

PENGELOMPOKAN TITIK GEMPA DI PULAU SULAWESI MENGGUNAKAN ALGORITMA ST-DBSCAN (*Spatio Temporal-Density Based Spatial Clustering Application with Noise*)

Denny Jales Manalu¹, Rita Rahmawati², Tatik Widiharih³

^{1,2,3}Departemen Statistika, Fakultas Sains dan Matematika, Universitas Diponegoro
dennyjales@gmail.com

ABSTRACT

Earthquake is a natural disaster which is quite serious in Indonesia, especially on Sulawesi Island. Earthquake is fearful because it can't be predicted when it will come, where it will come, and how strong the vibration, that often causes fatal damage and casualties. In effort to minimize losses caused by earthquake, it is necessary to divide areas which are easily affected by earthquake. One of the methods that can be used in dividing the area is by using the clustering technique. This research by using a clustering method with the ST-DBSCAN (Spatial Temporal-Density Based Spatial Clustering Application with Noise) algorithm on dataset of earthquake points in Sulawesi Island in 2019. This method by using the spatial distance parameters (Eps1 = 0.45), the temporal distance parameters (Eps2 = 7), and minimum number of cluster members (MinPts = 4), resulting in a total of 60 clusters with 8 large clusters and 216 noises

Keywords: Data mining, earthquake, clustering, spatio-temporal, ST-DBSCAN

1. PENDAHULUAN

Gempa bumi merupakan bencana alam yang cukup serius di Indonesia dikarenakan tidak dapat diperkirakan kapan datangnya, dimana tempatnya, dan seberapa kuat getarannya. Gempa bumi menciptakan gelombang seismik, yaitu gelombang yang bergerak melalui bumi akibat deformasi struktural, tekanan, atau tegangan sifat elastis kerak bumi. Gelombang ini membawa energi dan kemudian merambat ke segala arah di seluruh bagian bumi dan dapat direkam oleh seismograf (Siswawidjono, 1996).

Pulau Sulawesi merupakan salah satu kawasan seismik aktif di Indonesia yang mengakibatkan wilayah ini tergolong rawan terjadi gempa bumi. Pulau Sulawesi memiliki aktifitas kegempaan yang tinggi dikarenakan lokasinya terletak pada zona benturan ketiga lempeng tektonik dunia, yaitu lempeng Indo-Australia, lempeng Eurasia dan lempeng Pasifik. Ketiga lempeng ini memiliki pertemuan yang konvergen dan bertumbukan secara relatif.

Kepadatan penduduk dan infrastruktur sangat menentukan risiko wilayah yang telah dinyatakan rawan gempa bumi. Hal mendasar yang dapat dilakukan adalah pembagian wilayah-wilayah mana saja yang paling sering terjadi gempa bumi agar risiko korban jiwa dan infrastruktur dapat diminimalisir. Untuk memudahkan pembagian wilayah gempa dapat dilakukan dengan analisis *cluster* atau sering disebut dengan *clustering*.

Salah satu metode *clustering* yang dapat digunakan adalah ST-DBSCAN (*Spatio Temporal-Density Based Spatial Clustering Applications with Noise*). Algoritma ST-DBSCAN termasuk algoritma nonparametrik dalam *unsupervised learning* sehingga tidak memerlukan asumsi dalam pengerjaannya dan merupakan salah satu metode yang paling baik untuk menentukan *cluster* dari database *spasial* yang besar (Gaonkar & Swant, 2013). Penelitian sebelumnya oleh Trisnaningtyas (2014), melakukan analisis ST-DBSCAN dengan mengabaikan parameter jarak temporal (Eps2=0) dikarenakan data yang diteliti tidak termasuk data temporal. Pada penelitian ini menggunakan ketiga parameter dari ST-DBSCAN yaitu parameter jarak spasial (Eps1), parameter jarak temporal (Eps2), dan jumlah minimal anggota *cluster* (MinPts) karena didukung oleh data titik gempa bumi yang diteliti.

Berdasarkan uraian tersebut, penelitian ini bertujuan menerapkan algoritma ST-DBSCAN (*Spatio Temporal-Density Based Spatial Clustering Applications with Noise*) dalam membentuk *cluster* dari dataset titik gempa bumi Pulau Sulawesi pada tahun 2019 dan menentukan kualitas dari *cluster* dengan menggunakan *Silhouette Coefficient*.

