

Optimalisasi Pengelompokan Data Produksi Kelapa Sawit Menggunakan Algoritma K-Medoids

Fina Nasari¹

¹ Universitas Potensi Utama

Jl. K.L Yos Sudarso KM. 6.5 Tanjung Mulia Medan

finanasari@gmail.com

Intisari— Kelapa sawit merupakan salah satu tanaman yang banyak dikembangkan di area perkebunan di provinsi Riau. Luas area dan produksi kelapa sawit provinsi Riau menjadi salah satu yang tertinggi di Indonesia. Penyebaran perkebunan kelapa sawit yang hampir diseluruh kabupaten/kota di provinsi Riau, perlu adanya pengelompokan daerah berdasarkan jumlah produksi. Berdasarkan akurasi dan eksekusi algoritma k-medoids lebih baik dari algoritma k-means dalam pengelompokan data. Tujuan dari penelitian ini adalah mengelompokkan daerah produksi kelapa sawit dengan dalam 2 kelompok, 3 dan 5 kelompok. Dari hasil pengelompokan tersebut akan dilihat kelompok terbaik berdasarkan nilai DBI terkecil. Hasil penelitian untuk jenis 2 kelompok diperoleh hasil 7 data pada kelompok 1 dan 5 data pada kelompok 2. Untuk jenis 3 kelompok diperoleh hasil 5 data pada kelompok 1, 4 data pada kelompok 2 dan 3 data pada kelompok 3, sedangkan jenis 5 kelompok diperoleh hasil 3 data pada kelompok 1, 2 dan 4, 1 data pada kelompok 3 dan 2 data pada kelompok 4. Berdasarkan nilai DBI diperoleh jenis kelompok 2 menjadi jenis kelompok terbaik dengan nilai DBI -159796492242,667.

Kata kunci— clustering, K-Medoids, Nilai DBI, Kelapa Sawit.

Abstract— Oil palm is one of the plants that has been widely developed in plantation areas in Riau province. The area and production of oil palm in Riau province is one of the highest in Indonesia. The spread of oil palm plantations in almost all districts/cities in Riau province requires grouping areas based on the amount of production. based on the accuracy and execution of the k-medoids algorithm is better than the k-means algorithm in grouping data. The purpose of this research is to classify oil palm production areas into 2 groups, 3 and 5 groups. From the results of this grouping, the best group will be seen based on the smallest DBI value. The research results for type 2 groups obtained results of 7 data in group 1 and 5 data in group 2. For type 3 groups obtained results of 5 data in group 1, 4 data in group 2 and 3 data in group 3, while type 5 groups obtained results 3 data in groups 1, 2 and 4, 1 data in group 3 and 2 data in group 4. Based on the DBI value, it was found that group 2 was the best type of group with a DBI value of -159796492242.667.

Keywords— clustering, K-Medoids, DBI Value, Palm Oil.

I. PENDAHULUAN

Kelapa sawit merupakan salah satu tanaman yang banyak dibudidayakan dan sumber ekspor utama Indonesia. Kelapa sawit tersebar di pulau Sumatera, Kalimantan, Sulawesi dan Jawa. Produksi kelapa sawit di pulau Sumatera khususnya provinsi Riau menjadi salah satu dengan produksi kelapa sawit tertinggi. Sebaran perkebunan kelapa sawit di Riau hampir disetiap kabupaten. Untuk melihat daerah/kabupaten dengan hasil produksi tinggi dan rendah, perlu adanya pengelompokan daerah. Pengelompokan daerah berdasarkan produksi kelapa sawit.

Clustering adalah salah satu metode pengelompokan data. Algoritma pengelompokan data diantaranya algoritma k-means dan k-medoids. Jika dilihat dari akurasi, eksekusi dan kompleksitas waktu algoritma k-medoids lebih unggul dibandingkan dengan algoritma k-means berdasarkan pengujian dengan aplikasi berbasis java, hadoop dan hive[1].

Performa Algoritma k-means lebih baik dari pada algoritma k-medoids jika menggunakan tools rapid miner dengan menggunakan metode confusion matrix dengan parameter accuracy, recall dan precision[2].

