

Algoritme Genetika Untuk Optimasi *K-Means Clustering* Dalam Pengelompokan Data Tsunami

Dwi Anggraeni Kuntjoro¹, Budi Darma Setiawan², Rizal Setya Perdana³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹dwianggraenik59@gmail.com, ²s.budidarma@ub.ac.id, ³rizalespe@ub.ac.id

Abstrak

Tsunami adalah salah satu bencana alam yang paling mematikan yang menyebabkan kerusakan harta benda dan hilangnya nyawa. Tsunami sering terjadi secara tiba-tiba dan tidak dapat diprediksi. Kurangnya pengetahuan yang tepat mengakibatkan kerusakan yang besar dan memperburuk dampak dari tsunami. Pada penelitian ini mengimplementasikan optimasi Algoritme Genetika pada metode *K-Means* untuk pengelompokan data tsunami. Dengan mengoptimasi pusat *cluster* awal yang nantinya akan digunakan sebagai masukan untuk metode *K-Means*. Metode ini menghasilkan kinerja yang lebih optimal dibandingkan dengan metode *K-Means* konvensional karena titik pusat dioptimalkan dengan Algoritme Genetika. Hal ini dibuktikan dalam penelitian ini di mana nilai *fitness* didapatkan dari nilai Silhouette Coefficient untuk melihat seberapa cocok data dengan *cluster*. Representasi kromosom yang digunakan pada penelitian ini adalah *real code* untuk menginisialisasi nilai *centroid*. Metode *crossover* yang digunakan yaitu *extended intermediate crossover*. Metode mutasi yang digunakan yaitu *random mutation*. Sedangkan metode seleksi yang digunakan yaitu *elitism selection*. Berdasarkan hasil pengujian yang telah dilakukan, diperoleh parameter yang paling optimal yaitu pada populasi 50, generasi 70, serta kombinasi $Cr = 0.9$ dan $Mr = 0.1$ dengan *fitness* sebesar 0.995934.

Kata kunci: algoritme genetika (AG), AG-k-means.data tsunami, k-means.

Abstract

Tsunami is one of the most deadly disaster causing damage and loss of life and wealth. It happens in a sudden and unpredictable. Lack of awareness often leads to a great damage and worsening the impact of tsunami itself. This research implements genetic algorithm optimization into K-Means method for classify tsunami data. By optimizing the initial cluster center it will used as an input on K-Means method. The method result more optimal preference than the conventional K-Means method since the central point is optimized by genetic algorithm. It was proved on this research where fitness value resulted from Silhouette Coefficient to observe how suitable data with cluster. Chromosome representation used here is real code to initialize centroid value. Extended intermediate crossover applied for crossover method. For mutation method, random mutation is run here. Also for selection method it uses elitism selection. Based on testing result, the most optimum parameter accomplished are 50 population, 70 generation, and $Cr = 0.9$ and $Mr = 0.1$ combination with fitness value around 0.995934

Keywords: AG – k-means, genetic algorithm(AG), k-means, tsunami data.

1. PENDAHULUAN

Tsunami adalah gelombang raksasa yang disebabkan oleh gempa bumi atau letusan gunung berapi bawah laut. Terjadi di kedalaman laut, gelombang tsunami tidak secara dramatis meningkatkan ketinggian. Namun seperti perjalanan gelombang pedalaman, mereka membangun hingga ketinggian yang lebih tinggi dari ketinggian gelombang biasanya. Kecepatan

gelombang tsunami tergantung pada kedalaman laut daripada jarak dari sumber gelombang.

Gelombang tsunami dapat merambat ke segala arah. Di laut dalam, gelombang tsunami dapat merambat dengan kecepatan 500-1000 km per jam. Setara dengan kecepatan pesawat terbang. Ketinggian gelombang di laut dalam hanya sekitar 1 meter. Dengan demikian, laju gelombang tidak terasa oleh kapal yang sedang

berada di tengah laut. Ketika mendekati pantai, kecepatan gelombang tsunami menurun hingga sekitar 30 km per jam, namun ketinggiannya sudah meningkat hingga mencapai puluhan meter. Hantaman gelombang tsunami bisa masuk hingga puluhan kilometer dari bibir pantai. Kerusakan dan korban jiwa yang terjadi karena tsunami bisa diakibatkan karena hantaman air maupun material yang terbawa oleh aliran gelombang tsunami (Sugito, 2008).

