

# OPTIMASI JUMLAH CLUSTER METODE K-MEDOIDS BERDASARKAN NILAI DBI PADA PENGELOMPOKKAN DATA LUAS TANAMAN DAN PRODUKSI KELAPA SAWIT DI SUMATERA UTARA

Darmeli Nasution<sup>1</sup>, Donna N.M. Sirait<sup>2</sup>, Ivana Wardani<sup>3</sup>, Dwiyanto<sup>4</sup>

Universitas Pembangunan Panca Budi Medan<sup>1</sup>

Politeknik Penerbangan Medan<sup>2,3,4</sup>

Jl. Gatot Subroto KM. 4.5 Medan<sup>1</sup>

Jl. Penerbangan No. 85 Djamin Ginting Medan<sup>2,3,4</sup>

darmelinasution@gmail.com<sup>1</sup>, dehijo@gmail.com<sup>2</sup>, ivanawardani27@gmail.com<sup>3</sup>,  
ddwiyanto@gmail.com<sup>4</sup>

## **Abstract**

North Sumatra is the center of oil palm plantations in Indonesia. Palm oil is one of the leading commodities in North Sumatra province. Based on BPPS, North Sumatra has a very large land area of 440,000 ha and production yields of 7,000,000 tons. This study focuses on grouping the area of plantations and oil palm production in the districts/cities of North Sumatra province using the K-Medoids clustering method. In this grouping, several forms of the number of clusters are used to obtain a more optimal grouping. Each form of the number of clusters in this grouping is evaluated based on the DBI value. The smallest DBI value is the best form of the number of clusters in grouping plant area and oil palm production using K-Medoids. The results of this study indicate that the form of the number of clusters 2 is the best form of the number of clusters with a DBI value of 0.099.

**Keywords:** Data mining, Clustering, K-Medoids, Davies Bouldin Index.

## **Abstrak**

Sumatera Utara merupakan sentral perkebunan sawit di Indonesia. Kelapa sawit merupakan salah satu komoditas unggulan propinsi Sumatera Utara. Berdasarkan BPPS Sumatera Utara memiliki lahan yang sangat luas mencapai 440.000 ha dan hasil produksi mencapai 7.000.000 ton. Penelitian ini memfokuskan pada pengelompokkan luas tanaman dan produksi kelapa sawit di kabupaten/kota propinsi Sumatera Utara menggunakan metode K-Medoids clustering. Pada pengelompokkan ini menggunakan beberapa bentuk jumlah cluster untuk mendapatkan pengelompokkan yang lebih optimal. Masing-masing bentuk jumlah cluster pada pengelompokkan ini di evaluasi berdasarkan nilai DBI. Nilai DBI terkecil merupakan bentuk jumlah cluster terbaik pada pengelompokkan luas tanaman dan produksi kepala sawit menggunakan K-Medoids. Hasil penelitian ini menunjukkan bentuk jumlah cluster 2 merupakan bentuk jumlah cluster terbaik dengan nilai DBI 0,099.

**Kata kunci:** Data mining, Clustering, K-Medoids, Davies Bouldin Index.

## 1. PENDAHULUAN

Propinsi Sumatera Utara merupakan provinsi di Indonesia yang memiliki areal tanaman kelapa sawit yang luas. Luas tanaman kelapa sawit menurut data BPPS tahun 2020 mencapai 440.000 hektar dan produksi kelapa sawit mencapai 7 juta ton. Komoditas unggulan perkebunan di Sumatera Utara antara lain karet, kelapa sawit, kopi, kelapa, coklat, tembakau, dan tebu.

Perkenunan kelapa sawit dikelola oleh PTPN atau perusahaan swasta maupun perorangan. Luas tanaman dan produksi kelapa sawit menjadi perhatian khusus oleh pemerintah propinsi Sumatera Utara dalam rangka meningkatkan produksi kelapa sawit di propinsi Sumatera Utara. Oleh karena itu diperlukan sebuah informasi yang valid mengenai pengelompokan lahan tanama dan produksi kelapa sawit di kabupaten/kota di propinsi Sumatera Utara.

