

# GFK-12: IDENTIFIKASI PENYEBARAN GEMPA DI INDONESIA DENGAN METODE CLUSTERING

Refriani Pebria<sup>\*)</sup>, Bambang Heru I., dan Iwan Sugihartono

Universitas Negeri Jakarta, Alamat, Kota dan Kode Pos

<sup>\*)</sup> Email: [refriani.febria@gmail.com](mailto:refriani.febria@gmail.com)

## Abstrak

Identifikasi penyebaran gempa telah dilakukan dengan menggunakan metode *clustering* berdasarkan kekuatan gempa, kedalaman, lokasi dan elevasi. Metode *clustering* yang digunakan ialah metode EM, KMeans, dan Xmeans dengan empat buah proses *clustering*. Proses *clustering* pertama dilakukan berdasarkan keempat parameter, dengan menggunakan metode EM menghasilkan data gempa terbesar di daerah Laut Banda dengan lokasi 127.5747 BT dan 1.7943 LS, dan hasil dengan menggunakan metode KMeans dan XMeans terdapat di daerah pulau Sumatera dengan pusat gempa 95 BT dan 5 LU. Proses *clustering* kedua dilakukan berdasarkan parameter lokasi gempa, ketiga metode menghasilkan data gempa terbanyak terdapat di pulau Sumatera dengan pusat cluster 96 BT dan 4,5 LU. Pada proses *clustering* ketiga dilakukan dengan menggunakan parameter kekuatan gempa dan kedalaman gempa, ketiga metode menghasilkan data terbesar terletak di dekat permukaan bumi. Proses *clustering* terakhir dilakukan berdasarkan kekuatan, kedalaman, dan elevasi, secara umum ketiga metode ini menghasilkan data gempa terbesar terletak di daerah laut dengan kedalaman 42 km dengan kekuatan gempa 6 SR. Hasil tersebut menunjukkan bahwa metode *clustering* yang paling tepat digunakan untuk merepresentasikan penyebaran gempa yang sebenarnya adalah metode EM, karena dapat memperlihatkan penyebaran gempa dengan kemampuan data di setiap clusternya yang ditunjukkan oleh nilai standard deviasi pada setiap cluster.

## Abstract

*The spreads of earthquake has been identified by using Clustering methods which are depends on four parameters, i.e. magnitude, depth, location, and elevation. We used EM method, KMeans, and Xmeans as Clustering methods which is using four parameters. The first clustering step based on 4 parameters which used EM method. This method shows the highest earthquake happen at Banda's sea which is located 127.5747 east longitude and 1.7943 south latitudes. Whereas, KMeans and XMeans method predicts that the highest earthquake located at 95 east longitude and 5 north latitudes in northern Sumatera. The second clustering process was done based on location. This process which is using EM method, KMeans, and XMeans predicted that the most frequent earthquake happened at Sumatera Island which locates at 96 east longitude and 4.5 north latitudes. In the third clustering process, we used two parameters, i.e. magnitude and depth. According to three methods the results show the data's at earth surface. The last clustering process has been done based on magnitude, depth, and elevation. Generally, by using three methods, the results show the largest earth data's locate in sea area which has depth about 42 km and the magnitude is 6 scale Richter (SR). These all the clustering processes indicate that EM method is the most appropriate for representing actual the spread of earthquake. It's because of EM method can be showing the spread of earthquake with the most dense data's in every cluster based on deviation standard.*

**Keywords:** *The spread of earthquake, Clustering methods, EM, KMeans, XMeans.*

## 1. Pendahuluan

Indonesia merupakan negara yang paling rawan bencana alam di dunia demikian menurut *United Nations International Strategy for Disaster Reduction* (UNISDR; Badan PBB untuk Strategi Internasional Pengurangan Risiko Bencana) [1]. Karena Kepulauan Indonesia diapit oleh lempeng samudera yang menyebabkan lipatan dan tumbukan lempeng samudera terhadap lempeng darat yang menimbulkan terbentuknya deretan gunung. Pergeseran lempeng yang menumbuk lempeng lain, selain membentuk lipatan

dan gunung, juga menyebabkan terjadinya getaran pada kulit bumi dan mempengaruhi fenomena yang ada di permukaan bumi. Getaran-getaran yang disebabkan oleh tumbukan lempeng ini dikenal dengan gempa bumi [2]. Gempa bumi di Indonesia sering terjadi di beberapa daerah. Penyebaran gempa di Indonesia ini yang selanjutnya akan dikelompokkan agar dapat mempermudah kita membaca penyebaran gempa yang sejenis di Indonesia.

