

Optimasi Metode K-Means dan K-Medoids Berdasarkan Jumlah Cluster dan Nilai DBI Dalam Pengelompokan Produksi Kelapa Sawit Di Provinsi Riau

Optimization of K-Means and K-Medoids Methods Based on Number of Clusters and DBI Value in Grouping Oil Palm Production in Kampar Regency

Fina Nasari^{a,1}, Dahriani Hakim Tanjung^{b,2}, Fitri Handayani^{c,3}

^{a,c}Teknik Informatika, Politeknik Kampar, Bangkinang, Indonesia

^bSistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia
finanasari@poltek-kampar.ac.id¹, notashapire@gmail.com², fitrihandayani.fh888@gmail.com³

ABSTRAK

Kelapa sawit merupakan salah satu jenis sumber daya alam yang terkenal didunia. Perkebunan kelapa sawit terbesar didunia berada di Indonesia. Provinsi riau menjadi salah satu provinsi dengan produksi kelapa sawit tertinggi di Indonesia dengan luas kebun 2.8 juta Ha produksi 8.8 juta ton per tahun. Penyebaran kelapa sawit di provinsi riau hampir diseluruh kabupaten, sehingga perlu adanya pengelompokan daerah produksi kelapa sawit. Clustering menjadi salah satu metode yang dapat mengelompokkan data pada data yang sejenis. Proses clustering dapat menggunakan metode K-means atau K-medoids. Perlu adanya pengujian untuk melihat metode yang lebih optimal dalam proses cluster berdasarkan jumlah cluster dan nilai DBI untuk mendapatkan hasil pengelompokan daerah produksi kelapa sawit terbaik. Pengujian menggunakan Tools Rapi Miner. Jumlah Cluster yang digunakan dalam pengujian ini adalah 2, 3 dan 5. Hasil penelitian ini menunjukkan Jumlah Cluster 2 menjadi cluster terbaik dengan nilai DBI untuk metode K-medoid -159796492242,667 dan metode K-Means -82338884292,014. Metode K-Medoid menjadi metode cluster terbaik dengan cluster yang dihasilkan berupa 7 kabupaten pada kelompok jumlah produksi Tinggi dan 5 kabupaten masuk pada kelompok jumlah produksi Rendah.

Kata Kunci : Clustering, K-Means, K-Medoids, Nilai DBI, Kelapa Sawit

ABSTRACT

Oil palm is one of the most famous natural resources in the world. The world's largest oil palm plantations are in Indonesia. Riau Province is one of the provinces with the highest palm oil production in Indonesia with a plantation area of 2.8 million hectares producing 8.8 million ton per year. The spread of oil palm in Riau province is almost in all districts, so it is necessary to group the areas of oil palm production. Clustering is a method that can group data into similar data. The clustering process can use the K-means or K-medoids method. There needs to be a test to see which method is more optimal in the cluster process based on the number of clusters and the DBI value to get the best grouping results for oil palm production areas. Testing using Rapid Miner Tools. The number of clusters used in this test were 2, 3 and 5. The results of this study showed that the number of clusters 2 was the best cluster with DBI values for the K-medoids method -159796492242.667 and the K-Means method -82338884292.014. The K-Medoid method is the best cluster method with the resulting clusters in the form of 7 districts in the High production number group and 5 districts in the Low production number group.

Keywords : Clustering, K-Means, K-Medoids, DBI Value, Palm Oil

Disubmit: 22 Desember 2022

Info Artikel :
Direview: 10 Januari 2023

Diterima: 30 Januari 2023

Copyright © 2023 – InfoSys Journal. All rights reserved.

1. PENDAHULUAN

Negara penghasil kelapa sawit terbesar didunia adalah indonesia hal itu menjadikan kelapa sawit sebagai komoditas ekspor utama indonesia [1]. Berdasarkan data statistik kelapa sawit indonesia 2021 dari Badan Pusat Statistik, provinsi riau menjadi daerah dengan produksi CPO(*Crude Palm Oil*) tertinggi di Indonesia yaitu 20% dilanjutkan dengan Kalimantan Tengah 16%, kalimantan barat dan sumatera utara 12%, kalimantan timur 8% dan provinsi lainnya 32%. Jika dilihat dari luas area perkebunan kelapa sawit indonesia tahun 2021, provinsi Riau menjadi provinsi yang memiliki luas kebun kelapa sawit terbesar yaitu 2.86 juta Ha dilanjutkan dengan Kalimantan Barat 2.02 juta Ha, Kalimantan Tengah 1.82 juta Ha, Kalimantan Timur 1.31 juta Ha, Sumatera Utara 1.25 juta Ha, Sumatera Selatan 1.12 juta Ha dan Jambi 1.08 juta Ha. Berdasarkan data Direktorat Perusahaan Perkebunan Kelapa Sawit Provinsi Riau tahun 2020 terdapat 219 Perusahaan Besar Swasta, 19 Perusahaan Besar Negara yang tersebar di hampir seluruh kabupaten di provinsi riau.

