

Klasterisasi Usaha Perkebunan Perorangan di Kalimantan Barat Menggunakan Algoritma *K-Means* dengan Validasi Internal Klaster

Clustering of Individual Plantation Businesses in West Kalimantan Using the K-Means Algorithm with Internal Cluster Validation

Muthia Risti Fadhila^{1*}, Syari Azizah Rahim², Pinkkan Aulia³, Warsidah⁴, Rossie Wiedya Nusantara⁵,
Ahmadi⁶, Martini Pratiwi⁷, Retno Dian Ika Wati⁸, Ardi Atmaja⁹

^{1,2,3,4,5,6}Universitas Tanjungpura, Jl. Prof Dr. H. Hadari Nawawi, Pontianak;
^{7,8,9}BPS Provinsi Kalimantan Barat, Jl. Sutan Syahrir, Pontianak;
*Penulis Korespondensi. e-mail: h1091231049@student.untan.ac.id

ABSTRACT

This study aims to cluster plantation businesses in West Kalimantan Province based on similar commodity characteristics using the K-Means algorithm. The data used are sourced from the 2023 Agricultural Census by Statistics Indonesia (BPS), focusing on the Individual Farming Enterprises (UTP) for plantation crops by regency/city level. The clustering process was carried out with two to four clusters, and the quality of the clustering was evaluated using the Silhouette Index, the Davies-Bouldin Index, and the Dunn Index. The test results indicated that a two-cluster model was the best. The resulting clusters reveal clear regional differences in terms of dominant commodities and plantation business capacity. Cluster one is characterized by a high number of UTPs producing rubber and oil palm, while cluster two is characterized by a high number of UTPs producing coconut as well as relatively higher diversity in plantation commodities. These findings indicate that the structure of individual plantation businesses in West Kalimantan is heterogeneous and forms distinct regional clustering patterns. They also emphasize that policy formulation cannot be carried out uniformly but must consider the characteristics of each cluster.

Keywords: *K-Means, Internal Cluster Validation, Individual Farming Enterprises, Plantation Crops, 2023 Agricultural Census.*

ABSTRAK

Penelitian ini menganalisis pengelompokan usaha perkebunan di Provinsi Kalimantan Barat berdasarkan kesamaan karakteristik komoditas menggunakan algoritma *K-Means*. Data yang digunakan bersumber dari hasil Sensus Pertanian 2023 Badan Pusat Statistik, khususnya Usaha Pertanian Perorangan (UTP) tanaman perkebunan menurut kabupaten/kota. Proses klasterisasi dilakukan dengan jumlah klaster dua hingga empat, dan kualitas pengelompokan dievaluasi menggunakan *Silhouette Index*, *Davies-Bouldin Index*, dan *Dunn Index*. Hasil pengujian menunjukkan bahwa dua klaster merupakan model terbaik. Klasterisasi yang terbentuk memperlihatkan perbedaan yang jelas antarwilayah dalam hal dominasi komoditas dan kapasitas usaha perkebunan. Klaster satu dicirikan oleh tingginya jumlah UTP pada komoditas karet dan kelapa sawit, sementara klaster dua dicirikan oleh tingginya jumlah UTP pada komoditas kelapa serta keragaman komoditas perkebunan yang relatif lebih

tinggi. Temuan ini mengindikasikan bahwa struktur usaha perkebunan perorangan di Kalimantan Barat bersifat heterogen dan membentuk pola pengelompokan wilayah yang berbeda, serta menegaskan bahwa formulasi kebijakan tidak dapat dilakukan secara seragam, melainkan harus mempertimbangkan karakter tiap klaster.

Kata kunci: *K-Means*, Validasi Internal Klaster, Usaha Pertanian Perorangan, Tanaman Perkebunan, Sensus Pertanian 2023.