Pengelompokan menggunakan algoritma k-medoids mampu meningkatkan produksi panen kelapa sawit, dimana hasil pengelompokan secara manual dan menggunakan tools rapid miner memiliki hasil yang sama[3].

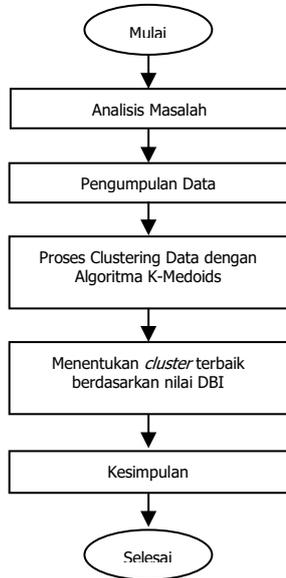
Tujuan dari penelitian ini adalah mengelompokkan data dengan menggunakan algoritma k-medoids dalam bentuk 2 kelompok, 3 dan 5 kelompok serta melihat kelompok terbaik berdasarkan nilai DBI.

II. METODOLOGI PENELITIAN

Tahapan penelitian yang akan dilakukan dimulai dari analisis masalah, mengumpulkan data-data yang dibutuhkan baik data primer ataupun data sekunder, melakukan proses pengelompokan data menggunakan algoritma k-medoids dengan menggunakan bantuan tools rapid miner, dilanjutkan dengan menentukan kelompok terbaik, kelompok terbaik

dipilih berdasarkan nilai DBI terkecil dan tahapan terakhir pengambilan keputusan dari penelitian yang dilakukan terkait hasil pengelompokan algoritma *k-medoids* serta kelompok terbaik berdasarkan nilai DBI.

Tahapan metode penelitian ini terlihat pada gambar 1 berikut ini:



Gambar 1. Tahapan Penelitian.

A. Analisis Masalah

Pada tahapan ini dilakukan analisis masalah terkait dengan penyebaran kelapa sawit yang menyeluruh diprovinsi riau, dan belum terpetakan daerah dengan produksi tinggi dan rendah. Penelitian sebelumnya telah melakukan pemetaan dengan menggunakan algoritma *k-means* clustering, perlu melakukan pengujian dengan menggunakan algoritma *k-medoids* untuk melihat hasil pengelompokannya.

B. Pengumpulan Data

Data yang digunakan pada penelitian ini ada data primer dan data sekunder. Dimana data primer berupa data produksi kelapa sawit provinsi riau untuk setiap kabupaten yang diperoleh dari BPS Riau. Sedangkan data sekunder dari penelitian ini berupa rujukan-rujukan dan studi literature berkaitan dengan pengelompokan produksi kelapa sawit, algoritma *clustering* khususnya menggunakan metode *k-medoids* dan performa dengan nilai *DBI*.

C. Proses Clustering Data Dengan Algoritma K-Medoids

Proses *clustering* yang akan dilakukan dengan menggunakan rapat miner. Rapat miner sendiri merupakan *tools* yang digunakan dalam proses analisis data.

Clustering mampu membagikan data menjadi beberapa group berdasarkan Kesamaan yang telah ditetapkan sebelumnya[4]. *clustering* mengelompokkan sejumlah data pada cluster yang sama dan data yang berbeda pada cluster yang berbeda[5][6].

Algoritma *k-medoids* merupakan salah satu algoritma *clustering* yang memiliki kemiripan dengan *k-means* dimana perbedaannya pada algoritma *k-medoids* menggunakan objek sebagai perwakilan(*medoids*)[2][7][8].