Dengan berkembangnya teknologi saat ini mudah untuk dapat mengumpulkan data-data statistik dari peristiwa tsunami yang pernah terjadi sebelumnya, mempelajari karakteristik tsunami tersebut dan melakukan pengelompokan data. *Clustering* dapat dilakukan secara cepat dan efisien. Tujuan dari metode *clustering* adalah data yang memiliki kesamaan ciri dikelompokkan menjadi satu *cluster* dan untuk data yang memiliki ciri yang berbeda dipisahkan ke dalam *cluster* yang berbeda. Data yang digunakan memiliki jumlah yang besar sehingga data tersebut perlu dilakukan *clustering* untuk melihat karakteristik pada tsunami. Sejumlah penelitian yang dibuat mengenai pengelompokan data, penelitian yang dilakukan oleh Khotimah et al., (2016) algoritme AG - *Kmeans* mampu menghasilkan tingkat pengelompokan *cluster* lebih baik dari pada *K-Means* konvensional. Sedangkan penelitian yang dilakukan oleh Kim dan Ahn, (2008) AG-*Kmeans* dapat menghasilkan segmentasi yang lebih baik dari pada algoritme pengelompokan tradisional seperti *K-means* sederhana dan SOM.

Pada algoritme *K-Means*, nilai *k* harus ditentukan di awal dengan mempertimbangkan ketidak miripan dalam mengelompokkan obyek yang ada. Selain menentukan jumlah *cluster*, parameter yang harus ditentukan adalah pusat *cluster* yang dilakukan secara acak. Semakin baik penentuan pusat *cluster*, maka semakin tepat dan cepat proses pengelompokan menggunakan *K-Means Clustering*. Tetapi karena pusat *cluster* (*centroid*) ditentukan secara acak, maka terkadang tingkat keakuratan kurang baik dan sering terjadi lokal optimum (Niknam dan Amiri, 2010). Optimum local dapat diatasi dengan menggunakan sebuah algoritme optimasi guna memperbaiki pusat *cluster* (*centroid*) yang pertama, sehingga mampu didapatkan pusat *cluster* yang paling optimal untuk perhitungan *K-Means*.

Algoritme Genetika (AG) adalah algoritme pencarian dan pengoptimalan berdasarkan prinsip evolusi alami, yang pertama kali diperkenalkan oleh John Holland pada tahun 1970. Algoritme genetika juga menerapkan strategi pengoptimalan dengan mensimulasikan evolusi spesies melalui seleksi alam. Algoritme genetika umumnya terdiri dari dua proses. Proses pertama pemilihan individu untuk produksi generasi berikutnya dan proses kedua adalah manipulasi individu terpilih untuk membentuk generasi berikutnya dengan teknik crossover dan mutasi. Mekanisme seleksi menentukan individu mana yang dipilih untuk reproduksi dan berapa banyak keturunan yang dihasilkan masing-masing individu. Prinsip utama strategi seleksi adalah semakin baik individu (Garg dan Mittal, 2014).

2. TSUNAMI

Tsunami adalah salah satu bencana alam yang paling merusak yang dikenal manusia. Untuk sebagian besar orang yang tinggal dekat dengan laut, tsunami adalah ancaman terbesar dari hidup mereka. Tsunami menyebabkan sungai dan jalur air lainnya yang berada di sekitaran pantai meluap. Air berlebihan ini dapat menghasilkan arus yang berbahaya dan dapat menyeret dan menenggelamkan orang. Sebuah tsunami memiliki semua efek destruktif yang dapat menambahkan kekuatan ombak yang dapat merusak daratan. Sebagian besar tsunami (hingga 75% dari semua kasus sejarah) dipicu oleh gempa bumi, tanah longsor dan gunung meletus (Das dan Baskey, 2012).

3. ALGORITMA GENETIKA (AG)

Algoritme Genetika (AG) merupakan salah satu tipe dari Algoritme Evolusi (EA) yang paling populer. AG mampu menyelesaikan berbagai masalah yang kompleks diberbagai bidang. AG sendiri akan memanipulasi kode yang di dapatkan dari parameter yang ada atau biasa disebut dengan kromosom. Dari beberapa titik dalam populasi dilakukan pencarian dengan menggunakan syarat dari tujuan. Operator acak digunakan karna mempunyai sifat teori peluang (Mahmudy, 2013).