PENDAHULUAN

Sektor perkebunan merupakan salah satu penopang penting perekonomian daerah di Provinsi Kalimantan Barat yang memiliki lahan perkebunan mencapai 2,78 juta hektare pada tahun 2023 (BPS, 2025). Perannya tidak hanya terlihat dari kontribusi terhadap nilai produksi dan ekspor, tetapi juga dari kemampuannya dalam menyerap tenaga kerja serta menggerakkan aktivitas ekonomi rumah tangga perdesaan. Keberadaan berbagai komoditas perkebunan, seperti kelapa sawit, karet, kakao, kelapa, kopi dan lada, menjadikan subsektor ini strategis dalam mendorong pertumbuhan ekonomi wilayah. Namun demikian, perkembangan usaha perkebunan perorangan di setiap kabupaten/kota menunjukkan karakteristik yang beragam. Perbedaan kondisi geografis, ketersediaan lahan, infrastruktur, serta kapasitas pelaku usaha menyebabkan variasi tingkat pengembangan dan dominasi komoditas di masing-masing wilayah.

Dalam konteks perencanaan pembangunan daerah, pemahaman terhadap perbedaan karakteristik usaha perkebunan antarwilayah menjadi hal yang penting. Analisis deskriptif semata sering kali belum mampu menggambarkan pola kesamaan dan perbedaan antarkabupaten/kota secara komprehensif. Beberapa penelitian terdahulu telah mengkaji aspek-aspek perkebunan di Kalimantan Barat, seperti analisis pengaruh luas lahan, umur, pengalaman bertani, serta curah hujan terhadap hasil produksi perkebunan (Afiantoro dkk., 2025), serta analisis komoditas unggulan subsektor pertanian di salah satu kabupaten Kalimantan Barat yakni Bengkayang (Natalia dkk., 2025). Namun, masih terdapat keterbatasan dalam pendekatan yang holistik dan berbasis klaster untuk memetakan karakteristik usaha perkebunan perorangan secara spasial. Padahal menurut Nafri dkk. (2025), memahami struktur dan pola distribusi usaha pertanian perorangan sangat penting, karena dapat memberikan wawasan berharga bagi pengambilan kebijakan pemerintah daerah demi meningkatkan produktivitas dan kesejahteraan petani.

Metode *K-Means* telah banyak diaplikasikan dalam berbagai studi klasterisasi sektor pertanian dan perkebunan, seperti pada penelitian oleh Margareta dkk. (2025) yang mengelompokkan usaha pertanian perorangan (UTP) tanaman pangan di Kalimantan Barat, serta studi Nasari dkk. (2023) yang melakukan klasterisasi produksi kelapa sawit di Provinsi Riau. Pemilihan algoritma *K-Means* sebagai metode klustering utama dalam penelitian ini didukung secara kuat oleh kajian pustaka terkini. Ikotun, Almutari, & Ezugwu (2021) secara tegas menyatakan bahwa keunggulan *K-Means* dibandingkan algoritma klustering mutakhir terletak pada kecepatan pemrosesan, aksesibilitas, kesederhanaan penggunaan, serta kemampuannya dalam menangani data tanpa label. Keunggulan ini menjadikan *K-Means* sebagai metode dasar (*baseline*) yang kokoh, yang tren penelitian saat ini justru mengarah pada penggabungannya dengan algoritma optimasi untuk menghasilkan segmentasi yang lebih baik (Hilyawan dkk., 2025). Popularitas *K-Means* selama dekade terakhir (2014–2024) semakin didorong oleh fleksibilitasnya dan efisiensinya dalam menangani kumpulan data berskala besar di berbagai bidang, seperti infrastruktur, transportasi, kesehatan, dan pendidikan (Ferdiansyah dkk., 2024). Lebih lanjut, untuk penelitian yang melibatkan data campuran (numerik dan kategorikal), pendekatan berbasis *K-Means* tetap menjadi pilihan utama karena kecepatan pemrosesan yang tinggi dan kemudahan implementasinya, yang sangat diinginkan dalam pemecahan masalah dunia nyata (Martarelli & Nagano, 2019). Dalam konteks Kalimantan Barat, pendekatan ini belum banyak diterapkan untuk analisis usaha perkebunan perorangan tingkat kabupaten/kota.

