

Optimasi K-Means dengan Algoritma Genetika untuk Target Pemanfaat Air Bersih Provinsi Riau

Taslim¹, Dafwen Toresa², Deny Jollyta³, Des Suryani⁴, Eka Sabna⁵

taslim_malano@gmail.com, dafwentoresa@gmail.com, deny.jollyta@lecturer.pelitaindonesia.ac.id,
des.suryani@eng.uir.ac.id, es3jelita@yahoo.com

^{1,2} Universitas Lancang Kuning, Pekanbaru

³ Institut Bisnis dan Teknologi Pelita Indonesia, Pekanbaru

⁴ Universitas Islam Riau, Pekanbaru

⁵ STMIK Hang Tuah, Pekanbaru

Informasi Artikel

Diterima : 30-01-2021

Direview : 09-02-2021

Disetujui : 13-03-2021

Kata Kunci

klasterisasi;k-means;
optimasi;genetika;
validitas

Abstrak

Air bersih merupakan hal yang penting dalam kehidupan manusia. Beberapa tindakan sudah dilakukan oleh pemerintah untuk memenuhi kebutuhan air bersih masyarakat Provinsi Riau. Salah satunya adalah dengan program penyediaan air minum dan sanitasi berbasis masyarakat. Sebelum melakukan kegiatan terkait pengadaan air bersih kepada masyarakat maka terlebih dahulu akan ditentukan target yang ingin dicapai untuk penyediaan air bersih ke depannya. Penelitian ini bertujuan untuk melakukan klusterisasi target pemanfaat air bersih menggunakan algoritma *k-means* dengan optimasi pada nilai *centroid* menggunakan algoritma genetika. *Average silhouette* number digunakan untuk mendapatkan jumlah kluster yang optimal yaitu sebanyak dua buah kluster. Hasil validitas kluster diukur menggunakan metode Davies Bouldin Index (DBI). Klusterisasi tanpa optimasi menghasilkan DBI sebesar 2.164763 dan hasil klusterisasi dengan optimasi genetika pada nilai *centroid* menghasilkan nilai DBI sebesar 2.06894.

Keywords

klasterization;k-means;
optimization;genetics;
validity

Abstract

Clean water is an important thing in human life. Several actions have been taken by the government to meet the clean water needs of the Riau province. One of them is the Community Based Drinking Water and Sanitation Provision program. Before carrying out activities related to the provision of clean water to the community, the targets to be achieved for the provision of clean water in the future will be determined. This study aims to klaster clean water beneficiary targets using the k-means algorithm with an optimization of the centroid value using a genetic algorithm. Average silhouette number is used to get the optimal number of klusters, which is two klusters. The results of klaster validity were measured using the Davies Bouldin Index (DBI) method where klasterization without optimization resulted in a DBI of 2.164763 and the results of klasterization by carrying out genetic optimization on the centroid value resulted in a DBI value of 2.06894.

A. Pendahuluan

Selaras dengan tujuan pemerintah dalam upaya terciptanya masyarakat yang berkeadilan dan makmur berdasarkan Pancasila dan UUD 1945 antara lain diwujudkan dengan pembangunan sarana air bersih. Akses berkelanjutan ke air bersih dalam jumlah yang cukup sangat penting untuk kesehatan manusia dan perkembangan sosial ekonomi [1]. Penebangan hutan dan pengalihan fungsinya menjadi areal pertanian monokultur mengakibatkan sumber air bersih di Provinsi Riau mulai berkurang. Berbagai hal sudah dilakukan oleh pemerintah provinsi untuk mengatasi kekurangan air bersih ini di Provinsi Riau. Salah satunya adalah dengan Penyediaan Air Minum dan Sanitasi Berbasis Masyarakat (PAMSIMAS), PAMSIMAS merupakan salah satu program pemerintah dengan dukungan Bank Dunia. Program ini dilaksanakan pada wilayah desa dan pinggiran kota yang belum memiliki sumber air bersih yang layak. Pada penelitian ini dilakukan klusterisasi untuk melihat target pemanfaat air bersih di Provinsi Riau tahun 2019.