Berdasarkan data *USDA* tahun 2022 terjadi penurunan produksi kelapa sawit, sehingga perlu adanya pengelompokan produksi kelapa sawit berdasarkan lokasi perkebunan kelapa sawit. Pengelompokan dilakukan dengan *clustering* dengan mengkombinasikan 2 metode yaitu *K-Medoid* dan *k-Means*. Hasil penelitian ini menjelaskan bahwa kombinasi kedua algoritma ini berfungsi dengan sangat baik, terdapat 72% produksi kelapa sawit dalam *Cluster* rendah, 20% dalam *cluster* sedang dan 2% dalam *cluster* tinggi[2]. *K-means* menggunakan pendekatan *partitioning* dengan menggunakan *centroid* dalam membuat *cluster*. Nilai *centroid* untuk awal iterasi dilakukan secara acak sedangkan untuk iterasi selanjutnya berdasarkan nilai rata-rata *cluster* pada iterasi tersebut[1]. *K-means* mampu mengelompokkan objek besar dan kecil dengan cepat namun proses penentuan nilai *centroid* awal yang dilakukan secara random memungkinkan ada *cluster* yang tidak memiliki anggota[3]. *K-medoids* menggunakan objek pada kumpulan objek dalam mewakili *cluster* dan cukup efisien untuk data set yang kecil. *K-medoid* lebih baik dari *k-means* karena *K-medoid* menemukan *k* sebagai objek yang representatif dalam meminimalkan jumlah ketidaksesuaian objek sedangkan *k-means* menggunakan jumlah jarak *euclidean distances* pada objek data[4].

Algoritma *K-means Clustering* mampu mengelompokkan produksi perkebunan kelapa sawit menjadi 3 *cluster* rendah, sedang dan tinggi. Terdapat 14 data masuk pada *cluster* rendah, 4 data masuk pada *cluster* sedang dan 3 data pada *cluster* tinggi. Pada *cluster* tinggi provinsi sumatera utara, riau dan kalimantan tengah dari 21 provinsi yang diuji[1]. Hasil analisis algoritma *k-means* menggunakan *Microsoft Excel* dan *tools rapid miner* memiliki nilai yang sama dalam pengelompokan data menjadi 3 *Cluster* yaitu Tinggi, Sedang dan rendah dengan 19 Blok pada *Cluster* Tinggi, 19 Blok *Cluster* Sedang dan 12 Blok *cluster* Rendah[5]. Algoritma *k-means Clustering* mampu membentuk 2 *cluster* yaitu tinggi dan rendah, terdapat 14 Blok Panen pada *cluster* tinggi dan 11 Blok panen pada *cluster* rendah[6].

Algoritma *K-medoids* mampu mengelompokkan data menjadi 2, 3 dan 5 *cluster*. Hasil *cluster* terbaik diperoleh pada nilai *DBI* terkecil untuk variabel luas tanaman dan produksi kelapa sawit menggunakan *K-medoid*. Hasil evaluasi *DBI* terbaik pada *Cluster* 2 dengan nilai *DBI* 0,099 terdapat 4 kabupaten masuk pada *cluster* tinggi, 29 kabupaten masuk pada *cluster* sedang[7].

Penelitian berfokus pada perbandingan *performance k-means* dan *k-medoids* yang dilihat berdasarkan nilai *DBI* serta menguji performa jumlah *cluster* 2, 3 dan 5 untuk metode *k-means* dan *k-medoids*. Performa terbaik *cluster* dilihat dari nilai *DBI* terkecil sehingga diperoleh data kelompok yang terbaik dari pengujian yang dilakukan.

2. METODE

Tahapan penelitian yang dilakukan dimulai dari pengumpulan data primer dan sekunder, dilanjutkan dengan membuat pemodelan untuk *k-means* dan *k-medoids*, setelah itu proses berlanjut dengan pengujian menggunakan *tools rapid miner* dan tahapan terakhir yang dilakukan adalah menganalisis hasil, *flowchart* tahapan penelitian terlihat pada gambar 1 berikut ini:



Gambar 1. Tahapan Penelitian

A. Analisis Masalah

Pada tahap ini dilakukan analisis permasalahan terkait dengan penyebaran daerah terhadap produksi kelapa sawit, untuk melihat daerah-daerah mana yang memiliki produksi kelapa sawit tinggi dan daerah mana yang memiliki produksi kelapa sawit rendah. Pengelompokan data dilakukan dengan teknik *clustering* dan untuk mendapatkan hasil *cluster* terbaik dilakukan dengan membandingkan 2 algoritma yaitu *k-means* dan *k-medoids*.

B. Pengumpulan Data

Data primer yang digunakan dalam penelitian ini adalah data produksi perkebunan provinsi riau tahun 2018-2019 yang diambil dari data website BPS riau melalui link: <https://riaubps.go.id/indicator/54/220/1/produksi-perkebunan.html>. Data produksi kelapa sawit yang digunakan berdasarkan data produksi kelapa sawit pada 12 kabupaten di provinsi riau. Pada tahap ini juga dilakukan pemilihan variable yang digunakan yaitu data produksi kelapa sawit untuk tahun 2018-2019.

Data sekunder dalam penelitian ini berupa literature review dari penelitian sebelumnya yang berkaitan dengan metode *k-means*, *k-medoid*, nilai *DBI*, produksi kelapa sawit hingga penggunaan *tools rapid miner*.

C. Pembuatan Model K-Means dan K-Medoids

1) Algoritma K-Means

Algoritma *k-means* adalah algoritma yang mengelompokkan data pada beberapa kelompok, data pada setiap kelompok memiliki karakteristik yang berbeda dengan kelompok lain[8]. Tahapan algoritma *k-means* adalah sebagai berikut:

- Menentukan jumlah cluster. Jumlah *cluster* yang akan diujicoba dalam penelitian ini adalah jumlah *cluster* 2, jumlah *cluster* 3 dan jumlah *cluster* 5.

- Membangkitkan *centroid* awal. untuk algoritma *k-means clustering* *centroid* awal dipilih secara acak sebanyak jumlah *cluster* yang digunakan.
- Menghitung jarak kemasing-masing *centorid* dengan menggunakan formula sebagai berikut:

$$d_{euclidean}(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Dimana:

X dan Y adalah nilai atribut dari 2 *record* [9].

