

IMPLEMENTASI ALGORITMA CLARA UNTUK DATA GEMPA BUMI DI INDONESIA

¹Isna Hidayatur Rifa, ²Hasih Pratiwi, ³Respatiwan

^{1,2,3}Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Sebelas Maret
e-mail: rifaisna@gmail.com

Abstrak

Gempa bumi adalah getaran yang terjadi di permukaan bumi secara tiba-tiba yang disebabkan oleh pergeseran lempeng bumi. Gempa bumi merupakan bencana alam yang sering terjadi di Indonesia. Hal ini dikarenakan Indonesia berada pada tiga lempeng besar dunia dan sembilan lempeng kecil lainnya yang saling bertemu di suatu wilayah membentuk tatanan lempeng yang kompleks. Dampak yang disebabkan oleh gempa bumi berbeda-beda bergantung pada magnitudo dan kedalamannya. Pada penelitian ini akan dilakukan pengelompokan data gempa bumi di Indonesia berdasarkan magnitudo dan kedalamannya menggunakan salah satu teknik data mining yaitu *clustering*. Metode *clustering* yang akan digunakan adalah algoritma *clustering large application* (CLARA). Algoritma CLARA dapat menangani data berukuran besar, dan lebih robust terhadap pencilan. Hasil penelitian menunjukkan bahwa jumlah *cluster* optimal berdasarkan magnitudo dan kedalaman terjadinya gempa adalah tiga *cluster*.

Kata Kunci: gempa bumi, data mining, *clustering*, algoritma CLARA

Abstract

Earthquake is a sudden vibration in the earth's surface caused by a shifting in the earth's plates. Earthquake is a natural disaster that often occur in Indonesia, since Indonesia is located on three largest plates in the world and nine other small plates which meet at one area that form a complex plate arrangement. Earthquake have various impact depending on the magnitude and the depth. The earthquake that commonly occur in Indonesia have various magnitude and depth. This research will classify earthquake data in Indonesia based on their magnitudes and their depths. This classification can be done by using one of the data mining techniques, namely clustering. Clustering method that will be used is clustering large application (CLARA) algorithm. The result shows that there are three clusters found as the optimal cluster based on the magnitude and the depth of the earthquake.

Keywords: earthquake, data mining, clustering, CLARA algorithm

PENDAHULUAN

Gempa bumi adalah getaran yang terjadi di permukaan bumi akibat pergerakan atau pergeseran lempeng bumi secara tiba-tiba. Indonesia menempati zona tektonik yang sangat aktif karena tiga lempeng besar dunia dan sembilan lempeng kecil lainnya saling bertemu di wilayah Indonesia dan membentuk jalur-jalur pertemuan lempeng yang kompleks (Bird, 2003). Menurut *United Nations International Strategy for Disaster Reduction* (UNISDR; Badan PBB untuk Strategi Internasional Pengurangan Risiko Bencana), Indonesia merupakan negara yang paling rawan bencana alam di dunia. Tingginya aktivitas kegempaan ini terlihat dari hasil pencatatan dimana dalam rentang waktu 1897-2009 terdapat lebih dari 14.000 kejadian gempa dengan magnitudo lebih dari 5,0. Gempa-gempa tersebut telah menyebabkan ribuan korban jiwa, keruntuhan dan kerusakan ribuan infrastruktur dan bangunan, serta dana trilyunan rupiah untuk rehabilitasi dan rekonstruksi (Irsyam dkk., 2010). Selama 2018 terjadi aktivitas gempa sebanyak 11.577 kali dalam berbagai magnitudo dan kedalaman, sedangkan pada tahun 2017, jumlah aktivitas gempa yang terjadi hanya 6.929 kali. Hal ini menunjukkan bahwa aktivitas gempa di Indonesia meningkat secara signifikan hanya dalam rentang waktu satu tahun. Untuk mengetahui pengelompokan pola gempa bumi di Indonesia dapat dilakukan dengan salah satu metode yaitu *data mining*.