2. TINJAUAN PUSTAKA

2.1 Gempa Bumi

Gempa bumi merupakan getaran yang terjadi secara tiba-tiba akibat adanya pelepasan energi dari pusat gempa. Energi ini disebut dengan gelombang seismik. Gempa bumi khususnya yang terjadi di Pulau Sulawesi tergolong gempa bumi yang merusak. Kejadian waktu dan tempat gempa bumi dapat dianalisis dengan melihat database titik gempa bumi yang sangat padat. *Data mining* dengan metode *clustering* merupakan salah satu analisis yang dapat digunakan untuk menganalisis data titik gempa bumi yang tujuannya menghasilkan pembagian wilayah rawan gempa.

2.2 Data Mining

Data mining dapat didefinisikan sebagai proses memiih, memeriksa, dan memodelkan data dalam jumlah besar untuk menemukan pola atau tren yang biasanya tidak disadari keberadaannya. Pengetahuan yang didapat dari proses *data mining* harus baru, mudah dimengerti, dan bermanfaat. Dalam *data mining*, data disimpan secara elektronik dan otomatis diproses oleh komputer menggunakan metode dan perhitungan tertentu (Pramadhani & Setiadi, 2014). Data mining berarti kegiatan yang melibatkan pengumpulan, penggunaan data historis untuk menemukan pola atau hubungan dalam dataset besar. Salah satu metode dalam data mining adalah analisis *cluster* atau sering disebut dengan *clustering*. Tujuan utama analisis *cluster* adalah untuk menggolongkan individu atau objek yang terkait menjadi jumlah yang lebih kecil. *Cluster* yang baik diharapkan memiliki kesamaan (homogenitas) di antara anggota suatu kelompok dan adanya perbedaan (heterogenitas) yang tinggi antar kelompok (Kuncoro, 2009)

2.3 Dissimilarity Matrix Jarak Euclidean

Matriks perbedaan (*dissimilarity matrix*) menyimpan kedekatan atau jarak untuk semua pasangan dari n objek. Jarak *Euclidean* adalah jarak antara dua objek data (i, j) dari n atribut bernilai numerik, yang dinyatakan sebagai $\mathbf{i} = x_{i1}, x_{i2}, \dots, x_{in}$ dan $\mathbf{j} = x_{j1}, x_{j2}, \dots, x_{jn}$ pada ruang dimensi n (R^n) (Han, et al., 2012):

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2}$$

untuk aspek spasial persamaan jarak *euclidean* menjadi (Han, et al., 2012):

$$d(i, j) = \sqrt{(x_{long_i} - x_{long_j})^2 + (x_{lat_i} - x_{lat_j})^2}$$

dimana : $long\ i = longitude$ data ke- $i, i=1, 2, \dots, n$

$lat\ j = latitude$ data ke- $j, j=1, 2, \dots, n$

untuk aspek temporal persamaan jarak *euclidean* dimodifikasi menjadi persamaan (Han, et al., 2012):

$$d(i, j) = |x_{tanggal_i} - x_{tanggal_j}|$$

dimana : $x_{tanggal}$ adalah objek pada kolom tanggal dan merupakan objek satu dimensi.

2.4 Algoritma ST-DBSCAN

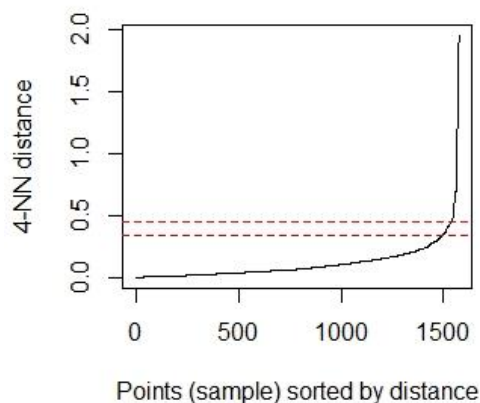
Tahapan dari algoritma ST-DBSCAN (Birant & Kut, 2007):

