



***CLUSTERING DAERAH RAWAN GEMPA DI SUMATRA BARAT
MENGUNAKAN METODE CLUSTERING LARGE APPLICATION DAN METODE
DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE***

JURNAL ILMIAH

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika



Oleh

DELLA MARSITA

B2A220002

PROGRAM STUDI STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS MUHAMMADIYAH SEMARANG

2021

**CLUSTERING DAERAH RAWAN GEMPA DI SUMATRA BARAT
MENGUNAKAN METODE CLUSTERING LARGE APPLICATION DAN METODE
DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE**

Della Marsita¹, Tiani Wahyu Utami², M. Al Haris³

¹²³Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang

Alamat e-mail : dellamarsita2@gmail.com

ABSTRAK

Gempabumi (*earthquake*) adalah peristiwa bergoncangnya bumi karena pergerakan/pergeseran lapisan batuan pada kulit bumi secara tiba-tiba akibat pergerakan lempeng-lempeng tektonik. Sumatra Barat merupakan salah satu daerah rawan gempa di Indonesia. Hal ini disebabkan karena letaknya yang berada pada jalur patahan Semangko, tepat di antara pertemuan dua lempeng benua besar, yaitu Eurasia dan Indo-Australia. Oleh karenanya, wilayah ini sering mengalami gempa bumi. Permasalahan yang dihadapi sampai saat ini yaitu belum ada teori yang tepat dan dapat digunakan untuk memprediksi kapan dan dimana gempa bumi akan terjadi. Algoritma *clustering* yang diterapkan dalam penelitian ini adalah metode CLARA dan DBSCAN. DBSCAN (*Density-Based Spatial Clustering Of Applications With Noise*) merupakan proses clustering yang dilakukan berdasarkan tingkat kedekatan/kerapatan jarak antar obyek dalam dataset..Sementara metode CLARA dibentuk karena ketidakefektifan Algoritma PAM yang bekerja efektif hanya untuk himpunan data kecil. Metode CLARA (*Clustering Large Applications*).menggunakan himpunan data sampel secara random. *Cluster* dibentuk menjadi 5 cluster dengan menggunakan Skala SIG BMKG. Dari kedua metode tersebut akan di buat perbandingan dengan menghitung *Coefficient Silhouette*. Metode CLARA mempunyai nilai *Coefficient Silhouette* sebesar 0,57 sementara metode DBSCAN 0,45. Maka metode yang cocok di gunakan adalah metode CLARA dengan daerah yang termasuk kerusakan berat (*heavy damage*) adalah Kepulauan Mentawai.

Kata Kunci: Gempa Bumi,Sumatra Barat, *Data Mining*, *Euclidean*, CLARA, DBSCAN, *Coefficient Shillhoute*.

ABSTRACT

Earthquake is an event that shakes the earth due to the sudden movement of rock layers in the earth's crust due to the movement of tectonic plates. West Sumatra is one of the earthquake-prone areas in Indonesia. This is due to its location on the Semangko fault line, right between the confluence of two large continental plates, namely Eurasia and Indo-Australia. Therefore, this area often experiences earthquakes. The problem faced so far is that there is no appropriate theory that can be used to predict when and where an earthquake will occur. One of the clustering algorithms applied in this research is the CLARA and DBSCAN methods. DBSCAN (Density-Based Spatial Clustering Of Applications With Noise) is a clustering process that is carried out based on the level of proximity/density of the distance Meanwhile, the CLARA method was formed due to the ineffectiveness of the PAM Algorithm which works effectively only for small data sets. The CLARA (Clustering Large Applications) method uses a random sample data set. The cluster is formed into 5 clusters using the BMKG SIG Scale. From the two methods, a comparison will be made by calculating the Silhouette Coefficient. The CLARA method has a Silhouette Coefficient value of 0.57 while the DBSCAN method is 0.45. So the suitable method to use is the CLARA method with the area which includes heavy damage is the Mentawai Islands. The CLARA (Clustering Large Applications) method uses a random sample data set. The cluster is formed into 5 clusters using the BMKG SIG Scale. From the two

methods, a comparison will be made by calculating the Silhouette Coefficient. The CLARA method has a Silhouette Coefficient value of 0.57 while the DBSCAN method is 0.45. So the suitable method to use is the CLARA method with the area which includes heavy damage is the Mentawai Islands. The CLARA (Clustering Large Applications) method uses a random sample data set. The cluster is formed into 5 clusters using the BMKG MMI Scale. From the two methods, a comparison will be made by calculating the Silhouette Coefficient. The CLARA method has a Silhouette Coefficient value of 0.57 while the DBSCAN method is 0.45. So the suitable method to use is the CLARA method with the area which includes heavy damage is the Mentawai Islands.

Keywords: Earthquake ,Sumatra Barat, Data Mining, Euclidean, CLARA, DBSCAN, Coefficient Shilloute.

PENDAHULUAN

Sumatra Barat merupakan salah satu daerah rawan gempa di Indonesia. Hal ini disebabkan karena Sumatra Barat berada pada jalur patahan Semangko dimana patahan ini merupakan pertemuan dua lempeng besar yaitu Eurasia dan Indo-Australia. Menurut Teori Elastic Rebound yang merupakan Seismolog Amerika (Bullen, 1965; Bolt, 1985) menjelaskan bahwa gempa bumi merupakan gejala alam yang di sebabkan oleh pelepasan energi regangan elastis batuan yang menyebabkan adanya deformasi batuan yang terjadi pada lapisan lithosfer.