Oleh karena itu, diperlukan pendekatan kuantitatif yang mampu mengelompokkan wilayah berdasarkan kesamaan karakteristik usaha perkebunan perorangan. Salah satu pendekatan yang dapat digunakan adalah analisis kluster, yang bertujuan untuk membentuk kelompok wilayah yang relatif homogen di dalam kluster dan heterogen antarkluster. Melalui pendekatan ini, pola pengelompokan wilayah dapat diidentifikasi secara lebih sistematis dan objektif.

Meskipun data Usaha Pertanian Perorangan (UTP) tanaman perkebunan telah tersedia secara rinci melalui Sensus Pertanian 2023 yang diselenggarakan oleh Badan Pusat Statistik, pemanfaatan data tersebut masih didominasi oleh penyajian statistik deskriptif. Kajian yang mengelompokkan kabupaten/kota berdasarkan kesamaan karakteristik usaha perkebunan perorangan dengan menggunakan metode klasterisasi kuantitatif masih relatif terbatas, khususnya di Provinsi Kalimantan Barat, sehingga belum mengeksplorasi secara mendalam karakteristik serta kondisi masing-masing kluster yang terbentuk. Selain itu, evaluasi kualitas hasil klasterisasi menggunakan ukuran validitas kluster, seperti *Silhouette Index*, *Davies-Bouldin Index*, dan *Dunn Index*, juga belum banyak diterapkan dalam analisis usaha perkebunan perorangan di tingkat wilayah. Sehingga, penelitian ini memberikan kontribusi baru berupa pemetaan kluster usaha perkebunan perorangan (UTP) tingkat kabupaten/kota Kalimantan Barat berbasis Sensus Pertanian 2023 dengan validasi kuantitatif *Silhouette Index*, *Davies-Bouldin Index*, dan *Dunn Index*, yang belum banyak dikaji pada penelitian sebelumnya.

Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengidentifikasi pola pengelompokan kabupaten/kota di Provinsi Kalimantan Barat berdasarkan karakteristik Usaha Pertanian Perorangan (UTP) tanaman perkebunan menggunakan algoritma *K-Means*. Keandalan hasil pengelompokan diperkuat melalui validasi internal menggunakan *Silhouette Index*, *Davies-Bouldin Index*, dan *Dunn Index* dalam menentukan jumlah kluster yang paling representatif. Melalui pengelompokan tersebut, penelitian ini diharapkan dapat mengungkap perbedaan karakteristik perkebunan antarwilayah sebagai dasar penyusunan kebijakan, penetapan prioritas pengembangan komoditas, dan optimalisasi potensi subsektor perkebunan di Kalimantan Barat.

METODOLOGI

Penelitian ini menggunakan data sekunder yang bersumber dari publikasi resmi Badan Pusat Statistik (BPS), yaitu “Hasil Pencacahan Lengkap Sensus Pertanian 2023 Usaha Pertanian Perorangan (UTP) Tanaman Perkebunan Tahap II”. Unit analisis dalam penelitian ini adalah kabupaten/kota di Provinsi Kalimantan Barat. Variabel yang digunakan dalam penelitian ini terdiri dari 10 variabel yaitu (X1) jumlah UTP cengkeh, (X2) jumlah UTP kakao, (X3) jumlah UTP karet, (X4) jumlah UTP kelapa sawit, (X5) jumlah UTP kelapa, (X6) jumlah UTP kemiri, (X7) jumlah UTP kopi, (X8) jumlah UTP lada, (X9) jumlah UTP pinang/jambe, dan (X10) jumlah UTP teh.

Penelitian ini menggunakan analisis kluster sebagai metode untuk mengelompokkan data. Klasterisasi berfungsi untuk mengidentifikasi kesamaan pola dan karakteristik antardata sehingga terbentuk kelompok-kelompok yang homogen (Margareta dkk., 2025). Hasil dari metode pada penelitian ini dievaluasi validitasnya menggunakan beberapa validasi kluster, yakni *Silhouette Index*, *Davies Bouldin Index*, dan *Dunn Index* untuk memperkuat validitas dari jumlah kluster yang digunakan. Namun, penelitian ini memiliki keterbatasan berupa jumlah unit analisis yang relatif kecil, yaitu hanya 14 kabupaten/kota di Kalimantan Barat.