Pengelompokan adalah merupakan pembelajaran mesin tanpa pengawasan yang umum, kumpulan data harus secara otomatis dipartisi ke dalam cluster, sehingga objek di dalam yang sama cluster lebih mirip, sedangkan objek dalam cluster yang berbeda lebih berbeda[1].

Data Mining adalah suatu kegiatan pengambilan informasi baru dari sekumpulan data besar yang digunakan sebagai bahan untuk mengambil keputusan atau bisa dikatakan sebagai proses mengumpulkan informasi-informasi penting dari sebuah data yang besar untuk kemudian diubah menjadi sebuah pengetahuan[2][3].

K-Means menggunakan nilai rata-rata (mean) sebagai pusat cluster. Algoritma K-Medoids memiliki kelebihan untuk mengatasi kelemahan pada pada algoritma K-Means yang sensitive terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data[4][5][6].

Penelitian ini berfokus pada pengelompokan data luas tanaman dan produksi kelapa sawit di propinsi Sumatera Utara menggunakan metode K-Medoids. Untuk menghasilkan pengelompokan yang optimal di lakukan percobaan dengan beberapa bentuk jumlah cluster. Penentuan bentuk jumlah cluster terbaik menggunakan evaluasi nilai DBI[7][8]. Bentuk jumlah cluster terbaik adalah bentuk jumlah cluster yang memiliki nilai DBI terkecil. Setelah ditemukan bentuk jumlah cluster terbaik, maka dilanjutkan dengan pengelompokan luas tanaman dan produksi kelapa sawit di propinsi Sumatera Utara.

## 2. METODOLOGI PENELITIAN

### 2.1. Data Penelitian

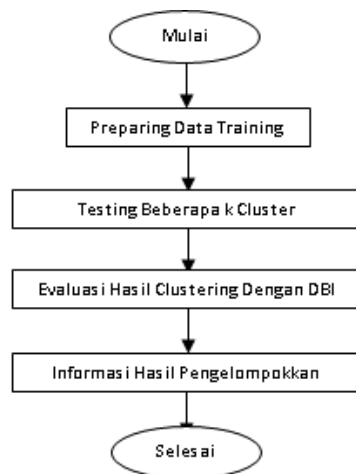
Data penelitian ini diambil dari Badan Pusat Statistik Sumatera Utara berupa data luas tanaman dan produksi kelapa sawit propinsi Sumatera Utara tahun 2019-2020.

Tabel 1. Data Mentah

No	L-2019	L-2020	P-2019	P-2020
1	0.00	0.00	0.00	0.00
2	18,570.00	19,081.52	306,172.73	315,129.09
3	6,586.00	6,590.00	71,677.27	78,831.82
4	3,609.00	3,642.00	36,390.91	42,290.91
5	35.00	36.00	263.64	331.82
6	1,129.00	1,129.00	11,977.27	14,000.00
7	35,460.00	35,527.00	505,372.73	532,600.00
8	77,147.00	77,243.00	1,622,468.18	1,631,013.64
9	30,257.00	30,327.00	512,095.45	520,518.18
10	569.00	575.00	3,559.09	3,690.91
11	1,711.00	1,718.00	27,622.73	27,677.27
12	14,076.00	14,122.00	217,372.73	224,595.45
13	47,174.00	47,198.00	758,718.18	764,222.73
14	931.00	945.00	3,127.27	3,740.91
15	373.00	374.00	1,763.64	2,686.36
16	1,337.00	1,346.00	2,286.36	2,331.82
17	0.00	0.00	0.00	0.00
18	12,882.00	13,821.00	219,340.91	227,845.45
19	9,923.00	9,977.00	131,322.73	138,763.64
20	27,769.00	27,845.00	339,345.45	347,286.36
21	34,644.00	34,652.00	515,231.82	521,672.73
22	42,922.00	43,009.00	637,304.55	682,302.73
23	72,113.00	72,137.00	1,083,036.36	1,117,481.82
24	0.00	0.00	0.00	0.00
25	0.00	0.00	0.00	0.00
26	0.00	0.00	0.00	0.00
27	0.00	0.00	0.00	0.00
28	0.00	0.00	0.00	0.00
29	0.00	0.00	0.00	0.00
30	0.00	0.00	0.00	0.00
31	0.00	0.00	0.00	0.00
32	98.00	105.00	536.36	736.36
33	0.00	0.00	0.00	0.00

## 2.2. Kerangka Penelitian

Untuk menghasilkan pengelompokan yang optimal, dilakukan dengan beberapa tahapan seperti yang ditampilkan pada gambar 1.