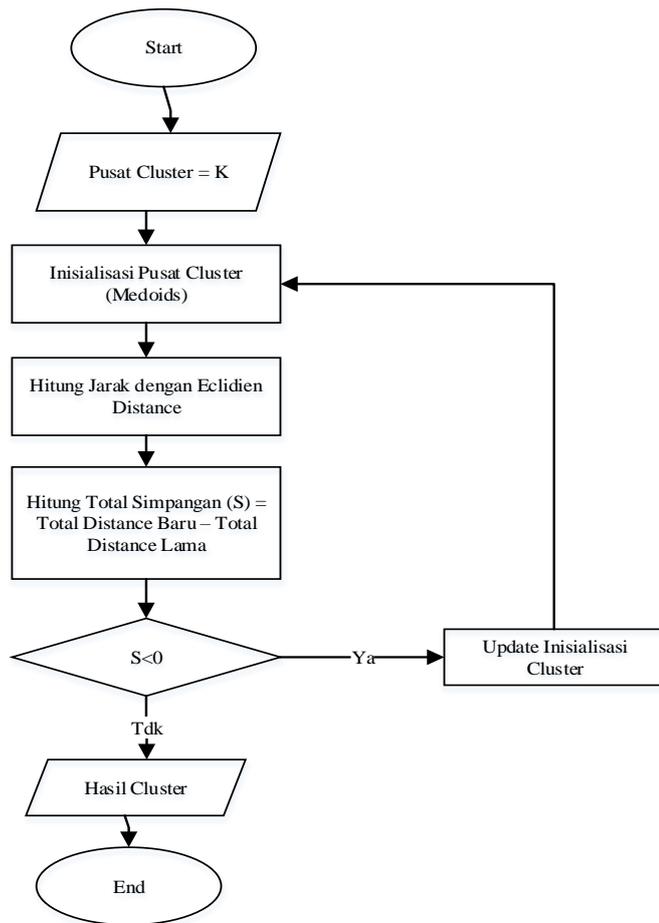
Tahapan algoritma *k-medoids* adalah sebagai berikut:

1. Inisialisasi pusat *cluster* sebanyak *k* (jumlah *cluster*)
2. Kumpulkan setiap data (objek) ke *cluster* terdekat dengan menggunakan rumus menghitung jarak *euclidian distance* berikut ini:

$$d(x,y) = \sqrt{\sum_{k=1}^n (x_i - x_i)^2}$$

3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoids* baru
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan *medoids* baru
5. Hitung total simpangan (*S*) dengan menghitung nilai total *distance* baru - total *distance* lama. jika $S < 0$, maka tukar objek dengan data *cluster* untuk menentukan sekumpulan *k* objek baru sebagai *medoids*
6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoids*, sehingga didapatkan *cluster* beserta anggota setiap *cluster*[8].

Gambar tahapan algoritma *k-medoids* dapat dilihat pada gambar 2.



Gambar. 2 Tahapan *K-Medoids* [8].

Tools yang digunakan untuk proses *k-medoids* menggunakan *Rapit miner* yang berupa aplikasi perangkat ilmu dengan pengembang perusahaan bernama sama dengan penyedia lingkungan integrasi dalam persiapan data pembelajaran mesin pembelajaran pengembangan teks dan animasi prediktif[9][10].

Sebelum diproses dalam *rapit miner*, dataset disimpan dalam bentuk *.xls* [9][10], selanjutnya data diimport kedalam *tools rapit miner* dan dibuat pemodelan *k-medoids* dengan jumlah *cluster* 2, 3 dan 5.

D. Menentukan Cluster Terbaik Berdasarkan Nilai DBI

Pada Tahap ini, Hasil *Cluster* yang diperoleh dari pengolahan menggunakan *tools rapit miner* selanjutnya dilihat bentuk *cluster* terbaik berdasarkan nilai DBI. Kelompok terbaik adalah kelompok yang memiliki nilai DBI terkecil.[9] [11].

Nilai *DBI* (*Davies Bouldin Index*) merupakan metode yang dapat digunakan untuk mengukur validitas jumlah *cluster* paling optimal, pendefinisian kohesi sebagai jumlah pendekatan data terhadap titik pusat *cluster* dari *cluster* yang diikuti[9][11][12].

E. Kesimpulan

Pada Tahap pengambilan kesimpulan dengan mengambil kesimpulan berdasarkan penelitian yang sudah dilakukan, kesimpulan *cluster* yang dihasilkan dan bentuk *cluster* terbaik berdasarkan nilai DBI.

