

# Perbandingan Kinerja Algoritma K-Means dan K-Medoids pada Pengelompokan Usaha Pertanian Perorangan Tanaman Pangan di Provinsi Kalimantan Barat

*Comparison of K-Means and K-Medoids Algorithm Performance on Individual Agricultural Business Clustering of Food Crop in Kalimantan Barat Province*

Tiara Margareta<sup>1\*</sup>, Neva Satyahadewi<sup>2</sup>, Retno Pertiwi<sup>3</sup>

<sup>1,2</sup>Universitas Tanjungpura, JL. Prof Dr. H. Hadari Nawawi, Pontianak;

<sup>3</sup>BPS Provinsi Kalimantan Barat, Jl. Sutan Syahrir, Pontianak;

\*Penulis Korespondensi. e-mail:- h1091221014@student.untan.ac.id

## ABSTRACT

*Kalimantan Barat Province as the third largest province in Indonesia, has diverse and promising potential, especially in agriculture. It is known based on the result of the 2023 Agricultural Census, 354,503 people in Kalimantan Barat are food crop subsector agricultural business (contributing 48.58 percent of the total individual agricultural holdings). It is necessary to classify the commodity of food crop to identify the potential commodities in each regencies/cities. The method used in this analysis are the K-Means and K-Medoids algorithms. The use of both methods aims to compare and determine the most suitable algorithm for this analysis. The data used is sourced from the result of the complete enumeration of the 2023 agricultural census with the variable is the number of individual agricultural holdings of food crop commodities (rice and secondary food crops). The purpose of this study is to classify the number of food crop agricultural holdings in Kalimantan Barat Province by regency/municipality. The analysis results showed that K-Means was the best algorithm used for this data, with a DBI of 0.637. Smaller than the K-Medoids algorithm, which had a DBI of 0.683. The result of the K-Means analysis identified four clusters of food crop commodity areas. The first cluster consist of five regencies/municipalities, the second cluster contains one, the third cluster includes six, the fourth cluster comprise two regencies/municipalities.*

*Keywords:* Food Crops, Non-Hierarchical, K-Means, K-Medoids, Davies Bouldin Index.

## ABSTRAK

Provinsi Kalimantan Barat sebagai provinsi terluas ketiga di Indonesia, memiliki potensi yang beragam dan menjanjikan khususnya di bidang pertanian. Berdasarkan hasil Sensus Pertanian 2023, terdapat 354.503 pelaku usaha subsektor pertanian tanaman pangan di Kalimantan Barat (menyumbang 48,58 persen dari total pelaku usaha pertanian). Perlu dilakukan klasifikasi komoditas tanaman pangan untuk mengetahui potensi komoditas tanaman pangan di masing-masing kabupaten/kota. Metode analisis yang digunakan dalam penelitian adalah algoritma *K-Means* dan *K-Medoids*. Penggunaan kedua metode ini bertujuan untuk membandingkan dan menentukan algoritma yang paling sesuai dalam analisis ini. Data yang digunakan bersumber dari hasil pencacahan lengkap Sensus Pertanian 2023 dengan variabelnya adalah jumlah usaha pertanian perorangan tanaman pangan (padi dan palawija).

Tujuan dari penelitian adalah mengelompokkan kabupaten/kota berdasarkan potensi komoditas tanaman pangan di Kalimantan Barat. Hasil analisis diperoleh *K-Means* sebagai algoritma terbaik untuk data ini, dengan DBI sebesar 0,637. Lebih kecil dibandingkan algoritma *K-Medoids* dengan DBI 0,683. Hasil analisis *K-Means* mengidentifikasi empat cluster area komoditas tanaman pangan. *Cluster* pertama terdiri dari lima kabupaten/kota, *cluster* kedua terdiri dari satu kabupaten/kota, *cluster* ketiga terdiri dari enam kabupaten/kota, dan *cluster* keempat terdiri dari dua kabupaten/kota.

Kata kunci: Tanaman Pangan, Non-Hierarki, *K-Means*, *K-Medoids*, Davies Bouldin Index.

## PENDAHULUAN

Ketahanan pangan nasional merupakan salah satu pilar kestabilan ekonomi bagi suatu negara (Sianipar & Tangkudung, 2020). Indonesia sebagai negara kepulauan dengan kekayaan sumber daya alam yang melimpah, menjadikan pertanian sebagai komponen penting dalam upaya untuk mendukung ketahanan pangan dan menyokong kestabilan ekonomi nasional. Isu kerawanan pangan menjadi isu yang krusial di Indonesia dan memerlukan strategi dan penanganan yang tepat. Pemerintah telah melakukan berbagai upaya dan sinergi yang dalam rangka menjaga ketahanan pangan nasional, salah satunya dengan ekstensifikasi lahan. Kalimantan Barat menempati posisi ketiga dalam hal luas wilayah provinsi di Indonesia, setelah Papua dan Kalimantan Tengah dengan total luas wilayah sebesar 147.307 km<sup>2</sup>. Bentangan wilayah yang sangat luas ini menjadikan wilayah Kalimantan Barat memperlihatkan peluang yang sangat menjanjikan terutama pada sektor pertanian.

