

Contents list available at [www.jurnal.unimed.ac.id](http://www.jurnal.unimed.ac.id)

## CESS (Journal of Computing Engineering, System and Science)

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



### Klastering Embung di Indonesia Berdasarkan Luas dan Volume Menggunakan Algoritma K-Means Clustering

### Clustering Reservoirs in Indonesia Based on Area and Volume Using the K-Means Algorithm

Shofinurdin<sup>1\*</sup>, Afifah Khaerani<sup>2</sup>, Arief Wibowo<sup>3</sup>

<sup>1,2,3</sup> Universitas Budiluhur, Jakarta, Indonesia

email: <sup>1</sup>[shofinurdin@gmail.com](mailto:shofinurdin@gmail.com), <sup>2</sup>[afifah.khaerani@gmail.com](mailto:afifah.khaerani@gmail.com), <sup>3</sup>[arief.wibowo@budiluhur.ac.id](mailto:arief.wibowo@budiluhur.ac.id)

#### ABSTRAK

Embung adalah cekungan yang digunakan untuk mengatur dan menampung aliran hujan dan meningkatkan kualitas air di badan air yang terkait. Embung juga digunakan untuk menjaga kualitas air tanah, mencegah banjir, menjaga estetika, dan mengairi. Terdapat 1.446 embung yang tersebar di seluruh wilayah Indonesia yang mempunyai luas dan volume yang beragam. Penelitian ini bertujuan untuk mengelompokkan embung di Indonesia berdasarkan luas dan volume menggunakan metode *data mining* klastering dengan algoritma K-Means. Sebelumnya, pengklasteran embung hanya dilakukan berdasarkan: tujuan pembangunan, penggunaan, aliran air, dan bahan pembentuknya. Hasil penelitian menunjukkan bahwa pengelompokkan embung yang terbaik adalah menjadi 4 klaster dengan rincian klaster kecil: 1.414 embung, klaster sedang: 24 embung, klaster tinggi: 7 embung dan klaster sangat tinggi: 1 embung.

**Kata Kunci:** Embung, Data Mining, Klastering, K-Means, Luas, Volume

#### ABSTRACT

An embung (Reservoir) is a basin used to control and collect rainfall runoff and improve the water quality in the associated water bodies. Embungs are utilized for preserving groundwater quality, preventing floods, enhancing aesthetics, and facilitating irrigation. There are 1,446 embungs scattered across the entire territory of Indonesia, varying in size and volume. This study aims to cluster the embungs in Indonesia based on their size and volume using the clustering data mining technique with the K-Means algorithm. Previously, embung clustering was only performed based on their purpose of construction, usage, water flow, and constituent materials. The research findings indicate that the optimal clustering of embungs consists of four clusters, with the following details: small cluster (1,414 embungs), medium cluster (24 embungs), high cluster (7 embungs), and very high cluster (1 embung).

**Keyword:** Reservoir, Data Mining, Clustering, K-Means, Size, Volume

\*Penulis Korespondensi:

email: [shofinurdin@gmail.com](mailto:shofinurdin@gmail.com)

## 1. PENDAHULUAN

Embung juga dikenal sebagai cekungan penampung, digunakan untuk mengatur dan menampung aliran air hujan dan meningkatkan kualitas air di badan air yang terkait, seperti sungai dan danau. Embung menampung air hujan di musim hujan dan kemudian digunakan pada musim kemarau. Embung digunakan untuk menjaga kualitas air tanah, mencegah banjir, mempertahankan keindahan, dan mengatur pengairan. Menurut Pusat Pendidikan dan Pelatihan Sumberdaya Air dan Konstruksi[1]. Embung adalah struktur yang digunakan untuk menyimpan air dan biasanya dibangun di daerah depresi di luar sungai. Setelah kolam embung menyimpan air di musim hujan, kelompok masyarakat hanya dapat menggunakan air selama musim kemarau untuk memenuhi kebutuhan mereka. Terdapat banyak embung di Indonesia dan tersebar di semua propinsi, menurut data yang diperoleh peneliti dari Direktorat Jenderal Pengendalian dan Penertiban Tanah dan Ruang Kementerian Agraria dan Tata Ruang/Badan Pertanahan Nasional pada tahun 2021 terdapat 1.800 embung.

Banyaknya embung yang ada di Indonesia tersebut telah mendorong peneliti untuk mencoba mengelompokkan embung tersebut dengan teknik klastering dalam data mining. Data mining sendiri adalah bidang ilmu untuk mengekstraksi pola yang menarik dari data dalam jumlah besar. Dengan kata lain, data mining adalah serangkaian proses yang dilakukan untuk mendapatkan suatu informasi yang bernilai dari sekumpulan data yang tidak bernilai. Jadi data mining juga bisa dikatakan sebagai pengolahan data menjadi suatu informasi yang akan memberikan pengetahuan [2].