Dalam memecahkan masalah AG menggunakan rangkaian tahapan dengan cara memetakan masalah menjadi kromosom string. Kromosom itu sendiri terdiri dari gen, dimana gen itu sendiri merupakan variabel untuk

mendapatkan solusi. Kemudian terdapat fungsi *fitness* untuk menilai kualitas dari individu itu sendiri apakah solusi tersebut telah optimal atau belum. Setelah melalui beberapa tahapan yang terdapat pada algoritme genetika, kromosom terbaik akan terpilih dan akan menjadi sebuah solusi seperti yang terjadi pada seleksi alami dari generasi ke generasi selanjutnya. Dari kromosom yang terbaik ini kemudian diuraikan menjadi sebuah solusi yang diharapkan mendekati optimum (Mahmudy, 2013).

3.1. Representasi Kromosom

Untuk mendapatkan solusi yang optimal AG memiliki banyak cara untuk merepresentasikan kromosom agar mendapatkan solusi yang optimal. Representasi solusi bisa di dapatkan dengan representasi *real-coded*, representasi permutasi dan representasi biner. Dalam penelitian ini untuk representasi solusi dipilih *Real-coded Genetic Algorithm* untuk merepresentasikan kromosom karena data yang akan digunakan memiliki nilai *real*.

Representasi kromosom menggunakan bilangan decimal untuk merepresentasikan solusi pada pengkodean *real* yang akan digunakan pada penelitian ini. Inisialisasi populasi akan dibangkitkan sesuai dengan kebutuhan dalam pemecahan masalah. Perkembangan populasi diatur dalam dua cara yaitu *crossover* dan mutasi.

3.2. Reproduksi Crossover

Menurut Saputro et al., (2015) proses dari reproduksi *crossover* dilakukan dengan cara menyilang individu secara acak untuk menghasilkan individu baru (*offspring*) atau anak. Salah satu metode *crossover* yang digunakan adalah metode *extended intermediate crossover*. *Extended intermediate crossover* dipilih karena menghasilkan *offspring* dari kombinasi dua individu. Misalkan terdapat kromosom *Parent* P_1 dan P_2 , maka pembangkitan *offspring* C_1 dan C_2 dapat dihitung dengan Persamaan 1

$$\begin{aligned} C_1 &= P_1 + \alpha (P_2 - P_1) \\ C_2 &= P_2 + \alpha (P_1 - P_2) \end{aligned} \tag{1}$$

Nilai α dapat dibangkitkan dengan cara acak pada *range* -0,25 sampai 1,25. Jumlah *offspring* yang didapatkan sebanyak *crossover rate* x banyak individu.

3.3. Reproduksi Mutasi

Menurut Saputro et al., (2015) proses dari reproduksi mutasi adalah melakukan modifikasi satu atau lebih gen dalam individu yang sama untuk membangkitkan individu baru. Fungsi dari mutasi adalah untuk menyediakan gen yang tidak ada dalam populasi awal serta menggantikan gen yang hilang dari populasi selama proses seleksi. Dengan *mutation rate* (*mr*) yang telah ditentukan pada setiap gen, proses mutasi dilakukan. Nilai dari *mutation rate* inilah yang menentukan jumlah terpilihnya gen untuk nantinya dilakukan mutasi dan menghasilkan *offspring* baru. Metode yang digunakan pada penelitian ini adalah *random mutation*. Metode mutasi ini memilih secara acak satu individu dari populasi. Kemudian dari proses tersebut mengurangi atau menambah nilai gen yang terpilih dengan bilangan random yang kecil. Misalkan terdapat domain variabel x_j adalah $[\min_{ij}, \max_{ij}]$ dan *offspring* yang didapatkan adalah $C[x'_1 \dots x'_n]$ maka nilai gen *offspring* dapat dibangkitkan melalui Persamaan 2.

$$x'_i = x_i + r (\max_i - \min_i) \tag{2}$$

Secara acak nilai r dibangkitkan pada *range* -0,1 sampai dengan 0,1. Jumlah *offspring* yang didapatkan sebanyak *mutation rate* x banyak individu.