Berdasarkan data hasil Sensus Pertanian 2023, dari 729.751 pelaku usaha pertanian perorangan di Provinsi Kalimantan Barat, sebanyak 354.503 orang atau 48,85 persen adalah pelaku usaha pertanian subsektor tanaman pangan (BPS Provinsi Kalimantan Barat, 2024). Meskipun jumlah usaha pertanian perorangan di Provinsi Kalimantan Barat menunjukkan angka yang besar, hasil Sensus Pertanian 2023 menunjukkan bahwa rata-rata luas lahan yang dikuasai usaha pertanian perorangan pada lahan sawah sebesar 5.044,76 m<sup>2</sup> dan lahan bukan sawah hanya sebesar 7.291,06 m<sup>2</sup>. Artinya, secara umum rata-rata luas lahan yang dikelola usaha pertanian perorangan pada lahan sawah dan bukan sawah masih kurang dari 1 hektar. Kondisi ini relatif jauh apabila dibandingkan dengan rata-rata luas lahan yang dikuasai tanaman tahunan yang mencapai 15.500,06 m<sup>2</sup>. Salah satu fokus pemerintah pada saat ini adalah dengan meningkatkan produksi tanaman pangan khususnya padi melalui program strategis seperti Perluasan Areal Tanam (PAT) dan pencetakan lahan sawah. Menyikapi hal tersebut, perlu dilakukan pengelompokan wilayah berdasarkan potensi usaha komoditas tanaman pangan.

Dalam analisis ini, identifikasi pengklasteran data dilakukan dengan pendekatan analisis *cluster*. Metode analisis *cluster* yang diterapkan dalam penelitian ini adalah metode *K-Means* dan *K-Medoids clustering*. *K-Means* merupakan metode pengelompokan yang menempatkan objek ke dalam *cluster* dengan melihat jarak terdekat ke pusat *cluster*. Algoritma *K-Medoids* atau yang dikenal pula sebagai *Partitioning Around Medoid* merupakan suatu metode pengelompokan data ke dalam sejumlah *cluster* tanpa memiliki hubungan yang bersifat hierarki antara satu sama lain (Dewangga et al., 2024). Sebagai varian pengembangan dari *K-Means clustering*, algoritma *K-Medoids* ini hadir bertujuan untuk menanggulangi kelemahan algoritma *K-Means* yang mencakup ketidakmampuannya menangani *outlier* secara efektif (Kamila et al., 2019).

Beberapa penelitian terdahulu yang telah membahas pemanfaatan metode *K-Means* dan *K-Medoids* dalam sejumlah kasus yang berbeda, di antaranya membandingkan kedua algoritma dalam analisis data *Covid-19* di Indonesia dan menemukan bahwa *K-Means* lebih unggul berdasarkan uji validitas SC dan DBI, sehingga lebih efektif untuk pengelompokan data dibandingkan dengan *K-Medoids* (Puspitasari et al., 2023). Penelitian lain juga membandingkan algoritma ini dalam

pengklasifikasian tesis mahasiswa menggunakan uji DBI, dengan hasil bahwa *K-Means* lebih cepat dan efisien daripada *K-Medoids* (Ramadhani et al., 2022). Selanjutnya, penelitian membahas pengelompokan Sekolah Dasar di Kabupaten Bojonegoro menggunakan tiga metode, yaitu *K-Means*, *K-Medoids*, dan *Random Clustering*. Hasilnya menunjukkan kinerja *K-Means* memberikan hasil terbaik, mengungguli metode lainnya (Nurdiansyah et al., 2023). Meskipun beberapa penelitian terdahulu menunjukkan bahwa *K-Means* seringkali menghasilkan *output* yang lebih unggul, penggunaan kedua metode dalam penelitian ini tetap penting karena tidak hanya ditujukan untuk memperoleh hasil pengelompokan, tetapi juga untuk membandingkan dan menganalisis kinerja kedua metode tersebut dalam konteks potensi usaha pertanian perorangan komoditas tanaman pangan di Kalimantan Barat.

Penelitian ini bertujuan untuk mengelompokkan potensi usaha komoditas tanaman pangan kabupaten/kota di Provinsi Kalimantan Barat dengan menggunakan variabel jumlah usaha pertanian perorangan per komoditas tanaman pangan. Pengelompokan ini dilakukan untuk memberikan acuan dalam pengelolaan pertanian pangan yang lebih baik sesuai dengan komoditas unggulan pada setiap kabupaten/kota di Provinsi Kalimantan Barat. Penelitian ini diharapkan menghasilkan informasi serta pemahaman yang jelas mengenai pengelompokan data usaha pertanian subsektor tanaman pangan (padi dan palawija) dari setiap komoditi agar dapat diketahui kabupaten/kota mana yang dikategorikan memiliki potensi yang besar. Dengan demikian, temuan dari penelitian ini dapat dimanfaatkan oleh pemerintah daerah maupun pihak yang berkepentingan lainnya dalam merancang kebijakan yang lebih efektif dan efisien, mendorong optimalisasi pemanfaatan lahan dan sumber daya, serta meningkatkan produktivitas sektor pertanian tanaman pangan yang selaras karakteristik dan potensi yang dimiliki oleh setiap wilayah.