Menurut Arief Wibowo[3] Pekerjaan *Data Mining* menggunakan algoritma yang spesifik untuk menyelesaikan fungsi yang berbeda. Algoritma mencari model yang paling sesuai dengan karakteristik dari data yang dijadikan pertimbangan. Salah satu metode data mining adalah klastering yaitu suatu metode data mining yang bersifat unsupervised (tanpa arahan), yang berarti bahwa metode ini dapat digunakan tanpa adanya latihan atau pelatihan, dan tidak memerlukan target output. Klastering adalah salah satu metode data mining yang mencari dan mengelompokkan data yang memiliki karakteristik yang mirip satu sama lain[4].

Algoritma yang lazim digunakan dalam klastering adalah algoritma K-Means yaitu metode analisis klaster non-hirarki, sering digunakan untuk membagi objek ke dalam satu atau lebih klaster atau kelompok objek berdasarkan karakteristiknya. Dengan demikian, objek dengan karakteristik yang sama dikelompokkan dalam satu klaster, dan objek dengan karakteristik yang berbeda dikelompokkan dalam klaster yang berbeda. [5]. Algoritma K-Means (KM) mengelompokkan N titik data ke dalam k klaster dengan meminimalkan jumlah jarak kuadrat antara setiap titik dan rata-rata klaster terdekat (centroid) [6]. Tahapan dalam algoritma K-Means terdiri dari penentuan jumlah klaster, menentukan pusat awal klaster dengan mengambil data secara acak, selanjutnya dilakukan perhitungan jarak dengan pusat klaster atau centroid menurut persamaan (1):

$$d(x,y)=|x-y|=\sqrt{\sum_{i=1}^n(x_i - y_i)^2} \quad (1)$$

Dengan demikian, penempatan data pada klaster didasarkan pada jarak nilai centroid terdekat. Penentuan pusat centroid baru dilakukan dengan menggunakan persamaan (2) untuk menentukan pusat klaster yang baru apabila centroid berubah:

$$\text{Pusat klaster baru} = \frac{x_1+x_2+x_3+\dots+x_n}{\text{jumlah } x} \quad (2)$$

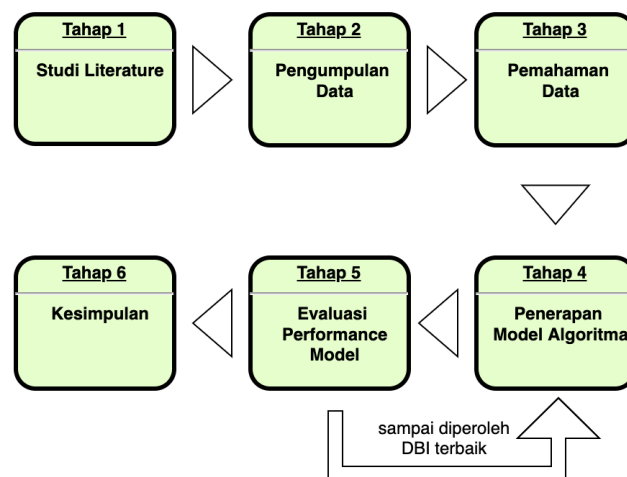
Untuk mengevaluasi apakah jumlah kluster dalam algoritma K-Means sudah tepat atau belum maka diukur dengan nilai Davies Bouldin Index (DBI) yaitu salah satu metode evaluasi internal, digunakan untuk mengevaluasi ketepatan jumlah kluster dalam algoritma K-Means. Nilai DBI ditentukan oleh metode pengelompokan yang bergantung pada nilai kohesi dan separasi, nilai kohesi adalah jumlah data terhadap centroid dari kluster yang diikuti, dan nilai separasi adalah jumlah data terhadap centroid dari kluster yang diikuti. [7]. Dengan nilai Davies Bouldin Index (DBI) yang lebih kecil (tidak negatif kurang dari 0), kluster yang dihasilkan lebih baik [8].

Sebelumnya, pengklasteran embung hanya dilakukan berdasarkan: tujuan pembangunan, penggunaan, aliran air, dan bahan pembentuknya, maka penelitian ini dilakukan bertujuan untuk mengelompokkan embung di Indonesia berdasarkan luas dan volume menggunakan metode *data mining* klustering dengan algoritma K-Means. Semoga hasil penelitian ini bisa memberikan kontribusi mengenai pengelompokan embung berdasarkan luas dan volume bagi praktisi dan akademisi di bidang hidrologi.

## 2. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode eksperimental dengan menggunakan data embung yang berjumlah 1.446 embung yang tersebar di seluruh wilayah Indonesia. Variabel yang digunakan adalah luas dan volume embung dengan mengesampingkan variable kedalaman karena dari data yang diperoleh semua kedalaman embung seragam 2,5 meter. Penelitian dilakukan dengan menggunakan laptop MacBook pro retina 2017 dengan spesifikasi Processor intel core i5-3210M CPU @ 2.50GHz, RAM 8 GB, system operasi MacOS Catalina, sedangkan tools yang dipakai adalah RapidMiner dan Microsoft excel.

Tahapan yang dilakukan adalah: studi literature, pengumpulan data, pemahaman data, penerapan model algoritma, evaluasi model, dan kesimpulan.