Pada bidang data mining, klusterisasi adalah topik penelitian yang penting dan berbagai macam aplikasi *unsupervised learning*. Klusterisasi adalah teknik yang membagi data menjadi ke dalam beberapa kluster yang bermakna. Salah satu algoritma pengelompokan yang paling populer adalah *k-means* [2]. Metode *k-means* bekerja dengan melakukan klusterisasi sejumlah data ke dalam sejumlah kluster dengan menghitung jarak terdekat setiap data ke *centroid* yang paling dekat [3], hasil dari kinerja metoda *k-means* sangat tergantung pada nilai *centroid* awal yang diberikan [3][4]. Agar *k-means* dapat bekerja lebih baik maka beberapa algoritma kecerdasan buatan telah diterapkan untuk mengoptimalkan hasil klusterisasi dari *k-means*. Kapil memperlihatkan bahwa perbandingan tingkat akurasi *k-means* dengan optimasi algoritma genetika lebih baik dibandingkan dengan *k-means* tanpa optimasi [5]. Asraa dalam penelitiannya terkait serangan jantung menggunakan algoritma genetika untuk meningkatkan hasil klasifikasi *k-means*, hasilnya memperlihatkan adanya peningkatan akurasi dari algoritma *k-means* dengan tambahan algoritma genetika [6]. Beberapa penelitian terkait klusterisasi pemanfaat air bersih juga sudah dilakukan oleh beberapa peneliti sebelumnya, dengan analisis *hierarchical kluster* Mayasari menemukan adanya *gap* atau jarak akses air bersih dan sanitasi antara daerah perkotaan dan pedesaan di Provinsi Lampung [7]. Savitri menggunakan algoritma *k-means klustering* untuk pengelompokan penyediaan akses air minum bersih dan air minum layak di Indonesia dan mendapatkan bahwa terdapat beberapa provinsi dengan penggunaan air minum bersih dan layak yang masih rendah [8].

Pada penelitian ini, akan dilakukan klusterisasi target pemanfaat air bersih di Provinsi Riau menggunakan algoritma *k-means* dengan terlebih dahulu melakukan optimasi pada nilai *centroid k-means* menggunakan algoritma genetika untuk peningkatan akurasi kluster. Data melibatkan sembilan kabupaten dan kota serta 282 dusun dan desa di Provinsi Riau baik itu desa baru maupun desa lama. Adapun data yang digunakan adalah data target pemanfaat air bersih PAMSIMAS tahun 2019. Penelitian ini bertujuan untuk melihat sejauh mana tingkat optimasi dari klusterisasi data pemanfaat air bersih di Provinsi Riau menggunakan algoritma optimasi genetika pada penentuan nilai *centroid k-means*.

B. Metode Penelitian

Pada penelitian ini, dilakukan optimasi pada nilai *centroid* awal *k-means* menggunakan algoritma genetika. Penelitian dimulai dari penyiapan data yang akan digunakan yaitu data pemanfaat air bersih provinsi Riau pada tahun 2019. Selanjutnya dilakukan normalisasi terhadap data yang ada menggunakan normalisasi *min-max*. Langkah selanjutnya adalah menentukan jumlah kluster yang ideal menggunakan metode *average silhouette number*. Setelah didapat jumlah kluster yang ideal, langkah berikutnya menentukan nilai *centroid* dengan terlebih dahulu melakukan optimasi terhadap nilai tersebut menggunakan algoritma genetika. Tahap terakhir adalah melakukan pengujian hasil menggunakan metode *davies bouldin indeks*.

1. *K-means*

K-means clustering merupakan sebuah metode *unsupervised* yang sudah umum digunakan dalam bidang data mining, metode ini merupakan sebuah cara untuk mengelompokkan objek yang sama ke dalam sebuah kluster yang berbeda [9][10][11]. *K-means* mempunyai keunggulan antara lain mudah untuk diimplementasikan, bisa melakukan klusterisasi untuk data yang besar dan stabil pada berbagai macam kasus [12]. Algoritma *k-means* membagi data kedalam kelompok-kelompok yang berbeda berdasarkan pada ukuran jarak antar objek misalnya dengan menggunakan *euclidean distance* [10]. Tujuan algoritma ini adalah untuk mengeksplorasi sebuah data dan menemukan pola dari data tersebut [13]. Misalkan $T = \{T_1, T_2, \dots, T_n\}$ akan dijadikan sebagai kumpulan dari data sampel yang akan dikelompokkan menjadi K kelompok, $C = \{C_1, i=1, \dots, K\}$. Algoritma *k-means* akan mengelompokkan objek ke dalam sejumlah K kelompok dengan cara mencari jarak terdekat antara objek dengan nilai *centroid* atau disebut juga dengan *sum-of-squared errors* (SSE) [3] [14], yang didefinisikan sebagai persamaan (1) berikut :