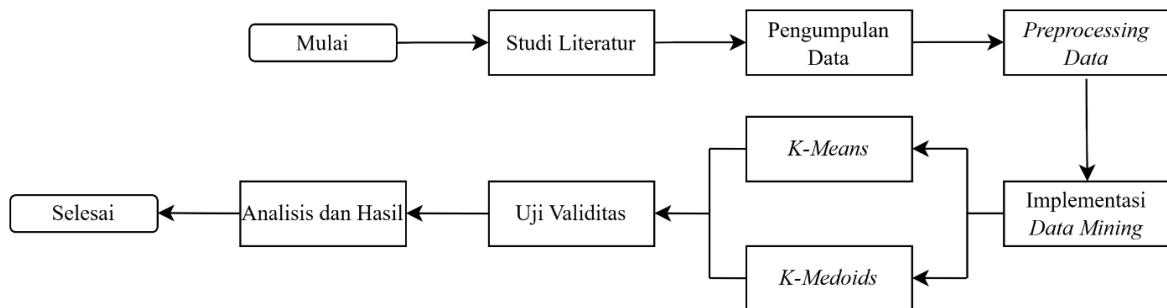
## METODOLOGI

Penelitian ini memanfaatkan data yang diperoleh dari hasil publikasi resmi Badan Pusat Statistik (BPS) “Hasil Pencacahan Lengkap Sensus Pertanian 2023 Usaha Pertanian Perorangan Tanaman Pangan Tahap II”. Variabel yang dipakai dalam penelitian ini yaitu berjumlah 6 variabel terdiri dari ( $X_1$ ) jumlah UTP padi ladang, ( $X_2$ ) jumlah UTP padi sawah, ( $X_3$ ) jumlah UTP jagung, ( $X_4$ ) jumlah UTP ubi kayu, dan ( $X_5$ ) jumlah UTP ubi jalar, dan ( $X_6$ ) jumlah UTP kacang tanah. Jumlah usaha pertanian perorangan tanaman padi meliputi padi ladang dan padi sawah, serta jumlah usaha pertanian perorangan tanaman palawija yang mencakup jagung, ubi kayu, ubi jalar, dan kacang tanah.

Penelitian ini menerapkan analisis *cluster* sebagai teknik dalam pengelompokan data. Klasterisasi merupakan teknik yang digunakan untuk menemukan kesamaan pola dan karakteristik pada data yang dibentuk dalam kelompok-kelompok (Setyadji et al., 2023). Proses pengolahan data pada penelitian ini menggunakan bantuan piranti lunak (*software*) RapidMiner. RapidMiner merupakan sebuah aplikasi yang digunakan dalam data mining untuk mendukung analisis data. *Software* tersebut dipilih karena tingkat efisiensi yang dihasilkan dalam analisisnya, dapat meningkatkan kualitas keputusan melalui analisis, mudah diakses, serta kompatibel dengan berbagai sumber data. RapidMiner mengidentifikasi pola-pola dalam data set dengan memadukan metode statistika, basis data, dan kecerdasan buatan. Hasil yang ditampilkan dari RapidMiner dapat divisualisasikan dalam bentuk grafik, sehingga RapidMiner dimanfaatkan untuk menerapkan teknik *data mining* dalam tahapan ekstraksi data (Rahmat et al., 2017). Hasil dari kedua metode dalam penelitian ini kemudian dilakukan evaluasi validitasnya dengan memanfaatkan *Davies Bouldin Index (DBI)*.

## Tahapan Penelitian

Penelitian ini mencakup sejumlah tahapan yang diikuti oleh peneliti guna memperoleh pemecahan dari suatu permasalahan secara ilmiah dan bersifat sistematis. Penelitian ini terdiri atas beberapa tahapan sistematis yang dijelaskan sebagai berikut:



Gambar 1. Tahapan Penelitian

Sumber: Penulis, 2024

## Metode Non Hierarki

Metode non hierarki merupakan pendekatan yang dimanfaatkan dalam proses pengelompokan objek ke dalam *cluster*, di mana jumlah *cluster* yang akan dibentuk dapat ditetapkan sebelumnya (Syafiyah et al., 2022). Dalam metode non-hierarki, terlebih dahulu dilakukan pemilihan sejumlah nilai cluster awal yang dibutuhkan, yaitu dua *cluster*, tiga *cluster*, dan seterusnya. Kemudian objek-objek data dialokasikan ke *cluster* yang sesuai dengan tidak mengikuti proses hirarki (Abidin et al., 2024). Kelebihan yang dimiliki metode *cluster* non-hierarki yaitu, pada metode ini dapat dilakukan analisis sampel dalam skala yang lebih besar secara efisien dan pendekatan ini hanya mempunyai sedikit kelemahan pada *outlier*.

### *K-Means clustering*

Algoritma *K-Means* digunakan sebagai teknik pengelompokan data tanpa tingkatan yang cepat sehingga efisien dalam penggunaannya (Aprizkiyandari et al., 2023). Metode klasterisasi non-hierarki *K-Means* diklasifikasikan sebagai teknik pengelompokan data berdasarkan kemiripan atribut yang digabungkan dalam satu kelompok data, sementara data yang lain yang dikelompokkan berdasarkan karakteristik data yang sesuai dimiliki oleh masing-masing data tersebut (Sari et al., 2023). Proses dalam penerapan algoritma *K-Means* untuk klasterisasi data dapat dijelaskan sebagai berikut (Septianingsih, 2022):

1. Lakukan proses *data preparation* juga standardisasi data jika memiliki ketidaksamaan ukuran pada masing-masing variabel.
2. Tentukan kuantitas *cluster*, di mana banyaknya *k* tidak boleh lebih dari banyaknya data.
3. Secara acak, tentukan *centroid* awal pada *cluster*. Tentukan banyaknya titik pusat *cluster* awal sesuai dengan banyaknya *cluster* yang akan dibentuk.
4. Hitung kedekatan antar objek melalui penerapan rumus *Eucledian Distance*.