Penelitian ini dilakukan untuk mengelompokkan data gempa bumi di Indonesia berdasarkan magnitudo dan kedalamannya menggunakan salah satu teknik data mining yaitu *clustering*. Berkhin (2006) mengatakan bahwa pengelompokan dapat dilihat sebagai pemodelan data yang menyediakan ringkasan data yang ringkas. Dalam penelitiannya Artatia dan Hakim (2015) melakukan pengelompokan dampak gempa bumi dari segi kerusakan fasilitas pada provinsi yang berpotensi gempa di Indonesia menggunakan *k-means clustering*. Han *et al.* (2012) menjelaskan bahwa algoritma *k-means*

sensitive terhadap pencilan. Ketika pencilan dikelompokkan ke dalam suatu *cluster*, maka pencilan tersebut dapat mempengaruhi nilai rata-rata dari *cluster*. Kaufman and Rousseeuw (1987) memperkenalkan metode *k-medoids* untuk mengatasi permasalahan pencilan pada metode *k-means*. Oleh karena itu, metode yang digunakan dalam penelitian ini adalah algoritma *k-medoids* yaitu *clustering large application* (CLARA). Algoritma CLARA bekerja efektif pada data berukuran besar dan *robust* terhadap pencilan karena menggunakan *medoid* sebagai pusat *cluster*. Penelitian ini diharapkan dapat membantu pemerintah dalam proses mitigasi bencana, khususnya gempa bumi.

METODE PENELITIAN

Data yang digunakan dalam penelitian ini adalah data gempa bumi di Indonesia tahun 1973-2017 yang diperoleh dari *United States Geological Survey* (USGS). USGS adalah sebuah agensi ilmiah dari pemerintah Amerika Serikat. Atribut yang digunakan dalam penelitian ini adalah magnitudo gempa dan kedalaman terjadinya gempa.

Gempa Bumi

Menurut Mustafa (2010), gempa bumi merupakan sebuah guncangan hebat yang menjalar ke permukaan bumi yang disebabkan oleh gangguan di dalam kulit bumi. Umumnya gempa bumi disebabkan dari pelepasan energi yang dihasilkan oleh tekanan yang dilakukan oleh lempengan yang bergerak. Semakin lama tekanan itu kian membesar dan akhirnya mencapai suatu keadaan dimana tekanan tersebut tidak dapat ditahan lagi oleh pinggiran lempengan. Pada saat itulah gempa bumi akan terjadi, yang energinya menjalar ke berbagai arah.

Data Mining

Data mining adalah proses mempekerjakan teknik *machine learning* untuk menemukan pola dan pengetahuan dari suatu data berukuran besar. Han dan Kamber (2006) mendefinisikan data mining sebagai proses menemukan pola yang menarik dari data yang berukuran besar. Menurut Hermawati (2013), data mining adalah suatu proses iteratif dan interaktif untuk menemukan pola atau model baru yang sah (sempurna), bermanfaat dan dapat dimengerti dalam suatu *database* yang sangat besar. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisis data yang berguna dan kemudian dapat dipelajari dengan lebih teliti. Dalam data mining terdapat beberapa teknik, yaitu asosiasi, klasifikasi, regresi, peramalan dan *clustering*. Teknik data mining yang digunakan untuk mengelompokkan data adalah metode *clustering*.

Clustering adalah pembentukan kelompok dimana suatu objek ditempatkan dalam kelompok yang sama dengan objek yang lain apabila memiliki hubungan, namun menempati kelompok yang berbeda apabila tidak memiliki hubungan. Tujuan dari *clustering* adalah meminimalkan jarak di dalam *cluster* dan memaksimalkan jarak antar *cluster*. Menurut Larose (2005), sebuah *cluster* adalah suatu kumpulan data yang mirip dengan lainnya atau ketidakmiripan data pada kelompok lain dan disimpulkan bahwa *clustering* merupakan usaha untuk mengidentifikasi kelompok objek yang mirip untuk membantu menemukan pola penyebaran dan pola hubungan dalam sekumpulan data yang besar. *Clustering* disebut sebagai *unsupervised learning* yaitu suatu teknik data mining yang tidak memerlukan adanya data latih terlebih dahulu. Algoritma *clustering* yang sering digunakan adalah algoritma *k-means* dan *k-medoids*. Menurut Septiana dan Djohan (2015), *k-means* adalah algoritma *clustering* yang bertujuan untuk membagi sejumlah pengamatan ke dalam *k* kelompok dimana himpunan setiap observasi milik kelompok dengan rata-rata terdekat. Ketika pencilan dikelompokkan ke dalam suatu *cluster*, maka pencilan tersebut dapat mempengaruhi nilai rata-rata dari *cluster*. Kaufman and Rousseeuw (1987) memperkenalkan metode *k-medoids* untuk mengatasi permasalahan pencilan pada metode *k-means*.