1. Menentukan nilai Eps dan MinPts
2. Hitung semua jarak euclidean antar objek berdasarkan aspek spasial dan temporal

3. Membentuk matriks jarak untuk semua pasangan dari n objek berdasarkan spasial dan temporal
4. Mulai dari titik pertama pertama kemudian mengambil semua titik pada aspek spasial dan aspek temporal dengan ketentuan:
 $A = \{x \mid x \leq \text{Eps1}, x \in \text{matriks jarak spasial}\}$
 $B = \{x \mid x \leq \text{Eps2}, x \in \text{matriks jarak temporal}\}$
5. Mengambil semua irisan dari aspek spasial dan aspek temporal dengan ketentuan:
 $A \cap B = \{x \mid x \in A \text{ dan } x \in B\}$
6. Jika jumlah objek pada irisan lebih kecil dari nilai MinPts, maka titik dinilai sebagai noise.
7. *Cluster* terbentuk jika titik memenuhi parameter Eps1, Eps2 dan MinPts.
8. Apabila titik p merupakan border point dan tidak terdapat titik lain pada irisan maka dilanjutkan ke titik selanjutnya dan *cluster* baru terbentuk.
9. Ulangi langkah 4 - 8 sampai semua titik terproses.
10. Jika dua *cluster* C1 dan C2 dekat satu sama lain, sebuah titik q mungkin termasuk dalam kedua *cluster* tersebut. Namun, algoritma ini akan menyatakan titik q sebagai bagian dari *cluster* yang pertama kali menemukannya

2.5 Penentuan Eps dan MinPts

Penentuan nilai parameter tersebut dilakukan secara *trial and error* menggunakan grafik *k-dist*. Grafik *k-dist* dibentuk dengan cara menghitung jarak antar objek kemudian diurutkan secara menurun dan plot nilai jarak terkecil ke-*k* dari setiap objek. Sumbu *x* menyatakan objek yang telah diurutkan, sedangkan sumbu *y* menyatakan nilai jarak dari objek. Titik dimana terjadi perubahan tajam pada *k-dist* akan digunakan sebagai Eps dan nilai *k* sebagai MinPts (Purwanto, 2012).



2.6 Pengukuran Kualitas Cluster

Salah satu ukuran untuk mengecek kualitas *cluster* adalah *Silhouette Coefficient*. Nilai $a(i)$ mewakili kepadatan *cluster* yang berisi objek *i*. Semakin kecil nilai $a(i)$, semakin padat *cluster* tersebut. Sementara itu, nilai $b(i)$ mendeskripsikan seberapa jauh objek *i* dari *cluster-cluster* lain. Jika $a(i)$ bernilai sangat kecil dan $b(i)$ bernilai sangat besar, maka *Silhouette Coefficient* dari *i* akan mendekati 1 yang berarti *cluster* sangat padat dan terpisah jauh dari *cluster-cluster* lain. Sebaliknya, jika $a(i)$ bernilai sangat besar dan $b(i)$ bernilai sangat kecil, maka *Silhouette Coefficient* dari *i* akan mendekati negatif 1 yang berarti *cluster* tidak padat dan berjarak sangat dekat dengan *cluster-cluster* lain (Suyanto, 2019)

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

dengan:

- $s(i)$: nilai *Silhouette Coefficient* dari i
 $a(i)$: rata-rata jarak objek i ke semua objek yang lain dalam *cluster* A
 $b(i)$: rata-rata jarak minimum dari objek i ke semua *cluster* lain (yang bukan *cluster* dari objek i)

3. METODOLOGI PENELITIAN

3.1. Data dan Variabel Penelitian

Data yang digunakan dalam penelitian ini adalah data gempa bumi di Pulau Sulawesi pada tahun 2019. Data diperoleh dari pusat *database* Badan Meteorologi, Klimatologi, dan Geofisika (BMKG). Data titik gempa bumi berjumlah 1575 dengan atribut tanggal, *longitude*, *latitude*, kekuatan gempa (*magnitudo*), dan kedalaman (*depth*).