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

Keterangan:

$d_{ij}$  = jarak antar objek  $i$  dan  $j$

$x_{ik}$  = nilai untuk objek  $i$  pada variabel ke- $k$

$x_{jk}$  = nilai untuk objek  $j$  pada variabel ke- $k$

$p$  = jumlah variabel yang dipantau

5. Tentukan setiap objek pada titik pusat terdekat.
6. Lakukan proses iterasi dari langkah ke-3 atau membangkitkan nilai *centroid* baru dengan persamaan:

$$C_k = \frac{1}{n_k} \sum d_i$$

Keterangan:

$n_k$  = banyaknya data dalam *cluster*.

$d_i$  = total jarak yang dihitung untuk masing-masing *cluster*.

7. Setelah nilai *centroid* dan anggota masing-masing *cluster* tidak berubah, maka diperoleh hasil *cluster* final.

### ***K-Medoids Clustering***

*K-Medoids clustering* adalah pengembangan dari *K-Means clustering* yang diterapkan guna mempartisi kumpulan data bertransformasi menjadi *cluster*. Pengelompokan data dengan metode ini diwakili oleh salah satu titik data dalam *cluster* di setiap *cluster*-nya. Titik-titik pusat ini dinamakan *medoid* dari masing-masing kelompok data. *Medoid* adalah entitas dalam sebuah *cluster* yang perbedaannya terletak pada rata-rata minimal antara titik dan semua anggota *cluster* yang lain (Gubu et al., 2021). Tahapan dalam menerapkan algoritma *K-Means* untuk *clustering* data adalah sebagai berikut (Sindi et al., 2020):

1. Tentukan jumlah  $k$  sebagai banyaknya jumlah *cluster* yang ingin dibentuk.
2. Lakukan perhitungan ukuran jarak setiap objek menggunakan persamaan *Euclidean Distance*, kemudian alokasikan setiap objek.
3. Tentukan satu objek secara acak dari tiap *cluster* sebagai *medoid* awal.
4. Lakukan perhitungan ulang jarak tiap objek terhadap *medoid* yang baru di setiap *cluster*.
5. Lakukan perhitungan selisih antara nilai total jarak baru menggunakan nilai total jarak yang lama untuk menentukan total simpangan ( $S$ ). Apabila  $S < 0$ , lakukan penggantian objek dengan data *cluster* untuk menghasilkan  $k$  kelompok baru sebagai *medoid*.
6. Ulangi tahap 3 sampai dengan 5 hingga tidak ada lagi perubahan *medoid*, dan hasil akhir adalah pembentukan *cluster* yang stabil.

### ***Davies Bouldin Index***

*Davies Bouldin Index* (DBI) adalah salah satu metode evaluasi yang berfungsi untuk mengukur kualitas hasil *clustering*. Penggunaan DBI sebagai indikator kinerja digunakan untuk mengevaluasi seberapa baik hasil pengelompokan dengan pembagian *cluster* berdasarkan perbedaan antar *cluster* dan persamaan dalam *cluster*, yang memberikan gambaran keseluruhan tentang kualitas pengelompokan (Awaliyah et al., 2024). Semakin kecil perolehan nilai DBI, maka semakin efektif pula hasil algoritma pengelompokan yang digunakan (Farahdinna et al., 2019). Langkah-langkah dalam perhitungan DBI adalah sebagai berikut (Ishak & Amiruddin, 2024):

1. Tentukan kesamaan antar *cluster*. Jika semakin rendah nilai kesamaan, maka semakin lebih baik *cluster* tersebut dipisahkan.
2. Lakukan perhitungan nilai DBI pada setiap *cluster*.

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

Keterangan:

$R_{ij}$  = nilai R pada *cluster* i dan j

$d_{ij}$  = jarak antar-*centroid* untuk *cluster* i dan j

$S_i$  dan  $S_j$  = nilai rata-rata jarak antara titik-titik *cluster* i dengan *centroid cluster* j

3. Tentukan hasil akhir dari perhitungan DBI.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max R_{ij}, i \neq j$$

Keterangan:

$R_{ij}$  = nilai R antara *cluster* i dan *cluster* j

$k$  = banyaknya *cluster*

## HASIL DAN PEMBAHASAN

### Pengumpulan dan *Preprocessing* Data

Dalam penelitian ini, variabel yang dianalisis mencakup jumlah usaha pertanian perorangan per komoditas tanaman pangan. Sebelum algoritma *K-Means* dan *K-Medoids* diterapkan, dilakukan tahap persiapan data. Rincian jumlah usaha pertanian perorangan per komoditas tanaman pangan disajikan dalam Tabel 1.