Sumatra barat khususnya merupakan daerah yang mempunyai aktifitas kegempaan yang cukup tinggi dengan kedalaman hiposenter kurang dari 50 km. Hal ini terbukti dengan banyaknya gempa yang terjadi di pulau Sumatra Barat. Beberapa gempa bumi yang cukup besar melanda sumatera barat adalah Gempa bumi 30 September 2009 dan Gempa bumi Kepulauan Mentawai 2010. Proses terjadinya gempa bumi sangat sulit untuk diamati secara langsung, sebab melibatkan interaksi yang sangat kompleks antara materi dan energi yang terdapat pada sistem besar aktif di bawah permukaan bumi. Parameter atau ukuran untuk menentukan besarnya sebuah gempa disebut *Magnitudo* serta di kedalaman berapa gempa itu terjadi atau bisa disebut *Depth*. Berkumpulnya data kejadian gempa bumi di Sumatra Barat dan berkembangnya ilmu pengetahuan dan teknologi memberikan kesempatan kepada peneliti untuk turut andil dalam pemecahan masalah yaitu dengan melakukan penerapan

Data Mining. Pemanfaatan data gempa bumi dapat dilakukan dengan mengelompokkan data berdasarkan Kabupaten/Kota yang sesuai dengan informasi yang ada pada data, sehingga dapat diketahui potensi daerah rawan gempa bumi yang ada di Sumatra Barat. Proses pengelompokkan ini dilakukan dengan menggunakan metode *clustering*.

Pada penelitian ini akan dilakukan clustering dengan menggunakan metode CLARA (Clustering Large Applications) dan DBSCAN (*Density-Based Spatial Clustering Of Applications With Noise*). Metode DBSCAN merupakan metode yang efektif untuk menangani outlier, karena pada wilayah tertentu aktivitas kegempaan dapat diam selama ratusan tahun namun tiba-tiba dapat terjadi dengan melepaskan energi besar sehingga dapat membuat data tersebut menjadi *noise/outlier*. Sementara pada metode CLARA, data yang digunakan merupakan data sampel acak dimana cluster dibangun dengan menghitung kedekatan yang dimiliki antara medoid dengan objek non-medoid. Maka dari itu, perlu dikaji metode manakah yang akan lebih efektif dan metode mana yang akan memberikan hasil cluster baik dengan melihat nilai *Coefficient Silhouette*.

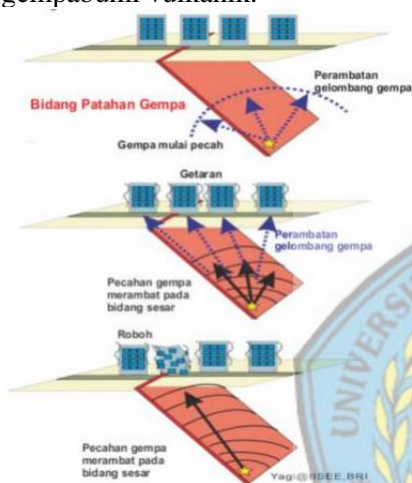
Tujuan dari penelitian ini akan mengelompokkan daerah rawan gempa di Sumatra Barat dengan menggunakan Skala SIG pada BMKG yang membagi daerah kerawanan menjadi 5 *cluster*. Setelah pembentukan *clustering* dengan metode CLARA dan DBSCAN akan disbanding

metode mana yang terbaik dengan menghitung *coefficient silhouette*.

TINJAUAN PUSTAKA

1. Gempa Bumi

Gempabumi (earthquake) adalah peristiwa bergetar atau bergoncangnya bumi karena pergerakan/pergeseran lapisan batuan pada kulit bumi secara tiba-tiba akibat pergerakan lempeng-lempeng tektonik. Gempabumi yang disebabkan oleh aktivitas pergerakan lempeng tektonik disebut gempabumi tektonik. Namun selain itu, gempabumi bisa saja terjadi akibat aktifitas gunung berapi yang disebut sebagai gempabumi vulkanik.



Pergeseran di kerak bumi memancarkan radiasi gelombang gempabumi hingga menimbulkan guncangan dan perubahan struktur batuan di permukaan

2. Kedalaman dan Kekuatan Gempabumi

Kedalaman dan Kekuatan Gempabumi Fowler (1990) mengklasifikasikan gempabumi berdasarkan kedalaman fokus (hypocentre) sebagai berikut: gempabumi dangkal (shallow) kurang dari 70 km, gempabumi menengah (intermediate) kurang dari 300 km, dan gempabumi dalam (deep) lebih dari 300 km atau 450 km. Gempabumi dangkal menimbulkan efek guncangan dan kehancuran yang lebih dahsyat dibanding gempabumi dalam. Ini karena sumber gempabumi lebih dekat ke permukaan bumi sehingga energi gelombangnya lebih besar. Karena pelemahan energi gelombang akibat perbedaan jarak sumber ke permukaan relatif kecil.

3. Skala Intensitas Gempa Bumi BMKG

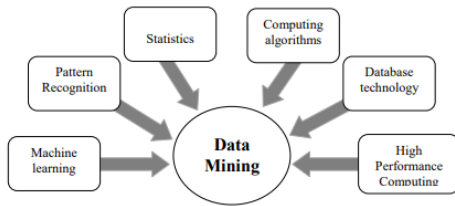
Skala yang biasa digunakan oleh BMKG adalah Skala MMI (Modified Mercalli Intensity). Skala Mercalli adalah satuan untuk mengukur kekuatan gempa bumi. Satuan ini diciptakan oleh seorang vulkanologis dari Italia yang bernama Giuseppe Mercalli pada tahun 1902. Skala Mercalli terbagi menjadi 12 pecahan berdasarkan informasi dari orang-orang yang selamat dari gempa tersebut dan juga dengan melihat serta membandingkan tingkat kerusakan akibat gempa bumi tersebut. Oleh itu skala Mercalli adalah sangat subjektif dan kurang tepat dibanding dengan perhitungan magnitudo gempa yang lain. Oleh karena itu, saat ini penggunaan Skala Richter lebih luas digunakan untuk mengukur kekuatan gempa bumi. Tetapi skala Mercalli yang dimodifikasi, pada tahun 1931 oleh ahli seismologi Harry Wood dan Frank Neumann masih sering digunakan terutama apabila tidak terdapat peralatan seismometer yang dapat mengukur kekuatan gempa bumi di tempat kejadian.