Variabel penelitian yang digunakan untuk penelitian ini yaitu *longitude*, *latitude*, dan tanggal terjadinya gempa. *Longitude* merupakan koordinat bujur geografi tempat terjadinya gempa, sedangkan *latitude* merupakan koordinat lintang geografi tempat terjadinya gempa

3.2. Langkah-langkah Penelitian

1. Melakukan *preprocessing data*
 - a. Perubahan format pada kolom tanggal
 - b. Reduksi data
2. Menentukan nilai parameter Eps dan MinPts menggunakan grafik observasi *k-dist*
3. Menghitung jarak *euclidean* antar objek berdasarkan aspek spasial (*longitude latitude*) dan berdasarkan aspek temporal (tanggal terjadinya gempa)
4. Melakukan analisis *cluster* dengan algoritma ST-DBSCAN:
5. Melakukan pengukuran kualitas *cluster* menggunakan Silhouette Coefficient

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$
6. Menganalisa pola *spatio-temporal*
7. Melakukan visualisasi dan interpretasi hasil dari terbentuknya *cluster-cluster* yang terbentuk dari metode ST-DBSCAN

4. HASIL DAN PEMBAHASAN

4.1. Preprocessing Data

Tahapan preprocessing data yang dilakukan adalah sebagai berikut:

1. Melakukan format pada kolom tanggal
 Pengubahan format ini dilakukan agar ketika terjadi penambahan data sebelum tanggal pertama data yang digunakan saat ini tidak menyebabkan variabel tanggal bernilai -1

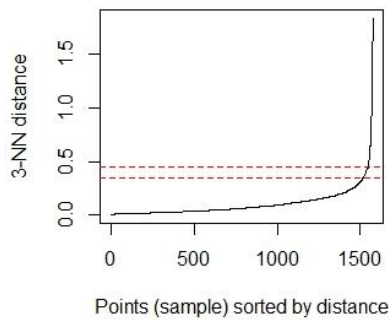
Tabel 1. Transformasi Format Tanggal menjadi Number

No	Tanggal (GMT)	Tanggal (number)
1	01/01/2019 (00:58:45)	43466
2	01/01/2019 (04:01:55)	43466
3	01/01/2019 (10:15:16)	43466
4	01/01/2019 (10:57:20)	43466
5	01/01/2019 (15:45:18)	43467
⋮	⋮	⋮
1975	31/12/2019 (09:50:44)	43830

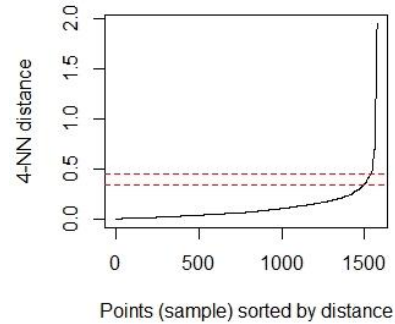
2. Reduksi data
 Reduksi dilakukan dengan memilih variabel yang akan digunakan. Variabel awal yaitu *longitude*, *latitude*, tanggal gempa, kekuatan gempa, dan kedalaman gempa. Pada tahap ini dilakukan reduksi data yang akan digunakan yaitu *longitude*, *latitude*, dan tanggal terjadinya gempa.

4.2. Penentuan Nilai Eps dan MinPts

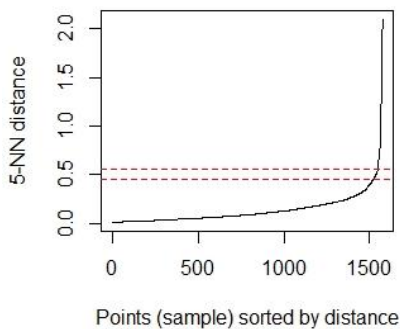
Pemilihan nilai Eps dan MinPts yang optimal dipilih berdasarkan grafik *k-dist*. Komputasi dilakukan untuk mendapatkan grafik *k-dist* pada nilai $k=3$, $k=4$, $k=5$, $k=7$.



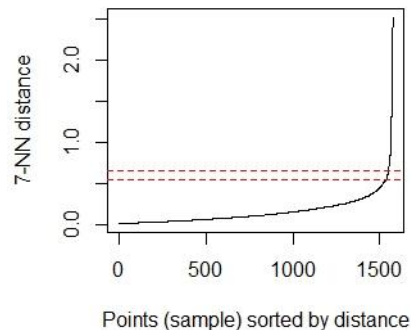
Gambar 1. *K-dist* dengan $k=3$



Gambar 2. *K-dist* dengan $k=4$



Gambar 3. *K-dist* dengan $k=5$



Gambar 4. *K-dist* dengan $k=7$

Dapat dilihat pada Gambar 1, Gambar 2, Gambar 3, dan Gambar 4, jika menggunakan MinPts=3 maka Eps1 yang optimal adalah 0,35 sampai dengan 0,45; jika menggunakan MinPts=4 maka Eps1 yang optimal adalah 0,35 sampai dengan 0,45; jika menggunakan MinPts=5 maka Eps1 yang optimal adalah 0,45 sampai dengan 0,55; jika menggunakan MinPts=7 maka Eps1 yang optimal adalah 0,55 sampai dengan 0,65.