$$j(C) = \sum_{i=1}^k \sum_{T_i \in C_i} (T_i - Z_i)^2 \quad (1)$$

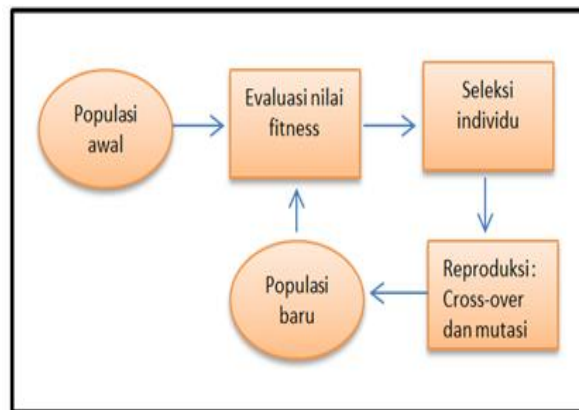
Tahap awal pada algoritma ini adalah memilih nilai *centroid* secara acak. Selanjutnya data data yang ada akan dimasukkan ke salah satu pusat kelompok yang telah ditentukan berdasarkan jarak terdekat antara data dengan nilai pusat kelompok. Posisi pusat kelompok akan dihitung kembali sampai semua data dikelompokkan kesetiap pusat kluster sampai terbentuk posisi pusat kelompok yang baru. Langkah ini akan terus berulang sampai tercipta kondisi konvergen. Salah satu kelemahan algoritma *k-means* ini sangat sensitif terhadap penentuan nilai *centroid* sehingga bisa menyebabkan nilainya terperangkap pada optimum lokal [14] [15][16]. Oleh sebab itu beberapa model klusterisasi untuk meningkatkan optimasi pada *k-means* telah dilakukan seperti algoritma *Tabu Search* [15], algoritma *Firefly* [14], algoritma optimasi *whale* [17], algoritma *bee colony* [18].

Pada penelitian ini jumlah dari kluster ditentukan terlebih dahulu menggunakan metoda *silhouette number*, sedangkan untuk menguji kualitas dan kluster data yang terbentuk dilakukan dengan uji validitas *Davies Bouldin Index*

(DBI). Pada metoda DBI ini hasil yang lebih minimum akan dianggap sebagai hasil yang lebih baik [19].

2. Algoritma Genetika

Algoritma genetika merupakan metode optimasi berdasarkan seleksi alam dan mekanisme genetika alami. Algoritma genetika bekerja dengan cara meniru prinsip genetika alami dan seleksi alam untuk membentuk prosedur pencarian dan optimasi [20]. Algoritma genetika menginisialisasi proses pencarian melalui menghasilkan sekelompok solusi acak yang mewakili populasi kromosom pada proses pencarian melalui evolusi, seleksi, *crossover*, dan mutasi untuk mendapatkan global optimal. Proses seleksi dilakukan dengan menyaring kromosom dengan nilai *fitness* yang lebih tinggi untuk menghasilkan keturunan dalam kumpulan gen generasi berikutnya. Selanjutnya, operator *crossover* melakukan pertukaran kandidat genetik, agar di dapat keturunan yang lebih baik [21]. Selama proses mutasi. Adapun siklus dari algoritma genetika dapat dilihat pada Gambar 1.



Gambar 1. Siklus Algoritma Genetika

3. Metode Davies Bouldin Index (DBI)

Validitas sebuah kluster perlu diuji untuk mendapatkan nilai kluster yang baik. Nilai validitas dilakukan dengan melakukan perbandingan antara dua kluster atau lebih sehingga bisa didapat kluster mana yang lebih baik. Pada penelitian ini untuk menguji nilai validitas digunakan metode DBI. Semakin kecil nilai DBI maka semakin bagus kluster yang dihasilkan [22]. Pada pengujian nilai DBI terdapat dua hal yaitu nilai kohesi yang digunakan untuk menghitung jumlah kemiripan data terhadap pusat kluster dan nilai separasi yaitu nilai jarak antara pusat kluster dari kluster. Kluster yang optimal adalah kluster yang memiliki nilai separasi yang tinggi dan nilai kohesi yang rendah. Adapun persamaan DBI dapat dilihat pada persamaan (2) berikut [23].

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{(i \neq j)} (R_{i,j}) \quad (2)$$

K merupakan jumlah *kluster* yang digunakan, untuk pengukuran rasio $(R_{i,j})(R_{i,j})$ diperoleh dari persamaan (3) yaitu.

$$(R_{i,j}) = \frac{SSW_i + SSW_j}{d_{C_i,C_j}} \quad (3)$$

d_{C_i,C_j} merupakan nilai separasi antar kluster atau *Sum of square between cluster* (SSB), *Sum of square within cluster* (SSW) adalah persamaan untuk menghitung matrik kohesi dalam sebuah *kluster* ke- i yang dihitung dengan menggunakan persamaan (4) berikut.

$$SSW_i = \frac{1}{m_i} \sum_{i=1}^{m_i} d(x_i, C_i) \quad (4)$$

m_i merupakan jumlah data dalam kluster ke i , C_i *centroid* kluster ke i , $d(x_j, c_i)$ merupakan jarak euclidean setiap data ke *centroid*.

C. Hasil dan Pembahasan

Data yang digunakan pada penelitian ini adalah data target pemanfaat air bersih yang dilakukan oleh Pamsimas Provinsi Riau pada Tahun 2019. Sebagian dari data yang digunakan dapat dilihat pada Tabel 1.