**Gambar 1.** Diagram langkah-langkah penelitian

### 2.1 Studi Literature

Tahap ini adalah tahap pengumpulan beberapa literatur yang membahas mengenai embung diantaranya adalah buku Teknik Bendungan karya Soedibyo diterbitkan oleh Pradnya Paramida. Menurut Soedibyo [9], embung dikategorikan ke dalam 4 kelompok yaitu: berdasarkan tujuan pembangunannya, berdasarkan penggunaannya, berdasarkan jalannya air dan berdasarkan material pembentuknya. Dalam buku tersebut embung belum dikategorikan berdasarkan luas dan volumenya, maka pada penelitian ini dilakukan

pengelompokan berdasarkan luas dan volume embung menggunakan pendekatan teknik *data mining*.

## 2.2 Pengumpulan Data

Data diperoleh dari Direktorat Jenderal Pengendalian dan Penertiban Tanah dan Ruang Kementerian Agraria dan Tata Ruang/Badan Pertanahan Nasional. Data berupa daftar embung yang tersebar di seluruh propinsi di Indonesia yang mempunyai luas dan volume. Dalam bentuk file Microsoft excel kemudian data dirapikan sehingga mudah dimengerti dan siap untuk dilakukan pemodelan.

## 2.3 Pemahaman Data

Dari data yang diperoleh terdiri dari 1.446 baris dan 6 kolom. Baris menunjukkan jumlah embung sedangkan kolom kolomnya terdiri dari "No", "Nama", "Propinsi", "Luas", "Kedalaman", dan "Volume". Kolom "No" menunjukkan nomor urut data dimulai dari 1 sampai dengan 1.446 yang menunjukkan bahwa jumlah data berjumlah 1.446 buah embung. Kolom "Nama" adalah kolom yang menunjukkan nama dari embung sesuai nomor urutnya. Kolom "Propinsi" menunjukkan nama propinsi yang ada di Indonesia tempat lokasi embung berada. Kolom "Luas" adalah kolom yang menunjukkan luas dalam satu satuan meter persegi (m<sup>2</sup>). Kolom "Kedalaman" menunjukkan kedalaman air embung dengan satuan meter (m), kolom ini seragam nilainya sebesar 2,5 m. Kolom "Volume" merupakan kolom yang berisi nilai volume air dalam satuan meter kubik (m<sup>3</sup>) nilainya diambil dari kolom "Luas" dikalikan dengan "Kedalaman". Secara rinci dapat dilihat dalam Tabel 1:

**Table 1.** Data Embung

No	Nama	Propinsi	Luas (m <sup>2</sup> )	Kedalaman (m)	Volume (m <sup>3</sup> )
1	Embung Lambadeuk	Aceh	63.889	2,5	159.749
2	Embung Lambeunot	Aceh	24.232	2,5	60.579
3	Embung Paya Gabus	Aceh	98.020	2,5	245.050
4	Embung Paya Cireh	Aceh	14.967	2,5	37.418
.....	.....	.....	.....	.....	.....
1.444	Waringin Sari Barat	Lampung	2.012	2,5	5.030
1.445	Way Sri Mulyo	Lampung	7.439	2,5	18.598
1.446	Way Tanjung Rejo	Lampung	5.798	2,5	14.495

## 2.4 Penerapan Algoritma

Dari data yang ada kemudian tahap selanjutnya adalah penerapan model algoritma dalam hal ini peneliti menggunakan tool rapid miner dan Algoritma yang dipakai adalah algoritma K-Means. Dipilihnya Algoritma K -means karena algoritma ini adalah algoritma klustering yang paling populer digunakan dan memiliki kelebihan yaitu sederhana dan mudah diimplementasikan [10]. Hasil dari penerapan algoritma dengan setiap parameter jumlah kluster yang berbeda kemudian dicatat untuk dievaluasi *performance*-nya.

## 2.5 Evaluasi Performance Model

Tahap ini adalah tahap dimana hasil penerapan algoritma dengan parameter jumlah kluster yang berbeda dibandingkan dengan nilai DBI (Davies Bouldin Index), dari hasil itu kemudian dievaluasi. Pertimbangan yang dinilai adalah jumlah DBI yang paling mendekati nilai nol dan jumlah kluster yang diterapkan apakah masuk akal atau tidak.

## 2.6 Kesimpulan

Tahap ini adalah tahap penarikan kesimpulan yang berupa pernyataan singkat tentang hasil analisis deskripsi atas beberapa kali hasil percobaan penerapan model yang dilakukan dan dievaluasi *performancenya*.