III. HASIL DAN PEMBAHASAN

A. Data Penelitian

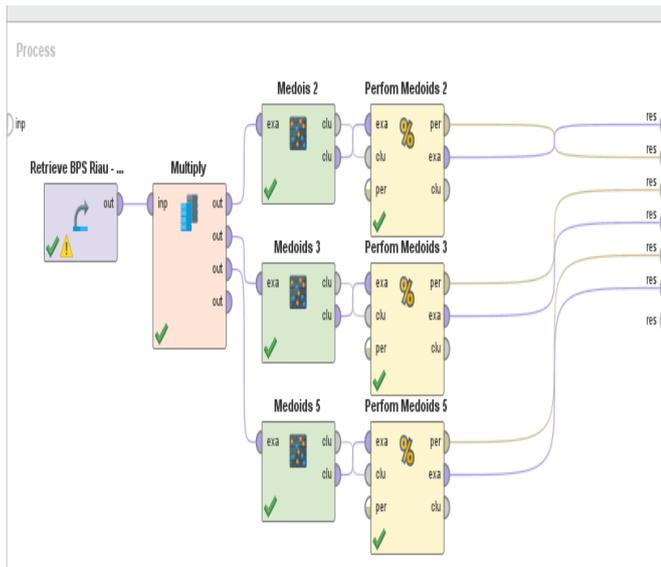
Data penelitian yang digunakan adalah data BPS provinsi Riau Terkait data Produksi Kelapa Sawit Tahun 2018-2019 [9], terlihat pada Tabel 1:

TABEL I
DATA PRODUKSI KELAPA SAWIT TAHUN 2018-2019

Kabupaten	2018	2019
Kuantan Singingi	452218	450804
Indragiri Hulu	286243	469273
Indragiri Hilir	733009	731009
Pelalawan	1339609	1339609
Siak	1193290	1098665
Kampar	1222465	955735
Rokan Hulu	1195460	1195460
Bengkalis	334066	334066
Rokan Hilir	813834	813832
Kepulauan Meranti	0	0
Pekanbaru	31219	36612
Dumai	82122	41195

B. Pemodelan Algoritma *K-medoids* Pada Rapit Miner

Bentuk pemodelan yang akan diuji pada penelitian ini adalah bentuk *k-medoids* dengan jumlah *cluster* 2, 3 dan 5 seperti terlihat pada gambar 3:



Gambar. 3 Pemodelan pengujian *k-medoids* dengan jumlah *cluster* 2, 3 dan 5

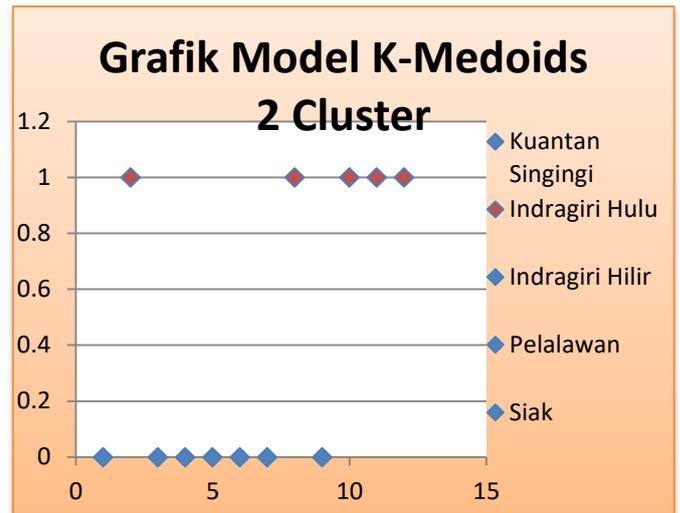
Gambar 3 menggambarkan bentuk pemodelan yang akan digunakan untuk menguji algoritma *k-medoids* yaitu pemodelan 2 *cluster*, 3 *cluster* dan 5 *cluster*. Sumber Data dihubungkan dengan fungsi *multiply* agar data dapat digunakan bersama dengan 3 model yang akan diuji, selanjutnya dihubungkan dengan setiap model *medoids* dan masing-masing model *medoids* dihubungkan dengan *cluster distance performance* untuk mengukur nilai DBI masing-masing model.