Tabel 1. Data Target Pemanfaat Air Bersih tahun 2019

No	Provinsi Kabupaten/Kota Kelurahan/Desa	Target Pemanfaat		Target Pemanfaatan SAM		Target Pemanfaat SAN			
		Dusun	Jiwa	Pipa		Akses Jamban KK	Akses Jamban Jiwa	CTPS Dusun	CTPS KK
				KK	Mis				
1	Banuaran, Kuantan Hilir	3	687	172	36	35	140	1	16
2	Ibul, Pucuk Rantau	2	600	100	53	36	144	1	24
3	Kampung Baru Ibul, Pucuk Rantau	2	600	112	27	40	160	1	53
4	Kepala Pulau, Kuantan Hilir	3	156	40	12	25	79	1	0
5	Kompe Berangin, Cerenti	3	796	199	11	44	175	1	8
----	-----	-----	-----	---	---	-----	-----	----	-----
----	-----	-----	-----	---	---	-----	-----	----	-----
28	Gurun Panjang, 0 Bukit Kapur	1	1353	193	105	0	0	0	0
28	Mekar Sari, 1 Dumai Selatan	1	771	31	72	0	0	0	0
28	Pangkalan Sesai, 2 Dumai Barat	2	2726	244	43	87	420	1	12

a. Normalisasi

Populasi terlebih dahulu dibangkitkan dengan melakukan normalisasi data. Dalam kegiatan data mining normalisasi merupakan proses penskalaan nilai atribut dari data sehingga bisa diubah pada *range* tersebut. Normalisasi dilakukan dengan metode *min-max* terhadap data-data yang akan diproses. Metode *min-max* merupakan metode normalisasi dengan melakukan transformasi linier terhadap data asli. Adapun persamaan normalisasi dapat dilihat pada persamaan (5).

$$\text{newData} = \frac{(\text{data} - \text{min}) \times (\text{newmax} - \text{newmin})}{(\text{max} - \text{min}) + \text{newmin}} \quad (5)$$

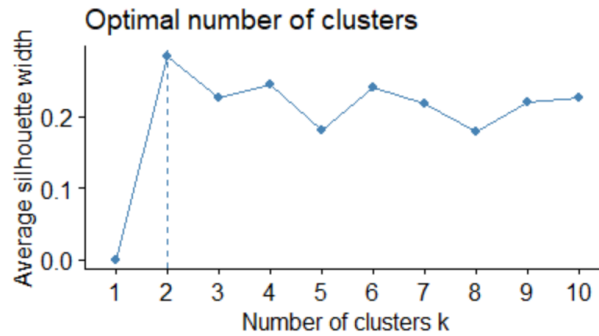
newData merupakan variabel yang digunakan untuk menampung hasil proses normalisasi, min adalah nilai minimum dari data per kolom, max merupakan nilai maksimum data per kolom, newMin adalah batas minimum yang diberikan sedangkan newMax merupakan batas maksimum yang diberikan. Adapun sebagian hasil dari normalisasi populasi data dapat dilihat pada Tabel 2.

Tabel 2. Normalisasi Populasi Data Pemanfaat Air Bersih tahun 2019

No	Provinsi Kabupaten/Kota Kelurahan/Desa	Target Pemanfaat		Target Pemanfaat SAM		Target Pemanfaat SAN			
		Dusun	Jiwa	Pipa		Akses Jamban KK	Akses Jamban Jiwa	CTPS Dusun	CTPS KK
				KK	Mis				
1	Banuaran, Kuantan Hilir	1,000	0,527	0,496	0,375	0,207	0,292	0,333	0,032
2	Ibul, Pucuk Rantau	0,500	0,444	0,288	0,552	0,214	0,301	0,333	0,048
3	Kampung Baru Ibul, Pucuk Rantau	0,500	0,444	0,323	0,281	0,241	0,338	0,333	0,105
4	Kepala Pulau, Kuantan Hilir	1,000	0,017	0,115	0,125	0,138	0,150	0,333	0,000
5	Kompe Berangin, Cerenti	1,000	0,632	0,573	0,115	0,269	0,373	0,333	0,016

b. Jumlah klaster

Untuk menentukan jumlah klaster yang terbaik maka digunakan metode *average silhouette number*. Hasil uji menggunakan R didapatkan hasil klaster yang terbaik adalah 2 klaster seperti terlihat pada Gambar 2.



Gambar 2. Hasil uji jumlah kluster dengan R

c. Optimasi dengan Algoritma Genetika

Untuk mendapatkan *centroid* yang optimal maka dilakukan optimasi dengan algoritma genetika terhadap dua kabupaten yaitu Kabupaten Bengkalis dan Kabupaten Kuantan Singingi, dimana hasil dari optimasi ini selanjutnya akan dijadikan sebagai nilai *centroid* kelompok pada algoritma *k-means*. Berikut tahapan optimasi pada Kabupaten Bengkalis.