- Pengelompokkan data *cluster*
- Apabila data cluster dihasilkan tidak sesuai dengan kelompok sebelumnya, maka proses kembali menghitung jarak ke masing-masing *centroid*, jika kelompok yang dihasilkan sudah sesuai maka proses berhenti. *Centroid* yang digunakan pada tahap selanjutnya adalah *centoroid* baru yang diperoleh dengan menggunakan *formula* berikut ini:

$$C = \frac{\sum m}{n}$$

Dimana:

C : *Centroid* data

m : anggota data yang termasuk kedalam *centorid* tertentu

n : jumlah data yang menjadi anggota *centroid* tertentu[9].

2) Algoritma K-Medoids

Algoritma *K-Medoid* menjadikan objek representatif sebagai titik acuan, bukan nilai rata-rata objek pada setiap *cluster*. Parameter input k dan jumlah *cluster* yang akan dipartisikan antara satu set n objek. Pengujian Menggunakan *Rapit Miner*[7].

- Inisialisasi *centroid cluster* sebanyak jumlah *cluster*
- Setiap data dikelompokkan pada *cluster* terdekat dengan menggunakan formula *ecludiece distance*
- Memilih acak sebuah objek pada setiap *cluster* sebagai *Medoids* baru
- Menghitung jarak pada setiap objek ke masing-masing *cluster* dengan nilai *medoid* baru
- Menghitung total simpangan(s) dengan nilai total *distance* baru hingga total *distance* lama. Jika $S < 0$, maka i tuka objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai *medoids*.
- Mengulangi step 3-5 hingga tidak terjadi perubahan *medoids*, sehingga diperoleh *cluster* setiap anggota[2][10].

D. Pengujian Menggunakan Rapit Miner

Rapit miner merupakan *tools* yang dapat digunakan untuk menganalisis *data mining*, *text mining* dan analisis *prediksi* yang bersifat *open sources*[1]. Penggunaan *rapit miner* memudahkan *user* dalam melakukan perhitungan data yang sangat banyak menggunakan operator-operator yang memiliki fungsi dalam meodifikasi data. Data dihubungkan pada *node* operator selanjutnya pada *node output* agar hasil dapat ditampilkan[4][11].

Pada tahap ini data set disimpan dalam bentuk *.xlsx* kemudian *diimport* ke dalam *rapit miner*. Data set tersebut selanjutnya dianalisis dengan menggunakan algoritma *k-means* dan *k-medoid* dengan melakukan pengujian dengan menggunakan jumlah *cluster* 2, 3 dan 5 dengan melihat performa nilai DBI setiap *algoritma* dan *cluster* yang digunakan.

E. Analisis Hasil

Pada tahap ini dilakukan analisis hasil pengujian *rapit miner*. Melihat hasil yang paling optimal dari *k-means* dan *k-medoid* dan jumlah cluster yang digunakan berdasarkan nilai *DBI* terkecil, semakin kecil nilai *DBI* maka semakin baik hasil pengklasterannya[7].

F. Kesimpulan

Tahap ini mengambil kesimpulan berdasarkan analisis hasil berupa algoritma dan jumlah *cluster* yang paling optimal dalam pengelompokkan data produksi kelapa sawit provinsi riau.

3. HASIL DAN PEMBAHASAN

A. Data Set

Penelitian ini menggunakan data BPS provinsi riau untuk data produksi perkebunan tahun 2018-2019 terlihat pada Tabel 1:

Tabel 1. Data Produksi Perkebunana Tahun 2018-2019

KabKota	Produksi Perkebunan (Ton)									
	Kelapa Sawit		Kelapa		Karet		Kopi		Kakao	
	2018	2019	2018	2019	2018	2019	2018	2019	2018	2019
Riau	7683535.00	7466260.00	392702.00	417172.00	373749.00	373726.00	3030.00	3019.00	2907.00	1602.00
Kuantan Singingi	452218.00	450804.00	1924.00	1924.00	83680.00	83983.00	-	5.00	-	660.00
Indragiri Hulu	286243.00	469273.00	-	250.00	49275.00	32306.00	-	44.00	-	117.00
Indragiri Hilir	733009.00	731009.00	335215.00	361348.00	4616.00	4616.00	-	242.00	1637.00	439.00
Pelalawan	1339609.00	1339609.00	15297.00	15297.00	43860.00	43860.00	-	656.00	-	83.00
Siak	1193290.00	1098665.00	1743.00	333.00	14249.00	31093.00	-	13.00	-	27.00
Kampar	1222465.00	955735.00	-	418.00	64857.00	65188.00	-	5.00	-	43.00
Rokan Hulu	1195460.00	1195460.00	-	474.00	54420.00	54420.00	-	95.00	-	110.00
Bengkalis	334066.00	334066.00	3262.00	3273.00	22453.00	22447.00	-	46.00	-	0.00
Rokan Hilir	813834.00	813832.00	4226.00	4227.00	22160.00	22161.00	-	1.00	-	114.00
Kepulauan Meranti	-	0.00	28781.00	29183.00	11941.00	12287.00	1883.00	1912.00	-	-
Pekanbaru	31219.00	36612.00	-	9.00	-	448.00	0.00	0.00	-	3.00
Dumai	82122.00	41195.00	-	436.00	1800.00	917.00	0.00	0.00	-	6.00

Sumber: <https://riau.bps.go.id/indicator/54/220/1/produksi-perkebunan.html>.

