Epsilon (Eps)* merupakan jarak maksimal antara dua titik yang diizinkan dalam satu cluster.
- Minimum Points (MinPts)* adalah jumlah minimal titik sehingga dapat dikatakan sebuah cluster.

Terdapat tiga jenis tipe titik dalam proses *clustering* dengan menggunakan algoritma DBSCAN, yakni:

- Core point* adalah titik yang berada dalam satu cluster dan memiliki titik tetangga sedikitnya sebanyak *MinPts*.
- Border point* adalah titik yang berada dalam suatu cluster namun jumlah titik tetangganya kurang dari *MinPts*.
- Noise point* adalah titik yang tidak termasuk anggota cluster manapun.

### 3. Hasil Penelitian

Hasil penelitian membahas mengenai proses analisis dengan menggunakan algoritma K-Means dan DBSCAN.

#### 3.1 Data Uji Clustering

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diambil dari USGS. Data yang diambil merupakan data gempa bumi di Indonesia pada tahun 2012 – 2021 dengan minimal *magnitude* 4,5 Mb dan kedalaman maksimal 300 km. Data tersebut digunakan untuk membandingkan pengelompokan menggunakan algoritma K-Means dan DBSCAN. Selain itu, data variabel *place*, *longitude*, dan *latitude* digunakan sebagai data pendukung guna mempermudah dalam visualisasi data. Langkah selanjutnya adalah melakukan *preprocessing* data. Dalam tahap *preprocessing* data dilakukan *data cleaning*, seleksi data, dan standarisasi data. Dalam penelitian ini, ditemukan beberapa *missing value* dalam data. Oleh karena itu, dilakukan tahap eliminasi data untuk data *missing value*, sehingga data penelitian sudah bersih. Tabel 1 menampilkan sampel data penelitian yang digunakan.

**Tabel 1.** Sampel Data Gempa Bumi di Indonesia Tahun 2012-2021

<i>Place</i>	<i>Latitude</i>	<i>Longitude</i>	<i>Depth</i>	<i>Mag</i>
55 km NE of Meulaboh, Indonesia	4,547	96,419	35	5.2
200 km SSW of Sofifi, Indonesia	-0,983	126,984	188	4.5
60 km NNW of Ende, Indonesia	-8,312	121,518	137	4.9
101 km WSW of Kendari, Indonesia	-4,427	121,72	35	4.5
55 km NE of Meulaboh, Indonesia	4,555	96,407	40	4.6

Dalam Tabel 1 terdapat nilai-nilai yang merupakan nilai asli yang didapatkan melalui website USGS. Tabel 1 merupakan sampel data yang digunakan dalam penelitian ini. Variabel yang digunakan dalam penelitian ini adalah *latitude*, *longitude*, *depth* (kedalaman), dan *magnitude*. Jumlah keseluruhan data yakni 1.922 data gempa bumi di Indonesia yang terjadi selama 2012-2021. Tabel 2 menampilkan sampel data untuk data yang telah di standarisasi

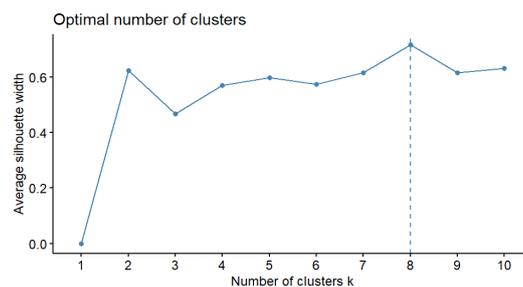
**Tabel 2.** Standarisasi Data Gempa Bumi di Indonesia Tahun 2012-2021

Place	Latitude	Longitude	Depth	Mag
55 km NE of Meulaboh, Indonesia	4,547	96,419	-0,475784012	2.16369748
200 km SSW of Sofifi, Indonesia	-0,983	126,984	0,317768028	-0.93136982
60 km NNW of Ende, Indonesia	-8,312	121,518	0,053250681	0.83724007
101 km WSW of Kendari, Indonesia	-4,427	121,72	-0,475784012	-0.93136982
55 km NE of Meulaboh, Indonesia	4,555	96,407	-0,449850939	-0.48921735

Data yang telah terstandarisasi memiliki jarak atau rentang nilai yang lebih kecil. Data yang telah di standarisasi juga memiliki rentang yang cukup sama antar-variabel. Hal tersebut bertujuan agar tidak adanya salah satu variabel yang saling mendominasi. Dapat dikatakan bahwa data telah siap untuk dilakukan analisis menggunakan metode K-Means dan DBSCAN.