Di dalam penelitian ini telah dilakukan identifikasi penyebaran gempa di Indonesia menggunakan metode *clustering* dengan tiga

algoritma diantaranya algoritma EM, KMeans, dan XMeans berdasarkan kekuatan, kedalaman, elevasi dan lokasi gempa. Kesimpulan dan saran penelitian dijelaskan secara sistematis.

## 2. Metode Eksperimen

### 2.1 Proses Pengumpulan Data

Data statistik penyebaran gempa yang diperoleh dari USGS dengan menggunakan Google Earth. Data statistik yang digunakan merupakan data gempa di sekitar wilayah Indonesia yang terletak diantara 6°LU-11°LS dan 95°BT-141°BT pada *eye altitude* sekitar 500 mil. Data yang digunakan diantaranya ialah kekuatan gempa, posisi gempa (dalam letak astronomi), kedalaman, episentrum.

### 2.2 Proses Pengolahan Data

Data yang telah diperoleh selanjutnya diolah menjadi sebuah data standard yang dapat digunakan dalam WEKA. WEKA menerima input data dalam format ARFF (*Attribute-Relation File Format*). ARFF adalah tipe file teks yang berisi berbagai *instance* data (catatan yang dijadikan objek) yang berhubungan dengan suatu set parameter data yang dideskripsikan dalam file tersebut. Parameter data yang digunakan ialah beberapa data yang telah disebutkan sebelumnya. Sehingga terdapat 4 parameter yang akan digunakan.

Data statistik yang telah dikumpulkan diubah ke dalam bentuk ARFF. Selanjutnya, dari format ARFF akan dideklarasikan setiap parameter yang digunakan. Kelima parameter tersebut digunakan dalam set data yang digunakan dalam proses data mining. Data statistik penyebaran gempa diolah dalam XAMPP untuk dapat membaca kelima parameter yang digunakan dengan menggunakan bahasa pemrograman PHP.

### 2.3 Proses Clustering

Setelah data diolah dalam bentuk .arff, maka selanjutnya data siap diolah dalam WEKA [WEKA 3.6]. WEKA adalah aplikasi *data mining* berbasis bahasa pemrograman Java dan dikembangkan oleh University of Waikato, New Zealand. WEKA dipilih sebagai salah satu perangkat lunak data mining karena WEKA bersifat *open source*, mudah digunakan serta memiliki implementasi algoritma yang cukup bervariasi. Didalam WEKA, terdapat tools untuk data pre-processing, classification, regression, *clustering*, association rules, dan visualisation. Algoritma *clustering* yang tersedia dalam WEKA diantaranya adalah EM, *K-means*, *X-means*, COBWEB, *Farthest First*, dan sebagainya. Berikut adalah tampilan utama dari WEKA.

### 2.4 Proses Visualisasi

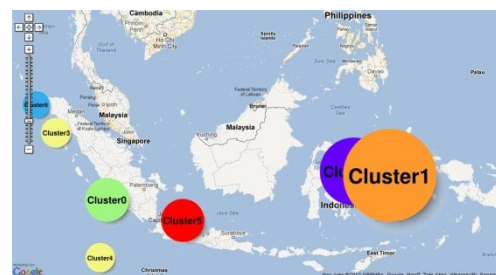
Hasil *clustering* dari masing-masing metode tersebut akan divisualisasikan berdasarkan hasil penyebaran gempa. Proses ini dilakukan berdasarkan jumlah *cluster* dari masing-masing metode yang digunakan sehingga dapat terlihat pegelompokan data berdasarkan parameter-parameter yang digunakan.