3.4. Evaluasi Hasil K-Means Tiap Individu

Evaluasi adalah tahapan dimana kromosom pada setiap generasi akan diukur nilai *fitness*nya untuk dapat menggambarkan kualitasnya. Dari nilai yang didapatkan kita dapat menentukan apakah kromosom tersebut sudah layak atau tidak menjadi calon solusi. Penentuan nilai *fitness* dapat dilakukan dengan menggunakan Persamaan 3.

$$\begin{aligned} a_{(i)} &= \frac{\sum D(i,j)}{|A|-1} \\ b_{(i)} &= \min(D(i, C)) \\ S_i &= \frac{(b_i - a_i)}{\max(a_i, b)} \end{aligned} \tag{3}$$

Hasil dikatakan bagus jika nilai mendekati 1 dan dikatakan jelek jika nilai mendekati -1.

3.5. Seleksi Elitism

Seleksi *Elitism* adalah seleksi yang digunakan untuk memilih individu baru sebanyak populasi dengan nilai *fitness* tertinggi

dari semua individu yang ada.

4. K-MEANS CLUSTERING

K-Means adalah salah satu metode pengelompokan data *non-hierarchical* yang berusaha untuk mempartisi data yang ada dalam bentuk satu atau lebih *cluster* (Hruschka dan Ebecken, 2003). Metode *K-Means* adalah proses pengelompokan yang telah dilemahkan karena penentuan pusat *cluster* awal. Algoritme *K-Means* mencari solusi lokal optimum untuk solusi awal guna memperbaiki hasil partisi (Wu dan Yang, 2002). Metode *non-heuristic* berbasis pendekatan lokal digunakan untuk pengelompokan *K-Means* dan membuktikannya melalui studi empiris. Jadi, *centroid clustering* awal yang baik dapat digunakan dengan teknik lainnya. *K-Means Clustering* meningkatkan *centroid clustering* yang optimal (Kumar et al., 2010).

Tahapan pada metode *K-Means* adalah:

1. Menentukan jumlah *cluster*
2. Menentukan nilai *centroid*

Untuk menentukan nilai *centroid* untuk awal iterasi, nilai *centroid* awal dilakukan secara acak. Berikut adalah rumus untuk menentukan nilai *centroid* awal.

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj} \tag{4}$$

Keterangan :

V_{ij} adalah *centroid*/ rata-rata *cluster* ke- i untuk variabel ke- j

N_i adalah jumlah data anggota *cluster* ke- i

i adalah indeks dari *cluster*

k adalah indeks dari data

j adalah indeks dari variabel

X_{kj} adalah nilai data ke- k yang terdapat di dalam *cluster* tersebut untuk variabel ke- j

3. Menghitung jarak antara titik pusat *cluster* (*centroid*) dengan titik tiap objek. Untuk menghitung jarak digunakan rumus *Euclidean Distance*. Berikut adalah rumus *Euclidean Distance*:

$$d_{ij} = \sqrt{\sum_{k=1}^p \{x_{ik} - x_{jk}\}^2} \tag{5}$$

Keterangan :

d_{ij} = jarak objek antara objek i dan j

P = dimensi data

x_{ik} = fitur ke- k data ke- i

x_{jk} = fitur ke- k data ke- j

4. Pengelompokan objek

Untuk menentukan anggota *cluster* digunakan perhitungan jarak minimum objek. Nilai yang di dapatkan dalam keanggotaan data pada *distance* matrik adalah 0 atau 1, dimana nilai 1 digunakan untuk data yang di alokasikan ke *cluster* dan nilai 0 untuk data yang di alokasikan ke *cluster* yang lain.

5. Kemudian kembali ke tahap 2, lakukan perulangan hingga di dapatkan *centroid* yang dihasilkan bernilai tetap dan anggota *cluster* tidak berubah dan berpindah ke *cluster* yang lain.

5. Deskripsi Data

Pada penelitian ini data yang digunakan adalah data sekunder. Data sekunder ini didapatkan dari situs resmi *database* historis yang ada dan informasi dari tsunami yang disediakan oleh organisasi pusat NOAA pacific peringatan tsunami. Data yang digunakan sebanyak 246 dengan 5 parameter. Parameter tersebut adalah *Focal depth*, *Primary magnitude*, *Latitude*, *Longitude* dan *Max water height*. Data yang digunakan adalah data tsunami yang terjadi pada tahun 1960-1990 dikarenakan pada tahun-tahun selanjutnya tidak memiliki data yang lengkap.