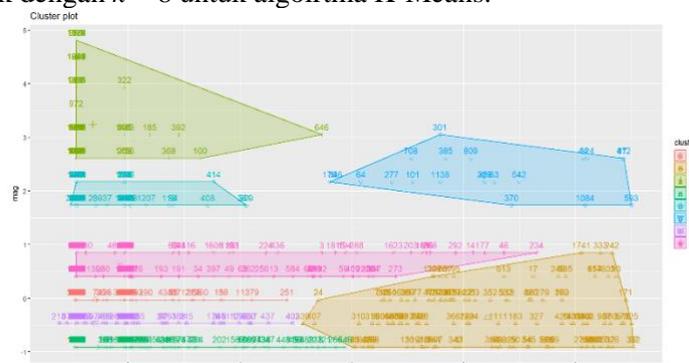
### 3.2 Clustering dengan Algoritma K-Means

Langkah awal dari algoritma K-Means adalah menentukan terlebih dahulu jumlah  $k$  (*cluster*) yang akan dibentuk pada data. Dalam menentukan nilai  $k$  diperlukan metode untuk menentukan jumlah  $k$  optimal. Metode *silhouette* merupakan salah satu metode pembangkit jumlah *cluster* yang dapat digunakan bersama dengan algoritma K-Means. Metode *silhouette* menampilkan informasi nilai  $k$  optimal. Pemilihan jumlah  $k$  optimal pada *silhouette* didapatkan dari nilai rata-ratanya. Semakin tinggi nilai rata-ratanya, maka semakin optimal. Pembangkitan nilai  $k$  optimal dengan metode *silhouette* terlihat dalam Gambar 1.



**Gambar 1.** Penentuan  $K$  Optimal dengan Metode *Silhouette*

Gambar 1 memperlihatkan hasil pembangkitan *cluster* optimal dengan menggunakan metode *silhouette*. Berdasarkan metode *silhouette*,  $k$  atau *cluster* optimal adalah 8. Selanjutnya, dilakukan proses visualisasi *cluster* yang terbentuk dengan  $k = 8$  untuk algoritma K-Means.



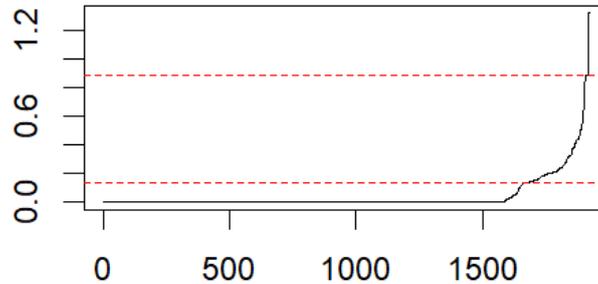
**Gambar 3.** Visualisasi *Cluster* dengan Algoritma K-Means untuk  $K = 8$

Gambar 3 memperlihatkan visualisasi bahwa algoritma K-Means berhasil meng-*clustering* data gempa bumi di Indonesia tahun 2012-2021 menjadi 8 *cluster* dimana *cluster* 1 terdiri dari 276 data, *cluster* 2

terdiri dari 111 data, *cluster* 3 terdiri dari 56 data, *cluster* 4 terdiri dari 513 data, *cluster* 5 terdiri dari 130 data, *cluster* 6 terdiri dari 21 data, *cluster* 7 terdiri dari 428 data, dan *cluster* 8 terdiri dari 387 data.

### 3.3 Clustering dengan Algoritma DBSCAN

Dalam algoritma DBSCAN, penentuan nilai *MinPts* dan *Eps* dilakukan dengan menggunakan K-NN (*K-Nearest Neighbors*). Dilakukan plot K-NN dengan  $k = 15$  untuk data gempa bumi di Indonesia tahun 2012-2021 terlihat dalam Gambar 4.



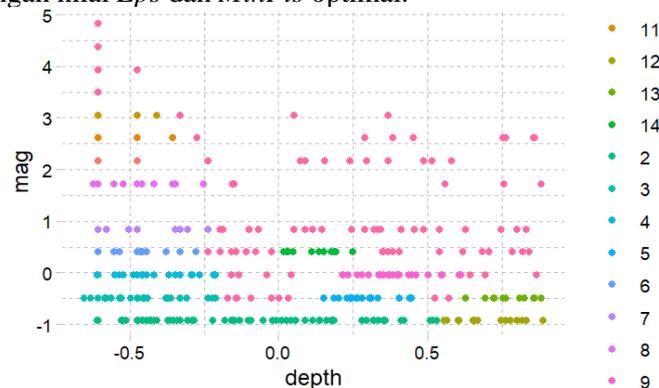
**Gambar 4.** Plot K-NN untuk  $K = 15$

Gambar 6 memperlihatkan bahwa terdapat sebuah patahan pada titik 0,13 dan 0,88 yang ditandai melalui garis bantu berwarna merah yang disebut dengan *knee*. Selanjutnya dilakukan percobaan untuk mengetahui nilai *Eps* optimal dengan mempertimbangkan nilai *silhouette coefficient*-nya. Percobaan dilakukan dengan memasukkan nilai *MinPts* dari 11 sampai 15 dan nilai *Eps* dari 0,13 dan 0,88.