Tahap Penelitian

Penelitian ini mencakup serangkaian tahapan yang ditempuh peneliti untuk memperoleh solusi terhadap suatu permasalahan secara ilmiah dan terstruktur. Adapun tahapan-tahapan penelitian tersebut dijelaskan sebagai berikut:

1. Data diambil dalam bentuk jumlah usaha perorangan per komoditas. Dilakukan *pre-processing data* seperti pengecekan *missing value*, duplikat data, *outlier*, uji multikolinearitas, serta standardisasi data. Pada penelitian ini, standardisasi data dilakukan menggunakan metode Z-Score.
2. Dilakukan klasterisasi dengan *K-Means*. Jumlah klaster (*k*) diuji pada nilai *k* = 2, 3, dan 4.
3. Validasi klaster yang dilakukan untuk menentukan jumlah klaster yang optimal dengan menggunakan *Silhouette Index*, *Davies Bouldin Index*, dan *Dunn Index*.
4. Profil klaster dianalisis berdasarkan nilai setiap variabel. Hasil klaster kemudian dikaitkan dengan kondisi riil dan potensi ekonomi daerah.

Metode Non Hirarki

Metode non-hirarki merupakan pendekatan yang digunakan untuk mengelompokkan objek ke dalam klaster, di mana jumlah klaster yang ingin dibentuk telah ditetapkan terlebih dahulu (Margareta dkk., 2025). Pada metode ini, peneliti menentukan jumlah klaster awal, misalnya dua, tiga, dan seterusnya, sebelum proses pengelompokan dimulai. Setelah itu, setiap objek data dialokasikan ke klaster yang paling sesuai tanpa mengikuti tahapan bertingkat seperti pada metode hirarki (Margareta dkk., 2025). Keunggulan metode klaster non-hirarki adalah kemampuannya menganalisis sampel berukuran besar secara efisien serta memiliki kerentanan yang relatif rendah terhadap keberadaan *outlier*.

Algoritma K-Means

Algoritma *K-Means* merupakan algoritma yang digunakan untuk mengelompokkan data ke dalam sejumlah klaster, di mana setiap klaster memiliki karakteristik yang berbeda satu sama lain (Nasari dkk., 2023). Adapun tahapan dalam algoritma *K-Means* adalah sebagai berikut:

1. Menentukan jumlah klaster. Jumlah klaster yang akan diuji coba dalam penelitian ini adalah jumlah klaster 2, jumlah klaster 3 dan jumlah klaster 4.
2. Membangkitkan *centroid* awal. Untuk algoritma *K-Means clustering*, *centroid* awal dipilih secara acak sebanyak jumlah klaster yang digunakan.
3. Menghitung jarak ke masing-masing *centroid* 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* (Nasari dkk., 2023).

4. Pengelompokkan data klaster.
5. Apabila data klaster 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 *centroid* baru yang diperoleh dengan menggunakan formula berikut ini:

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

Dimana:

C: *Centroid* data

m: anggota data yang termasuk ke dalam *centroid* tertentu

n: jumlah data yang menjadi anggota *centroid* tertentu (Nasari dkk., 2023).

Standardisasi Z-Score

Menurut Wahidah dkk. (2023), *Z-Score* merupakan nilai baku atau standar nilai dari suatu data. Rumus dari *Z-Score* yakni sebagai berikut.

$$z_i = \frac{x_i - \bar{x}}{S}$$

dimana,

z_i : nilai *Z-Score* ke- i

x_i : nilai atau objek ke- i

$\bar{x}$: rata-rata sampel

S : standar deviasi

Silhouette Index

Silhouette Index merupakan salah satu metode evaluasi yang digunakan untuk menilai kualitas hasil klusterisasi. Indeks ini menilai seberapa baik setiap objek terkelompok dalam kluster yang tepat. Nilai skor mendekati +1 menunjukkan bahwa objek berada dalam kluster yang sangat sesuai, sedangkan nilai mendekati -1 menunjukkan kemungkinan salah kluster (Rhomaningtias dkk., 2025). Menurut Putra dkk. (2024), *Silhouette Index* memiliki persamaan yakni sebagai berikut.