Pemilihan nilai Eps2 yaitu 7 hari, 14 hari, 21 hari dan 30 hari, sedangkan untuk parameter Eps1 yang akan digunakan adalah nilai ambang batas atas, nilai ambang batas bawah, dan nilai tengah dari grafik *k-dist*. Nilai Eps1 yang dipilih pada penelitian ini dapat dilihat pada Tabel 3.

Tabel 2. Nilai Eps1 yang Diteliti

MinPts	Eps1
3	0,35
	0,4
	0,45
4	0,35
	0,4
	0,45
5	0,45
	0,5
	0,55
7	0,55
	0,6
	0,65

4.3. Jarak Euclidean

Dalam tahapan pembentukan *cluster* diperlukan menghitung jarak antar objek. Jarak antar objek dihitung menggunakan jarak euclidean. Jarak euclidean dihitung berdasarkan

aspek spasial (*longitude & latitude*) dan aspek temporal (tanggal gempa) dapat dilihat pada Tabel 4.

Tabel 3. Perhitungan Jarak Euclidean

(i,j)	d(i,j) spasial	d(i,j) temporal
(1,1)	0	0
(1,2)	4,583	0
(1,3)	0,816	0
(1,4)	3,089	0
(1,5)	3,045	1
(1,6)	4,555	1
(1,7)	0,769	1
(1,8)	3,429	2
⋮	⋮	⋮
(1574,1575)	2,991	0
(1575,1575)	0	0

4.4. Pembentukan Cluster

Hasil pembentukan *cluster* dari 48 pasangan parameter Eps1, Eps2, dan MinPts pada data titik gempa bumi Pulau Sulawesi menggunakan algoritma ST-DBSCAN dengan menggunakan bantuan software R. Setelah mendapatkan hasil *cluster* dari 48 pasangan selanjutnya dipilih parameter dengan nilai *Silhouette Coefficient* tertinggi dari tiap MinPts pada Tabel 5.

Tabel 4. Parameter Terbaik dari Tiap MinPts

MinPts	Eps1	Eps2	Jumlah Cluster	Jumlah Noise	Silhouette Spasial	Silhouette Temporal
3	0,45	21	33	152	0,0077	-0,5769
4	0,45	7	60	216	0,0569	-0,5723
5	0,5	21	19	195	0,0363	-0,3449
7	0,55	21	11	249	0,0827	-0,1105

Pada Tabel 5 dapat dilihat bahwa parameter Eps1=0,55; Eps2=21 dan MinPts=7 mendapatkan nilai *Silhouette Coefficient* tertinggi tetapi memiliki jumlah noise yang paling banyak jumlah *cluster* yang paling sedikit, maka untuk parameter terbaik pada penelitian ini dipilih parameter Eps1=0,45; Eps2=7; dan MinPts=4 karena pada parameter ini jumlah *cluster* dan noise lebih stabil dengan jumlah *cluster* 60 dan noise sebanyak 216 titik.