6. ALGORITME GENETIKA UNTUK OPTIMASI CENTROID AWAL PADA K-MEANS

Metode yang diimplementasikan adalah AG - *K-Means*. *Cluster* ditentukan kemudian inialisasi parameter. Inialisasi parameter adalah banyak individu (*popSize*), banyak generasi AG, *crossover rate*, dan *mutation rate*. Gambar 1 adalah diagram alir dari perancangan algoritme AG - *K-Means*.

Gambar 1 merupakan tahap pertama dari metode yang diimplementasikan adalah

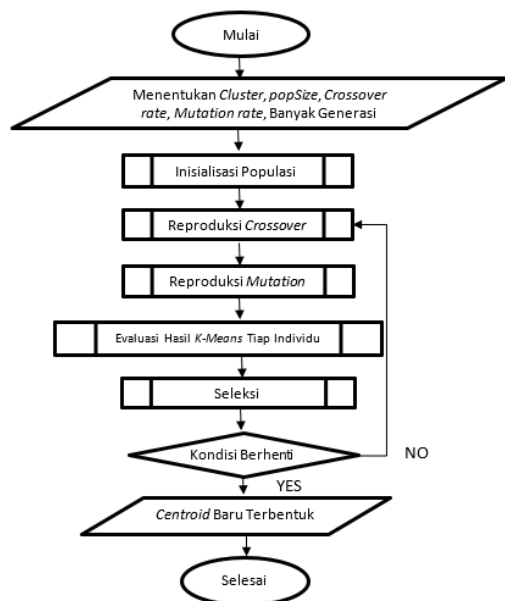
menentukan inialisasi populasi kemudian menentukan jumlah *cluster*. Inialisasi populasi yaitu menghasilkan beberapa individu awal. Pada penelitian ini, banyak *cluster* ditentukan setelah melakukan inialisasi populasi. Panjang kromosom sebanyak 5.

Reproduksi *crossover* dilakukan secara acak memilih 2 individu awal dan membangkitkan secara acak nilai *alpha*. Banyaknya *offspring* yang dihasilkan sebanyak *crossover rate* * ukuran populasi (*popSize*)

Reproduksi *mutation* dilakukan dengan memilih 1 individu secara acak untuk selanjutnya diproses dan membangkitkan secara acak nilai *r*. Banyaknya *offspring* yang dihasilkan sebanyak *mutation rate* * ukuran populasi (*popSize*).

Evaluasi hasil *k-means* tiap individu adalah proses perhitungan nilai *fitness* dari setiap individu. Pada tahapan ini nilai *fitness* digantikan dengan nilai Silhouette Coefficient sesuai dengan Persamaan 3.

Seleksi *Elitism* yaitu suatu proses untuk memilih individu terbaik dari sejumlah individu sebanyak ukuran populasi (*popSize*) yang didapatkan dari individu *parent* maupun *offspring*.



Gambar 1. Diagram Alir Perancangan Algoritme.

7. HASIL DAN PEMBAHASAN

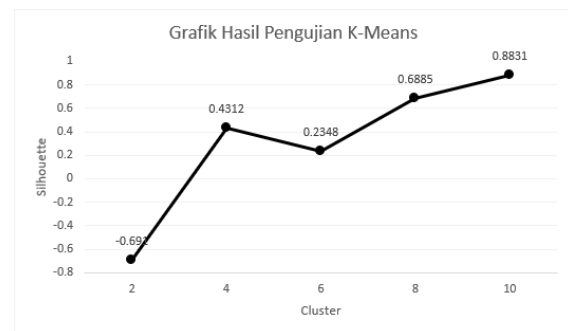
Berikut ini adalah hasil dari implementasi

AG - *K-Means* beserta pembahasan dari pengujian yang telah dilakukan.

7.1. Pengujian *K-Means*

Dilakukan pengujian ini bertujuan untuk mengetahui *cluster* dengan hasil paling optimal. Dari hasil pengujian tersebut kemudian akan dilakukan analisis yang digunakan untuk mendapatkan kesimpulan dari penelitian ini. Hasil dan analisis pengujian *K-Means* dapat dilihat pada Gambar 2.