1. Membangkitkan populasi awal

Populasi awal akan dibangkitkan secara acak sesuai dengan parameter sebagai berikut, ukuran populasi 18, probabilitas crossover 0,5 dan probabilitas mutasi 0,2. Proses pembangkitan kromosom dengan nilai acak dapat dilihat pada Tabel 3.

Tabel 3. Proses Pembangkitan Kromosom Dengan Nilai Acak

Proses Pembangkitan kromosom dengan nilai acak								
Populasi dengan 18 individu								
k1	0,5	0,225	0,105	0,694	0,000	0,000	0,000	0,051
k2	0	0,212	0,275	0,746	0,000	0,000	1,000	0,000
k3	0	0,072	0,537	0,373	0,000	0,000	0,000	0,364
k4	0	0,033	0,052	0,000	0,000	0,000	0,400	0,344
k5	0	0,013	1,000	0,373	0,000	0,000	0,200	1,000
---	-----	-----	-----	-----	-----	-----	-----	-----
k18	0,5	0,067	0,463	0,336	0,000	0,000	0,600	0,454

2. Evaluasi Fitness

Selanjutnya dilakukan perhitungan evaluasi *fitness*

$$fitness[i] += \sum nilaiGen[i] \quad (6)$$

Fitness[i] adalah variabel untuk menampung nilai *fitness* yang ke *i* yang didapat dari sigma *nilaiGen* yang ke *i*.

3. Seleksi Kromosom

Untuk seleksi kromosom dilakukan dengan metode *roulette wheel*, dimana nilai yang dipilih adalah nilai yang terbesar.

4. Crossover

Peluang *crossover* ditetapkan 0,5, dengan proses *crossover* dilakukan dengan pindah silang satu titik. Hasil dari *crossover* dapat dilihat pada Tabel 4.

Tabel 4. Tabel hasil crossover

R1	0,500	0,225	0,105	0,694	0,000	0,000	0,000	0,051
R2	0,500	0,225	0,105	0,694	0,000	0,000	0,000	0,051
R3	0,500	0,225	0,105	0,694	0,000	0,000	0,000	0,051
R4	0,500	0,225	0,105	0,694	0,000	0,000	1,000	0,000
R5	0,000	0,212	0,275	0,746	0,000	0,000	1,000	0,000
---	----	-----	-----	-----	-----	-----	-----	-----
R18	0,000	0,128	0,312	0,746	0,000	0,000	0,000	0,262

5. Mutasi

Mutasi dilakukan dengan menghitung panjang total gen pada populasi yaitu popsize dikali dengan jumlah gen pada setiap kromosom, sedangkan posisi gen yang dimutasi dilakukan dengan membangkitkan bilangan acak antara 1 sampai dengan panjang total gen. Nilai pm ditentukan sebesar 0,1, maka jumlah terdapat 14 yang akan dimutasi dan dipilih secara acak. Hasil dari proses mutasi gen dapat dilihat pada Tabel 5.