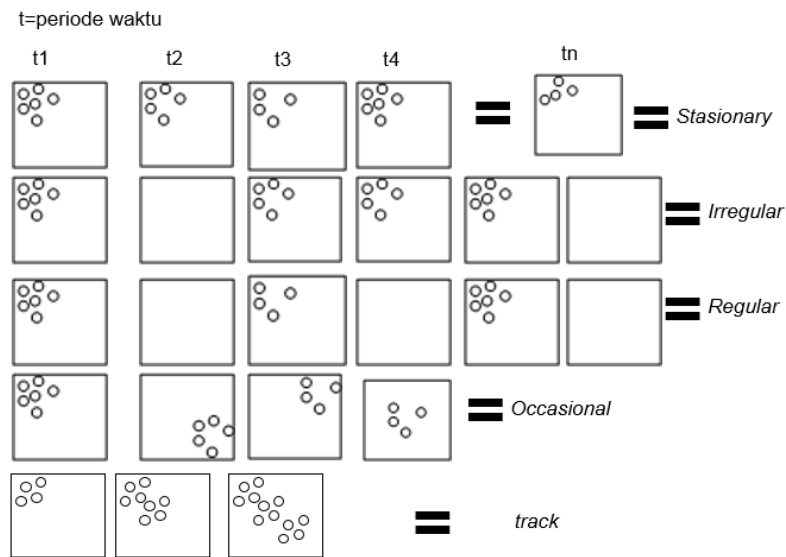
4.5. Analisa Pola Spatio-Temporal

Pola spatio-temporal terbagi menjadi 4 yaitu *stasioneri*, *reappearing*, *oscasional*, dan *track*. Menurut Pöelitz dan Andrienko (2010), *cluster* besar dan dapat dianalisis polanya adalah *cluster* yang memiliki jumlah titik minimal berjumlah 30, sedangkan *cluster* yang memiliki jumlah titik kurang dari 30 merupakan *cluster* kecil dan polanya tidak dapat terdeteksi. Berdasarkan parameter terbaik Eps1=0,45; Eps2=7; dan MinPts=4 terdapat 8 *cluster* besar yang memiliki jumlah titik lebih dari 30 (Tabel 6)

Tabel 5. Cluster Besar

Cluster ke-	Jumlah titik
1	67
2	125
4	68
11	50
14	174
19	90
25	40
48	52

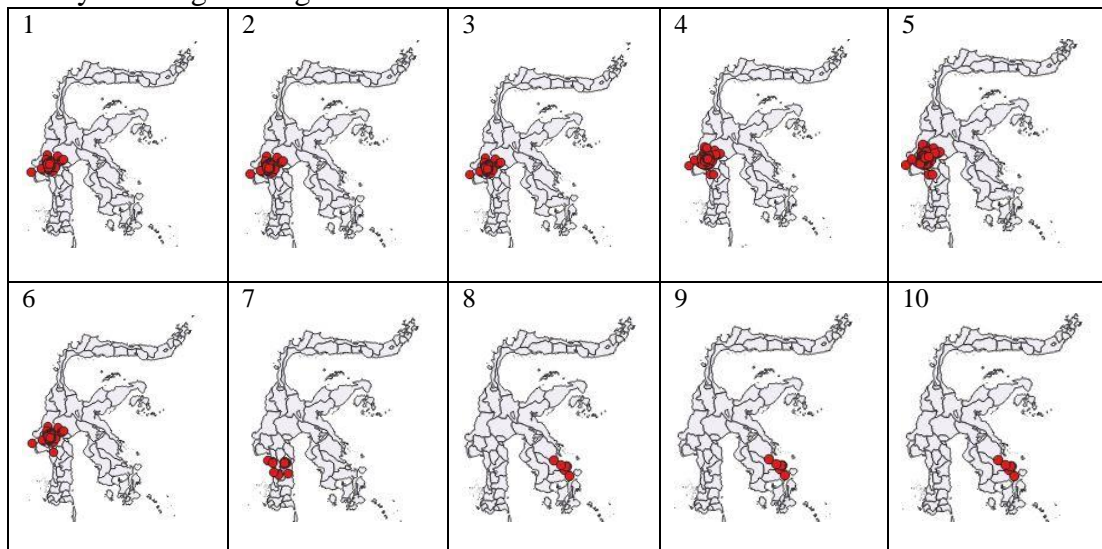
Tiap *cluster* akan dianalisis polanya, ilustrasi pola dapat dilihat pada Gambar 5.