Gambar 1. Kerangka Penelitian

### 2.3. Metode K-Medoids

K-Medoids menggunakan objek representatif sebagai titik acuan, bukan mengambil nilai rata-rata dari objek dalam setiap kluster. Algoritma mengambil parameter input k, jumlah kluster yang akan dipartisi di antara satu set n objek. Tahapan dari K- medoids adalah:

Tahap 1: Pilih titik awal K. Titik ini adalah medoid yang berfungsi sebagai kandidat dan dimaksudkan untuk menjadi poin paling sentral dari kelompok tersebut.

Tahap 2: Kemudian setiap objek yang tersisa dikelompokkan dengan objek representatif yang paling mirip.

Tahap 3: Kemudian mengganti salah satu medoid dengan medoid lain yang secara kualitas lebih baik dan dilakukan secara iterative C[7].

## 3. HASIL DAN PEMBAHASAN

### 3.1. Perhitungan Algoritma K-Medoids Dengan Jumlah Cluster 3

Setelah data dinormalisasi, dilanjutkan dengan proses pengelompokkan dengan algoritma K-Means.

a. Penentuan nilai Centroid

Tabel 2. Nilai Centroid Iterasi-1

Centroid	L-2019	L-2020	P-2019	P-2020
C1	0.0005	0.0005	0.0034	0.0002
C2	0.4596	0.4599	6.5426	0.3265
C3	0.9347	0.9339	14.0212	0.6851

b. Menghitung jarak dari Centroid

Perhitungan jarak menggunakan Euclidian Distance iterasi-1 dengan rumus :

$$D_{(i,f)} = \sqrt{(X_{1i} - X_{1f})^2 + (X_{2i} - X_{2f})^2 + \dots + (X_{ki} - X_{kf})^2}$$

Berikut adalah perhitungannya pada iterasi-1.

$$D_{x1,c1} =$$

$$\sqrt{(0.0000 - 0.0005)^2 + (0.0000 - 0.0005)^2} + (0.0000 - 0.0034)^2 + (0.0000 - 0.0002)^2$$

$$= 0.0035$$

$$D_{x2,c1} =$$

$$\sqrt{(0.2407 - 0.0005)^2 + (0.2407 - 0.0005)^2} + (3.9638 - 0.0034)^2 + (0.1932 - 0.0002)^2$$

$$= 3.9800$$

Dan seterusnya sampai dengan  $D_{x33,c1}$ . Kemudian dilanjutkan dengan menghitung jarak dari Centroid ke-2 :

$$D_{x1,c2} =$$

$$\sqrt{(0.0000 - 0.4596)^2 + (0.0000 - 0.4596)^2} + (0.0000 - 6.5426)^2 + (0.0000 - 0.3265)^2$$

$$= 6.5749$$

$$D_{x2,c2} =$$

$$\sqrt{(0.2407 - 0.4596)^2 + (0.2407 - 0.4596)^2} + (3.9638 - 6.5426)^2 + (0.1932 - 0.3265)^2$$

$$= 2.6041$$

Dan seterusnya sampai dengan  $D_{x33,c2}$ . Kemudian dilanjutkan dengan menghitung jarak dari Centroid ke-3 :

$$D_{x1,c3} =$$

$$\sqrt{(0.0000 - 0.9347)^2 + (0.0000 - 0.9339)^2} + (0.0000 - 14.0212)^2 + (0.0000 - 0.6851)^2$$

$$= 14.0833$$

$$D_{x2,c3} =$$

$$\sqrt{(0.2407 - 0.9347)^2 + (0.2407 - 0.9339)^2} + (3.9638 - 14.0212)^2 + (0.1932 - 0.6851)^2$$

$$= 10.1065$$

Perhitungan dilanjutkan sampai dengan  $D_{x33,c3}$ .