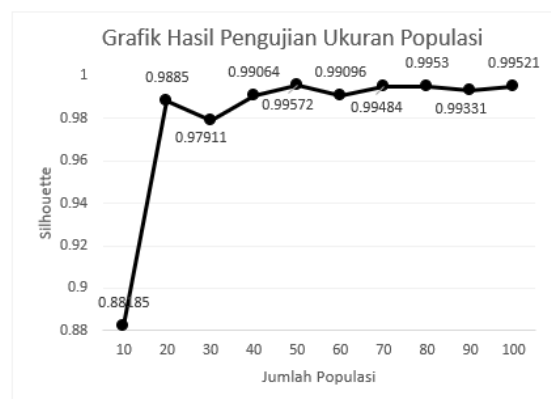
Dari grafik dapat dilihat bahwa jumlah *cluster* dapat mempengaruhi nilai Silhouette Coefficient. Untuk *cluster* 10 memiliki nilai Silhouette tertinggi sebesar 0,8831 dibandingkan dengan nilai Silhouette lainnya. Sedangkan nilai Silhouette terendah terdapat pada *cluster* 2 dengan nilai Silhouette -0,691.



Gambar 2. Grafik Pengujian *K-Means*

7.2. Pengujian Ukuran Populasi

Dilakukannya uji coba adalah untuk mengetahui ukuran populasi yang dapat menghasilkan *centroid* yang paling optimal pada *K-Means* untuk menentukan potensi tsunami. Uji coba ini dilakukan sebanyak 10 kali dengan menggunakan *Cr* (*crossover rate*) = 0.3 dan *Mr* (*mutation rate*) = 0.2 dengan jumlah generasi sebanyak 100. Jumlah populasi yang akan di uji merupakan nilai kelipatan 10, mulai dari 10



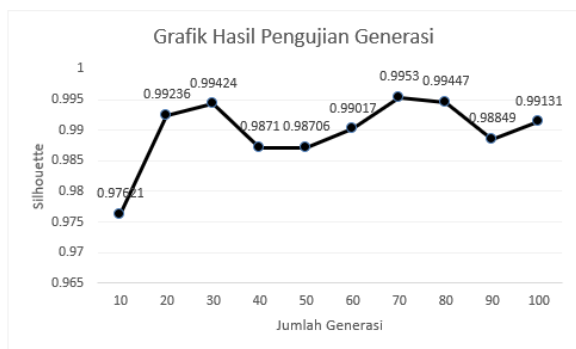
sampai dengan 100. Hasil pengujian ukuran populasi dijabarkan pada Gambar 3.

Dari grafik dapat dilihat bahwa jumlah *popSize* dapat mempengaruhi nilai *fitness*. Untuk *popSize* 50 memiliki nilai *fitness* tertinggi sebesar 0.99572 dibandingkan dengan nilai *fitness* lainnya. Sedangkan nilai *fitness* terendah terdapat pada *popsize* 10 dengan nilai *fitness* sebesar 0.88185.

Gambar 3. Grafik Pengujian Ukuran Populasi

7.3. Pengujian Banyak Generasi

Pengujian banyak generasi dilakukan untuk mengetahui banyak generasi yang dapat menghasilkan solusi optimal pada penelitian ini.



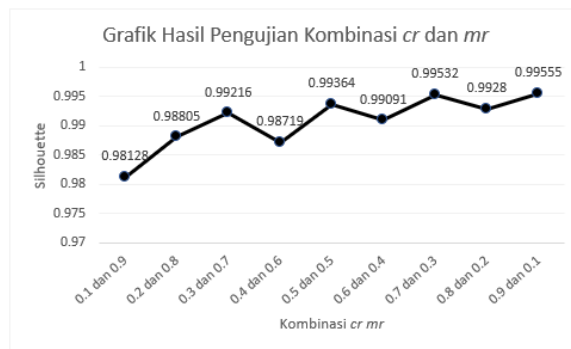
Gambar 4. Grafik Pengujian Banyak Generasi

Pengujian dilakukan 10 kali pada setiap banyak generasi dengan banyak generasi 10 sampai dengan 100. Pengujian ini dilakukan dengan ukuran populasi 50, nilai *cr* 0,5 , nilai *mr* 0,5, jumlah *cluster* 4. Hasil pengujian banyak generasi akan dijabarkan pada Gambar 4. Dari grafik tersebut dapat dilihat bahwa jumlah generasi dapat mempengaruhi nilai *fitness*. Untuk generasi 70 memiliki nilai *fitness* tertinggi sebesar 0.9953 dibandingkan dengan nilai *fitness* lain. Sedangkan nilai *fitness* terendah terdapat pada generasi 10 dengan nilai *fitness* 0,97621.