Algoritma *K-medoids*

K-medoids adalah algoritma *clustering* yang bertujuan untuk mengelompokkan objek-objek menjadi beberapa *cluster* dengan *medoid* sebagai ukuran pusatnya. *K-medoids* merupakan algoritma *partitioned clustering* selain *k-means*. Menurut Han *et al.* (2012), algoritma *k-means* mempunyai kelemahan terhadap data pencilan karena menggunakan nilai rata-rata sebagai *centroid* sehingga tidak *robust* terhadap pencilan, sedangkan algoritma *k-medoids* menggunakan *medoid* sebagai *centroid*

sehingga lebih *robust* terhadap pencilan. Algoritma *k-medoids* yang sering digunakan adalah *partitioning around medoids* (PAM) dan *clustering large application* (CLARA). Algoritma PAM bekerja efektif untuk data berukuran kecil, sehingga kurang efektif untuk data yang berukuran besar. Untuk bekerja dengan *dataset* yang besar, algoritma berbasis *sampling* yang disebut CLARA dapat digunakan.

Ukuran Kemiripan

Untuk mengukur kemiripan antara dua objek digunakan ukuran jarak. Semakin besar nilai jarak maka semakin besar perbedaan antar kedua objek. Menurut Johnson and Wichern (2002), ukuran kemiripan yang digunakan dalam *clustering* adalah jarak Euclidean dan jarak manhattan. Jarak Euclidean merupakan jarak terpendek antara dua titik. Jarak Euclidean ditulis sebagai

$$d_{euc}(x_{ij}, c_{kj}) = \sqrt{\sum_{j=1}^p \sum_{i=1}^n (x_{ij} - c_{kj})^2} \quad (1)$$

dengan x_{ij} adalah objek pada pengamatan ke- i pada variabel ke- j , c_{kj} adalah pusat *cluster* ke- k pada variabel ke- j , p adalah banyaknya variabel yang diamati, dan n adalah banyaknya pengamatan.

Sementara itu, jarak Manhattan merupakan pengukuran yang dihasilkan berdasarkan perhitungan jarak selisih antara dua buah objek dengan cara tegak lurus. Jarak manhattan ditulis sebagai

$$d_{man}(x_{ij}, c_{kj}) = \sum_{j=1}^p \sum_{i=1}^n |x_{ij} - c_{kj}| \quad (2)$$

dengan $d_{man}(x_{ij}, c_{kj})$ adalah jarak manhattan antara pengamatan ke- i variabel ke- j ke pusat *cluster* ke- k pada variabel ke- j .

Tahap Analisis

Tahapan analisis dalam penelitian ini adalah

1. menentukan jumlah kelompok k yang ingin dibentuk menggunakan metode *silhouette*,
2. melakukan pengelompokan data dengan algoritma CLARA:
 - a. memilih secara acak pusat *cluster* awal (*medoid*) sebanyak k ,
 - b. menghitung jarak setiap objek terhadap masing-masing *medoid* pada tiap *cluster* dan menempatkan tiap objek tersebut ke *medoid* terdekat,
 - c. menghitung total jaraknya,
 - d. memilih kandidat *medoid* baru,
 - e. menghitung jarak setiap objek terhadap kandidat *medoid* baru dan menempatkan tiap objek tersebut ke kandidat *medoid* terdekat,
 - f. menghitung total jarak,
 - g. menghitung selisih total jarak (S), yaitu selisih dari total jarak pada kandidat *medoid* baru dengan total jarak pada *medoid* lama,
 - h. jika diperoleh nilai $S < 0$, maka kandidat *medoid* tersebut menjadi *medoid* baru,
 - i. ulangi langkah d sampai h hingga tidak terjadi perubahan *medoid*,
3. menampilkan hasil pengelompokan menggunakan algoritma CLARA.