c. Menghitung nilai jarak terdekat dan total jarak terdekat

Table 3. Jarak Centroid Iterasi ke-1

C1	C2	C3	Jarak Terpendek	Hassil Cluster
0.0035	6.5749	14.0833	0.0035	C1
3.9800	2.6041	10.1065	2.6041	C2
0.9335	5.6398	13.1482	0.9335	C1

0.4730	6.0996	13.6080	0.4730	C1
0.0002	6.5714	14.0798	0.0002	C1
0.1532	6.4185	13.9270	0.1532	C1
6.5795	0.3265	7.5157	0.3265	C2
21.0726	14.5167	7.0554	7.0554	C3
6.6571	0.3442	7.4380	0.3442	C2
0.0438	6.5280	14.0364	0.0438	C2
0.3559	6.2159	13.7243	0.3559	C2
2.8259	3.7516	11.2582	2.8259	C2
9.8681	3.3201	4.2494	3.3201	C2
0.0407	6.5329	14.0414	0.0407	C1
0.0204	6.5515	14.0599	0.0204	C1
0.0355	6.5430	14.0515	0.0355	C1
0.0035	6.5749	14.0833	0.0035	C1
2.8501	3.7278	11.2341	2.8501	C1
1.7085	4.8658	12.3739	1.7085	C1
4.4243	2.1645	9.6645	2.1645	C2
6.7046	0.3447	7.3898	0.3447	C2
8.2952	1.7638	5.8103	1.7638	C2
14.0965	7.5398	0.6851	0.6851	C3
0.0035	6.5749	14.0833	0.0035	C1
0.0035	6.5749	14.0833	0.0035	C1
0.0035	6.5749	14.0833	0.0035	C1
0.0035	6.5749	14.0833	0.0035	C1
0.0035	6.5749	14.0833	0.0035	C1
0.0035	6.5749	14.0833	0.0035	C1
0.0035	6.5749	14.0833	0.0035	C1
0.0035	6.5749	14.0833	0.0035	C1
0.0035	6.5749	14.0833	0.0035	C1
0.0038	6.5678	14.0762	0.0038	C1
Total Jarak Terdekat			28.0879	

Table 4. Jarak Centroid Iterasi ke-2

<b>C1</b>	<b>C2</b>	<b>C3</b>	<b>Jarak Terpendek</b>	<b>Hasil Cluster</b>
0.0000	6.6529	21.0523	0.0000	C1
3.9834	2.6811	17.0756	2.6811	C2
0.9370	5.7184	20.1185	0.9370	C1
0.4765	6.1779	20.5778	0.4765	C1
0.0035	6.6494	21.0488	0.0035	C1
0.1567	6.4966	20.8962	0.1567	C1
6.5830	0.3511	14.4859	0.3511	C2
21.0760	14.4354	1.0000	1.0000	C3
6.6605	0.3191	14.4043	0.3191	C2
0.0473	6.6061	21.0056	0.0473	C2
0.3594	6.2939	20.6934	0.3594	C2
2.8293	3.8295	18.2278	2.8293	C2

9.8716	3.2418	11.2055	3.2418	C2
0.0440	6.6111	21.0107	0.0440	C1
0.0239	6.6295	21.0290	0.0239	C1
0.0385	6.6213	21.0211	0.0385	C1
0.0000	6.6529	21.0523	0.0000	C1
2.8536	3.8053	18.2033	2.8536	C1
1.7120	4.9443	19.3441	1.7120	C1
4.4278	2.2470	16.6375	2.2470	C2
6.7080	0.3321	14.3592	0.3321	C2
8.2987	1.6901	12.7763	1.6901	C2
14.0999	7.4626	7.0177	7.0177	C3
0.0000	6.6529	21.0523	0.0000	C1
0.0000	6.6529	21.0523	0.0000	C1
0.0000	6.6529	21.0523	0.0000	C1
0.0000	6.6529	21.0523	0.0000	C1
0.0000	6.6529	21.0523	0.0000	C1
0.0000	6.6529	21.0523	0.0000	C1
0.0000	6.6529	21.0523	0.0000	C1
0.0000	6.6529	21.0523	0.0000	C1
0.0000	6.6529	21.0523	0.0000	C1
0.0000	6.6529	21.0523	0.0000	C1
0.0072	6.6458	21.0452	0.0072	C1
Total Jarak Terdekat			28.3690	