Tabel 5. Hasil proses mutasi gen

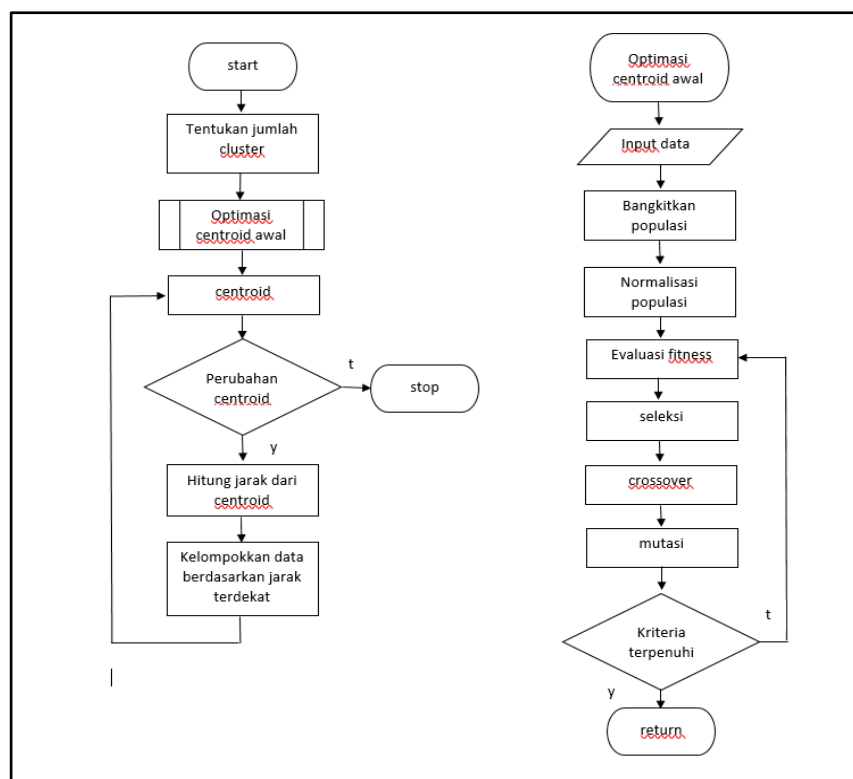
	G1	G2	G3	G4	G5	G6	G7	G8	fitness
K1	0,500	0,225	0,105	0,000	0,694	0,000	0,000	0,051	1,575
K2	0,225	0,500	0,105	0,694	0,000	0,000	0,000	0,051	1,575
K3	0,500	0,225	0,694	0,105	0,000	0,000	0,000	0,051	1,575
K4	0,000	0,225	0,105	0,694	0,000	0,000	1,000	0,500	2,524
K5	0,000	0,212	0,275	0,746	0,000	0,000	1,000	0,000	2,233
K6	0,000	0,212	0,275	0,000	0,746	0,000	0,400	0,344	1,977
K7	0,072	0,000	0,537	0,373	0,000	0,000	0,000	0,364	1,346
K8	0,000	0,033	0,052	0,000	0,000	0,400	0,000	0,344	0,829
K9	0,000	0,033	0,052	0,000	0,000	0,400	0,000	0,344	0,829
K10	0,033	0,000	0,052	0,000	0,000	0,000	0,400	0,344	0,829
K11	0,000	0,033	0,052	0,000	0,000	0,000	0,000	0,400	0,485
K12	0,000	0,013	1,000	0,000	0,373	0,000	0,200	1,000	2,587
K13	0,009	0,000	0,000	0,545	0,000	0,000	0,400	0,007	0,961
K14	0,000	0,000	0,009	0,545	0,000	0,000	0,400	0,007	0,961
K15	0,000	0,009	0,000	0,545	0,000	0,000	0,400	0,007	0,961
K16	0,000	0,100	0,236	1,000	0,000	0,000	0,000	0,051	1,386
K17	0,000	0,100	0,236	1,000	0,000	0,000	0,400	0,000	1,735
K18	0,000	0,128	0,312	0,746	0,000	0,000	0,000	0,262	1,448

Dari hasil proses mutasi gen nilai fitness terbesar didapat pada kromosom ke 12 dengan nilai sebesar 2,587. Selanjutnya kromosom dengan nilai terbesar ini akan dijadikan sebagai salah satu pusat kelompok pada algoritma k Means. Dengan cara yang sama selanjutnya dilakukan proses optimasi dengan algoritma genetika pada kabupaten Kuantan Singingi. Adapun hasil dari optimasi tersebut dapat dilihat pada Tabel 6.

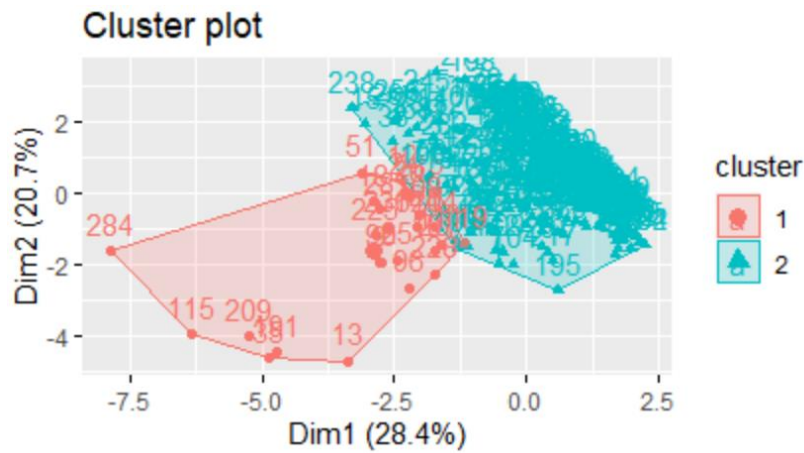
Tabel 6. Hasil optimasi algoritma genetika dua kabupaten

	G1	G2	G3	G4	G5	G6	G7	G8	Fitness
Kab. Bengkalis	0,000	0,013	1,000	0,000	0,373	0,000	0,200	1,000	2,587
Kab. Kuantan Singingi	1,000	0,718	0,777	0,917	0,228	0,546	0,333	0,000	4,519

Hasil dari *k-means* sangat tergantung pada nilai *centroid* yang menyebabkan *k-means* sering terjebak pada optimum lokal, dimana titik pusat akhir yang dihasilkan tidak benar-benar menjadi pusat kluster yang sesungguhnya. Pada penelitian ini nilai *centroid* akan diambil dari hasil optimasi diatas untuk melihat sejauh mana pemakaian algoritma genetika dapat memberikan pengaruh terhadap kluster dari *k-means*. Adapun *flowchart* penentuan nilai awal *centroid* dengan algoritma genetika dapat dilihat pada Gambar 3.