$$S_{(i)} = \frac{b_i + a_i}{\max(a(i), b(i))},$$

$$b(i) = \min d(i, c), \text{ dan } a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j)$$

dimana,

S_{max} : jumlah jarak maksimum dalam kluster dan antarkluster

S_{min} : jumlah jarak minimum dalam kluster dan antarkluster

Davies Bouldin Index

Davies Bouldin Index (DBI) merupakan salah satu metode evaluasi yang digunakan untuk menilai kualitas hasil klusterisasi. Indeks ini berfungsi sebagai indikator kinerja untuk melihat sejauh mana pembentukan kluster sudah optimal, dengan mempertimbangkan tingkat kemiripan antaranggota dalam satu kluster serta perbedaan antarkluster (Awaliyah dkk., 2024). Nilai DBI yang lebih kecil menunjukkan bahwa algoritma pengelompokan bekerja dengan lebih efektif (Farahdinna dkk., 2019). Adapun tahapan perhitungan DBI adalah sebagai berikut.

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

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

Dimana:

R_{ij} : nilai R pada kluster i dan j

d_{ij} : jarak antara *centroid* untuk kluster i dan j

s_i dan s_j : nilai rata-rata jarak antara titik-titik kluster i dengan *centroid* kluster j

3. Tentukan hasil akhir dari perhitungan DBI.

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

Dimana:

R_{ij} : nilai R antara kluster i dan kluster j

k = banyaknya kluster

(Ishak & Bengnga, 2024).

Dunn Index

Dunn Index merupakan salah satu metode evaluasi yang digunakan untuk menilai kualitas hasil klasterisasi. Indeks ini mengukur rasio antara jarak minimum antarklaster dan diameter maksimum dalam klaster. Semakin tinggi nilai indeks, semakin baik kualitas pemisahan antarklaster (Rhomaningtias dkk., 2025).

HASIL DAN PEMBAHASAN

Pengumpulan dan *Preprocessing* Data

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

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

Kabupaten/ Kota	Tanaman Perkebunan									
	Cengkeh (X_1)	Kakao (X_2)	Karet (X_3)	Kelapa Sawit (X_4)	Kelapa (X_5)	Kemiri (X_6)	Kopi (X_7)	Lada (X_8)	Pinang/ Jambe (X_9)	Teh (X_{10})
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Sambas	74	17	30.756	30.531	18.784	18	326	2.831	2.500	1
Bengkayang	290	40	14.961	12.806	2.850	538	20	3.278	611	0
Landak	3	12	24.787	48.375	52	28	15	1.465	57	0
Mempawah	4	42	3.556	6.056	7.664	4	75	79	2.909	0
Sanggau	4	21	27.326	56.681	110	0	9	2.514	121	0
Ketapang	1	4	13.969	43.138	1.515	1	18	10	252	0
Sintang	2	40	55.639	23.935	369	11	26	566	225	0
Kapuas Hulu	1	5	20.636	3.140	190	0	168	25	14	1
Sekadau	1	18	23.118	26.991	148	2	16	194	104	0
Melawi	0	3	25.367	7.853	295	1	8	40	55	0
Kayong Utara	160	2	2.601	9.290	4.013	6	510	120	941	2
Kubu Raya	0	2	12.218	16.091	16.198	10	581	261	7.076	0
Pontianak	2	1	30	286	186	0	9	7	74	0
Singkawang	61	111	2.123	1.003	1.045	69	4	105	233	0

Sumber: BPS Provinsi Kalimantan Barat, diolah.