Tabel 1. Data Jumlah Usaha Pertanian Perorangan Tanaman Pangan dan Palawija menurut Kabupaten/Kota di Kalimantan Barat dan Komoditasnya (Unit Usaha) Tahun 2023

Kabupaten/Kota	Tanaman Pangan Padi (Unit)			Tanaman Palawija (Unit)		
	Padi Ladang	Padi Sawah	Jagung	Ubi kayu	Ubi jalar	Kacang Tanah
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sambas	544	57.392	251	394	219	51
Bengkayang	10.775	12.143	7.784	401	760	365
Landak	12.755	35.021	3.185	224	26	198
Mempawah	2.006	10.723	454	1.187	163	94
Sanggau	28.709	22.149	429	364	45	534
Ketapang	16.209	8.769	390	1.312	287	46
Sintang	34.323	4.447	2.234	1.257	135	118
Kapuas Hulu	18.605	2.838	278	697	105	83
Sekadau	19.873	11.516	254	934	24	29
Melawi	16.433	2.266	117	1,654	218	121
Kayong Utara	0	5.788	21	672	115	74
Kubu Raya	110	17.998	466	2.053	372	171
Pontianak	123	253	25	375	60	6
Singkawang	0	3.429	323	796	159	116

Sumber: Badan Pusat Statistik (2024)

### Analisis Deskriptif

Tabel di bawah ini menunjukkan ringkasan statistik jumlah usaha pertanian perorangan di Provinsi Kalimantan Barat berdasarkan hasil Sensus Pertanian 2023.

Tabel 2. Nilai Minimum, Rata-Rata dan Nilai Maksimum Jumlah Usaha Pertanian Perorangan Tanaman Pangan di Kalimantan Barat Menurut Komoditasnya (Unit Usaha) Tahun 2023

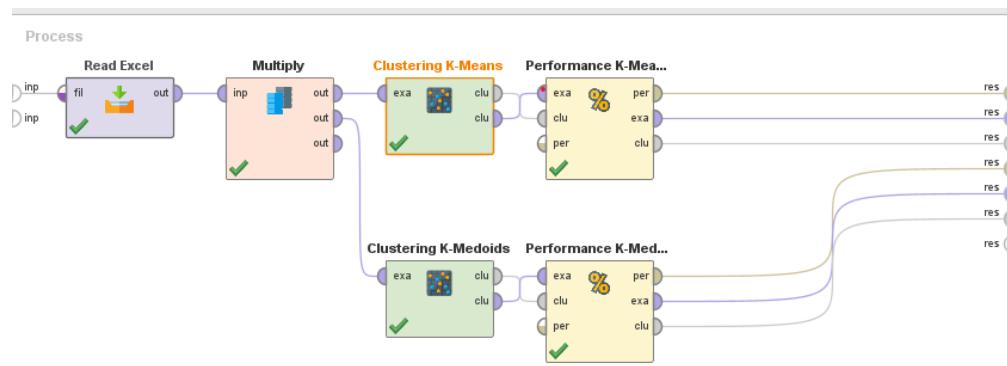
<b>Variabel</b>	<b>Nilai Minimum</b>	<b>Rata-rata</b>	<b>Nilai Maksimum</b>
(1)	(2)	(3)	(4)
Padi Ladang	0,000	11.461,786	34.323,000
Padi Sawah	253,000	13.909,429	57.392,000
Jagung	21,000	1.157,929	7.784,000
Ubi kayu	224,000	880,000	2.053,000
Ubi jalar	24,000	192,000	760,000
Kacang Tanah	6,000	143,286	534,000

Sumber: Badan Pusat Statistik (diolah), 2024

Berdasarkan hasil Sensus Pertanian 2023, jumlah unit usaha pertanian perorangan terbanyak pada subsektor tanaman pangan berada pada komoditas padi sawah dengan jumlah 194.372 unit. Rata-rata jumlah unit usaha komoditas padi sawah per kabupaten/kota sebesar 13.909 unit dengan jumlah unit usaha terbanyak berada pada Kabupaten Sambas dengan 57.392 unit usaha. Rata-rata jumlah unit usaha terkecil pada penelitian ini terdapat pada komoditas kacang tanah dengan rata-rata kabupaten/kota hanya 143 unit usaha. Secara lebih rinci, sebaran jumlah unit usaha pertanian menunjukkan variasi yang cukup mencolok antarwilayah. Kabupaten Sambas, Kabupaten Sanggau, dan Kabupaten Landak cenderung memiliki unit usaha pertanian yang jumlahnya melebihi kabupaten/kota lainnya terutama komoditas padi ladang dan padi sawah. Sementara itu, daerah dengan lahan pertanian yang lebih terbatas, seperti Kota Pontianak dan Kota Singkawang memiliki unit usaha pertanian yang relatif rendah. Hal ini mengindikasikan bahwa potensi usaha pertanian di Provinsi Kalimantan Barat sangat bergantung pada keadaan geografis, luas wilayah, dan dukungan infrastruktur di masing-masing wilayah.

### **Implementasi *Data Mining***

Setelah melalui tahap *preprocessing*, data diproses menggunakan algoritma *K-Means* dan *K-Medoids*, kemudian divalidasi dengan perhitungan *Davies Bouldin Index* (DBI) untuk mendapatkan *cluster* yang paling tepat. Berikut ini merupakan desain proses yang dilakukan pada *software* RapidMiner dalam menganalisis data menggunakan algoritma *K-Means* serta algoritma *K-Medoids*.