HASIL DAN PEMBAHASAN

Deskripsi Data

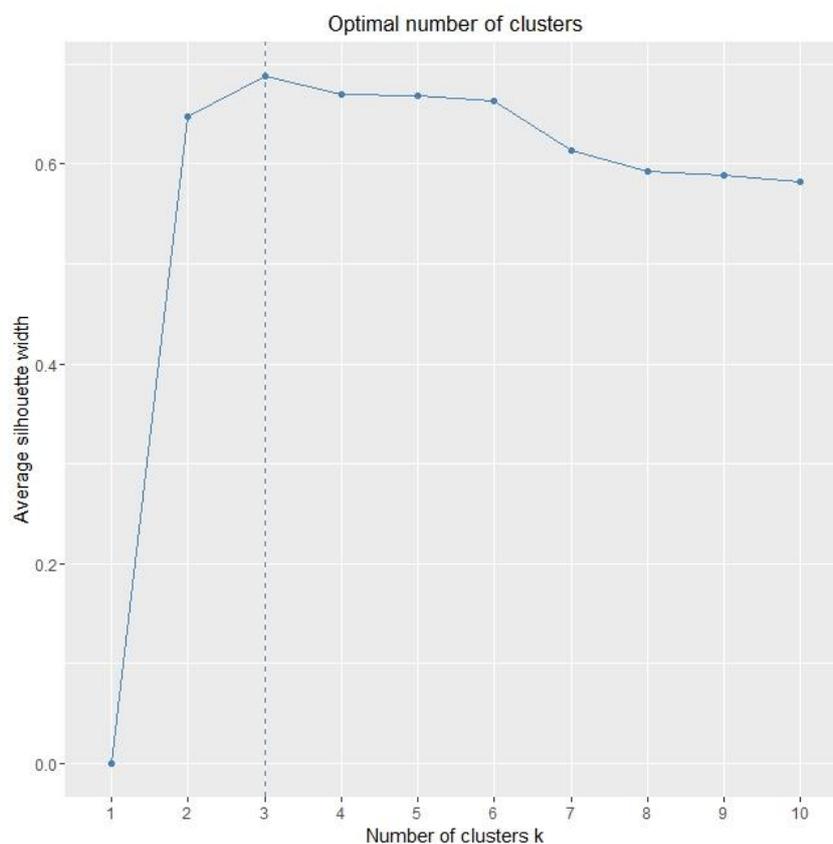
Data yang digunakan adalah data gempa bumi di Indonesia yang didapatkan dari web *United States Geological Survey* (USGS). Jumlah total data sebanyak 7095 dengan dua atribut yaitu kedalaman (*depth*) dan magnitudo, dengan ketentuan kedalaman maksimal 70 meter di bawah permukaan laut. Nama baris diambil dari kolom *place* pada data USGS. Sebagian dari data gempa bumi di Indonesia ditunjukkan pada Tabel 1.

Tabel 1. *Head Data Gempa Bumi di Indonesia*

	depth (km)	Mag (mb)
0km W of Komerda, Indonesia	28	6,47
100km S of Kotaagung, Indonesia	45,93	5,1
100km S of Sungaipenuh, Indonesia	55,3	5,84
100km SSE of Bengkulu, Indonesia	50,48	5
100km SSW of Cigarogol, Indonesia	27	6,09
100km WSW of Sibolga, Indonesia	36	6,09

Analisis Algoritma CLARA

Langkah pertama yang dilakukan untuk *clustering* adalah mencari jumlah *cluster* optimal untuk data gempa bumi di Indonesia. Metode *silhouette* dapat digunakan untuk mencari jumlah *cluster* optimal. Jumlah *cluster* optimal untuk data gempa bumi di Indonesia berdasarkan magnitudo dan kedalamannya menggunakan *software R* ditunjukkan pada Gambar 1.



Gambar 1. Jumlah *Cluster* Optimal

Gambar 1 menunjukkan bahwa untuk data gempa bumi di Indonesia didapatkan jumlah *cluster* optimal adalah tiga *cluster* menggunakan metode *silhouette*. Hal ini berarti untuk langkah selanjutnya akan dilakukan *clustering* gempa bumi di Indonesia menjadi tiga *cluster* menggunakan algoritma CLARA.