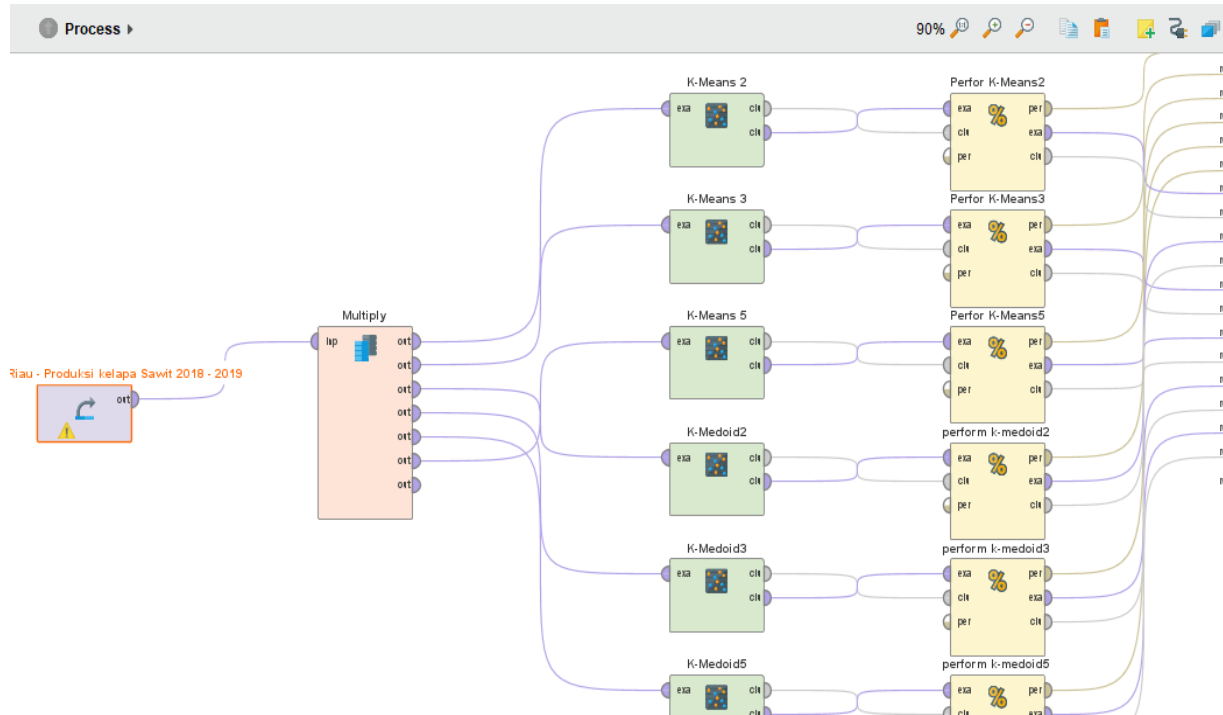
Fokus penelitian yang dilakukan adalah pengelompokkan produksi kelapa sawit tahun 2018 dan 2019. Terdapat 12 kabupaten pada provinsi riau, dimana hampir seluruh kabupaten pada provinsi riau memiliki perkebunan sawit dan telah memproduksi kelapa sawit. Data *selection* yang digunakan terlihat pada Tabel.2 :

Tabel 2. Data Produksi Kelapa Sawit Provinsi Riau 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

Tahap selanjutnya proses pengujian menggunakan *rapid miner* dengan pemodelan seperti yang terlihat pada gambar 2 berikut ini:



Gambar. 2 Pemodelan pengujian *k-means* dan *k-medoids*

Pemodelan yang digunakan dalam proses pengujian seperti yang terlihat pada gambar 2 dimana data set yang digunakan dihubungkan dengan *multiply* agar data yang sama dapat digunakan untuk pengujian dengan model yang berbeda-beda. Model yang diujikan ada 6 model yaitu *algoritma k-means* dengan jumlah *cluster* 2, 3 dan 5 dan *algoritma k-medoids* dengan jumlah *cluster* 2,3 dan 5.

B. Pengujian K-Means

Pengujian algoritma *k-means* yang dilakukan dengan menggunakan 3 model, yaitu model 2 *cluster*, 3 *cluster* dan 5 *cluster*. Berdasarkan 3 model tersebut akan dilihat hasil pengelompokkan untuk setiap model.

a) Pengujian K-means untuk jumlah cluster 2

Berdasarkan pengujian dengan pemodelan yang terlihat pada gambar 2 diperoleh hasil pengujian *algoritma k-means* dengan jumlah *cluster* 2 seperti yang terlihat pada table 3 berikut:

Tabel 3. Hasil Pengujian Algoritma K-means jumlah cluster 2

No.	Kabupaten/Kota	Kluster	2018	2019
1	Kuantan Singingi	cluster_1	452218.0	450804.0
2	Indragiri Hulu	cluster_1	286243.0	469273.0
3	Indragiri Hilir	cluster_0	733009.0	731009.0
4	Pelalawan	cluster_0	1339609.0	1339609.0
5	Siak	cluster_0	1193290.0	1098665.0
6	Kampar	cluster_0	1222465.0	955735.0
7	Rokan Hulu	cluster_0	1195460.0	1195460.0
8	Bengkalis	cluster_1	334066.0	334066.0
9	Rokan Hilir	cluster_0	813834.0	813832.0
10	Kepulauan Meranti	cluster_1	0.0	0.0
11	Pekanbaru	cluster_1	31219.0	36612.0
12	Dumai	cluster_1	82122.0	41195.0

Berdasarkan hasil pengujian seperti yang terlihat pada table. 3 diperoleh hasil, terdapat 6 items pada *cluster* 0 dan 6 items pada *cluster* 1. Dimana *cluster* 0 mengidentifikasi *cluster* dengan jumlah produksi tinggi sedangkan *cluster* 1 mengidentifikasi jumlah produksi rendah.