Dari grafik dapat dilihat bahwa jumlah generasi dapat mempengaruhi nilai *fitness*. Untuk generasi 70 memiliki nilai *fitness* tertinggi sebesar 0,9953 dibandingkan dengan nilai *fitness* lain. Sedangkan nilai *fitness* terendah terdapat pada generasi 10 dengan nilai *fitness* 0,97621.

7.4. Pengujian Kombinasi *cr* dan *mr*

Pengujian kombinasi *cr* dan *mr* dilakukan untuk mengetahui kombinasi *cr* dan *mr* agar dapat menghasilkan solusi optimal pada penelitian ini.



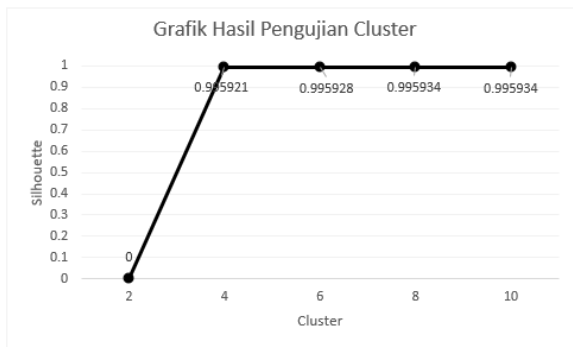
Gambar 5. Grafik Pengujian Kombinasi *cr* dan *mr*

Pengujian dilakukan 10 kali dengan nilai *cr* dan *mr* 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, dan 0.9. Pengujian dilakukan untuk 70 generasi dengan ukuran populasi 50. Hasil pengujian kombinasi crossover rate dan mutation rate dijabarkan pada Gambar 5.

Dari grafik dapat dilihat bahwa, grafik cenderung naik dari kombinasi pertama sampai kombinasi 3 kemudian turun pada kombinasi 4 dan naik lagi pada kombinasi 9. Kombinasi *cr* dan *mr* dengan 0,9 dan 0,1 memiliki nilai rata-rata *fitness* yang tertinggi dibandingkan dengan kombinasi *cr* dan *mr* lainnya dengan pengujian yang sama sebesar 0,9955. Maka untuk pengujian selanjutnya digunakan kombinasi *cr* dan *mr* dengan nilai 0,9 dan 0,1.

7.5. Pengujian Banyak Cluster

Pengujian banyak *cluster* dilakukan untuk mengetahui *cluster* mana yang dapat menghasilkan solusi optimal pada penelitian ini. Pengujian ini dilakukan dengan ukuran populasi 50, generasi 70 , nilai *cr* dan *mr* 0,9 dan 0,1. Hasil pengujian banyak *cluster* dijabarkan pada Gambar 6.

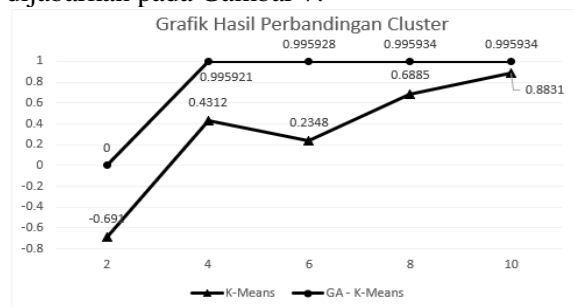


Gambar 6. Grafik Pengujian Cluster

Dari grafik dapat dilihat bahwa jumlah cluster dapat mempengaruhi nilai Silhouette Coefficient. Untuk cluster 8 dan 10 memiliki nilai Fitness tertinggi sebesar 0,995934. Sedangkan nilai Fitness terendah terdapat pada cluster 2 dengan nilai fitness 0.

7.6. Perbandingan AG - K-Means

Perbandingan nilai antara Silhouette Coefficient K-Means dengan Silhouette Coefficient AG - K-Means. Hasil perbandingan dijabarkan pada Gambar 7.



Gambar 7. Grafik Perbandingan AG-K-Means

Dari grafik dapat menunjukkan bahwa Silhouette Coefficient yang dihasilkan oleh AG - K-Means memiliki nilai yang lebih tinggi dibandingkan dengan Silhouette Coefficient K-Means. Untuk memudahkan membaca hasil pengujian, terdapat grafik perbandingan nilai Silhouette Coefficient. Dapat dilihat bahwa metode AG - K-Means dapat mempengaruhi nilai fitness.