Nilai S= 28.3690-28.0879= 0.2811. Maka perhitungan algoritma K-Medoids berhenti.

### 3.2. Evaluasi Jumlah Cluster Berdasarkan Nilai DBI

Untuk melihat seberapa optimal jumlah cluster pada metode K-Medoids, peneliti melakukan evaluasi berdasarkan nilai DBI. Semakin kecil nilai DBI maka semakin baik hasil pengklasterannya. Peneliti menggunakan 8 bentuk yaitu jumlah cluster 2, jumlah cluster 3, jumlah cluster 4, jumlah cluster 5, jumlah cluster 6, jumlah cluster 7, jumlah cluster 8, dan 9 jumlah cluster. Adapun rumus mencari nilai DBI adalah sebagai berikut :

a. 1. Mencari SSW dengan rumus :

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i)$$

Keterangan :

$m_i$  = Jumlah data dalam cluster ke-i

X = Data dalam cluster

D(x,c) = Jarak data ke centroid

$x_j$  = Data pada cluster tersebut

$c_i$  = Centroid cluster ke-i

2. Mencari SSB dengan rumus :

$$SSB_{ij} = d(c_i, c_j)$$

Keterangan :

$c_i$  = Cluster 1

$c_j$  = Cluster lainnya

$d(c_i, c_j)$  = Jarak antara centroid satu dengan lainnya

3. Mencari Rasio dengan rumus :

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

Keterangan :

$R_{ij}$  = Rasio antar cluster

$SSW_i$  = Cluster 1

$SSW_j$  = Cluster 2

$SSB_{ij}$  = Separasi dari cluster 1 dan 2

4. Mencari DBI dengan rumus :

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{ij})$$

Keterangan :

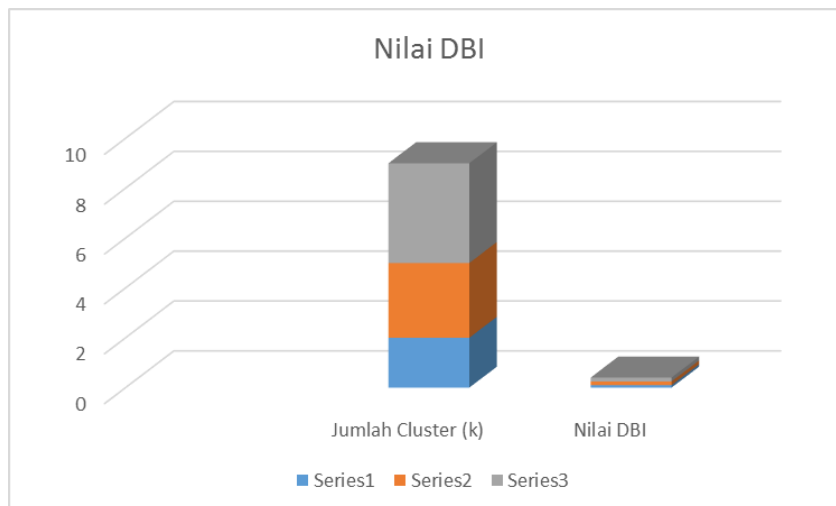
$K$  = Cluster yang ada

$R_{i,j}$  = Rasio antara cluster i dan j

$SSW_i$  = Dicari rasio antar cluster yang terbesar

### 3.3. Hasil Nilai DBI Dari Beberapa bentuk jumlah cluster K-Medoids Clustering Dan Pengelompokkan Jumlah Cluster Optimal