Skala SIG BMKG	Deskripsi Sederhana	Deskripsi Rinci	Skala MMI	PGA (gal)
I	TIDAK DIRASAKAN (<i>Non Felt</i>)	Tidak dirasakan atau dirasakan hanya oleh beberapa orang tetapi terakam oleh alat	I-III	<2.9
II	DIRASAKAN (<i>Felt</i>)	Dirasakan oleh orang banyak tetapi tidak menimbulkan kerusakan. Benda-benda ringan yang digantung berayun dan jendela kaca bergetar	III-VI	2.9 - 88
III	KERUSAKAN RINGAN (<i>Slight Damage</i>)	Bagian non struktur bangunan mengalami kerusakan ringan, seperti retak rambut pada dinding, genteng berdeser ke bawah dan sebagian berjatuhan	VI	89-167
IV	KERUSAKAN SEDANG (<i>Moderate Damage</i>)	Banyak retakan terjadi pada dinding bangunan sederhana, sebagian roboh, kaca pecah. Sebagian plester dinding lenas. Hampir sebagian besar genteng bergeser ke bawah atau jatuh. Struktur bangunan mengalami kerusakan ringan sampai sedang.	VII-VIII	168-564
V	KERUSAKAN BERAT (<i>Heavy Damage</i>)	Sebagian besar dinding bangunan seraman roboh. Struktur bangunan mengalami kerusakan berat. Rel kereta api melengkung.	IX-XII	> 564

4. Definisi Data Mining

Data mining dikenal sejak tahun 1990-an, ketika adanya suatu pekerjaan yang memanfaatkan data menjadi suatu hal yang lebih penting dalam berbagai bidang, seperti marketing dan bisnis, sains dan teknik, serta seni dan hiburan. Sebagian ahli menyatakan bahwa data mining merupakan suatu langkah untuk menganalisis pengetahuan dalam basis data atau biasa

disebut *Knowledge Discovery in Database (KDD)*. *Data mining* merupakan proses untuk menemukan pola data dan pengetahuan yang menarik dari kumpulan data yang sangat besar.



Tahapan proses data mining ada beberapa yang sesuai dengan proses KDD adalah sebagai berikut:

a. Data Cleaning (Pembersihan Data)

Data cleaning (Pembersihan data) adalah proses yang dilakukan untuk menghilangkan noise pada data yang tidak konsisten atau bisa disebut tidak relevan.

b. Data Integration (Integrasi data)

Integrasi data merupakan proses penggabungan data dari berbagai database sehingga menjadi satu database baru. Data yang diperlukan pada proses data mining tidak hanya berasal dari satu database tetapi juga dapat berasal dari beberapa database.

c. Data Selection (Seleksi Data)

Tidak semua data yang terdapat dalam database akan dipakai, karena hanya data yang sesuai saja yang akan dianalisis dan diambil dari database.

d. Data Transformation (Transformasi Data)

Transformasi data merupakan proses perubahan data dan penggabungan data ke dalam format tertentu. Data mining membutuhkan format data khusus sebelum diaplikasikan.

Untuk melakukan normalisasi, salah satu rumus yang dapat digunakan yaitu

$$X = \frac{x - \min(x_i, x_j)}{\max(x_i, x_j) - \min(x_i, x_j)}$$

X = nilai setelah normalisasi

x = nilai pada atribut y

min (x_i, x_j) = nilai terkecil pada atribut y

max (x_i, x_j) = nilai terbesar pada atribut y

e. Evaluasi Pola (Pattern Evaluation)

Evaluasi pola bertugas untuk mengidentifikasi pola-pola yang menarik ke dalam knowledge based yang ditemukan. Pada tahap ini dihasilkan polapola yang khas dari

model klasifikasi yang dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

f. Presentasi Pengetahuan (Knowledge Presentation)

Knowledge presentation merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan atau informasi yang telah digali oleh pengguna. Tahap terakhir dari proses data mining adalah memformulasikan keputusan dari hasil analisis yang didapat.

5. Ukuran Kemiripan

Untuk mengukur kemiripan antara dua objek digunakan ukuran jarak. Semakin besar nilai jarak maka semakin besar perbedaan antar kedua objek. Jarak Euclidean merupakan jarak terpendek antara dua titik. Jarak Euclidean ditulis sebagai

$$d_{\text{euc}}(X_{ij}, C_{kj}) = \sqrt{\sum_{j=1}^p \sum_{i=1}^n (X_{ij} - C_{kj})^2}$$

X_{ij} = Objek pengamatan ke-i pada variable ke-j

C_{kj} = Pusat Cluster ke- k pada variable ke-j

p = Banyaknya variabel yang diamarti

n = Banyaknya pengamatan

6. CLARA (Clustering Large Application)

Pada kasus dengan dataset yang besar, PAM atau K-Medoids tidak akan berjalan dengan baik. Maka metode berbasis sampling yang disebut dengan CLARA (*Clustering Large Application*) yang di perkenalkan oleh (Kaufman and Rousseeuw, 1990) bisa di gunakan. Dalam hal ini CLARA menggunakan sample dataset secara acak. Dalam banyak kasus, sample yang semakin besar akan bekerja dengan baik, sehingga setiap objek memiliki probabilitas yang sama untuk dipilih sebagai sample. Objek-objek yang dipilih menjadi pusat kluster (medoids) akan cenderung mirip dengan yang sudah dipilih dari seluruh dataset. CLARA membuat clustering dari banyak sampel secara acak dan menghasilkan clustering terbaik sebagai outputnya. Kompleksitas dalam menghitung medoids pada sample acak adalah sebagai berikut:

$O(k s_2 + k(n-k))$
 s = Ukuran Sampel
 k = Jumlah Cluster
 n = Total Objek.
 = Total Objek.