**Tabel 4.** Perbandingan Nilai *MinPts* dan *Eps*

<i>Eps</i>	<i>MinPts</i>	Jumlah Cluster	Jumlah Noise	<i>Silhouette Coefficient</i>
0,13	11	13	94	0,70
	12	14	102	0,73
	13	12	139	0,72
	14	11	160	0,71
	15	10	176	0,71
0,88	11	2	0	0,67
	12	3	0	0,65
	13	3	0	0,65
	14	3	0	0,65
	15	3	0	0,65

Tabel 4 menunjukkan bahwa kombinasi nilai *Eps* 0,13 dengan *MinPts* 12 memiliki nilai *silhouette coefficient* paling besar, yakni 0,73. Oleh karena itu, digunakan kombinasi *Eps* dan *MinPts* di 0,13 dan 12 sebagai kombinasi optimal. Setelah mendapatkan nilai *Eps* dan *MinPts* optimal, dilakukan visualisasi untuk titik *cluster* dengan nilai *Eps* dan *MinPts* optimal.



**Gambar 5.** Visualisasi Cluster dengan Algoritma DBSCAN

Dalam Gambar 5 terlihat dari visualisasi bahwa algoritma DBSCAN berhasil untuk *clustering* data gempa bumi di Indonesia tahun 2012-2021 menjadi 14 *cluster*, dimana *cluster* 1 terdiri dari 40 data, *cluster* 2 terdiri dari 531 data, *cluster* 3 terdiri dari 422 data, *cluster* 4 terdiri dari 273 data, *cluster* 5 terdiri dari 16 data, *cluster* 6 terdiri dari 176 data, *cluster* 7 terdiri dari 174 data, *cluster* 8 terdiri dari 87 data, *cluster* 9 terdiri dari 23 data, *cluster* 10 terdiri dari 21 data, *cluster* 11 terdiri dari 16 data, *cluster* 12 terdiri dari 17 data, *cluster* 13 terdiri dari 12 data, *cluster* 14 terdiri dari 12 data, dan 102 data adalah *noise*.

### 3.4 Perbandingan Algoritma Clustering

Setelah dilakukan analisis dengan menggunakan algoritma K-Means dan DBSCAN, dilakukan perbandingan hasil untuk mendapatkan metode lebih baik dalam *clustering* data gempa bumi di Indonesia tahun 2012-2021. Perbandingan hasil untuk metode lebih baik dilakukan dengan membandingkan nilai *silhouette coefficient* dari masing-masing algoritma. Plot visualisasi dari rata-rata nilai *silhouette coefficient* setiap *cluster* yang terbentuk dengan menggunakan kedua algoritma ditampilkan pada Gambar 6 dan Gambar 7.



**Gambar 6.** Plot *Silhouette Coefficient* Algoritma K-Means

Gambar 6 menampilkan plot dari rata-rata nilai *silhouette coefficient* setiap *cluster* yang terbentuk dengan menggunakan algoritma K-Means. Nilai rata-rata yang dihasilkan dari plot tersebut adalah 0,70 yang menandakan bahwa terdapat ikatan yang cukup kuat antar-objek di dalam *cluster*. Akan tetapi, di dalam *cluster* dua terdapat *silhouette coefficient* yang bernilai negatif, hal tersebut menandakan terdapat data yang masuk ke dalam *cluster* yang salah.



**Gambar 7.** Plot *Silhouette Coefficient* Algoritma DBSCAN

Gambar 7 menampilkan plot dari rata-rata nilai *silhouette coefficient* setiap *cluster* yang terbentuk dengan menggunakan algoritma DBSCAN. Nilai rata-rata yang dihasilkan dari plot tersebut adalah 0,73, yang menandakan bahwa terdapat ikatan yang sangat kuat antar-objek di dalam *cluster*. Akan tetapi, di dalam *cluster* satu dan dua terdapat *silhouette coefficient* yang bernilai negatif, hal tersebut menandakan terdapat data yang masuk ke dalam *cluster* yang salah.