C. Hasil Pengujian K-Medoids Dengan Jumlah Cluster 2

Hasil pengujian model *k-medoids* jumlah *cluster* 2 terlihat pada Tabel 2 dan gambar 4 berikut:

TABEL II
HASIL PENGUJIAN K-MEDOIDS 2 CLUSTER

Kabupaten/Kota	2018	2019	Kluster
Kuantan Singingi	452218	450804	cluster_0
Indragiri Hulu	286243	469273	cluster_1
Indragiri Hilir	733009	731009	cluster_0
Pelalawan	1339609	1339609	cluster_0
Siak	1193290	1098665	cluster_0
Kampar	1222465	955735	cluster_0
Rokan Hulu	1195460	1195460	cluster_0
Bengkalis	334066	334066	cluster_1
Rokan Hilir	813834	813832	cluster_0
Kepulauan Meranti	0	0	cluster_1
Pekanbaru	31219	36612	cluster_1
Dumai	82122	41195	cluster_1



Gambar. 4 Grafik Model *k-medoids* 2 Cluster

Hasil pengujian berdasarkan tabel 2 dan gambar 4 terdapat 7 data pada *cluster* 0 dan 5 data pada *cluster* 1. Data yang masuk pada *cluster* 0 adalah Kuantan Singingi, Indragiri Hilir, Pelalawan, Siak, Kampar, Rokan Hulu dan Rokan Hilir. Data yang masuk pada *cluster* 1 adalah Indragiri Hulu, Bengkalis, Kepulauan Meranti, Pekanbaru dan Dumai.

D. Hasil Pengujian K-Medoids Dengan Jumlah Cluster 3

Hasil pengujian model *k-medoids* untuk jumlah *cluster* 3 terlihat pada Tabel 3 dan gambar 5 berikut:

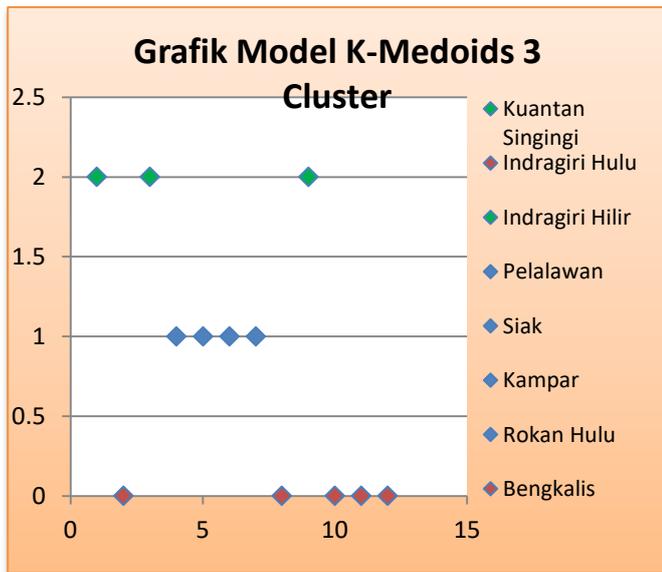
TABEL III
HASIL PENGUJIAN K-MEDOIDS 3 CLUSTER

Kabupaten/Kota	2018	2019	Kluster
Kuantan Singingi	452218	450804	cluster_2
Indragiri Hulu	286243	469273	cluster_0
Indragiri Hilir	733009	731009	cluster_2
Pelalawan	1339609	1339609	cluster_1
Siak	1193290	1098665	cluster_1
Kampar	1222465	955735	cluster_1
Rokan Hulu	1195460	1195460	cluster_1
Bengkalis	334066	334066	cluster_0
Rokan Hilir	813834	813832	cluster_2
Kepulauan Meranti	0	0	cluster_0
Pekanbaru	31219	36612	cluster_0
Dumai	82122	41195	cluster_0

Berdasarkan tabel 3 dan gambar 5 diperoleh hasil 5 data pada *cluster* 0, 4 data pada *cluster* 1 dan 3 data pada *cluster* 2. Daerah yang pada *cluster* 0 adalah Indragiri Hulu, Bengkalis, Kepulauan Meranti, Pekanbaru dan Dumai. Daerah yang masuk pada *cluster* 1 adalah Pelalawan, Siak, Kampar dan

Rokan Hulu. Daerah yang masuk pada *cluster 2* adalah Kuantan Singingi, Indragiri Hilir dan Rokan Hilir.

cluster 3 sedangkan Kepulauan Meranti, Pekanbaru dan Dumai masuk pada *cluster 4*.