Gambar 2. Desain alur *clustering K-Means* dan *K-Medoids* menggunakan RapidMiner

Source: RapidMiner, 2024

Proses pengelompokan data dilakukan dengan menggunakan metode *K-Means* dan *K-Medoids* dilakukan pada *cluster*  $k = 2$  sampai dengan  $k = 4$ . Batasan jumlah *cluster* ditetapkan sampai  $k = 4$  karena jumlah *cluster* yang terlalu banyak maka dapat menyulitkan interpretasi hasil klasterisasi. Pembatasan jumlah *cluster* ini juga dapat membuat proses pengambilan kesimpulan terkait sifat atau pola yang ditemukan pada setiap *cluster* menjadi lebih mudah karena jumlah *cluster* yang kecil dan terfokus dapat menghasilkan informasi yang lebih bermakna.

Tabel 3 dan tabel 4 menunjukkan hasil klasterisasi melalui penerapan algoritma *K-Means* dan *K-Medoids*. Nilai  $k$  pada tabel menunjukkan jumlah *cluster* yang ingin dibentuk. Hasil klasterisasi ini menunjukkan bagaimana kedua algoritma dapat mengelompokkan data dengan cara yang berbeda, tergantung pada jumlah *cluster* yang digunakan, yang mempengaruhi distribusi data pada setiap *cluster*. Pemilihan jumlah *cluster* memiliki dampak langsung pada bagaimana data akhirnya dikelompokkan dan tersebar di dalam *cluster*.

Tabel 3. Hasil Klasterisasi Algoritma *K-Means*

Jumlah k	<i>Cluster</i>			
	1	2	3	4
(1)	(2)	(3)	(4)	(5)
$k = 2$	12	2	-	-
$k = 3$	2	6	6	-
$k = 4$	5	1	6	2

Sumber: Badan Pusat Statistik (diolah), 2024

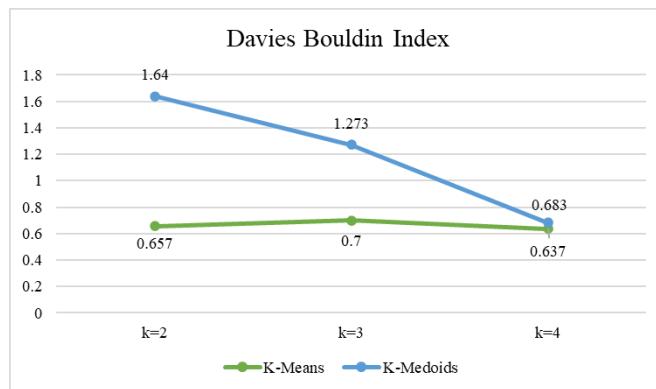
Tabel 4. Hasil Klasterisasi Algoritma *K-Medoids*

Jumlah k	<i>Cluster</i>			
	1	2	3	4
(1)	(2)	(3)	(4)	(5)
$k = 2$	6	8	-	-
$k = 3$	4	7	3	-
$k = 4$	1	2	6	5

Sumber: Badan Pusat Statistik (diolah), 2024

### Uji Validitas

Uji Validitas menggunakan *Davies Bouldin Index* (DBI) diterapkan guna melihat tingkat kualitas hasil clustering data yang dihasilkan oleh algoritma pengelompokan *K-Means* dan *K-Medoids*. Gambar 2 menunjukkan hasil nilai DBI pada algoritma *K-Means* dan *K-Medoids*. Hasil DBI memanfaatkan pengukuran kohesi yang menunjukkan seberapa erat hubungan dalam masing-masing *cluster*, serta diikuti dengan pengukuran separasi yang menunjukkan perbedaan antara anggota *cluster* yang berbeda.



Gambar 2. Perbandingan DBI *K-Means* dan *K-Medoids*

Sumber: Badan Pusat Statistik (diolah), 2024

Hasil pengujian setiap *cluster* dengan menggunakan metode *K-Means* dan *K-Medoids* disajikan pada Tabel 5.

Tabel 5. Hasil Uji Validitas *Davies Bouldin Index*

Cluster (1)	<i>K-Means</i>		<i>K-Medoids</i>
	(2)	(3)	
$k = 2$	0,657		1,640
$k = 3$	0,700		1,273
$k = 4$	<b>0,637</b>		<b>0,683</b>

Sumber: Badan Pusat Statistik (diolah), 2024

Berdasarkan hasil pengujian validitas *Davies Bouldin Index* (DBI), performa terbaik diperoleh dari algoritma *K-Means* dengan nilai DBI terkecil, yaitu 0,637 pada percobaan  $k = 4$ . Mengacu pada prinsip DBI, semakin rendah nilai DBI atau mendekati nol, maka kualitas pengelompokan data dianggap semakin baik atau mendekati kondisi optimal.

### Interpretasi Hasil *Cluster*

Setelah melakukan percobaan analisis *cluster* menggunakan algoritma *K-Means* dan *K-Medoids*, algoritma *K-Means* terbukti menjadi metode pengklasteran terbaik. Analisis *cluster* *K-Means* ini melakukan pengelompokan kabupaten/kota di Provinsi Kalimantan Barat ke dalam empat *cluster*.

Tabel 6. Rata-rata Variabel Berdasarkan *Cluster* Pada Algoritma *K-Means*

Variabel	Cluster 1	Cluster 2	Cluster 3	Cluster 4
(1)	(2)	(3)	(4)	(5)
Padi Ladang	21.088,6	544	2.169	20.732
Padi Sawah	5.967,2	57.392	8.389	28.585
Jagung	654,6	251	1.512,167	1.807
Ubi kayu	1.170,8	394	914	294
Ubi jalar	153,8	219	271,5	35,5
Kacang Tanah	79,4	51	137,667	366

Sumber: Badan Pusat Statistik (diolah), 2024

Dari tabel 6 serta hasil pembentukan *cluster* dengan algoritma *K-Means*, hasil pengelompokan diperoleh empat *cluster* dan diberikan kategori pada masing masing *cluster*. Berikut ini pembagian kabupaten/kota berdasarkan karakteristik usaha pertanian perorangan tanaman pangan pada *cluster*.