Berikut adalah hasil nilai DBI dari K-Medoids Clustering dengan beberapa jumlah cluster :



Gambar 2. Nilai DBI dari Beberapa Jumlah Cluster

Hasil pengelompokkan luas tanaman dan produksi kelapa sawit metode K-Medoids dengan bentuk jumlah jumlah cluster optimal dapat dilihat pada gambar 3 berikut ini :





Gambar 3. Hasil Pengelompokan Cluster Optimal

Dari hasil evaluasi berdasarkan nilai DBI dari bentuk masing-masing jumlah cluster didapatkan nilai DBI sebagai berikut :

Table 5. Nilai DBI dari Masing-Masing Bentuk Jumlah Cluster

No	Bentuk	Nilai DBI
1	Jumlah Cluster 2	0.099
2	Jumlah Cluster 3	0.149
3	Jumlah Cluster 4	0.155

Dari tabel diatas maka bentuk jumlah cluster terbaik adalah jumlah cluster 2 terdiri dari tinggi, tinggi, dan rendah. Dari bentuk jumlah cluster terbaik kita dapatkan hasil pengelompokan luas tanaman dan produksi kelapa sawit di Sumatera Utara yaitu cluster sangat tinggi : Asahan, Langkat, Labuhanbatu Selatan dan Labuhanbatu Utara. Sedangkan cluster rendah : 29 kabupaten/kota di Sumatera Utara selain dari 4 kabupaten yang masuk ke cluster tinggi.

#### 4. SIMPULAN

Hasil dari penelitian ini adalah sebuah perbandingan dari jumlah cluster berdasarkan nilai DBI pada pengelompokan luas tanama dan produksi kelapa sawit di propinsi Sumatera Utara. Nilai DBI terkecil adalah adalah jumlah cluster 2 metode K-Medoids Clustering yaitu : 0.099 . Hasil pengelomppokkan dari bentuk jumlah 2 cluster metode K-Medoids Clustering pada pengelompokan luas tanaman dan produksi kelapa sawit di propinsi Sumatera Utara adalah cluster tinggi terdiri dari Asahan, Langkat, Labuhanbatu Selatan dan Labuhanbatu Utara. Sedangkan cluster rendah adalah 29 kabupaten/kota di Sumatera Utara selain dari 4 kabupaten yang masuk ke cluster tinggi.

#### DAFTAR PUSTAKA

- [1] E. Schubert and P. J. Rousseeuw, *Fast and eager k-medoids clustering:  $O(k)$  runtime improvement of the PAM, CLARA, and CLARANS algorithms*, vol. 101, no. 124020371. 2021.
- [2] H. Effendi, A. Syahril, S. Prayoga, and W. D. Hidayat, "Penerapan Metode K-Means Clustering Untuk Pengelompokan Lahan Sawit Produktif Pada PT Kasih Agro Mandiri," *Teknomatika*, vol. 11, no. 02, pp. 117–126, 2021.
- [3] M. S. Mustafa, M. R. Ramadhan, and A. P. Thenata, "Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *Creat. Inf. Technol. J.*, vol. 4, no. 2, p. 151, 2018.
- [4] B. Wira, A. E. Budianto, and A. S. Wiguna, "Implementasi Metode K-Medoids Clustering untuk Mengetahui Pola Pemilihan Program Studi," *J. Terap. Sains Teknol.*, vol. 1, no. 3, pp. 54–69, 2019.
- [5] H. Ningrum, E. Irawan, and M. R. Lubis, "Implementasi Metode K-Medoids Clustering Dalam Pengelompokan Data Penyakit Alergi Pada Anak," *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.)*, vol. 6, no. 1, p. 130, 2021.
- [6] S. Asmiatun, "Penerapan Metode K-Medoids Untuk Pengelompokan Kondisi Jalan Di Kota Semarang," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 6, no. 2, pp. 171–180, 2019.
- [7] R. D. Ramadhani and D. J. Ak, "Evaluasi K-Means dan K-Medoids pada Dataset Kecil," *Semin. Nas. Inform. dan Apl.*, no. September, pp. 20–24, 2017.