Gambar. 5 Grafik Jumlah Cluster 3

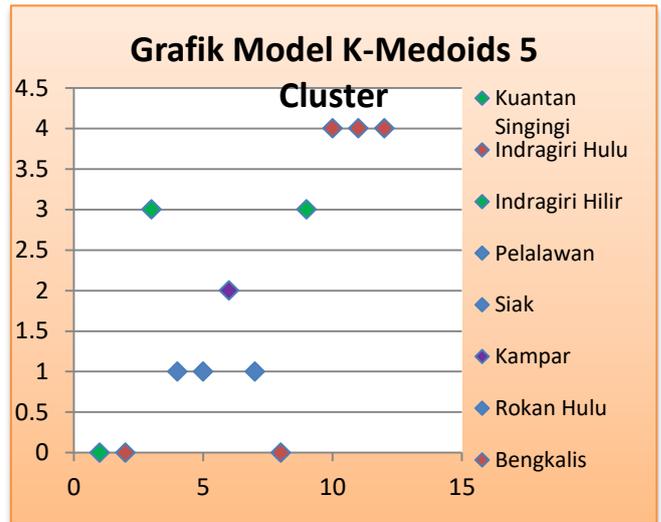
E. Hasil Pengujian K-Medoids Dengan Jumlah Cluster 5

Hasil pengujian model *k-medoids* untuk jumlah *cluster 3* terlihat pada Tabel 4 dan gambar 6 berikut:

TABEL IV
HASIL PENGUJIAN K-MEDOIS 5 CLUSTER

Kabupaten/Kota	2018	2019	Kluster
Kuantan Singingi	452218	450804	cluster_0
Indragiri Hulu	286243	469273	cluster_0
Indragiri Hilir	733009	731009	cluster_3
Pelalawan	1339609	1339609	cluster_1
Siak	1193290	1098665	cluster_1
Kampar	1222465	955735	cluster_2
Rokan Hulu	1195460	1195460	cluster_1
Bengkalis	334066	334066	cluster_0
Rokan Hilir	813834	813832	cluster_3
Kepulauan Meranti	0	0	cluster_4
Pekanbaru	31219	36612	cluster_4
Dumai	82122	41195	cluster_4

Berdasarkan tabel 4 dan gambar 6 diperoleh hasil terdapat 3 data pada *cluster 0*, 1 dan 4. Terdapat 1 data pada *cluster 2* dan 2 data pada *cluster 3*. Daerah Kuantan Singingi, Indragiri Hulu dan Bengkalis masuk pada *cluster 0*. Daerah Pelalawan, Siak dan Rokan Hulu masuk pada *cluster 1*. Kampar masuk pada *cluster 2*. Daerah Indragiri Hilir dan Rokan Hilir masuk pada



Gambar. 6 Grafik Jumlah Cluster 5

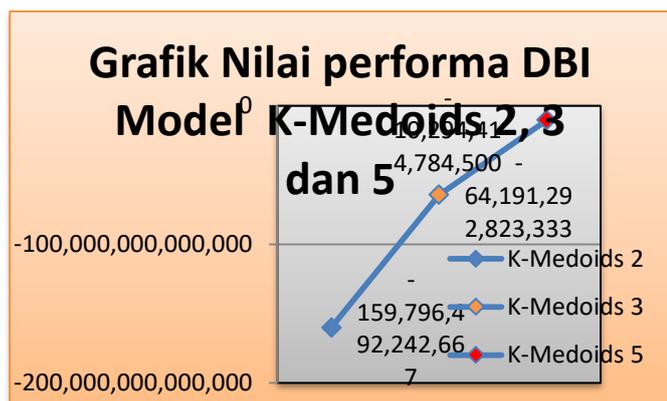
F. Hasil Pengukuran Performance Nilai DBI

Hasil pengukuran *performance* setiap model yang diuji berdasarkan Nilai DBI terlihat pada tabel 5 berikut:

TABEL V
HASIL PERFORMANCE NILAI DBI

Pengujian	Nilai performa
K-Medoids 2	-159.796.492.242.667
K-Medoids 3	-64.191.292.823.333
K-Medoids 5	-10.294.414.784.500