## 3. Hasil dan Pembahasan

Data yang dikumpulkan merupakan data gempa dari awal tahun 2000 hingga bulan April 2012 yaitu sebanyak 328 data yang terdiri dari beberapa parameter, diantaranya adalah: kekuatan gempa, kedalaman, lokasi, dan elevasi dengan masing-masing parameter merupakan tipe data numerik. Setelah itu, dilakukan *clustering* dengan menggunakan berbagai algoritma yang tersedia. *Clustering* ini dilakukan dengan 4 buah proses.

Pada proses pertama dilakukan klusterisasi berdasarkan keempat parameter, sedangkan pada proses kedua dilakukan klusterisasi berdasarkan satu parameter, yakni lokasi terjadinya gempa, proses ketiga dilakukan klusterisasi berdasarkan dua parameter, yakni kekuatan gempa dan kedalaman gempa, serta proses keempat dilakukan klusterisasi berdasarkan tiga parameter, yakni kekuatan gempa, kedalaman gempa, dan elevasi. Setiap hasil *clustering* dari setiap algoritma dan proses akan dibandingkan serta dianalisis. Sehingga, diperoleh suatu kesimpulan data penyebaran gempa maupun algoritma-algoritma yang digunakan.

### 3.1 Proses Clustering 1



Gb. 1 Hasil Visualisasi Metode EM proses 1



Gb. 2 Hasil Visualisasi Metode KMeans proses 1



Gb. 3 Hasil Visualisasi Metode XMeans proses 1



Gb. 6 Hasil Visualisasi Metode XMeans proses 2

Hasil visualisasi menunjukkan data gempa setiap cluster dengan menggunakan parameter lokasi terjadinya gempa. Besar lingkaran menunjukkan nilai standard deviasi dari parameter lokasi tersebut. Sedangkan warna tiap cluster menunjukkan intensitas tiap cluster. Semakin tinggi frekuensi warna maka semakin besar intensitas data. Identifikasi menggunakan metode EM dengan empat parameter menghasilkan data yang paling besar di daerah Laut Banda. Namun, kemampuan yang paling besar terdapat pada data gempa di bagian Sumatera bagian Utara. Hal tersebut dikarenakan terdapat cukup banyak data namun dengan besar cluster yang kecil, sedangkan bila menggunakan metode KMeans dan XMeans data yang dihasilkan hampir sama. Selanjutnya kedua metode tersebut memberikan informasi bahwa jumlah gempa yang paling besar adalah di pulau Sumatera bagian Utara.

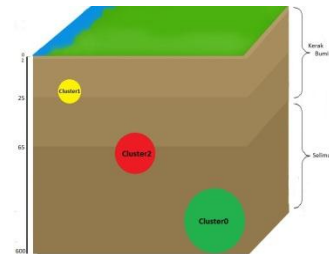
Hasil *clustering* dengan menggunakan metode EM dan KMeans menghasilkan data lokasi cluster yang hampir sama. Namun, bila dibandingkan dengan metode XMeans, data cluster yang dihasilkan sedikit berbeda. Metode XMeans menghasilkan dua buah cluster di daerah Sumatera bagian Utara. Berdasarkan ketiga metode dapat dijelaskan bahwa cluster yang memiliki kemampuan yang paling besar ialah data di daerah pulau Sumatera bagian Utara.

### 3.2 Proses Clustering 2

### 3.3 Proses Clustering 3



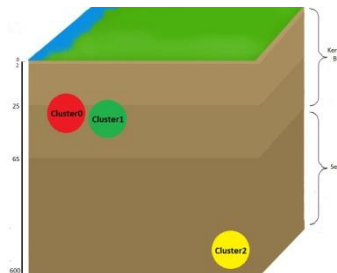
Gb. 4 Hasil Visualisasi Metode EM proses 2



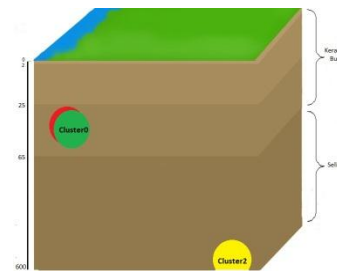
Gb. 7 Hasil Visualisasi Metode EM proses 3