- b) Pengujian *K-means* untuk jumlah *cluster* 3
Hasil pengujian pemodelan *algoritma k-means* dengan jumlah *cluster* 3 sesuai yang terlihat pada pemodelan gambar 2 dapat dilihat pada table. 4 berikut:

Tabel 4. Hasil Pengujian *Algoritma K-means* jumlah *cluster* 3

No.	Kabupaten/Kota	Kluster	2018	2019
1	Kuantan Singingi	cluster_0	452218.0	450804.0
2	Indragiri Hulu	cluster_1	286243.0	469273.0
3	Indragiri Hilir	cluster_0	733009.0	731009.0
4	Pelalawan	cluster_2	1339609.0	1339609.0
5	Siak	cluster_2	1193290.0	1098665.0
6	Kampar	cluster_2	1222465.0	955735.0
7	Rokan Hulu	cluster_2	1195460.0	1195460.0
8	Bengkalis	cluster_1	334066.0	334066.0
9	Rokan Hilir	cluster_0	813834.0	813832.0
10	Kepulauan Meranti	cluster_1	0.0	0.0
11	Pekanbaru	cluster_1	31219.0	36612.0
12	Dumai	cluster_1	82122.0	41195.0

Berdasarkan hasil pengujian seperti yang terlihat pada Tabel 4. Diperoleh hasil terdapat 3 items pada *cluster* 0, 5 Items pada *cluster* 1 dan 4 items pada *cluster* 2. *Cluster* 0 mengidentifikasi jumlah produksi sedang, *cluster* 1 mengidentifikasi jumlah produksi rendah dan *cluster* 2 mengidentifikasi jumlah produksi tinggi.

- c) Pengujian *K-means* untuk jumlah *cluster* 5

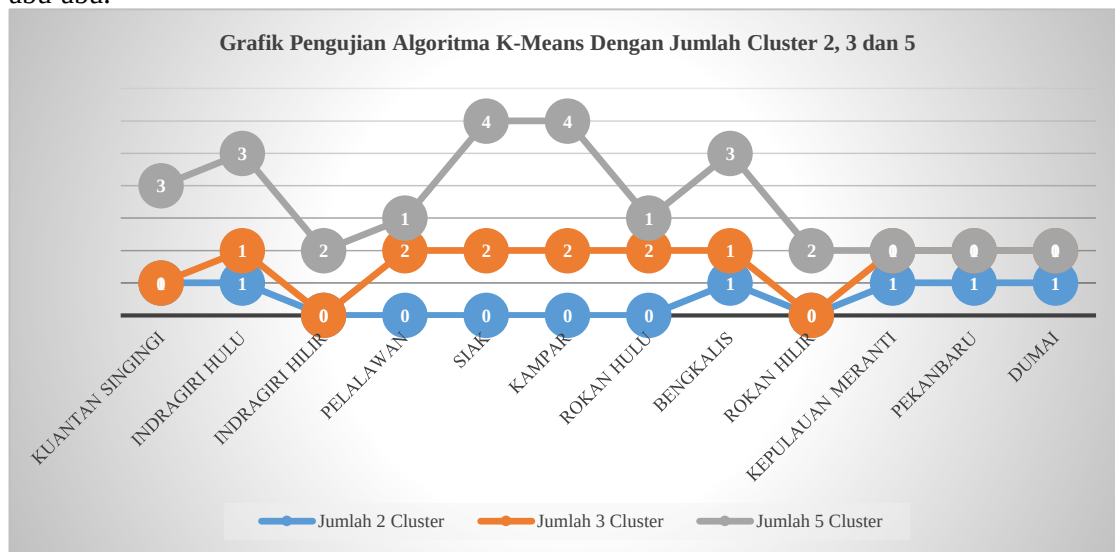
Pengujian pemodelan *algoritma k-means* dengan jumlah *cluster* 5 dapat sesuai dengan pemodelan pada gambar 2, hasil pengujian dapat dilihat pada table. 5 berikut:

Tabel 5. Hasil Pengujian *Algoritma K-means* jumlah *cluster* 5

No.	Kabupaten/Kota	Kluster	2018	2019
1	Kuantan Singingi	cluster_3	452218.0	450804.0
2	Indragiri Hulu	cluster_3	286243.0	469273.0
3	Indragiri Hilir	cluster_2	733009.0	731009.0
4	Pelalawan	cluster_1	1339609.0	1339609.0
5	Siak	cluster_4	1193290.0	1098665.0
6	Kampar	cluster_4	1222465.0	955735.0
7	Rokan Hulu	cluster_1	1195460.0	1195460.0
8	Bengkalis	cluster_3	334066.0	334066.0
9	Rokan Hilir	cluster_2	813834.0	813832.0
10	Kepulauan Meranti	cluster_0	0.0	0.0
11	Pekanbaru	cluster_0	31219.0	36612.0
12	Dumai	cluster_0	82122.0	41195.0

Berdasarkan tabel 5 hasil pengujian *k-means* untuk jumlah 5 *cluster* terdapat 3 items pada *cluster* 0, 2 items pada *cluster* 1, 2 dan 4, 3 items pada *cluster* 3. *Cluster* 0 mengidentifikasi jumlah produksi rendah, *cluster* 1 mengidentifikasi jumlah produksi yang tinggi, *cluster* 2, 3 mengidentifikasi jumlah produksi menengah dan *cluster* 4 mengidentifikasi jumlah produksi sangat tinggi.