Pada algoritma PAM mencari K-Medoids terbaik di antara dataset, tetapi CLARA mencari k-medoids terbaik di antara sample dataset yang terpilih. CLARA tidak bisa menghasilkan clustering yang baik jika medoids terbaik yang di disampelkan sangat jauh dari k-medoids terbaik. Jika objek merupakan salah satu dari k-medoids terbaik tetapi tidak dipilih selama proses sampling maka CLARA tidak akan pernah menghasilkan clustering terbaik (Han, Kamber and Pei, 2012).

7. DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*)

DBSCAN merupakan salah satu algoritma clustering yang mengelompokkan obyek ke dalam cluster dimana kategori atau *class* data ujinya belum diketahui (*unsupervised learning*). DBSCAN adalah algoritma pengelompokan yang didasarkan pada kepadatan (density) data. Konsep kepadatan yang dimaksud dalam DBSCAN adalah jumlah data yang berada dalam radius Eps (ϵ) dari setiap data. Jika jumlah data dalam radius ϵ lebih dari atau sama dengan MinPts (jumlah minimal data dalam radius ϵ), data tersebut masuk dalam kategori kepadatan yang diinginkan, jumlah data dalam radius tersebut termasuk data itu sendiri (Prasetyo, 2012). Metode ini melakukan pengelompokan dengan baik pada data berkepadatan tinggi, dan bisa menemukan bentuk sembarang kelompok dengan baik. DBSCAN dapat memisahkan data berkepadatan tinggi dan data berkepadatan rendah karena algoritma ini berbasis kepadatan data.

8. *Silhouette Index (SI/ Coefficient Silhouette)*

SI dapat digunakan untuk memvalidasi baik sebuah data, cluster tunggal (satu cluster dari sejumlah cluster), atau bahkan keseluruhan cluster. Untuk menghitung nilai SI dari

sebuah data ke- i , ada 2 komponen yaitu a_i dan b_i . a_i adalah rata-rata jarak data ke- i terhadap semua data lainnya dalam satu cluster, sedangkan b_i^j didapatkan dengan menghitung rata-rata jarak data ke- i terhadap semua data dari cluster yang lain tidak dalam satu cluster dengan data ke- i , kemudian diambil yang terkecil (Tan et al, 2006, dalam Prasetyo, 2014).

METODE PENELITIAN

1. Sumber data

Data yang digunakan merupakan data sekunder tentang data gempa bumi diantaranya kekuatan gempa dan kedalaman gempa itu terjadi pada daerah tingkat Kota/Kabupaten di Sumatra Barat. Data gempa bumi di dapatkan pada data yang sudah tersedia di *Incorporated Research Institutions for Seismology (IRIS)* yang dihimpun dari tanggal 1 Januari 2010 hingga tanggal 31 Desember 2020.

2. Variabel Penelitian

Variabel dari penelitian ini yaitu Kekuatan Gempa (Magnitudo) dan Kedalaman Gempa (Depth).

3. Teknik Analisis Data

Teknik analisis data pada penelitian ini antara lain:

1. Melakukan analisis deskriptif terhadap data tentang kejadian gempa bumi di Sumatra Barat dari tahun 2010 sampai dengan 2020.
2. Melakukan analisis data menggunakan tahapan KDD (*Knowledge Discovery Database*).
3. Melakukan tahapan *clustering* menggunakan metode CLARA yaitu:
 - a. memilih secara acak pusat cluster awal (medoid) sebanyak k ,
 - b. menghitung jarak setiap objek terhadap masing-masing medoid pada tiap *cluster*
 - c. menempatkan tiap objek tersebut ke medoid terdekat,
 - d. menghitung total jaraknya,
 - e. memilih kandidat medoid baru,
 - f. menghitung jarak setiap objek terhadap kandidat medoid baru dan menempatkan tiap
 - g. objek tersebut ke kandidat medoid terdekat,
 - h. menghitung total jarak,
 - i. menghitung selisih total jarak (S), yaitu selisih dari total jarak pada

- kandidat medoid baru dengan total jarak pada medoid lama,
- j. jika diperoleh nilai $S < 0$, maka kandidat medoid tersebut menjadi medoid baru,
 - k. ulangi langkah e sampai i hingga tidak terjadi perubahan medoid,
4. Melakukan tahapan *clustering* menggunakan metode DBSCAN yaitu:
 - a. Menentukan dua parameter penting, yaitu *Eps* (ϵ) dan jumlah data tetangga minimal untuk membentuk kelompok (*MinPts*).
 - b. melakukan pengelompokan data dengan algoritma DBSCAN.
 - c. Pilih titik awal r secara acak.
 - d. Inisialisasi parameter input *MinPts* dan *Eps*.
 - e. Hitung *Eps* atau semua jarak densitas terjangkau terhadap r menggunakan jarak euclidean seperti persamaan.
 - f. Jika jumlah titik yang memenuhi *Eps* lebih dari *MinPts* maka titik r adalah titik pusat (*core*) dan cluster terbentuk.
 - g. Ulangi langkah d – e hingga semua titik diproses. Jika r adalah titik border dan tidak ada titik yang densitas terjangkau terhadap r , maka proses dilanjutkan ke titik yang lain.
 5. Melakukan pengujian hasil pengelompokan akan diuji tingkat validitasnya menggunakan *Indeks Validitas Silhouette* untuk menentukan jumlah cluster yang terbaik pada metode DBSCAN sekaligus membandingkan *Indeks Validitas Silhouette* pada metode CLARA.
 6. Melakukan pemetaan daerah rawan gempa di Sumatra Barat berdasarkan metode terbaik.