Gb. 5 Hasil Visualisasi Metode KMeans proses 2



Gb. 8 Hasil Visualisasi Metode KMeans proses 3

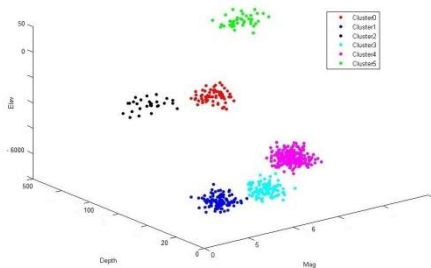


Gb. 9 Hasil Visualisasi Metode XMeans proses 3

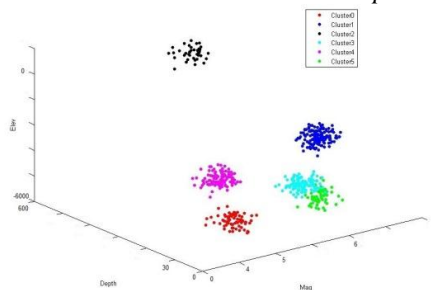
Visualisasi gempa hasil *clustering* proses ketiga merupakan visualisasi gempa secara aktual. Hasil visualisasi ini menggunakan nilai kedalaman gempa sebagai titik acuan untuk menggambarkan hasil *clustering*. Semakin ke bawah *cluster* yang

digambarkan maka semakin dalam pusat terjadinya gempa. Besar lingkaran menunjukkan perbandingan nilai deviasi dari masing-masing *cluster*. Sedangkan warna disetiap *cluster* menunjukkan intensitas. Artinya, semakin tinggi frekuensi warna maka semakin besar intensitas data. Berdasarkan hasil *clustering* pada proses ini terlihat bahwa gempa secara garis besar dikelompokkan ke dalam dua bagian yaitu, gempa yang memiliki pusat dekat dengan permukaan dan jauh di bawah permukaan bumi.

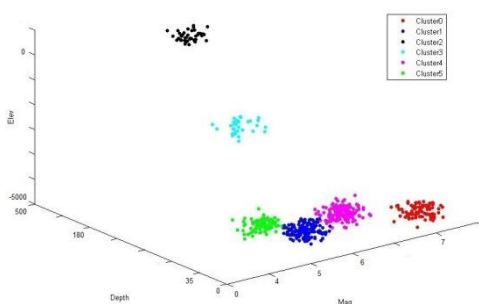
### 3.4 Proses *Clustering* 4



Gb. 10 Hasil Visualisasi Metode EM proses 4



Gb. 11 Hasil Visualisasi Metode KMeans proses 4



Gb. 12 Hasil Visualisasi Metode XMeans proses 4

Ketiga metode tersebut memberikan informasi bahwa data gempa terbesar terletak di daerah laut dengan kedalaman gempa yang dekat dengan permukaan bumi yaitu berkisar 42 km dengan kekuatan gempa berkisar 6 SR. Sedangkan, data terkecil berada di daerah laut dengan kedalaman gempa yang jauh berada di bawah permukaan bumi dengan kekuatan gempa berkisar 6,2 SR.

## 4. KESIMPULAN

Berdasarkan hasil penelitian, maka dapat disimpulkan bahwa:

1. Metode *clustering* yang paling tepat digunakan untuk merepresentasikan penyebaran gempa ini adalah metode EM.
2. Hasil *clustering* menggunakan metode EM menghasilkan data gempa yang paling banyak terjadi di daerah Laut Banda dengan lokasi gempa 127.5747 BT dan 1.7943 LS. Sedangkan dengan metode KMeans dan XMeans data gempa yang paling banyak terjadi ialah di daerah pulau Sumatera bagian Utara dengan pusat gempa gempa sekitar 95 BT dan 5 LU.
3. Proses *clustering* dari ketiga metode dengan parameter lokasi memberikan informasi data gempa yang paling banyak ialah di daerah pulau Sumatera bagian Utara dengan pusat cluster gempa sekitar 96 BT dan 4.5 LU.
4. proses *clustering* dari ketiga metode dengan menggunakan parameter kedalaman dan kekuatan menghasilkan data gempa terbesar terletak di bagian dekat dengan permukaan kulit bumi dan yang jauh dengan permukaan bumi. Namun, data gempa yang terletak dengan permukaan bumi terbagi kembali dalam 2 buah cluster, sehingga gempa banyak terdapat dari jarak yang dekat dengan permukaan.
5. Proses *clustering* keempat menggunakan parameter kekuatan gempa, kedalaman gempa, dan elevasi menghasilkan data gempa terbesar terletak di daerah laut dengan kedalaman gempa yang dekat dengan permukaan kulit bumi yaitu berkisar 42 km dan dengan kekuatan gempa berkisar 6 SR. sedangkan data yang terkecil terletak di daerah laut dengan kedalaman gempa yang jauh berada di bawah permukaan kulit bumi dengan kekuatan gempa berkisar 6,2 SR.

## DAFTAR ACUAN

- [1] Briceno, Salvano. "Perkataan menjadi Tidakan: Panduan untuk implementasi Kerangka Kerja Hyogo." [http://www.unisdr.org/files/594\\_BahasaHFA.pdf](http://www.unisdr.org/files/594_BahasaHFA.pdf). Diakses pada tanggal 02 Desember 2011, pk. 15.00 WIB.
- [2] Irsyam, Masyhur, dkk. "Ringkasan Hasil Studi Tim Revisi Peta Gempa di Indonesia 2010." [www.unisdr.org/files/14654\\_AIFDR.pdf](http://www.unisdr.org/files/14654_AIFDR.pdf). Diakses pada tanggal 02 Desember 2011, pkl. 17.00 WIB.

- [3] Jardin, dkk. “Clustering Methods”. <http://users.ics.tkk.fi/sami/thesis/node9.html>. Diakses pada tanggal 05 Desember 2011, pk. 09.00 WIB.
- [4] Keutapati, Engkon. “Pengertian Gempa dan Macam-Macam Gempa”. <http://www.scribd.com/document/downloads/direct/37685391?extension=pdf&ft=1325780355&lt=1325783965&uahk=o2UOJv6JvUxIwz7WFbrp8fyct1U>. Diakses pada tanggal 05 Desember 2011, pk. 14.00 WIB.
- [5] Manthey, B., Röglin, H. Improved smoothed analysis of the  $k$ -means method. *In: Proceedings of the 20th Annual Symposium on Discrete Algorithms*, pp. 461–470 (2009)
- [6] MM, Milenova, BL, “O-Cluster: Mengatur Scalable Clustering Besar Data Dimensi Tinggi”, Oracle Data Mining Technologies, Van De Graaff 10 Drive, Burlington, MA 01803.
- [7] Moore, Andrew: “K-means and Hierarchical Clustering - Tutorial Slides” <http://www-2.cs.cmu.edu/~awm/tutorials/kmeans.html>. Diakses pada tanggal 05 Desember 2011, pk. 14.00 WIB.
- [8] Mustapha, Nurwati, dkk “Expectation Maximization Clustering Algorithm for User Modeling in Web Usage Mining Systems”. *European Journal of Scientific Research ISSN 1450-216X Vol.32 No.4 (2009)*, pp.467-476.
- [9] Pelleg, Dan, dkk. “X-means : Extending K-means with Efficient Estimation of the Number of Cluster”. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213 USA.
- [10] Rubin, D.B dkk. “Maximum Likelihood from Incomplete Data via the EM Algorithm”. *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 39, No. 1. (1977), pp.1-38.
- [11] Sanjoy, D. (1999) Learning Mixtures of Gaussians, *IEEE Symposium on Foundations of Computer Science (FOCS)*.
- [12] Vattani, A.  $k$ -means requires exponentially many iterations even in the plane. *In: Proceedings of the 25th Annual Symposium on Computational Geometry*, pp. 324–332. (2009)
- [13] W., Forgy. E. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *In: Biometric Society Meeting, Riverside, California, 1965. Abstract in Biometrics*, vol. 21, p. 768 (1965)