Berdasarkan nilai DBI yang terlihat pada tabel 5 diperoleh model *cluster* dengan performa terbaik yaitu model *k-medoids 2* dengan nilai DBI -159.796.492.242.667. *cluster* terbaik diperoleh dari model dengan nilai DBI terendah. Model terbaik kedua adalah *k-medoids 3* dengan nilai -64.191.292.823.333 dan model terbaik ke-tiga adalah *k-medoids 5* dengan nilai DBI -10.292.414.784.500, grafik nilai performa DBI terlihat pada gambar 7 berikut:



Gambar. 7 Grafik Nilai Performa DBI

IV. KESIMPULAN

Bersarkan hasil pengujian ketiga model *k-medoids* yaitu model *k-medoids 2*, *k-medoids 3* dan *k-medoids 5* diperoleh performa terbaik pada model *k-medoids 2* dengan nilai DBI -159.796.492.242.667. Dilanjutkan dengan model *k-medoids 3* dengan nilai DBI -64.191.292.823.333. Model *k-medoids 5* menjadi model dengan performa terendah dengan nilai DBI -10.294.414.784.500.

Daerah hasil pengelompokkan berdasarkan model terbaik yaitu dibagi menjadi 2 kelompok, dengan kelompok produksi tertinggi berada pada daerah Kuantan Singingi, Indragiri Hilir, Pelalawan, Siak, Kampar, Rokan Hulu dan Rokan Hilir. Daerah dengan Produksi Tendah yaitu Indragiri Hulu, Bengkalis, Kepulauan Meranti, pekanbaru dan Dumai.

REFERENSI

- [1] L. K. W. and B. Nurhayati, N. S. Sinatrya, "This research's goal is to find out the better performance algorithm between K-Means and K-Medoids algorithm. The performance of both algorithm are compared by testing data using Java-based application, Hadoop, and Hive. Comparison was conducted in terms," *2018 6th Int. Conf. Cyber IT Serv. Manag.*, doi: 10.1109/CITSM.2018.8674251.
- [2] A. upi Fitriyadi, "Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional," *Kilat*, vol. 10, no. 1, pp. 157–168, 2021, doi: 10.33322/kilat.v10i1.1174.
- [3] L. Agustini and I. O. Kirana, "Pengelompokan Data Janjang Panen Kelapa Sawit Menggunakan Algoritma K-Medoids Pada PT SIR MANDAU," *J. Mach. Learn. Data Anal.*, vol. 01, no. 01, pp. 36–44, 2022.
- [4] I. M. Pulungan, S. Saifullah, M. Fauzan, and A. P. Windarto, "Implementasi Algoritma K-Means Clustering dalam Menentukan Blok Tanaman Sawit Paling Produktif," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 338, 2019, doi: 10.30645/senaris.v1i0.39.
- [5] F. Nasari, "ALGORITMA K-MEANS CLUSTERINGDALAM PENYEBARAN PENYAKIT DIARE".
- [6] A. Nofiar, S. Defit, and Sumijan, "Penentuan Mutu Kelapa Sawit Menggunakan Metode K-Means Clustering," *J. KomtekInfo*, vol. 5, no. 3, pp. 1–9, 2019, doi: 10.35134/komtekinf.v5i3.26.
- [7] R. Gustrianda and D. I. Mulyana, "Penerapan Data Mining Dalam Pemilihan Produk Unggulan dengan Metode Algoritma K-Means Dan K-Medoids," *J. Media Inform. Budidarma*, vol. 6, no. 1, p. 27,

- 2022, doi: 10.30865/mib.v6i1.3294.
- [8] D. Marlina, N. Lina, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 4, no. 2, p. 64, 2018, doi: 10.24014/coreit.v4i2.4498.
- [9] F. Nasari and U. P. Utama, "Pengelompokan Daerah Produksi Kelapa Sawit Menggunakan Algoritma K-Means Clustering," vol. 4, no. 2, pp. 34–38, 2021.
- [10] D. F. Pasaribu, I. S. Damanik, E. Irawan, Suhada, and H. S. Tambunan, "Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 11–20, 2021, doi: 10.37148/bios.v2i1.17.
- [11] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021, [Online]. Available: www.bps.go.id
- [12] A. Bates and J. Kalita, "Counting Clusters in Twitter Posts," in *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*, 2016, doi: 10.1145/2905055.2905295.