HASIL DAN PEMBAHASAN

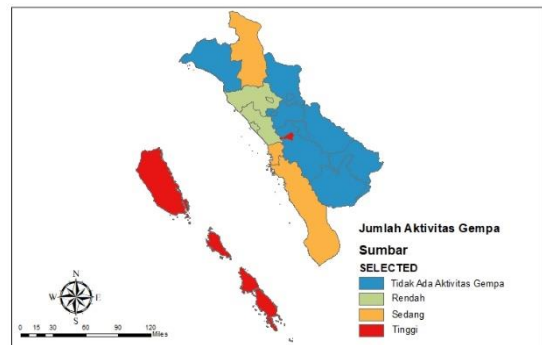
1. Statistik Deskriptif Variabel

Variabel	Minimum	Maksimum	Rata-rata
Mag	3	6,3	4,54
Depth	7,40	98	39,24

Pada variabel kekuatan gempa (*Mag*) menunjukkan bahwa kekuatan terkecil yang terjadi di Sumatra Barat berada pada 3 SR dan kekuatan terkecil nya berada di angka 6,3 SR. Rata-rata kejadian pada kekuatan gempa di sekitar 4,54. Sementara itu untuk variabel kedalaman gempa (*Depth*) terjadi paling minimum pada kedalaman 7,40 km

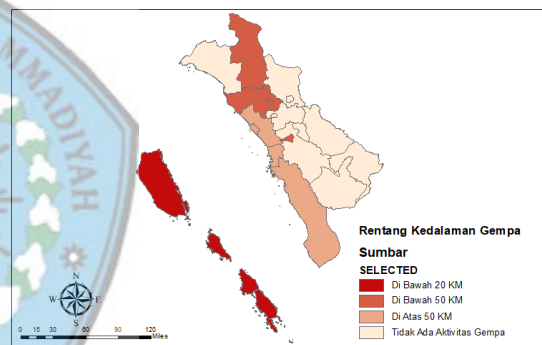
dan untuk kedalaman maksimal berada di 98 km dengan rata-rata pada kedalaman gempa berada disekitar 39,24.

Peta Sumatera Barat Berdasarkan Daerah Kerawanan Menurut Jumlah Aktivitas Gempa



Pada gambar diatas menunjukkan peta Sumatra Barat berdasarkan daerah kerawanan gempa dengan jumlah aktivitas gempa dari tingkat aktivitas tidak ada sampai dengan aktivitas kegempaan yang tinggi

Peta Sumatera Barat Berdasarkan Daerah Kerawanan Menurut Titik Kedalaman Gempa Bumi



Pada gambar diatas menunjukkan peta Sumatra Barat berdasarkan daerah kerawanan gempa titik kedalaman gempa dari yang terendah sampai dengan titik kedalaman yang tinggi, Pada dasarnya dengan titik kedalaman gempa yang rendah akan membuat gempa berpotensi tsunami dan memiliki potensi kerusakan lebih parah.

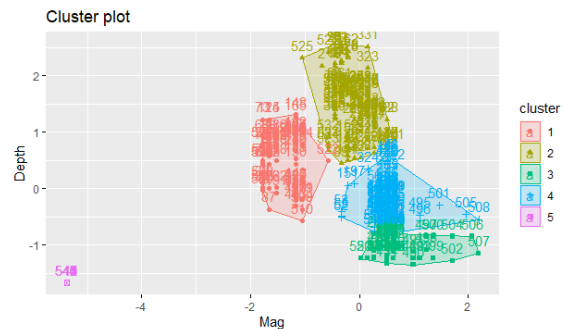
2. Clustering Menggunakan Metode CLARA (Clustering Large Application)

Data yang terhimpun pada data kejadian perhari dari tanggal 1 Januari 2010 sampai dengan 31 Desember 2020 tercatat bahwa ada 539 kejadian gempa. Gempa yang terjadi beragam di antara 3 SR sampai dengan 6,3 SR yang berpusat kejadian itu terjadi di Sumatra Barat. Titik kejadian gempa tersebut di akan dikelompokkan berdasarkan Kabupaten/Kota dimana titik lokasi pusat gempa. Hasil clustering dengan menggunakan metode CLARA yang telah dilakukan didapatkan bahwa dari 5 cluster yang terbentuk terdapat daerah yang termasuk daerah

rawan gempa dilihat berdasarkan Kekuatan (Mag) terbesar yang berada pada cluster tersebut.

No	Kab/Kota	Cluster	Maksimal Kekuatan Gempa	Kedalaman Terendah
1	Kab. Agam	1	3,6 Mag	34 Km
2	Kota Bukittinggi	1	3,6 Mag	34 Km
3	Kota Padang	2	4,7 Mag	49 Km
4	Kota Pariaman	2	4,7 Mag	49 Km
5	Kota Payakumbuh	4	4,2 Mag	40 Km
6	Kota Sawahlunto	5	0	0
7	Kota Solok	1	3,6 Mag	34 Km
8	Kota Padang Panjang	5	0	0
9	Kab. Pesisir Selatan	2	4,7 Mag	49 Km
10	Kab. Solok	5	0	0
11	Kab. Tanah Datar	1	3,6 Mag	34 Km
12	Kab. Padang Pariaman	2	4,7 Mag	49 Km
13	Kepulauan Mentawai	3	6,3 Mag	7,4 Km
14	Kab. Sijunjung	5	0	0
15	Kab. Solok Selatan	5	0	0
16	Kab. Dharmasraya	5	0	0
17	Kab. Pasaman Barat	5	0	0
18	Kab. Pasaman	2	4,7 Mag	49 Km
19	Kab. Lima Puluh Kota	5	0	0

termasuk daerah yang di kategorikan kerusakan berat (*heavy damage*). Pada cluster 4 memiliki kekuatan maksimal berada pada 4,2 Mag yang diartikan bahwa pada daerah tersebut kemungkinan terjadi gempa akan berada di titik 4,2 Daerah yang termasuk cluster 4 termasuk daerah yang di kategorikan kerusakan ringan (*slight damage*). Pada cluster 5 merupakan daerah yang tidak pernah menjadi titik pusat gempa. Daerah yang termasuk cluster 5 termasuk daerah yang di kategorikan tidak dirasakan (*non felt*).