Tabel 8. Hasil Klasifikasi Data Menggunakan Pendekatan *K-Means*

Cluster	Anggota	Karakteristik
(1)	(2)	(3)
1	Kabupaten Ketapang, Kabupaten Sintang, Kabupaten Kapuas Hulu, Kabupaten Sekadau, dan Kabupaten Melawi	Memiliki kecenderungan merepresentasikan jumlah usaha pertanian perorangan padi ladang. Padi sawah memiliki kontribusi cukup signifikan, meskipun jauh lebih kecil dibandingkan padi ladang. Usaha pertanian perorangan tanaman lain seperti jagung, ubi kayu, ubi jalar, dan kacang tanah relatif rendah.
2	Kabupaten Sambas	Memiliki gambaran dominasi yang sangat menonjol pada jumlah usaha pertanian perorangan padi sawah. UTP padi ladang, jagung, ubi kayu, dan lainnya tidak sampai 10 persen dari total usaha pada komoditas tersebut. Kabupaten Sambas memiliki potensi untuk mempertahankan posisinya sebagai lumbung beras Kalimantan Barat.
3	Kabupaten Bengkayang, Kabupaten Mempawah, Kabupaten Kayong Utara, Kabupaten Kubu Raya, Kota Pontianak, dan Kota Singkawang	Tidak ada komoditas yang paling mendominasi dan fokus jumlah usaha komoditas yang sedang padi ladang, padi sawah, dan jagung.
4	Kabupaten Landak dan Kabupaten Sanggau	Memiliki keseimbangan di antara komoditas padi ladang dan padi sawah. Akan tetapi, jumlah usaha tanaman pangan lain relatif minim.

Sumber: Badan Pusat Statistik (diolah), 2024

## KESIMPULAN DAN SARAN

Penelitian dengan komparasi dua algoritma *data mining* yaitu *K-Means* dan *K-Medoids* menghasilkan kesimpulan yang menunjukkan metode algoritma terbaik dalam mengelompokkan jumlah usaha pertanian perorangan subsektor tanaman pangan di Provinsi Kalimantan Barat yaitu algoritma *K-Means* dengan nilai DBI terkecil pada algoritma ini sebesar 0,637. Adapun jumlah *cluster* terbaik yang dihasilkan dari metode *K-Means* yaitu sebanyak empat *cluster*. *Cluster* pertama dengan satu kabupaten di mana memiliki representasi usaha padi sawah yang tinggi. *Cluster* kedua dengan satu kabupaten/kota dengan dominasi usaha pertanian tanaman padi sawah dalam skala besar. *Cluster* ketiga dengan enam kabupaten dengan usaha pertanian dengan diversifikasi produksi, tidak ada komoditas yang dominan. *Cluster* keempat dengan dua kabupaten di mana memiliki keseimbangan dalam jumlah usaha pertanian padi ladang dan padi sawah.

Setelah melakukan analisis dan mengetahui hasil dari penggunaan algoritma *K-Means* dan *K-Medoids* untuk klasifikasi usaha pertanian perorangan subsektor tanaman pangan (padi dan palawija) di Provinsi Kalimantan Barat, hasil penelitian ini diharapkan dapat mendorong penelitian berikutnya untuk dapat melakukan analisis *cluster* menggunakan algoritma yang lain, seperti DBSCAN, *Hierarchical Clustering*, dan lain-lain. Dapat juga menggunakan variabel dari subsektor lain hasil pencacahan Sensus Pertanian 2023. Selain itu, pengujian dalam pemilihan jumlah *cluster* yang tepat dapat menggunakan uji validitas yang lain seperti *Sum Squared Error (SSE)*, *Silhouette Coefficient (SC)*, dan lain sebagainya.