## 3. HASIL DAN PEMBAHASAN

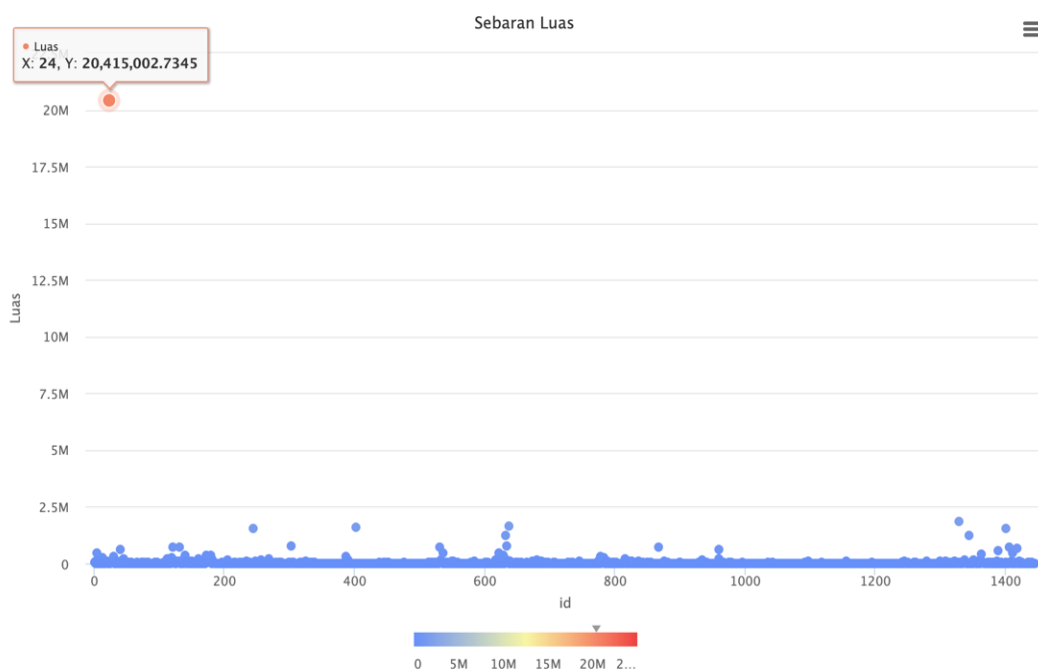
### 3.1 Deskriptif Statistik

Secara statistik deskriptif data yang ada dapat dilihat nilainya dalam Tabel 2:

**Tabel 2.** Deskriptif Statistik

Kolom	Min	Max	Average	Deviation
Luas	47	20.415.003	50.575	550.880
Volume	17	51.037.507	126.437	1.377.200

Dari tabel 2 terlihat bahwa luas dan volume terkecil dengan nilai 247 m<sup>2</sup> dan 617 m<sup>3</sup>. Luas dan volume ini terdapat pada data ke-910 yaitu Embung Sabri yang terletak di propinsi Nusa Tenggara Barat. Sedangkan luas dan volume tertinggi dengan nilai 20.415.003 m<sup>2</sup> dan 51.037.507 terdapat pada data ke-24 yaitu embung Talang Bersemi yang terletak di propinsi Riau. Untuk memberikan gambaran sebaran datanya dapat dilihat dalam grafik scatter plot sebagai berikut:

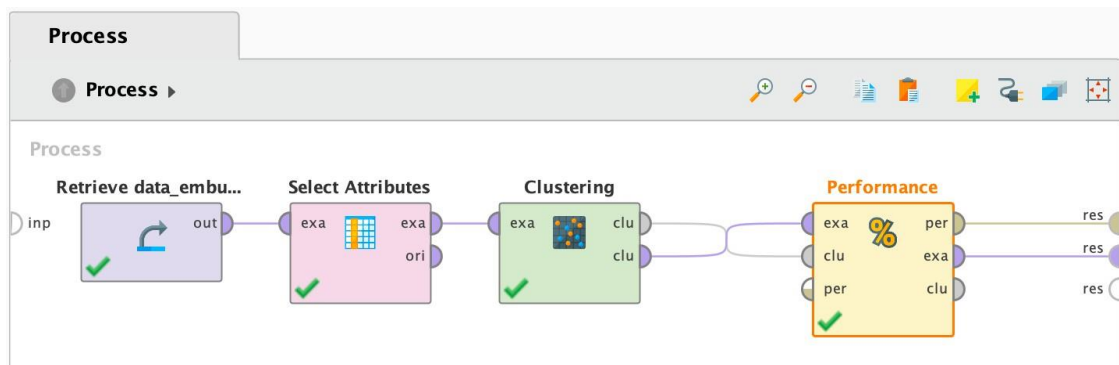


**Gambar 2.** Sebaran Luas

Tampak dalam gambar 2 terdapat 1 data yang terpisah dari data lainnya yaitu data ke-24 dengan nilai luas 20.415.003 m<sup>2</sup>. Secara visual data ini sudah bisa dipastikan data outlier atau pencilon yaitu data observasi yang muncul dengan nilai-nilai ekstrim, baik secara univariat ataupun multivariat. Yang dimaksud dengan nilai-nilai ekstrim ini adalah nilai yang sangat berbeda atau jauh dari sebagian besar nilai lain dalam kelompoknya [11]. Dalam penelitian ini data outlier tersebut dibiarkan dan tidak diganti atau dihilangkan karena apabila dihilangkan maupun diganti nilainya akan merubah data yang ada dan tidak mencerminkan keadaan sebenarnya.