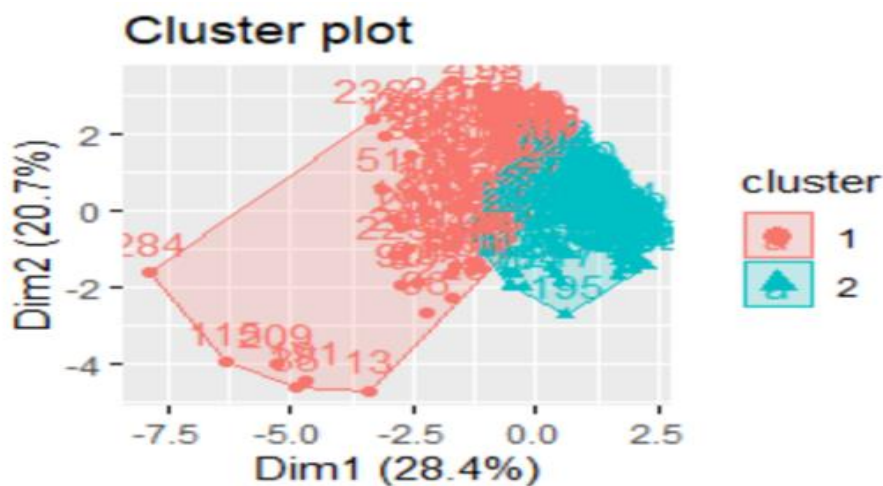
**Gambar 3.** Flowchart Optimasi Centroid K Means

Percobaan pertama dengan pemilihan kluster secara acak menghasilkan sebaran data seperti terlihat pada Gambar 4.



Gambar 4. Klasterisasi Data tanpa Optimasi

Selanjutnya hasil percobaan dengan melakukan optimasi genetika pada nilai *centroid* didapatkan hasil klaster sebagai seperti terlihat pada Gambar 5.



Gambar 5. Klasterisasi Data dengan optimasi genetika pada nilai centroid

Dari hasil uji validitas klaster menggunakan DBI didapatkan hasil yaitu untuk klasterisasi tanpa optimasi *centroid* didapat nilai DBI sebesar 2.164763 sedangkan klasterisasi dengan optimasi *centroid* dengan algoritma genetika didapat hasil DBI sebesar 2.06894.

D. Simpulan

K-means merupakan sebuah algoritma yang sederhana dan efektif sehingga membuatnya menjadi salah satu metode yang populer pada bidang analisa klasterisasi data, meskipun pemilihan *centroid* secara acak menjadi salah satu kelemahan pada algoritma ini. Penelitian ini bertujuan untuk melakukan optimasi pusat *centroid k-means* menggunakan algoritma genetika, dari hasil uji coba terlihat bahwa melakukan optimasi pada nilai *centroid* menghasilkan klaster yang lebih baik dibandingkan dengan klasterisasi *k-means* tanpa optimasi *centroid*. Hal

ini terlihat dari nilai validitas kluster yang lebih baik jika dibandingkan dengan nilai DBI pemilihan *centroid k-means* secara acak.