7.7. Analisis Hasil Clustering

Analisis rata-rata hasil dari clustering pada tiap parameter. Hasil dari clustering dapat dilihat pada tabel 1.

Tabel 1. Rata-rata Analisis Hasil Clustering

Cluster	Focal Depth	Primary Magnitude	Latitude	Longitude	Maximum Height Water
Cluster 1	17.28	5.67	48.20	148.37	66.6
Cluster 2	14.56	7.31	38.36	-64.19	95.40
Cluster 3	41.24	6.78	33.60	-193.80	67.16
Cluster 4	58.23	6.65	49.19	174.72	66.14
Cluster 5	30.77	6.08	-12.29	150.82	70.37
Cluster 6	53.25	6.98	30.07	146.88	67.81
Cluster 7	26.31	6.64	-19.44	116.68	72.69
Cluster 8	30.30	6.61	31.95	170.82	67.08

Dari analisis dapat disimpulkan bahwa pada cluster 1 tsunami biasanya terjadi pada kedalaman 17,28 kilometer dengan ketinggian air 66,6 meter di atas permukaan laut dan terjadi pada kisaran wilayah 48,20 lintang utara 148,37 bujur timur. Begitu juga pada cluster-cluster selanjutnya.

8. KESIMPULAN DAN SARAN

Kesimpulan yang didapatkan melalui hasil uji coba yang telah dilakukan mengenai optimasi centroid K-Means menggunakan Algoritme Genetika untuk pengelompokan tsunami adalah sebagai berikut:

1. Penggunaan algoritme genetika untuk optimasi centroid mampu mengatasi permasalahan yang ada pada algoritme K-Means biasa dengan memberikan hasil centroid yang lebih optimal. Dengan pusat cluster yang optimal memberikan hasil yang lebih baik pada kualitas cluster menggunakan silhouette coefficient .
2. Pengelompokan data tsunami menggunakan metode AG-K-Means menggunakan metode Silhouette Coefficient menghasilkan nilai validasi yang lebih baik dari metode K-Means biasa. Untuk 10 kali percobaan dengan menggunakan algoritma genetika dan K-Means menghasilkan nilai Silhouette Coefficient sebesar 0,9959. Sedangkan untuk K-Means biasa menghasilkan Silhouette Coefficient sebesar 0,8831.
3. Parameter terbaik yang memberikan hasil optimal pada metode AG- K-Means adalah banyak popSize sejumlah 50, banyak generasi 70 dan kombinasi Cr= 0,9 Mr = 0,1.

9. DAFTAR PUSTAKA*Recognition*, 35(10): 2267–2278.

Das, S. & Baskey, B.B. 2012. Design Of An Embedded System For The Detection Of Tsunami. *Rourkela*.

Garg, R. & Mittal, S. 2014. Optimization by Genetic Algorithm in Data Mining. 4: (4): 534-8.

Hruschka, E.R. & Ebecken, N.F.. 2003. A Genetic Algorithm For Cluster Analysis.

Khotimah, B.K., Irhamni, F. & Sundarwati, T. 2016. A Genetic Algorithm For Optimized Initial Centers K-Means Clustering In SMEs. 90.

Kim, K. & Ahn, H. 2008. A recommender system using GA K-means clustering in an online shopping market. *Expert Systems with Applications*, 34: 1200–1209.

Kumar, A., Sabharwal, Y. & Sen, S. 2010. Linear Time Approximation Scheme For Clustering Problem In Any Dimensions.

Mahmudy, W.. 2013. Algoritma Evolusi. *Malang: Program Teknologi Informasi Dan Ilmu Komputer. Universitas Brawijaya*.

Niknam, T. & Amiri, B. 2010. An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis. *Appl. Soft Comput.*, 10: 183–197.

Saputro, H.A., Mahmudy, W.F. & Dewi, C. 2015. Implementasi Algoritma Genetika Untuk Optimasi Penggunaan Lahan Pertanian.

Sugito, N.T. 2008. *Tsunami*. Bandung: Jurusan Pendidikan Geografi, Universitas Pendidikan Indonesia. Tersedia di: <http://file.upi.edu/Direktori/FPIPS/JUR._PEND._GEOGRAFI/198304032008012-NANIN_TRIANA_SUGITO/TSUNAMI.pdf>

Wu, K.-L. & Yang, M.-S. 2002. Alternative c-means clustering algorithms. *Pattern*