Perolehan nilai *silhouette coefficient* kedua algoritma memiliki perbedaan yang tidak signifikan. Apabila dilihat dari rata-rata nilai *silhouette coefficient* kedua algoritma, maka algoritma DBSCAN merupakan metode lebih baik untuk *clustering* data gempa bumi. Akan tetapi dilihat dari plot *silhouette coefficient* kedua algoritma, kedua algoritma memiliki *silhouette width* yang bernilai negatif, hal tersebut

menunjukkan adanya kesalahan dalam proses *clustering*. Nilai *error* dalam *silhouette width* algoritma K-Means lebih sedikit dibandingkan dengan algoritma DBSCAN, oleh karena itu algoritma K-Means ditentukan sebagai algoritma yang lebih baik dalam *clustering* data gempa bumi di Indonesia tahun 2012-2021.

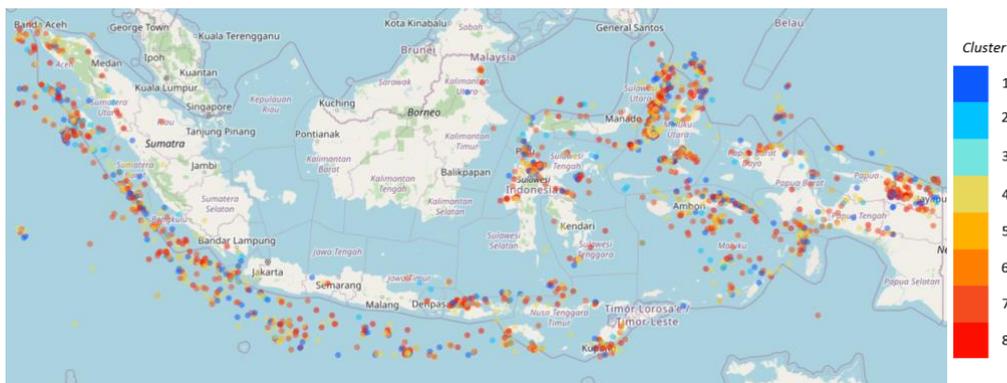
### 3.5 Clustering Gempa Bumi di Indonesia

Setelah disimpulkan bahwa algoritma K-Means merupakan algoritma lebih baik untuk *clustering* data gempa di Indonesia, selanjutnya dilakukan deskripsi data gempa bumi dari hasil *clustering*.

**Tabel 5.** Deskripsi Cluster

Cluster	Variabel	Statistik			Anggota Cluster
		Mean	Min	Maks	
1	Latitude	-2,5657	-10,8850	5,019	276
	Longitude	120,46	94,18	140,71	
	Depth	19,72	10	119	
	Mag	4,7	4,7	4,7	
2	Latitude	-2,8163	-10,834	4,745	111
	Longitude	117,38	94,55	140,73	
	Depth	222,5	127	298	
	Mag	4,625	4,5	4,9	
3	Latitude	-3,7181	-10,768	4,1851	56
	Longitude	120,37	96,89	140,18	
	Depth	18,57	10	137	
	Mag	5,443	5,3	5,8	
4	Latitude	-2,7956	-10,9995	5,7719	513
	Longitude	120,6	94	140,7	
	Depth	22,33	9	149	
	Mag	4,5	4,5	4,5	
5	Latitude	-3,1159	-10,901	5,373	130
	Longitude	120,62	94,62	140,63	
	Depth	19,82	7	98	
	Mag	5,132	5,1	5,2	
6	Latitude	-3,149	-10,653	5,375	21
	Longitude	117,69	97,46	139,02	
	Depth	220,8	141	297	
	Mag	5,229	5,1	5,4	
7	Latitude	-2,714	-10,983	5,924	428
	Longitude	120,2	94,24	140,78	
	Depth	20,1	1	122	
	Mag	4,6	4,6	4,6	
8	Latitude	-2,8012	-10,849	5,6641	387
	Longitude	119,86	94,13	140,59	
	Depth	32,39	10	248	
	Mag	4,85	4,8	4,9	

Dari hasil deskripsi data pada Tabel 5, selanjutnya dilakukan pemetaan dengan menggunakan variabel *latitude* dan *longitude* dari hasil *clustering* menggunakan algoritma K-Means.