Pada gambar diatas menunjukkan hasil Clustering dengan menggunakan metode CLARA Pada cluster 5 tersebut merupakan daerah yang sama sekali tidak terjadi gempa bumi dan pada akhirnya membentuk cluster nya sendiri. Pada plot gambar pun terlihat bahwa pembentukan cluster dilihat berdasarkan kedekatan antar objek dimana pada metode CLARA yaitu dengan menghitung jarak Euclidean dan mencari nilai minimum antar objeknya.

3. Clustering Menggunakan Metode DBSCAN(*Density Based Spatial Clustering of Applications with Noise*)

Pada tahap ini dilakukan beberapa skenario eksperimen berdasarkan kombinasi masukan dua parameter yaitu eps dan minPts. Hasil eksperimen yang menerapkan beberapa skenario eksperimen dapat dilihat pada tabel dibawah ini:

Pada metode CLARA membagi cluster atas 5 bagian yang masing masing cluster menunjukkan kekuatan gempa terbesar di masing-masing cluster. Pada cluster 1 memiliki kekuatan maksimal berada pada 3,6 Mag yang diartikan bahwa pada daerah tersebut kemungkinan terjadi gempa akan berada di titik maksimum 3,6 Mag cluster 1 termasuk daerah yang di kategorikan dirasakan (*felt*). Pada cluster 2 memiliki kekuatan maksimal berada pada 4,7 Mag kerusakan sedang (*moderate damage*. Pada cluster 3 memiliki kekuatan maksimal berada pada 6,3 Mag Daerah yang termasuk cluster 3

No	Eps	MinPts	Jumlah Cluster	Jumlah Noise	Average sillhouette width
1	1	1	30	0	0
2	1	2	19	11	0,05
3	1	3	18	13	0,07
4	1	4	15	28	0,09
5	1	5	13	46	0,05
6	1,25	1	13	0	0,02
7	1,25	2	10	3	0,17
8	1,25	3	10	3	0,17

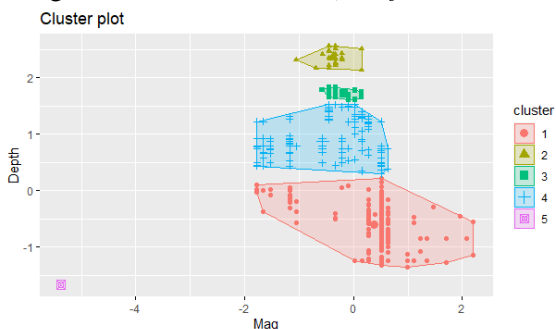
9	1,25	4	10	4	0,21	7	Kota Solok	4	4,9 Mag	49 Km
10	1,25	5	10	14	0,21	8	Kota Padang Panjang	5	0	0
11	1,5	1	7	0	0,09	9	Kab. Pesisir Selatan	4	4,9 Mag	49 Km
12	1,5	2	6	1	0,09	10	Kab. Solok	5	0	0
13	1,5	3	6	1	0,09	11	Kab. Tanah Datar	1	6,3 Mag	7,4 Km
14	1,5	4	7	1	0,11	12	Kab. Padang Pariaman	3	4,6 Mag	76 Km
15	1,5	5	7	5	0,3	13	Kepulauan Mentawai	1	6,3 Mag	7,4 Km
16	1,75	1	5	0	0,45	14	Kab. Sijunjung	5	0	0
17	1,75	2	5	0	0,45	15	Kab. Solok Selatan	5	0	0
18	1,75	3	5	0	0,45	16	Kab. Dharmasraya	5	0	0
19	1,75	4	6	0	0,45	17	Kab. Pasaman Barat	5	0	0
20	1,75	5	6	0	0,45	18	Kab. Pasaman	2	4,2 Mag	81 Km
21	2	1	3	0	0,4	19	Kab. Lima Puluh Kota	5	0	0
22	2	2	3	0	0,4					

Tahap evaluasi cluster dilakukan dengan menggunakan Average Silhouette width. Skor nilai *Average Silhouette width* antara 0 sampai dengan 1, lebih besar dan mendekati 1 menunjukkan kualitas cluster yang lebih baik. Hasil tabel tersebut terlihat bahwa nilai yang menunjukkan *Noise* paling kecil dan memiliki nilai *Average silhouette width* yang tinggi adalah nilai Eps 1.75 dengan MinPts sebesar 3. Hasil Eps dan MinPts yang ditentukan menghasilkan 5 cluster dengan nilai *Average silhouette width* sebesar 0,45. Maka dari itu ditetapkan dalam penelitian ini akan menggunakan Eps 1.75 dan MinPts 3 dengan hasil cluster 5 dan tidak terdapat daerah yang masuk kedalam noise sehingga menghasilkan *clustering* sebagai berikut.

No	Kab/Kota	Cluster	Maksimal Kekuatan Gempa	Kedalaman Terendah
1	Kab. Agam	1	6,3 Mag	7,4 Km
2	Kota Bukittinggi	1	6,3 Mag	7,4 Km
3	Kota Padang	2	4,2 Mag	81 Km
4	Kota Pariaman	3	4,6 Mag	76 Km
5	Kota Payakumbuh	1	6,3 Mag	7,4 Km
6	Kota Sawahlunto	5	0	0