Tabel 1 menunjukkan hasil Sensus Pertanian 2023 yakni Usaha Pertanian Perorangan (UTP) tanaman perkebunan di Kalimantan Barat untuk 11 komoditas, dengan pola umum bahwa kelapa sawit merupakan komoditas yang paling dominan di banyak kabupaten, terutama di Sanggau, Ketapang, dan Sekadau. Komoditas karet juga memiliki skala usaha besar, khususnya di Sintang dan Sanggau, sementara kopi tersebar di banyak wilayah tetapi dengan jumlah unit usaha yang lebih kecil. Beberapa wilayah seperti Sambas, Landak, Sintang, dan Kubu Raya memiliki keragaman komoditas yang tinggi karena hampir semua komoditas diusahakan, meskipun dalam jumlah yang bervariasi. Sebaliknya, kota Pontianak dan Singkawang menunjukkan skala usaha perkebunan yang kecil karena karakteristiknya sebagai wilayah perkotaan. Komoditas tertentu seperti teh, pinang/jambe, dan cengkeh hanya diusahakan dalam jumlah sangat terbatas dan tidak merata antarwilayah. Secara keseluruhan, tabel ini memperlihatkan bahwa struktur perkebunan Kalimantan Barat sangat dipengaruhi oleh dominasi kelapa sawit dan karet, sementara komoditas lain hanya berkembang signifikan pada kabupaten tertentu.

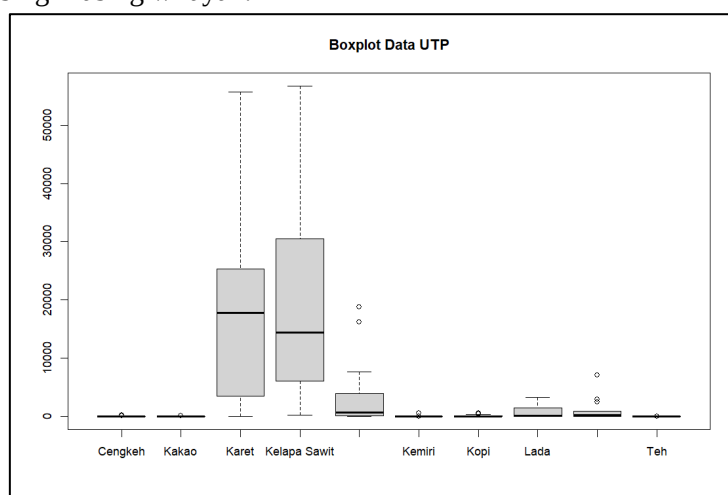
Tabel di bawah ini menunjukkan ringkasan statistik jumlah usaha pertanian perorangan (UTP) di Provinsi Kalimantan Barat berdasarkan hasil Sensus Pertanian 2023.

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

Variabel	Nilai Minimum	Rata-Rata	Nilai Maksimum
(1)	(2)	(3)	(4)
Cengkeh	0	43,07	290
Kakao	1	22,71	111
Karet	30	18.363,36	55.639
Kelapa Sawit	286	20.441,14	56.681
Kelapa	52	3.815,64	18.784
Kemiri	0	49,14	538
Kopi	4	127,50	581
Lada	7	821,07	3.278
Pinang/Jambe	14	1.083,71	7.076
Teh	0	0,28	2

Sumber: BPS Provinsi Kalimantan Barat, diolah.

Tabel 2 menunjukkan hasil Sensus Pertanian 2023 dan telah diolah yang menyajikan nilai minimum, rata-rata, serta maksimum jumlah Usaha Pertanian Perorangan (UTP) tanaman perkebunan di Kalimantan Barat menurut jenis komoditas. Secara umum, kelapa sawit dan karet merupakan komoditas dengan skala usaha terbesar, terlihat dari nilai maksimum yang sangat tinggi, masing-masing mencapai 56.681 dan 55.639 unit usaha, jauh di atas komoditas lainnya. Sementara itu, komoditas seperti teh, cengkeh, dan kemiri memiliki nilai minimum nol, menunjukkan bahwa beberapa wilayah tidak mengusahakannya sama sekali. Rata-rata unit usaha tertinggi terdapat pada kelapa sawit (20.441,14) dan karet (18.363,36), sedangkan rata-rata terendah terdapat pada teh (0,28). Data ini menggambarkan perbedaan signifikan antarkomoditas, dengan beberapa komoditas menjadi basis utama perkebunan di banyak kabupaten, sementara yang lain berkembang lebih terbatas dan hanya muncul pada wilayah tertentu. 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.