**Gambar 8.** Hasil Pemetaan *Clustering* dengan Algoritma K-Means

Dari Gambar 8 terlihat bahwa data gempa bumi di Indonesia berhasil dikelompokkan mejadi 8 *cluster*. *Cluster-cluster* yang terbentuk menyebar hampir di seluruh Indonesia. Dari titik hasil pemetaan *clustering*, dapat disimpulkan bahwa terdapat beberapa daerah yang perlu melakukan langkah mitigasi secara matang karena besarnya potensi bencana yang akan terjadi, seperti daerah yang termasuk dalam *cluster* 3 dan *cluster* 5 dimana daerah tersebut memiliki rata-rata kekuatan gempa sebesar 5,443 dan 5,132 Mb yakni berkekuatan besar dan juga memiliki rata-rata kedalaman gempa yang dangkal.

#### 4. Kesimpulan dan Saran

Data gempa bumi di Indonesia tahun 2012-2021 berhasil dilakukan *clustering* menggunakan algoritma K-Means dan DBSCAN. Dari hasil analisis, algoritma K-Means berhasil membangkitkan 8 *cluster* dengan menggunakan metode *silhouette*, sedangkan algoritma DBSCAN berhasil membangkitkan 14 *cluster* dan 102 data dikategorikan sebagai *noise* dengan menggunakan nilai *MinPts* dan *Eps* optimal sebesar 12 dan 0,13. Nilai *silhouette coefficient* yang dihasilkan K-Means sebesar 0,70, sedangkan DBSCAN 0,73. Dalam analisis ini K-Means mampu melakukan *clustering* lebih baik dibandingkan dengan DBSCAN karena memiliki *cluster error* yang lebih rendah. Saran yang dapat disampaikan untuk penelitian selanjutnya adalah untuk menambahkan algoritma lain, sehingga dapat menunjukkan hasil *clustering* yang lebih optimal.

#### 5. Daftar Pustaka

- [1] D. P. S. Ningsih, I. Rahmawati, R. Aprianti, S. Wulan, V. P. Giena, and Y. Elvira, Penyuluhan tentang Gempa Bumi dengan Media Leaflet pada Masyarakat di Kelurahan Malabero Kota Bengkulu, *Jurnal Pengabdian Kepada Masyarakat Wahana Usada*, vol 4, no. 2, p 96-104. 2022.
- [2] A. Sabtaji, Statistik Kejadian Gempa Bumi Tektonik Tiap Provinsi di Wilayah Indonesia selama 11 Tahun Pengamatan (2009-2019), *Buletin Meteorologi, Klimatologi, Dan Geofisika*, vol. 1, no. 7, pp 31-46. 2020.
- [3] G. Pasau, and A. Tanauma, Pemodelan Sumber Gempa di Wilayah Sulawesi Utara sebagai Upaya Mitigasi Bencana Gempa Bumi, *Jurnal Ilmiah Sains*, vol. 11, no. 2, p 202-209. 2011.
- [4] E. D. Sikumbang, Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori, *Jurnal Teknik Komputer*, vol 4, no. 1. 2018.
- [5] Risman, Syaripuddin, and Suyitno, Implementasi Metode Dbscan Pada Pengelompokan Kabupaten/Kota di Pulau Kalimantan Berdasarkan Indikator Kesejahteraan Rakyat, *Prosiding Seminar Nasional Matematika, Statistika, Dan Aplikasinya*, vol 1, p 22-28. 2019.
- [6] A. Amalia, U. Harmoko, and G. Yuliyanto, Clustering of Seismicity in the Indonesian Region for the 2018-2020 Period using the DBSCAN Algorithm, *Journal of Physics and Its Applications*, Vol. 4, Issue 1. 2021.
- [7] S. Srigantina, A. H. Nasyuha, and S. Suharsil, Implementasi Data Mining Dalam Pengelompokan Data Transaksi Penjualan Kosmetik di WN Kosmetik Dengan Menggunakan Metode K-Means Clustering, *Jurnal Cyber Tech*, vol. 1, no. 12. 2022.

- [8] M. Ester, H. P. Kriegel, J. Sander, and X. Xu, A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Knowledge Discovery and Data Mining*, vol. 96, no. 34, p 226-231. 1996.
- [9] E. Prasetyo, *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Andi, 2021.
- [10] Ediyanti, M. N. Mara, and N. Satyahadewi, Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis, *Buletin Ilmiah Mat. Stat. Dan Terapannya*, vol. 2 no. 2, p 133-136. 2013.
- [11] R. R. A. Rahman, and A. W. Wijayanto, Pengelompokan Data Gempa Bumi Menggunakan Algoritma DBSCAN, *Jurnal Meteorologi Dan Geofisika*, vol. 22, no. 1, p 31-38. 2021.