Hasil akhir *clustering* yang diperoleh dengan menggunakan algoritma CLARA adalah *cluster* pertama memiliki pusat *cluster* pada objek “Savu Sea 12” yaitu berada di Laut Sawu, tenggara Kecamatan Pulau Ende, Kabupaten Ende, Provinsi Nusa Tenggara Timur dengan titik koordinat 9°09'50.4"S 121°49'15.6"E. *Cluster* kedua memiliki pusat pada objek “southern Sumatra, Indonesia 9”

yaitu berada di Samudra Hindia, barat daya Pantai Carocok, Kecamatan Painan, Kabupaten Pesisir Selatan, Provinsi Sumatra Barat dengan titik koordinat $1^{\circ}34'22.8''S$ $100^{\circ}04'12.0''E$. *Cluster* ketiga memiliki pusat *cluster* pada objek “Kepulauan Alor, Indonesia 11” yaitu berada di Desa Menetwati, Kecamatan Alor Tengah Utara, Kabupaten Alor, Provinsi Nusa Tenggara Timur dengan titik koordinat $8^{\circ}15'00.0''S$ $124^{\circ}40'48.0''E$. Hasil *clustering* data gempa bumi di Indonesia dengan menggunakan algoritma CLARA dapat dilihat pada Tabel 2.

Tabel 2. Pusat *Cluster* Data Gempa Bumi di Indonesia

	Depth (km)	Mag (mb)	cluster size
Savu Sea 12	33	5.2	4314
Southern Sumatra, Indonesia 9	50	5.1	1678
Kepulauan Alor, Indonesia 11	10	5.2	1103

Tabel 2 menunjukkan bahwa kedalaman dan magnitudo untuk pusat *cluster* pertama adalah 33 kilometer di bawah permukaan laut dan 5,2mb. *Cluster* kedua memiliki kedalaman 50 kilometer di bawah permukaan laut dan magnitudo 5,1mb, sedangkan *cluster* ketiga memiliki kedalaman 10 kilometer di bawah permukaan laut dan magnitudo 5,2mb. *Cluster* pertama memiliki anggota sebanyak 4314 objek, sedangkan *cluster* kedua dan ketiga memiliki anggota sebanyak 1678 dan 1103 objek.

Tabel 3. Hasil *Clustering* Data Gempa Bumi di Indonesia

	Depth (km)	Mag (mb)	cluster
0km W of Komerda, Indonesia	28	6,5	1
100km S of Kotaagung, Indonesia	45,9	5,1	2
100km S of Sungaipenuh, Indonesia	55,3	5,8	2
100km SSE of Bengkulu, Indonesia	50,5	5	2
100km SSW of Cigarogol, Indonesia	27	6,1	1
100km WSW of Sibolga, Indonesia	36	6,1	1
101km ENE of Amahai, Indonesia	16,1	5	3
101km SSW of Ogotua, Indonesia (1)	44,1	5,3	2
101km SSW of Ogotua, Indonesia (2)	58	5,3	2
102km NNW of Boroko, Indonesia	43,3	5,8	2
103km NNW of Dobo, Indonesia	34,0	5,1	1
104km ENE of Luwuk, Indonesia	28	5,9	1
104km SE of Amahai, Indonesia	25,8	5	1
104km SSW of Kute, Indonesia	35,2	5,3	1
105km SE of Amahai, Indonesia (1)	10	5,2	3
105km SE of Amahai, Indonesia (2)	10	5,8	3
105km WNW of Laiwui, Indonesia	37,3	5	1
106km N of Maurole, Indonesia	8,6	5	3
107km ENE of Poso, Indonesia	11,0	5	3
107km NNW of Atambua, Indonesia	18,6	5,9	3

Tabel 3 menunjukkan 20 objek awal hasil *clustering* data gempa bumi di Indonesia berdasarkan kedalaman dan magnitudo menggunakan algoritma CLARA. Dapat dilihat bahwa “0km W Komerda, Indonesia” dengan kedalaman 29 kilometer di bawah permukaan laut dan magnitudo 6,5mb berada pada *cluster* pertama, “100km S of Kotaagung, Indonesia” dengan kedalaman 45,93 kilometer di bawah permukaan laut dan magnitudo 5,1mb berada pada *cluster* kedua, “100km S of Sungaipenuh, Indonesia”

dengan kedalaman 55,3 kilometer di bawah permukaan laut dan magnitudo 5,8mb berada pada *cluster* kedua, dan seterusnya sampai “107km NNW of Atambua, Indonesia” dengan kedalaman 18,6 kilometer di bawah permukaan laut dan magnitudo 5,9mb berada pada *cluster* ketiga.

SIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa data gempa bumi di Indonesia tahun 1973-2017 berdasarkan kedalaman dan magnitudo dapat dikelompokkan menjadi tiga *cluster* dengan pusat *cluster* pertama, kedua, dan ketiga secara berurutan berada di Laut Sawu, Samudra Hindia, dan Desa Menetwati. Jumlah anggota masing-masing *cluster* secara berurutan adalah 4314, 1678, dan 1103 objek. Untuk penelitian selanjutnya diharapkan dapat dilakukan implementasi algoritma *clustering* yang lainnya pada data gempa bumi di Indonesia untuk dibandingkan dengan hasil *clustering* menggunakan algoritma CLARA.

DAFTAR PUSTAKA

- Artatia, H., & Hakim, R.B.F. (2015). Pengelompokan Dampak Gempa Bumi dari Segi Kerusakan Fasilitas pada Provinsi yang Berpotensi Gempa di Indonesia Menggunakan K-means Clustering. *Prosiding Seminar Nasional Matematika dan Pendidikan Matematika UMS 2015*. ISBN: 978.602.361.002.0.
- Berkhin P. (2006). A Survey of Clustering Data Mining Techniques. *Kogan J., Nicholas C., Teboulle M. (eds) Grouping Multidimensional Data*. Springer, Berlin, Heidelberg. Diakses pada tanggal 3 Maret 2019 dari https://doi.org/10.1007/3-540-28349-8_2.
- Bird, P. (2003). An updated digital model of plate boundaries. *Geochemistry Geophysics Geosystem Journal*. 4(3): 1-52. doi:10.1029/2001GC000252.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publisher, San Francisco.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining : Concepts and Techniques*, Third Edition. USA: Elsevier Inc. Waltham.
- Hermawati, FA. 2013. *DATA MINING*. Surabaya: CV Andi Offset.
- Irsyam, M., Sengara, I.W., Aldiamar, F., Widiyantoro, S., Triyoso, W., Natawidjaja, D.H., Kertapati, E., Meilano, I., Suhardjono., Asrurifak, M., & Ridwan, M. (2010). *Ringkasan Hasil Studi Tim Revisi Peta Gempa di Indonesia 2010*. Diakses pada tanggal 18 Maret 2018 dari https://www.preventionweb.net/files/14654_AIFDR.pdf.
- Johnson, R. A., & Wichern, D.W. (2002). *Applied Multivariate Analysis 5th edition*. New Jersey: Prentice Hall.
- Kaufman, L., & Rousseeuw, P.J. (1987). *Clustering By Means of Medoids*. Diakses pada tanggal 12 November 2018 dari https://www.researchgate.net/publication/243777819_Clustering_by_Means_of_Medoids.
- Larose, D.T. (2005). *Discovering Knowledge in Data : An Introduction to Data Mining*. New Jersey: John Willey and Sons. Inc.
- Mustafa, Badrul. (2010). Analisis Gempa Nias dan Gempa Sumatra Barat dan Kesamaannya yang Tidak Menimbulkan Tsunami. *Jurnal Ilmu Fisika Universitas Andalas*. Diakses pada tanggal 29 Maret 2019 dari <http://jif.fmipa.unand.ac.id/index.php/jif/article/view/46/34>.
- Septiana,L., & Djohan, N. (2015). Analisis Perbandingan Algoritma K-means Clustering dan Expectation-Maxmimation (EM) untuk Klasifikasi Butir Beras. *Jurnal Teknik dan Ilmu Komputer*. 4(15): 245-253.
- United States Geological Survey. (2018). Search Earthquake Catalog. Diakses pada tanggal 9 Oktober 2018 dari <https://earthquake.usgs.gov/earthquakes/search/>.