Pada tabel diatas menunjukkan hasil clustering dengan menggunakan metode DBSCAN menunjukkan hasil dengan menjadi 5 cluster. Pada cluster 1 memiliki kekuatan maksimal berada pada 6,3 Mag yang diartikan bahwa pada daerah tersebut kemungkinan terjadi gempa akan berada di titik maksimum 6,3 Mag dengan kategori kerusakan berat (*heavy damage*). Pada cluster 2 memiliki kekuatan maksimal berada pada 4,2 Mag yang diartikan bahwa pada daerah tersebut kemungkinan terjadi gempa akan berada di titik 4,2 Mag. Daerah yang termasuk cluster 2 termasuk daerah yang di kategorikan dirasakan (*felt*). Pada cluster 3 memiliki kekuatan maksimal berada pada 4,6 Mag yang diartikan bahwa pada daerah tersebut kemungkinan terjadi gempa akan berada di titik 4,6 Mag. Daerah yang termasuk cluster 3 termasuk daerah yang di kategorikan kerusakan ringan (*slight damage*). Pada cluster 4 memiliki kekuatan maksimal berada pada 4,9 Mag yang diartikan bahwa pada daerah tersebut kemungkinan terjadi gempa akan berada di titik 4,9 Mag. Daerah yang termasuk cluster 4 termasuk daerah yang di kategorikan kerusakan sedang (*moderate damage*). Pada cluster 5 merupakan daerah yang tidak pernah menjadi titik pusat gempa. Daerah

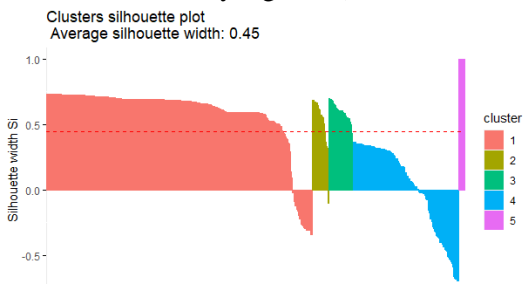
yang termasuk cluster 5 termasuk daerah yang di kategorikan tidak dirasakan (*non felt*).



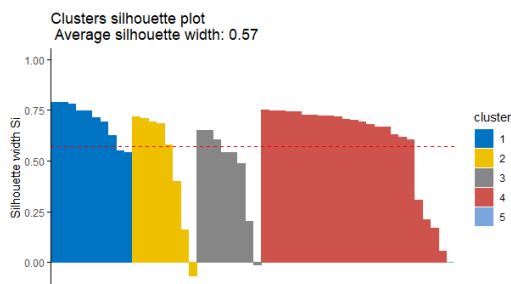
Pada plot cluster DBSCAN pada akhirnya membentuk tingkatan cluster dengan memperhatikan besaran kedua atribut yaitu kedalaman dan kekuatan. Pada gambar 4.5 menunjukan bahwa pada metode DBSCAN membentuk cluster berdasarkan kerapatan objek. Pola pembentukan antara metode CLARA dan metode DBSCAN cukup terlihat jelas. Pada metode CLARA membentuk pola berdasarkan kedekatannya dimana data dengan kedalaman dan kekuatannya tidak memiliki tingkatan atau urutan yang pasti di antara tiap cluster sementara pada metode DBSCAN pola yang terbentuk langsung memiliki urutan dan pola yang terbentuk pada akhirnya memiliki tingkatan.

4. Validasi Clustering dan Penentuan Metode Clustering Terbaik

Tahap evaluasi cluster dilakukan dengan menggunakan Average Silhouette width atau Coefficient Silhouette. Skor nilai Average Silhouette width antara 0 sampai dengan 1, lebih besar dan mendekati 1 menunjukkan kualitas cluster yang lebih baik, yaitu memiliki jarak inter-cluster (jarak antar satu cluster dengan cluster lainnya) tinggi dan jarak intracluster (jarak antar obyek dengan obyek lainnya di dalam satu cluster yang sama) rendah.

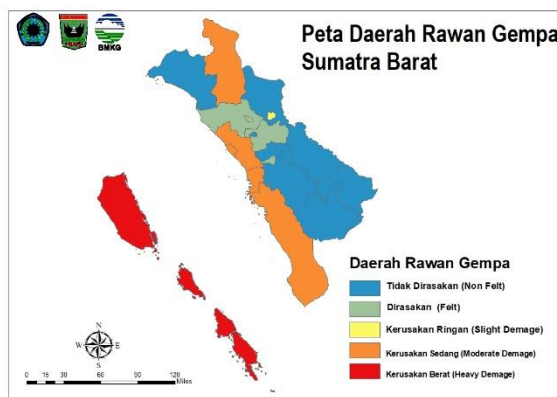


Validasi Clustering DBSCAN



Validasi Clustering CLARA

Pada gambar diatas memperlihatkan hasil validasi clustering dengan menggunakan *coefficient sillhoutte* dan menggunakan grafik validasi cluster. Pada metode DBSCAN menunjukan hasil *Average Silhouette Width* sebesar 0,45 sementara pada metode CLARA menunjukan hasil *Average Silhouette Width* sebesar 0,57. Maka untuk menentukan metode pengelompokkan pada data kejadian gempa bumi di Sumatra Barat dilihat dari nilai yang memberikan hasil yang lebih tinggi. Pada nilai tersebut terlihat bahwa metode CLARA memiliki nilai yang lebih tinggi dibandingkan dengan metode DBSCAN. Maka dari itu dipilih metode CLARA lebih cocok untuk data kejadian gempa bumi di Sumatra Barat dengan menghasilkan 5 cluster. Pada metode sillhoutte, nilai yang menunjukan nilai berada di bawah 0 atau minus menunjukan bahwa titik ataupun data tersebut tidak cocok untuk berada pada cluster tersebut. Pada metode DBSCAN menunjukan nilai yang berada pada nilai minus cukup banyak dan itu berada pada cluster 1, cluster 2 dan cluster 4, sementara pada metode CLARA hanya menunjukan pada cluster 2 dan itu pun tidak sebanyak pada metode DBSCAN. Maka dari penjelasan tersebut lah yang pada akhirnya membuat metode CLARA lebih baik dibandingkan dengan metode DBSCAN.



KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan diperoleh beberapa kesimpulan sebagai berikut:

1. Pada pengelompokan daerah rawan gempa di Sumatra Barat dengan menggunakan metode *clustering* CLARA (*Clustering Large Application*) membentuk 5 *cluster* dengan daerah yang merupakan daerah dengan kerusakan berat (*heavy damage*) adalah Kepulauan Mentawai.
2. Pada pengelompokan daerah rawan gempa di Sumatra Barat dengan menggunakan metode *clustering* DBSCAN (*Density Based Spatial Clustering Of Applications With Noise*) membentuk 5 *cluster* dengan daerah dengan kerusakan berat (*heavy damage*) adalah Kab. Agam, Kota Bukittinggi, Kota Payakumbuh, Kab. Tanah Datar dan Kepulauan Mentawai.
3. Setelah dilakukan pengelompokan menggunakan 2 metode yaitu CLARA dan DBSCAN, dilakukan validasi *clustering* menggunakan *coefficient silhouette*. Pada metode CLARA memberikan nilai *Average Silhouette width* sebesar 0,57 sedangkan pada metode DBSCAN memberikan nilai *Average Silhouette width* sebesar 0,45. Maka dapat disimpulkan bahwa metode CLARA lebih cocok atau lebih baik digunakan untuk *clustering* daerah rawan gempa di Sumatra Barat.

DAFTAR PUSTAKA

- Akrnam, dkk. 2006. *Studi Karakteristik Fisis Sumber Geopabumi Di Wilayah Sumatera Barat*. Jurusan Fisika Fakultas Matematika Dan Ilmu Pengetahuan Alam Universitas Negeri Padang
- Curry, J. R. G.G Shor Jr, R.W, Rairr and Henry. 1977. *Seismic Refraction and Reflection Studies of Crustal Structure of Eastren Sunda and Westren Sunda Arcs*. *Jurnal of Geophysis Research*, Vol. 82, 2479-2489.
- Ginting Grachella Yowanda Br. 2018. *Clara (Clustering Large Application) Pada Data Simulasi Trivariat 1000 Objek*. Bandung; Universitas Pendidikan Indonesia
- repository.upi.edu
|perpustakaan.upi.ed
- Helilintar Risa, Risky Aswi Ramadhani, dan Siti Rochana. 2017. *DATA MINING K-Nearest Neighbor*. Kediri; Fakultas Teknik Universitas Nusantara PGRI Kediri.
- Kurmiati, Dilla. 2021. *Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm Klasterisasi Daerah Rawan Gempa Bumi di Indonesia Menggunakan Algoritma K-Medoids*. MALCOM: Indonesian Journal of Machine Learning and Computer Science Journal Homepage: <https://journal.irpi.or.id/index.php/malcom> Vol. 1 Iss. 1 April 2021, pp: 47-57 P-ISSN: 0000 0000 E-ISSN: 0000 0000
- Mc.Caffray, R, P. Molnor, S.W. Roecher and Y.S.Joyodiwirjo. 1985. *Microearthquake Seismicity and Fault Plane Solution Related to Are Conitnent Collision in the Eastern Sunda- are Indonesia*. *Jurnal Of Geophysis Research*, Vol. 90, pp 4511-4528
- Mohamad Bentar Cahyadahrena. 2015. *Deteksi Pencilan Pada Data Titik Panas Menggunakan Clustering Berbasis Medoids*. Departemen Ilmu Komputer Fakultas Matematika Dan Ilmu Pengetahuan Alam Institut Pertanian Bogor
- Muslim, Much Aziz, dkk. 2019. *Data Mining Algoritma C4.5 Disertai contoh kasus dan penerapannya dengan program computer*. Semarang: Universitas Dian Nuswantoro
- Novi Shella, 2020. *Clustering Status Desa Menggunakan Metode Dbscan (Density-Based Spatialclustering Of Application With Noise)*. Pekanbaru: Fakultas Sains Dan Teknologi

Universitas Islam Negeri Sultan Syarif
Kasim Riau Pekanbaru

Posavec, M, D. Taylor, th van euwan and A
Spector. 1973. *Tectonic Controls of
Volcaism and Complex Movement
Alang the Sumatram Fault System.*
GSociaty Malaysia, Bulleten , pp-
43-60

Rifa Isna Hidayatur dkk, 2019.
*.Implementasi Algoritma Clara Untuk
Data Gempa Bumi Di Indonesia.*
Program Studi Statistika, Fakultas
Matematika dan Ilmu Pengetahuan
Alam Universitas Sebelas Maret.
Seminar Nasional Penelitian
Pendidikan Matematika (Snp2m) 2019

Sari ,Betha Nurina dan Aji Primajaya. 2019.
*Penerapan Clustering Dbscan Untuk
Pertanian Padi Di Kabupaten
Karawang .Jurnal Informatika dan
Komputer (JIKO) – Vol. 4, No. 1,
Februari 2019 28*

Silitonga ,Parasian. 2016. *Analisis Pola
Penyebaran Penyakit Pasien
Pengguna Badan Penyelenggara
Jaminan Sosial (Bpjs) Kesehatan
Dengan Menggunakan Metode
Dbscan Clustering (Studi Kasus
Rumah Sakit Umum Pusat Haji
Adam Malik Medan) . Jurnal TIMES ,
Vol. V No 1 : 36-39 , 2016 ISSN :
2337 - 3601 36 Fakultas Ilmu
Komputer Unika St. Thomas S.U*

Wardianti , Muhammad Kasim Aidid dan
Muhammad Nusrang. 2019.
*Pengelompokan Kabupaten/Kota
Provinsi Sulawesi Selatan dan Barat
Berdasarkan Angka Partisipasi
Pendidikan SMA/SMK/MA
Menggunakan K-Medoid dan
CLARA. VARIANSI: Journal of
Statistics and Its Application
on Teaching and Research ISSN 2684-
7590 (Online) Vol. 1 No. 3
(2019), 48-65*