Gambar 1. Boxplot Data UTP Tanaman Perkebunan Menurut Kab/Kota Kalimantan Barat Tahun 2023

Sumber: BPS Provinsi Kalimantan Barat, diolah.

Selanjutnya, dilakukan tahap pengecekan *missing* dan duplikat data. Pada data Jumlah Usaha Pertanian Perorangan (UTP) Tanaman Perkebunan menurut Kabupaten/Kota di Kalimantan Barat dan Komoditasnya (Unit Usaha) Tahun 2023, tidak ditemukannya *missing* ataupun duplikat data, sehingga bisa lanjut ke proses selanjutnya yakni deteksi *outlier*. Berdasarkan hasil visualisasi boxplot, diperoleh bahwa terdapat beberapa variabel yang terindikasi outlier.

Tabel 3. Hasil Deteksi *Outlier*

Variabel	Indikasi <i>Outlier</i>	Keputusan
(1)	(2)	(3)
X_1	Ya	Dipertahankan
X_2	Ya	Dipertahankan
X_3	Tidak	Dipertahankan
X_4	Tidak	Dipertahankan
X_5	Ya	Dipertahankan
X_6	Ya	Dipertahankan
X_7	Ya	Dipertahankan
X_8	Tidak	Dipertahankan
X_9	Ya	Dipertahankan
X_{10}	Ya	Dipertahankan

Sumber: BPS Provinsi Kalimantan Barat, diolah.

Meskipun terdapat 2 variabel yang terindikasi *outlier*, nilai ekstrem yang terindikasi tersebut tetap dipertahankan dalam penelitian ini. Keputusan tersebut diambil dengan mempertimbangkan bahwa nilai-nilai tersebut mencerminkan kondisi riil dan keragaman karakteristik antarkabupaten/kota di Kalimantan Barat, sehingga penghapusan *outlier* berpotensi menghilangkan informasi penting dalam pembentukan klaster wilayah.

Selanjutnya, uji multikolinearitas dilakukan untuk mengetahui hubungan linear antarvariabel yang digunakan dalam penelitian ini.

Tabel 4. Uji Multikolinearitas Data Jumlah Usaha Pertanian Perorangan (UTP) Tanaman Perkebunan menurut Kabupaten/Kota di Kalimantan Barat dan Komoditasnya (Unit Usaha) Tahun 2023

Variabel	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
X_1	1,000	0,204	-0,216	-0,211	0,096	0,852	0,162	0,542	-0,052	0,331
X_2	0,204	1,000	-0,097	-0,230	-0,115	0,273	-0,338	0,067	-0,130	-0,276
X_3	-0,216	-0,097	1,000	0,474	-0,051	-0,085	-0,192	0,312	-0,179	-0,142
X_4	-0,211	-0,230	0,474	1,000	-0,031	-0,135	-0,169	0,466	-0,120	-0,201
X_5	0,096	-0,115	-0,051	-0,031	1,000	-0,040	0,718	0,259	0,839	0,238
X_6	0,852	0,273	-0,085	-0,135	-0,040	1,000	-0,166	0,598	-0,072	-0,147
X_7	0,162	-0,338	-0,192	-0,169	0,718	-0,166	1,000	-0,062	0,736	0,636
X_8	0,542	0,067	0,312	0,466	0,259	0,598	-0,062	1,000	-0,020	-0,019
X_9	-0,052	-0,130	-0,179	-0,120	0,839	-0,072	0,736	-0,020	1,000	0,003
X_{10}	0,331	-0,276	-0,142	0,201	0,238	-0,147	0,636	-0,019	0,003	1,000

Sumber: BPS Provinsi Kalimantan Barat, diolah.