Hasil pengujian untuk setiap model *k-means* terlihat pada gambar 3 untuk jumlah *cluster* 2 terlihat garis biru, untuk *cluster* 3 terlihat pada grafik orange dan *cluster* 5 terlihat pada grafik abu-abu.



Gambar. 3 Grafik hasil pengujian pada setiap model pengujian *k-means*

C. Pengujian *K-Medoids*

Berdasarkan gambar 2, pemodelan yang digunakan untuk pengujian *algoritma k-medoids* menggunakan 3 jenis *cluster* yaitu dengan model 2 *cluster*, 3 *cluster*, 5 *cluster*.

a) Pengujian *K-medoid* dengan jumlah *cluster* 2

Hasil pengujian *k-medoids* dengan jumlah *cluster* 2 dapat dilihat pada tabel 6 berikut:

Tabel 6. Hasil Pengujian *Algoritma K-medoids* jumlah *cluster* 2

No.	Kabupaten/Kota	Kluster	2018	2019
1	Kuantan Singingi	cluster_0	452218.0	450804.0
2	Indragiri Hulu	cluster_1	286243.0	469273.0
3	Indragiri Hilir	cluster_0	733009.0	731009.0
4	Pelalawan	cluster_0	1339609.0	1339609.0
5	Siak	cluster_0	1193290.0	1098665.0
6	Kampar	cluster_0	1222465.0	955735.0
7	Rokan Hulu	cluster_0	1195460.0	1195460.0
8	Bengkalis	cluster_1	334066.0	334066.0
9	Rokan Hilir	cluster_0	813834.0	813832.0
10	Kepulauan Meranti	cluster_1	0.0	0.0
11	Pekanbaru	cluster_1	31219.0	36612.0
12	Dumai	cluster_1	82122.0	41195.0

Berdasarkan tabel 5 diperoleh hasil terdapat 7 data *cluster* 0 dan 5 data *cluster* 1. *Cluster* 0 mengidentifikasi jumlah produksi Tinggi dan *cluster* 1 mengidentifikasi jumlah produksi Rendah.

b) Pengujian *K-medoid* dengan jumlah *cluster* 3

Hasil pengujian *k-medoids* dengan jumlah *cluster* 3 terlihat pada tabel 7 berikut:

Tabel 7. Hasil Pengujian *Algoritma K-medoids* jumlah *cluster* 3

No.	Kabupaten/Kota	Kluster	2018	2019
1	Kuantan Singingi	cluster_2	452218.0	450804.0
2	Indragiri Hulu	cluster_0	286243.0	469273.0
3	Indragiri Hilir	cluster_2	733009.0	731009.0
4	Pelalawan	cluster_1	1339609.0	1339609.0
5	Siak	cluster_1	1193290.0	1098665.0
6	Kampar	cluster_1	1222465.0	955735.0
7	Rokan Hulu	cluster_1	1195460.0	1195460.0
8	Bengkalis	cluster_0	334066.0	334066.0
9	Rokan Hilir	cluster_2	813834.0	813832.0

10	Kepulauan Meranti	cluster_0	0.0	0.0
11	Pekanbaru	cluster_0	31219.0	36612.0
12	Dumai	cluster_0	82122.0	41195.0

Berdasarkan tabel. 7 hasil pengujian *k-medoid* dengan jumlah *cluster* 3 dimana terdapat 5 items pada *cluster* 0, 4 items pada *cluster* 1 dan 3 items pada *cluster* 2. *Cluster* 0 mengidentifikasi jumlah produksi rendah, *cluster* 1 mengidentifikasi jumlah produksi tinggi dan *cluster* 2 mengidentifikasi jumlah produksi sedang.

c) Pengujian *K-medoid* dengan jumlah *cluster* 5

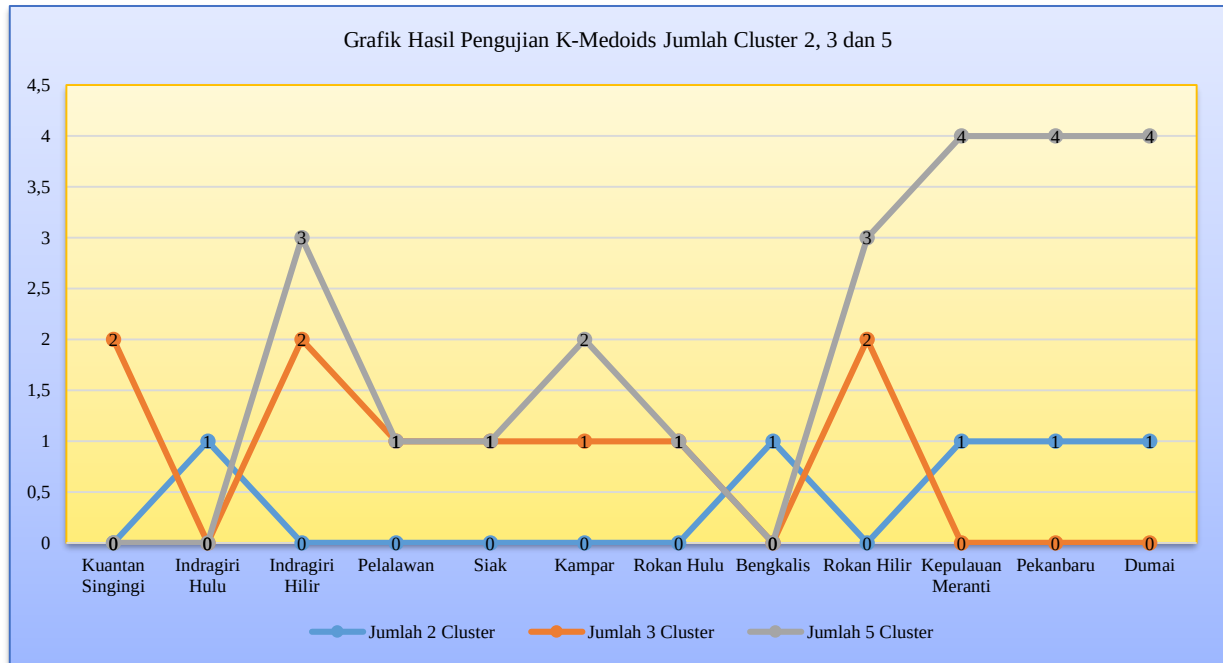
Hasil pengujian *k-medoids* dengan menggunakan jumlah *cluster* 5 dapat dilihat pada tabel 7 berikut:

Tabel 8. Hasil Pengujian *Algoritma K-medoids* jumlah *cluster* 5

No.	Kabupaten/Kota	Kluster	2018	2019
1	Kuantan Singingi	cluster_0	452218.0	450804.0
2	Indragiri Hulu	cluster_0	286243.0	469273.0
3	Indragiri Hilir	cluster_3	733009.0	731009.0
4	Pelalawan	cluster_1	1339609.0	1339609.0
5	Siak	cluster_1	1193290.0	1098665.0
6	Kampar	cluster_2	1222465.0	955735.0
7	Rokan Hulu	cluster_1	1195460.0	1195460.0
8	Bengkalis	cluster_0	334066.0	334066.0
9	Rokan Hilir	cluster_3	813834.0	813832.0
10	Kepulauan Meranti	cluster_4	0.0	0.0
11	Pekanbaru	cluster_4	31219.0	36612.0
12	Dumai	cluster_4	82122.0	41195.0

Berdasarkan tabel 8 diperoleh hasil terdapat 3 items pada *cluster* 0, 1 dan 4, 1 items pada *cluster* 2 dan 2 items pada *cluster* 3.

Hasil pengujian *k-medoids* untuk setiap model terlihat pada gambar 4, dimana untuk jumlah *cluster* 2 terlihat pada garis biru, untuk jumlah *cluster* 3 terlihat pada garis orange dan untuk jumlah *cluster* 5 terlihat pada garis abu-abu.



Gambar. 4 Grafik hasil pengujian pada setiap model pengujian *k-medoids*

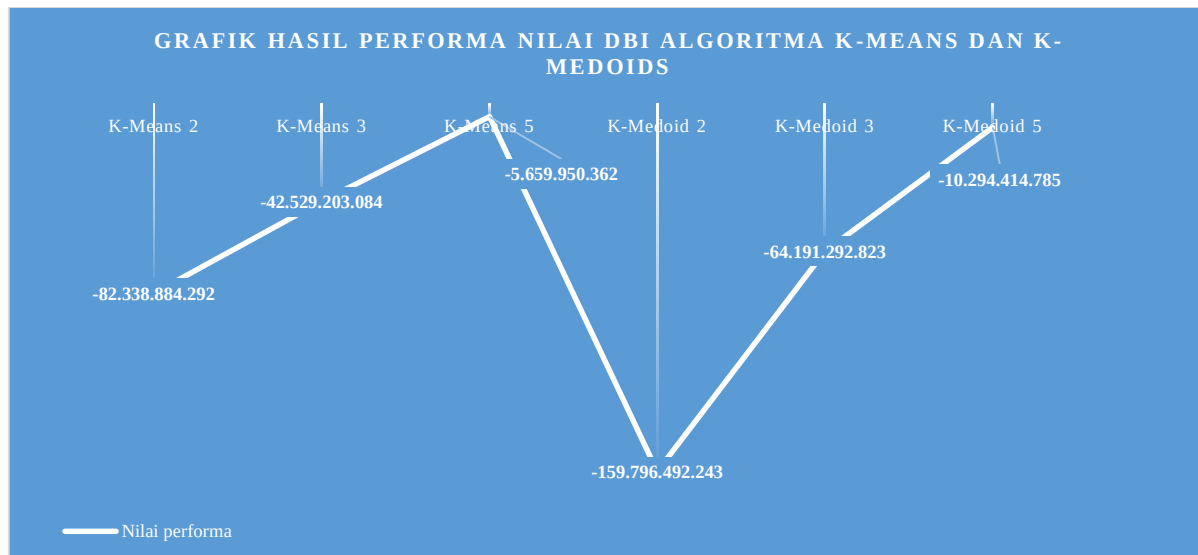
D. Analisis Nilai DBI

Pengujian *k-means* dan *k-medoid* dengan jumlah *cluster* 2, 3 dan 5 dianalisis dengan hasil performa nilai *DBI* untuk masing-masing *cluster*. Pengukuran nilai *DBI* menggunakan *tools rapid miner*, dengan hasil nilai *DBI* pada setiap *cluster* seperti terlihat pada tabel 9 berikut:

Tabel 9. Nilai DBI Pengujian *K-Means* dan *K-Medoids*

Pengujian	Nilai performa
K-Means 2	-82.338.884.292,014
K-Means 3	-42.529.203.084,185
K-Means 5	-5.659.950.362,375
K-Medoid 2	-159.796.492.242,667
K-Medoid 3	-64.191.292.823,333
K-Medoid 5	-10.294.414.784,500
Nilai DBI Terendah	-159.796.492.242,667

Grafik hasil performa nilai *DBI* terlihat pada gambar 5 berikut ini:



Gambar. 5 Grafik performa nilai DBI pada masing-masing model pengujian

Hasil yang terbaik diperoleh dari nilai DBI terendah[7], berdasarkan tabel 9 dan grafik pada gambar 5 terlihat bahwa urutan hasil performa nilai DBI terendah dimulai dari model *k-medoids* dengan jumlah *cluster* 2 yang memiliki nilai -159.796.792.242,667 dilanjutkan dengan model *k-means* jumlah *cluster* 2 dengan nilai -82.338.884.292,014 dilanjutkan dengan model *k-medoids* jumlah *cluster* 3 dengan nilai DBI -64.191.292.823,333 dilanjutkan dengan model *k-means* jumlah *cluster* 3 dengan nilai DBI -41.529.203.084,185 dilanjutkan dengan model *k-medoids* jumlah *cluster* 5 dengan nilai DBI -10.294.414.785 dan urutan terakhir dengan model *k-means* jumlah *cluster* 5 dengan nilai DBI -5.659.950.362,375. Berdasarkan nilai DBI tersebut model terbaik yaitu menggunakan *algoritma k-medoids* dengan jumlah *cluster* 2.

4. KESIMPULAN

Berdasarkan pengujian *algoritma k-means* dan *k-medoids* dengan model 2 *cluster*, 3 *cluster*, 5 *cluster* diperoleh hasil *cluster* 2 menjadi model *cluster* terbaik jika dibandingkan dengan model *cluster* 3 dan 5 karena memiliki nilai *DBI* terkecil yaitu *k-means* -82.338.884.292,014 dan *k-medoids* -159.796.492.242,667. Sedangkan untuk *algoritma* terbaik dalam pengelompokan data produksi kelapa sawit yaitu *K-medoids* dengan nilai DBI 159.796.492.242,667. Hasil pengelompokan yang terbaik terdapat 7 data pada *cluster 0* untuk kelompok jumlah produksi tinggi dan 5 data pada *cluster 1* untuk kelompok jumlah produksi rendah. Kabupaten yang masuk pada *cluster 0* atau produksi Tinggi adalah Kuantan seingingi, Indragiri Hilir, Pelalawan, Siak, Kampar, Rokan Hulu dan Rokan Hilir dan kabupaten yang masuk pada *cluster 1* atau produksi rendah adalah Indragiri Hulu, Bengkalis, Kepulauan Meranti, Pekanbaru dan Dumai.

UCAPAN TERIMA KASIH

Terima Kasih kepada Program studi D3 Teknik Informatika Politeknik Kampar dan S1 Sistem Informasi Universitas Potensi Utama sebagai tempat penulis mengabdikan diri menjadi Pendidik.

REFERENSI

- [1] Haryadi, D. (2021). Penerapan Algoritma K-Means Clustering Pada Produksi Perkebunan Kelapa Sawit Menurut Provinsi. *Journal of Informatics and Communication Technology (JICT)*, 3(1), 50-64.
- [2] Pratama, F. H., Triayudi, A., & Mardiani, E. (2022). DATA MINING K-MEDOIDS DAN K-MEANS UNTUK PENGELOMPOKAN POTENSI PRODUKSI KELAPA SAWIT DI INDONESIA. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 7(4), 1294-1310.
- [3] Sari, W. E., Muslimin, M., Franz, A., & Sugiartawan, P. (2022). Deteksi Tingkat Kematangan Tandan Buah Segar Kelapa Sawit dengan Algoritme K-Means. *SINTECH (Science and Information Technology) Journal*, 5(2), 154-164.
- [4] Agustini, L., & Kirana, I. O. (2022). Pengelompokan Data Janjang Panen Kelapa Sawit Menggunakan Algoritma K-Medoids Pada PT SIR MANDAU. *Journal of Machine Learning and Data Analytics*, 1(1), 36-44.
- [5] Lili, A., & Widodo, S. (2022). Pengelompokan Hasil Panen Kelapa Sawit Dalam Produksi Per Blok Menggunakan Algoritma K-Means. *Journal of Machine Learning and Data Analytics*, 1(1), 45-54.
- [6] Pasaribu, D. F., Damanik, I. S., Irawan, E., & Tambunan, H. S. (2021). Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat. *BIOS: Jurnal Teknologi Informasi Dan Rekayasa Komputer*, 2(1), 11-20.
- [7] Nasution, D., Sirait, D. N., Wardani, I., & Dwiyanto, D. (2022). OPTIMASI JUMLAH CLUSTER METODE K-MEDOIDS BERDASARKAN NILAI DBI PADA PENGELOMPOKKAN DATA LUAS TANAMAN DAN PRODUKSI KELAPA SAWIT DI SUMATERA UTARA. *KLIK-KUMPULAN JURNAL ILMU KOMPUTER*, 9(2), 381-390.
- [8] Diansyah, T. M., Tanjung, A. F., & Rismayanti, R. (2019). Pemanfaatan Algoritma K-Means Clustering Sebagai Pengamanan Pencurian Buah Kelapa Sawit Se-Distrik Tandun PT. Perkebunan Nusantara V. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 3(4), 351-356.
- [9] Nasari, F., & Sianturi, C. J. M. (2016). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat. *Cogito Smart Journal*, 2(2), 108-119.
- [10] Arief, N., Damanik, I. S., & Irawan, E. (2021). Penerapan Algoritma K-Medoids Dalam Mengelompokkan Tingkat Kasus Kejahatan di Setiap Provinsi. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 2(3), 111-116.
- [11] Ardiansyah, D. (2019). Algoritma c4. 5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner. *Jurnal Inkofar*, 1(2).