Gambar 5. Ilustrasi Pola Spatio-Temporal (Purwanto, 2012)

Cluster dibagi menjadi per periode, karena parameter terbaik pada penelitian ini dengan nilai $Eps=7$, maka periodenya tujuh hari untuk perbandingan lokasi persebaran titik gempa suatu periode dengan periode lainnya. Sebagai contoh pada *cluster 2* didapatkan 10 periode (Gambar 6)

Pada periode 1 sampai 7 titik gempa berpusat pada kabupaten Mamasa, selanjutnya pada periode 8 sampai 10 titik gempa berada pada Kota Kendari. Dapat dilihat bahwa titik gempa hanya berpindah saat periode 7 ke periode 8, periode lain titik gempa tidak berpindah posisi maka dapat dikatakan *cluster 2* mempunyai pola cenderung *stasioner*. Analisis 8 *cluster* besar pada penelitian ini menunjukkan bahwa tipe pola *spatio-temporal* yang paling banyak muncul adalah tipe pola *oscasional*. Terdapat 5 pola *oscasional* dan memiliki pusat *clusternya* masing-masing.



Gambar 6. Pola Titik Gempa Cluster 2

Kabupaten Mamasa, Kabupaten Poso, dan Laut di antara Sulawesi Utara dan Sulawesi Selatan adalah daerah yang paling banyak menjadi pusat *cluster* titik gempa. Dengan adanya analisis ini diharapkan pemerintah daerah dapat memberikan tindakan yang tepat guna meminimalisir kerugian baik materil ataupun korban jiwa di daerah yang rawan gempa. Sebagai contoh, untuk daerah Kabupaten Mamasa dan Kabupaten Poso dapat memperkuat

kontruksi fasilitas umum dan bangunan-bangunan vital mengikuti standar kualitas bangunan tahan gempa.

5. KESIMPULAN

Berdasarkan penentuan nilai parameter menggunakan k-dist didapatkan nilai Eps1 optimal untuk MinPts=3 adalah 0,35 sampai dengan 0,45; untuk MinPts=4 adalah 0,35 sampai dengan 0,45; untuk MinPts=5 adalah 0,45 sampai dengan 0,55; untuk MinPts=7 adalah 0,55 sampai dengan 0,65. Setelah melakukan *clustering* 48 pasangan parameter Eps1, Eps2, dan MinPts didapatkan parameter terbaik yaitu Eps1=0,45; Eps2=7; dan MinPts=4 menghasilkan 60 *cluster* dan 216 noise dengan *Silhouette Coefficient* Spasial sebesar 0,0569 dan *Silhouette Coefficient* Temporal sebesar -0,5723. Pusat titik gempa paling banyak berada di Kabupaten Mamasa, Kabupaten Poso, dan Laut di antara Sulawesi Utara dan Sulawesi Selatan. Tipe pola *spatio-temporal* yang paling banyak ditemukan adalah tipe pola *oscasional*.

DAFTAR PUSTAKA

- Birant, D. & Kut, A., 2007. ST-DBSCAN: An Algorithm for Clustering Spatial-temporal data. *Data and Knowledge Engineering*, pp. 60:208-221.
- Gaonkar, M. N. & Sawant, K., 2013. AutoEPsDBSCAN. DBSCAN with Eps Automatic for Large, pp. 2:2319-2526.
- Han, J., Kamber, M. & Pei, J., 2012. *Data Mining Concept & Techniques*. Waltham: Elsevier.
- Poelitz, C. & Andrienko, N., 2010. Finding Arbitrary Shaped Cluster with Related Extents in Space and Time. IEEE-VGTC Symposium on Visualization.
- Pramadhani, A. E. & Setiadi, T., 2014. Data Mining untuk Klasifikasi Prediksi Penyakit ISPA dengan Algoritma Decision Tree. *Jurnal Sarjana Teknik Informatika*, pp. 831-839.
- Purwanto, B. U. B. Y., 2012. Spatial Hotspots Clustering of Forest and Land Fires using DBSCAN and ST-DBSCAN. Scientific Repository (IPB).
- Santosa, B., 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Shearer, M. P., 2009. *Introduction of Seismology*. In: Cambridge: Cambridge University Press.
- Siswowidjoyo, S., 1996. *Pengantar Seismologi Gunungapi dan Hubungannya dengan Kegiatan Gunungapi*, Bandung: Pusat Vulkanologi dan Mitigasi Bencana Geologi.
- Suyanto, D., 2019. *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Penerbit Informatika.
- Trisnaningtyas, A., 2014. *Pengelompokan Data Indeks Pembangunan Manusia di Pulau Jawa Dengan Algoritme ST-DBSCAN dan Bahasa Pemrograman Python*.