### 3.2 Pemodelan

Uji coba pemodelan dilakukan dengan menggunakan RapidMiner sebagai tool data mining yang banyak digunakan karena bersifat free dan mudah digunakan. Pemodelan dilakukan dengan menerapkan beberapa modul operator sesuai fungsinya yaitu operator “retrieve”, operator “select attributes”, operator “Klastering” dan operator “performance”. Secara detail dapat dilihat dalam gambar 3.



Gambar 3. Diagram Pemodelan

Operator “retrieve” berfungsi untuk meload data ke dalam design dengan parameternya yang berisi data embung dalam bentuk excel. Setelah berhasil diload kemudian dengan menggunakan operator “select attributes” dipilih kolom yang diinginkan sebagai variable yang akan diproses dalam pemodelan. Kolom-kolom yang dipilih adalah kolom “Luas” dan “Volume” yang dimasukkan dalam parameternya. Operator “Klastering” berfungsi untuk memroses data kedalam pemodelan. Secara default Klastering ini menggunakan algoritma K-Means karena diambil dari kelompok segmentation dan dipilih K-Means, parameter utamanya yang harus diisi adalah jumlah klaster yang akan dipakai. Rapidminer akan menentukan secara otomatis titik centroid awalnya, sehingga tidak ada nilai parameter centroid yang perlu diisi. Kemudian operator “Performance” digunakan untuk melihat hasil evaluasi dari hasil penerapan algoritma dengan parameter jumlah klaster yang diinginkan. Parameter yang perlu dipilih adalah parameter “main criterion” yang diisi dengan “Davies Bouldin” dan “maximize perlu dipilih untuk mendapatkan nilai positif DBI-nya.

Setelah operator-operator diisi dengan parameter yang sesuai selanjutnya design ini diproses dengan mengklik tombol “Run Process Locally” atau menggunakan tombol keyboard “f11”. Selanjutnya hasil dari pemrosesan akan tampil dalam tab “result” baik hasil klasteringnya maupun hasil evaluasinya yang dapat dilihat dalam gambar 4.

Row No.	id	cluster	Luas	Vol
1	1	cluster_0	63899.481	159748.704
2	2	cluster_5	24231.624	60579.059
3	3	cluster_0	98019.880	245049.701
4	4	cluster_5	14967.021	37417.553
5	5	cluster_6	445991.250	1114978.125
6	6	cluster_5	16801.406	42003.515
7	7	cluster_5	44335.806	110839.515
8	8	cluster_0	92671.262	231678.154
9	9	cluster_5	26819.092	67047.730
10	10	cluster_5	5612.932	14032.329
11	11	cluster_5	33268.324	83170.809

Gambar 4. Hasil Pemodelan

### 3.3 Hasil Percobaan

Percobaan pemodelan dilakukan sebanyak 6 kali dengan mengisi parameter jumlah kluster dari 2 sampai dengan 7. Dari hasil percobaan itu diperoleh nilai centroid dan DBI yang berbeda-beda, secara rinci dapat dilihat dalam Tabel 3.

Table 3. Hasil uji coba

Percobaan	Jumlah kluster	Score DBI	Rincian Kluster	centroid	
				Luas	Volume
ke-1	2	0,002	C_0 : 1.445 item	C_0 : 36.482	C_0 : 91.204
			C_1 : 1 item	C_1 : 20.415.003	C_1 : 51.037.507
Ke-2	3	0,289	C_0 : 1.427 item	C_0 : 24.233	C_0 : 60.584
			C_1 : 1 item	C_1 : 20.415.003	C_1 : 51.037.507
			C_2 : 18 item	C_2 : 1.007.497	C_2 : 2.518.742
Ke-3	4	0,262	C_0 : 1.414 item	C_0 : 21.038	C_0 : 52.596
			C_1 : 1 item	C_1 : 20.415.003	C_1 : 51.037.507
			C_2 : 7 item	C_2 : 1.511.649	C_2 : 3.779.123
			C_3 : 24 item	C_3 : 516.092	C_3 : 1.290.229
Ke-4	5	0,292	C_0 : 1.387 item	C_0 : 17.900	C_0 : 44.751
			C_1 : 1 item	C_1 : 20.415.003	C_1 : 51.037.507
			C_2 : 7 item	C_2 : 1.551.649	C_2 : 3.779.123
			C_3 : 11 item	C_3 : 686.672	C_3 : 1.716.681
			C_4 : 40 item	C_4 : 243.837	C_4 : 609.593
Ke-5	6	0,297	C_0 : 126 item	C_0 : 15.764	C_0 : 39.411
			C_1 : 1 item	C_1 : 20.415.003	C_1 : 51.037.507
			C_2 : 5 item	C_2 : 1.630.700	C_2 : 4.076.750
			C_3 : 11 item	C_3 : 612.226	C_3 : 1.530.565
			C_4 : 2 item	C_4 : 163.854	C_4 : 409.636
			C_5 : 1.282 item	C_5 : 1.214.023	C_5 : 3.035.057
Ke-6	7	0,305	C_0 : 126 item	C_0 : 95.515	C_0 : 238.788
			C_1 : 1 item	C_1 : 20.415.003	C_1 : 51.037.507