## DAFTAR PUSTAKA

- Abidin, M. A. H., Nugroho, S. R., & Wulandari, S. P. (2024). Pengelompokan Kabupaten/Kota Berdasarkan Indeks Pembangunan Manusia Provinsi Jawa Timur Tahun 2023 Menggunakan Analisis Klaster. *Jurnal Cakrawala Akademika (JCA)*, 1(4), 1143–1157. <https://doi.org/10.70182/JCA.v1i4.10>
- Aprizkiyandari, S., Satyahadewi, N., Pratama, A. N., Rivaldo, R., Nurdiansyah, S. I., & Helena, S. (2023). Implementasi K-Means Cluster untuk Menentukan Persebaran Tingkat Pengangguran. *Empiricism Journal*, 4(2), 400–406. <https://doi.org/10.36312/ej.v4i2.1518>
- Awaliyah, L., Rahaningsih, N., & Dana, R. D. (2024). Implementasi Algoritma K-Means dalam Analisis Klaster Korban Kekerasan di Provinsi Jawa Barat. *Jurnal Mahasiswa Teknik Informatika*, 8(1), 188–195. [https://doi.org/https://doi.org/10.36040/jati.v8i1.8332](https://doi.org/10.36040/jati.v8i1.8332)
- BPS Provinsi Kalimantan Barat. (2024). *Hasil Pencacahan Lengkap Sensus Pertanian 2023 - Tahap II: Usaha Pertanian Perorangan (UTP) Tanaman Pangan Provinsi Kalimantan Barat*.
- Dewangga, H., Khairil, & Yati Beti, I. (2024). Penerapan Metode K-Medoid dalam Pengelompokan Data Pertanian di Dinas Tanaman Pangan Hortikultura dan Perkebunan Provinsi Bengkulu. *Jurnal Media Infotama*, 20(2), 501–509. <https://doi.org/https://doi.org/10.37676/jmi.v20i2.6459>
- Farahdinna, F., Nurdiansyah, I., Suryani, A., & Wibowo, A. (2019). Perbandingan Algoritma *K-Means* dan *K-Medoids* dalam Klasterisasi Produk Asuransi Perusahaan Nasional. *Jurnal Ilmiah FIFO*, 11(2), 208–214. <https://doi.org/10.22441/fifo.2019.v11i2.010>
- Gubu, L., Rosadi, D., & Abdurakhman. (2021). Pembentukan Portofolio Saham Menggunakan Klustering Time Series K-Medoid dengan Ukuran Jarak Dynamic Time Warping. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 13(2), 35–46. <https://doi.org/https://doi.org/10.34123/jurnalasks.v13i2.295>

- Ishak, R., & Amiruddin. (2024). Clustering Prestasi Akademik Lulusan Menggunakan Metode K-Means. *Jambura Journal of Electrical and Electronics Engineering*, 6(1), 76–81. <https://doi.org/https://doi.org/10.37905/jjee.v6i1.23967>
- Kamila, I., Khairunnisa, U., & Mustakim. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119–125. <https://doi.org/http://dx.doi.org/10.24014/rmsi.v5i1.7381>
- Nurdiansyah, D., Saidah, S., & Cahyani, N. (2023). Data Mining Study for Grouping Elementary Schools in Bojonegoro Regency Based on Capacity and Educational Facilities. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 17(2), 1081–1092. <https://doi.org/10.30598/barekengvol17iss2pp1081-1092>
- Puspitasari, N., Lempas, G., Hamdani, H., Haviuddin, H., & Septiarini, A. (2023). Perbandingan Algoritma K-Means dan Algoritma K-Medoids Pada Kasus Covid-19 di Indonesia. *Building of Informatics, Technology and Science (BITS)*, 4(4). <https://doi.org/10.47065/bits.v4i4.2994>
- Rahmat, C. T. I. B., Gafar, A. A., Fajriani, N., Ramdani, U., Uyun, F. R., Purnamasar P, Y., & Ransi, N. (2017). Implementasi K-Means Clustering pada RapidMiner untuk Analisis Daerah Rawan Kecelakaan. *Seminar Nasional Riset Kuantitatif Terapan 2017*, 58–63.
- Ramadhani, S., Azzahra, D., & Z, T. (2022). Comparison of K-Means and K-Medoids Algorithms in Text Mining based on Davies Bouldin Index Testing for Classification of Student's Thesis. *Digital Zone: Jurnal Teknologi Informasi Dan Komunikasi*, 13(1), 24–33. <https://doi.org/10.31849/digitalzone.v13i1.9292>
- Sari, L. P., Fanani, A., & Asyhar, A. H. (2023). Analisis Perbandingan Pengelompokan Kota di Indonesia Berdasarkan Indikator Inflasi Tahun 2021 dengan Metode Ward dan K-Means. *Jurnal Sains Matematika Dan Statistika*, 9(2), 108–118. <https://doi.org/10.24014/jsms.v9i2.21100>
- Septianingsih, A. (2022). Analisis K-Means Clustering pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni. *Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, 3(1). <https://doi.org/10.46306/lb.v3i1>
- Setyadji, A. E. S., Wibowo, A. P., Matthew D, I. G. N. A., Pratama, R. B., Masyhuda, T. A., Sinaga, Y. A. A., Purwanti, E., & Werdiningsih, I. (2023). Analisis Klaster Data Pasien Diabetes untuk Identifikasi Pola dan Karakteristik Pasien. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(3), 172–182. <https://doi.org/10.47233/jtekstis.v5i3.828>
- Sianipar, B., & Tangkudung, A. G. (2020). Tinjauan Ekonomi, Politik dan Keamanan Terhadap Pengembangan Food Estate di Kalimantan Tengah Sebagai Alternatif Menjaga Ketahanan Pangan di Tengah Pandemi Covid-19. *Jurnal Keamanan Nasional*, 6(2), 235–248.
- Sindi, S., Ningse, W. R. O., Sihombing, I. A., Zer, P. P. P. A. N. W. F. I. I. R. H., & Hartama, D. (2020). Analisis Algoritma K-Medoids Clustering dalam Pengelompokan Penyebaran COVID-19 di Indonesia. *Jurnal Teknologi Informasi*, 4(1).
- Syafiyah, U., Asrafi, I., Wicaksono, B., Puspitasari, D. P., & Sirait, F. M. (2022). Analisis Perbandingan Hierarchical dan Non-Hierarchical Clustering pada Data Indikator Ketenagakerjaan di Jawa Barat Tahun 2020. *Seminar Nasional Official Statistics*, 2022(1), 803–812. <https://doi.org/https://doi.org/10.34123/semnasoffstat.v2022i1.1221>