E. Referensi

- [1] A. V. Kulinkina *et al.*, "Piped water consumption in Ghana: A case study of temporal and spatial patterns of clean water demand relative to alternative water sources in rural small towns," *Sci. Total Environ.*, vol. 559, pp. 291–301, 2016, doi: 10.1016/j.scitotenv.2016.03.148.
- [2] P. Vora and B. Oza, "A Survey on K-mean Klastering and Particle Swarm Optimization," *Int. J. Sci. Mod. Eng.*, vol. 1, no. 3, pp. 24–26, 2013, [Online]. Available: <http://citeseerx.ist.psu.edu/messages/downloadexceeded.html>.
- [3] P. Fränti and S. Sieranoja, "How much can k-means be improved by using better initialization and repeats?," *Pattern Recognit.*, vol. 93, pp. 95–112, 2019, doi: 10.1016/j.patcog.2019.04.014.
- [4] S. K. Majhi and S. Biswal, "Optimal kluster analysis using hybrid K-Means and Ant Lion Optimizer," *Karbala Int. J. Mod. Sci.*, vol. 4, no. 4, pp. 347–360, 2018, doi: 10.1016/j.kijoms.2018.09.001.
- [5] S. Kapil, M. Chawla, and M. D. Ansari, "On K-means data klastering algorithm with genetic algorithm," *2016 4th Int. Conf. Parallel, Distrib. Grid Comput. PDGC 2016*, pp. 202–206, 2016, doi: 10.1109/PDGC.2016.7913145.
- [6] A. A. Hussein, "Improve The Performance of K-means by using Genetic Algorithm for Classification Heart Attack," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 2, p. 1256, 2018, doi: 10.11591/ijece.v8i2.pp1256-1261.
- [7] Mayasari, T. R., "Klastering Akses Air Bersih Dan Sanitasi Layak (Klastering of Clean Water access and Worth Sanitation in District / City Lampung Province)," *Fungsional Stat. Pertama BPS Kabupaten Pesawaran*, pp. 563–572, 2019.
- [8] H. S. Savitri, "Implementasi Importance Performance Analysis dan Algoritma K-means untuk Wilayah Indonesia," *J. Mat. Stat. dan Komputasi*, vol. 15, no. 1, p. 97, 2018, doi: 10.20956/jmsk.v15i1.4428.
- [9] Z. Long, G. Xu, J. Du, H. Zhu, T. Yan, and Y. Yu, "Flexible Subspace Klastering : A Joint Feature Selection and K-Means Klastering Framework," *Big Data Res.*, vol. 23, p. 100170, 2021, doi: 10.1016/j.bdr.2020.100170.
- [10] P. Govender and V. Sivakumar, *Application of k-means and hierarchical klastering techniques for analysis of air pollution: A review (1980–2019)*, vol. 11, no. 1. Turkish National Committee for Air Pollution Research and Control, 2020.
- [11] S. S. Yu, S. W. Chu, C. M. Wang, Y. K. Chan, and T. C. Chang, "Two improved k-means algorithms," *Appl. Soft Comput. J.*, vol. 68, pp. 747–755, 2018, doi: 10.1016/j.asoc.2017.08.032.
- [12] S. Chakraborty and S. Das, "Simultaneous variable weighting and determining the number of klusters—A weighted Gaussian means algorithm," *Stat. Probab. Lett.*, vol. 137, pp. 148–156, 2018, doi: 10.1016/j.spl.2018.01.015.
- [13] H. Kim, H. K. Kim, and S. Cho, "Improving spherical k-means for document klastering: Fast initialization, sparse centroid projection, and efficient kluster labeling," *Expert Syst. Appl.*, vol. 150, p. 113288, 2020, doi: 10.1016/j.eswa.2020.113288.

- [14] H. Xie *et al.*, "Improving K-means clustering with enhanced Firefly Algorithms," *Appl. Soft Comput. J.*, vol. 84, p. 105763, 2019, doi: 10.1016/j.asoc.2019.105763.
- [15] Y. Lu, B. Cao, C. Rego, and F. Glover, "A Tabu search based clustering algorithm and its parallel implementation on Spark," *Appl. Soft Comput. J.*, vol. 63, pp. 97–109, 2018, doi: 10.1016/j.asoc.2017.11.038.
- [16] A. Vitoria and O. B. P. Lezama, "Improvements for determining the number of clusters in k-means for innovation databases in SMEs," *Procedia Comput. Sci.*, vol. 151, no. 2018, pp. 1201–1206, 2019, doi: 10.1016/j.procs.2019.04.172.
- [17] J. Nasiri and F. M. Khiyabani, "A whale optimization algorithm (WOA) approach for clustering," *Cogent Math. Stat.*, vol. 5, no. 1, pp. 1–13, 2018, doi: 10.1080/25742558.2018.1483565.
- [18] S. Phon-Amnuaisuk and T. W. Au, "Computational Intelligence in Information Systems: Proceedings of the Fourth INNS Symposia Series on Computational Intelligence in Information Systems (INNS-CIIS 2014)," *Adv. Intell. Syst. Comput.*, vol. 331, 2015, doi: 10.1007/978-3-319-13153-5.
- [19] S. Kapoor, I. Zeya, C. Singhal, and S. J. Nanda, "A Grey Wolf Optimizer Based Automatic Clustering Algorithm for Satellite Image Segmentation," *Procedia Comput. Sci.*, vol. 115, pp. 415–422, 2017, doi: 10.1016/j.procs.2017.09.100.
- [20] T. Bhoskar, O. K. Kulkarni, N. K. Kulkarni, S. L. Patekar, G. M. Kakandikar, and V. M. Nandedkar, "Genetic Algorithm and its Applications to Mechanical Engineering: A Review," *Mater. Today Proc.*, vol. 2, no. 4–5, pp. 2624–2630, 2015, doi: 10.1016/j.matpr.2015.07.219.
- [21] M. Jahandideh-Tehrani, "A comparison of particle swarm optimization and genetic algorithm for daily rainfall-runoff modelling a case study for Southeast Queensland, Australia.pdf," *Optim. Eng.*, 2020, doi: 10.1007/s11081-020-09538-3.
- [22] X. Feng, X. Zhang, and Y. Xiang, "An inconsistency assessment method for backup battery packs based on time-series clustering," *J. Energy Storage*, vol. 31, no. August, 2020, doi: 10.1016/j.est.2020.101666.
- [23] D. A. I. C. Dewi and D. A. K. Pramita, "Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.