C_2 : 5 item	C_2 : 1.630.700	C_2 : 4.076.750
C_3 : 11 item	C_3 : 686.672	C_3 : 1.716.681
C_4 : 2 item	C_4 : 1.214.023	C_4 : 3.035.057
C_5 : 1.282 item	C_5 : 12.704	C_5 : 31.760
C_6 : 19 item	C_6 : 329.467	C_6 : 823.668

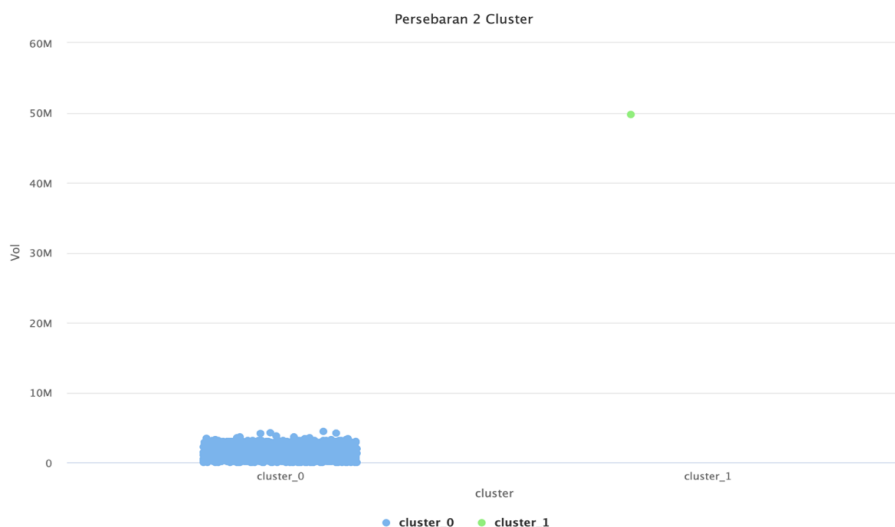
### 3.3 Pembahasan

Dari hasil percobaan dapat dilihat bahwa nilai terkecil DBI diperoleh pada percobaan ke-1 yaitu 0,002 dengan jumlah klaster 2 kemudian percobaan ke-3 yaitu 0,262 dengan jumlah klaster 4. Lebih lanjut dapat dilihat dalam gambar grafik 5.



**Gambar 5.** Score DBI

Nampak dalam grafik score DBI pada percobaan ke-1 senilai 0,002 kemudian naik pada percobaan ke-2 menjadi 0,289 dan menurun menjadi 0,262 pada percobaan ke-3, selanjutnya naik terus pada percobaan ke-4 sampai dengan percobaan ke-6 masing-masing senilai 0,292, 0,297 dan 0,305. Dari hasil tersebut nilai DBI terendah pada percobaan ke-1 senilai 0,002 dengan jumlah klasternya sebanyak 2 buah. Dalam table 3 terlihat bahwa percobaan ke-1 klaster 0 terdiri dari 1.445 embung dan klaster 1 sebanyak 1 embung. Jika dilihat persebaran datanya menggunakan grafik scatter plot akan tampak sebagai berikut:

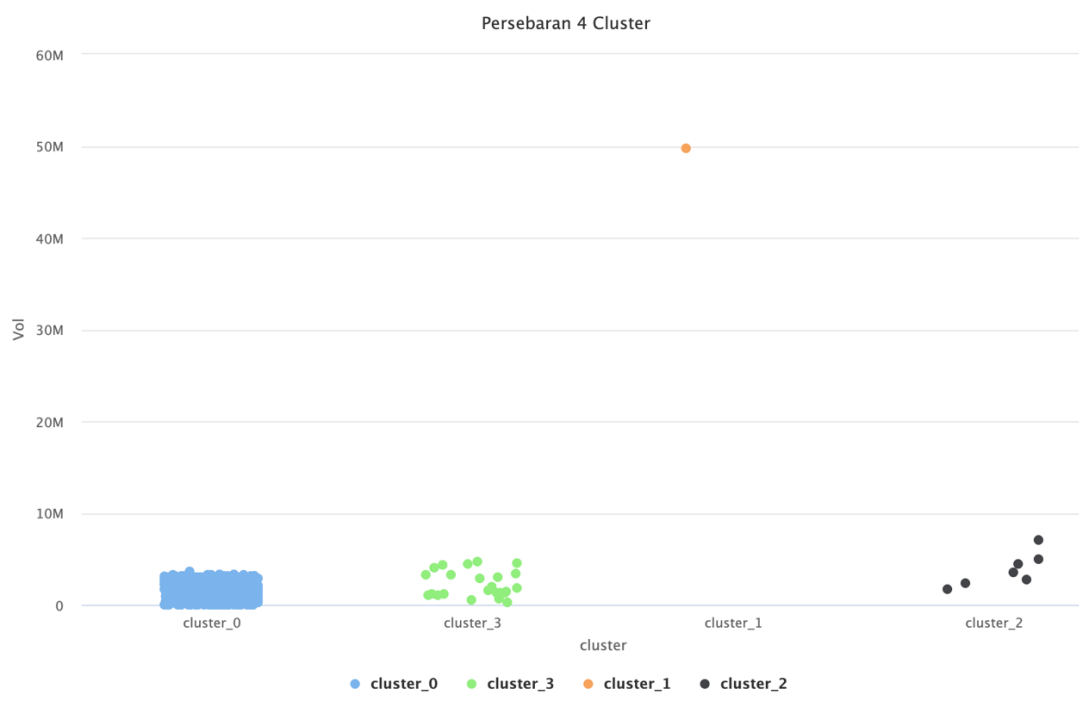


**Gambar 6.** Persebaran Data 2 Klaster



Nampak dalam gambar 6 bahwa klaster 1 hanya terdiri dari 1 buah embung yaitu data ke-24 embung Talang Bersemi, dimana data ke-24 ini adalah data outlier sebagaimana disebutkan sebelumnya. Dengan kata lain pembagian kedalam 2 klaster ini walaupun nilai DBI nya rendah (0,002) namun tidak mencerminkan keragaman sebenarnya karena hanya dibagi atas data outlier dan bukan outlier.

Nilai DBI yang terendah selanjutnya adalah 0,262 yaitu pada percobaan ke-3 dengan membagi data menjadi 4 klaster dimana dalam table 3 hasil ujicoba dirinci dengan klaster 0 terdiri dari 1.427 item dengan centroid luas 21.038 m<sup>2</sup> dan volume 52.596 m<sup>3</sup>, klaster 1 terdiri dari 1 item dengan centroid luas 20.415.003 dan volume 51.037.507, klaster 2 terdiri dari 7 item dengan centroid luas 1.511.649 m<sup>2</sup> dan volume 3.779.123 m<sup>3</sup> dan klaster 3 terdiri dari 24 item dengan centroid luas 516.092 m<sup>2</sup> dan volume 1.290.229. Jika dilihat persebaran datanya menggunakan scatter plot 2 dimensi dapat dilihat dalam gambar grafik 7.



**Gambar 7.** Persebaran 4 Klaster

Tampak dalam gambar 7 persebaran data dalam 4 klaster lebih terlihat beragam daripada dalam pembagian 2 klaster walaupun masih ada jumlah data yang hanya 1 item pada klaster 1. Sebagaimana dalam pembagian menjadi 2 klaster, klaster 1 berisi data outlier yaitu data ke-24 embung talangsari.

Peneliti berpendapat bahwa pembagian menjadi 4 klaster ini adalah pembagian yang terbaik karena disamping mempunyai nilai DBI yang paling kecil senilai 0.262 setelah pembagian menjadi 2 klaster, jumlah klasternya pun sudah mencerminkan keragaman dibandingkan dengan pembagian menjadi 2 klaster. Lebih lanjut pembagian menjadi klaster 4 ini dapat dilihat dalam table 4:

**Table 4.** Rincian 4 Klaster

Klaster	Centroid		Rentang Luas		Rentang Volume	
	Luas	Volume	Min	Max	Min	Max
C_0	21.038	52.596	246.931	267.723	617.327	669.306
C_1	20.415.003	51.037.507	20.415.003	20.415.003	51.037.507	51.037.507
C_2	1.511.649	3.779.123	1.210.518	1.835.189	3.026.296	4.587.972
C_3	516.092	1.290.229	287.357	761.949	718.393	1.904872

Dengan memperhatikan hasil percobaan klastering peneliti mengusulkan untuk pembagian embung menjadi 4 kategori dilihat dari luas dan volumenya yang diurutkan dari terkecil yaitu:

- Embung kategori Kecil (c\_0) dengan rentang luas 246.931m<sup>2</sup> sampai dengan 267.723 m<sup>2</sup>
- Embung kategori Sedang (c\_3) dengan rentang luas 287.357 m<sup>2</sup> sampai dengan 761.949 m<sup>2</sup>
- Embung kategori Besar (c\_2) dengan rentang luas 1.210.518m<sup>2</sup> sampai dengan 1.835.189m<sup>2</sup>
- Embung kategori Sangat Besar (c\_1) dengan luas 20.415.003 m<sup>2</sup>

#### 4. KESIMPULAN

Dari data 1.446 embung terdapat data *outlier* yaitu data ke 24 embung talangsari terletak di Propinsi Riau yang mempunyai luas 20.415.003 m<sup>2</sup> dan volume 51.037.507m<sup>3</sup>. Data outlier ini selalu membentuk klaster tersendiri di setiap ujicoba penerapan algoritma dengan parameter jumlah klaster yang berbeda. Jumlah klaster menurut peneliti yang paling baik adalah 4 klaster yang mempunyai nilai DBI 0,262 yang di peroleh pada percobaan ke-3. Empat klaster yang terbentuk adalah klaster 0 terdiri dari 1.414 embung yang mempunyai centroid (21.038, 52.596), klaster 1 terdiri dari 1 embung dengan centroid (20.415.003, 51.037.507), klaster 2 terdiri dari 7 embung yang dengan centroid (1.511.649, 3.779.123) dan klaster 3 terdiri dari 24 embung centroidnya (516.092, 1.290.229). Peneliti mengusulkan pembagian embung berdasarkan luas dan volumenya secara berurutan adalah embung kategori Kecil (C\_0) dengan rentang luas 246.931m<sup>2</sup> sampai dengan 267.723 m<sup>2</sup>, embung kategori Sedang (C\_3) dengan rentang luas 287.357 m<sup>2</sup> sampai dengan 761.949 m<sup>2</sup>, embung kategori Besar (C\_2) dengan rentang luas 1.210.518m<sup>2</sup> sampai dengan 1.835.189m<sup>2</sup>, embung kategori Sangat Besar (C\_1) dengan luas 20.415.003 m<sup>2</sup>. Dengan telah menghasilkan empat klaster embung maka penelitian ini telah menjawab tujuan penelitian untuk membagi embung berdasarkan luas dan volumenya. Penelitian ini hanya menggunakan data embung yang mempunyai data luas dan volume saja sebanyak 1.446 embung. Penelitian selanjutnya diharapkan menggunakan data yang lebih lengkap dan menggunakan variabel yang lebih banyak agar mendapatkan hasil pengelompokan yang lebih tepat.

#### REFERENSI

- [1] Arsyad, "Modul Pengantar Perencanaan Embung," hal. 1, 2017, [Daring]. Tersedia pada:  
[https://simantu.pu.go.id/epel/edok/45a22\\_04.\\_Modul\\_4\\_Pengantar\\_Perencanaan\\_Embung.pdf](https://simantu.pu.go.id/epel/edok/45a22_04._Modul_4_Pengantar_Perencanaan_Embung.pdf)

- [2] E. Dewi, S. Mulyani, F. Mulady, D. Ramadhan, A. Ariyantono, dan D. Ramdani, "Estimasi Harga Jual Mobil Bekas Menggunakan Metode Regresi Linier Berganda," *e-Jurnal JUSITI (Jurnal Sist. Inf. dan Teknol. Informasi)*, vol. 9, no. 1, hal. 1–8, 2020, doi: 10.36774/jusiti.v9i1.649.
- [3] A. Wibowo dan E. Winarko, "Paper Review: Data Mining Twitter," *Maint. Cult. Herit. Through Inf. Technol. a Smart Futur.*, no. November 2014, hal. 1–10, 2014, [Daring]. Tersedia pada: [https://www.researchgate.net/publication/329207488\\_Paper\\_Review\\_Data\\_Mining\\_Twitter](https://www.researchgate.net/publication/329207488_Paper_Review_Data_Mining_Twitter)
- [4] B. Santosa, "Data mining teknik pemanfaatan data untuk keperluan bisnis," *Yogyakarta Graha Ilmu*, vol. 978, no. 979, hal. 756, 2007.
- [5] M. S. Nawawi, F. Sembiring, dan A. Erfina, "Implementasi Algoritma K-Means Clustering Menggunakan Orange Untuk Penentuan Produk Busana Muslim Terlaris," *Semin. Nas. Teknol. Inf. dan Komunikasi-2021*, vol. 4, no. 1, hal. 789–797, 2021, [Daring]. Tersedia pada: [http://prosiding.unipma.ac.id/index.php/SENATIK/article/view/1837%0Ahttp://prosiding.uni\\_pma.ac.id/index.php/SENATIK/article/viewFile/1837/1723](http://prosiding.unipma.ac.id/index.php/SENATIK/article/view/1837%0Ahttp://prosiding.uni_pma.ac.id/index.php/SENATIK/article/viewFile/1837/1723)
- [6] P. Fränti dan S. Sieranoja, "How much can k-means be improved by using better initialization and repeats?," *Pattern Recognit.*, vol. 93, hal. 95–112, 2019, doi: 10.1016/j.patcog.2019.04.014.
- [7] W. Romadhona, B. Indarmawan Nugroho, dan A. Alim Murtopo, "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means," *J. Minfo Polgan*, vol. 11, no. 2, hal. 100–104, 2022, doi: 10.33395/jmp.v11i2.11797.
- [8] A. F. Solikin, Kusri, dan F. W. Wibowo, "Evaluasi Cluster Data Interkomparasi Anak Timbangan Dengan Algoritma Self Organizing Maps Cluster Evaluation Weighing Intercomparison Data with Self Organizing Maps Algorithm," *Sisfotenika*, vol. 11, no. 2, hal. 208–219, 2021.
- [9] I. Soedibyo, *Teknik Bendungan*, vol. 1. Jakarta: Pradnya Paramita, 1993.
- [10] Widiarina dan R. S. Wahono, "Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam Pemetaan Nasabah Potensial," *J. Intell. Syst.*, vol. 1, no. 1, hal. 33– 35, 2015.
- [11] Sabarudin *et al.*, "Efektivitas Pemberian Edukasi secara Online melalui Media Video dan Leaflet terhadap Tingkat Pengetahuan Pencegahan Covid-19 di Kota Baubau," *J. Farm. Galen. (Galenika J. Pharmacy)*, vol. 6, no. 2, hal. 309–318, 2020, doi: 10.22487/j24428744.2020.v6.i2.15253.