Dari hasil uji multikolinearitas, terlihat bahwa terdapat pasangan variabel yang menunjukkan korelasi relatif tinggi yakni dengan nilai korelasi di atas ambang batas umum sebesar 0,8. Dengan demikian, variabel tersebut dieliminasi agar data memenuhi asumsi dalam metode *k-means* dan menghasilkan

proses klustering yang lebih baik. Setelah variabel dieliminasi, data dengan variabel terbaru di uji kembali multikolinearitasnya. Sehingga didapat seperti tabel 5 dibawah.

Tabel 5. Uji Multikolinearitas Setelah Hapus Variabel X_6 & X_9 Data Jumlah Usaha Pertanian Perorangan (UTP) Tanaman Perkebunan menurut Kabupaten/Kota di Kalimantan Barat dan Komoditasnya (Unit Usaha) Tahun 2023

Variabel	X_1	X_2	X_3	X_4	X_5	X_7	X_8	X_{10}
(1)	(2)	(3)	(4)	(5)	(6)	(8)	(9)	(11)
X_1	1,0000	0,2049	-0,2167	-0,2115	0,0963	0,1628	0,5425	0,3314
X_2	0,2049	1,0000	-0,0973	-0,2300	-0,1151	-0,3386	0,0675	-0,2766
X_3	-0,2167	-0,0973	1,0000	0,4741	-0,0517	-0,1926	0,3128	-0,1426
X_4	-0,2115	-0,2300	0,4741	1,0000	-0,0314	-0,1697	0,4660	-0,2015
X_5	0,0963	-0,1151	-0,0517	-0,0314	1,0000	0,7188	0,2591	0,2386
X_7	0,1628	-0,3386	-0,1926	-0,1697	0,7188	1,0000	-0,0623	0,6364
X_8	0,5425	0,0675	0,3128	0,4660	0,2591	-0,0623	1,0000	-0,0199
X_{10}	0,3314	-0,2766	-0,1426	0,2015	0,2386	0,6364	-0,0199	1,0000

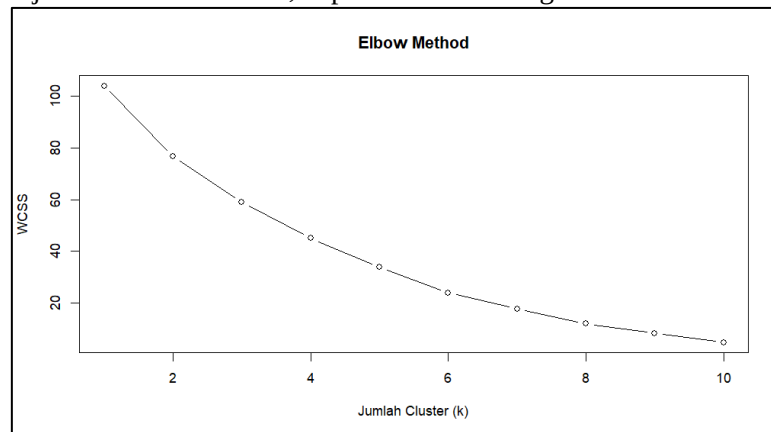
Sumber: BPS Provinsi Kalimantan Barat, diolah.

Berdasarkan hasil pengujian setelah dieliminasi, nilai korelasi seluruh variabel masih berada di bawah ambang batas umum sebesar 0,8. Dengan demikian, dapat disimpulkan bahwa tidak terdapat indikasi multikolinearitas yang serius pada variabel penelitian. Sehingga seluruh variabel tetap dipertahankan dan digunakan pada tahap analisis kluster. Tahap terakhir yakni standardisasi data menggunakan metode z-score.

Dalam penyusunan kluster perlu memperhatikan jarak. Metode penghitungan yang digunakan pada penelitian ini adalah *Euclidean distance*. Pemilihan metode ini didasarkan pada tujuan analisis kluster dengan menjadikan besaran nilai setiap variabel sebagai penjelas karakteristik kabupaten/kota di Kalimantan Barat.

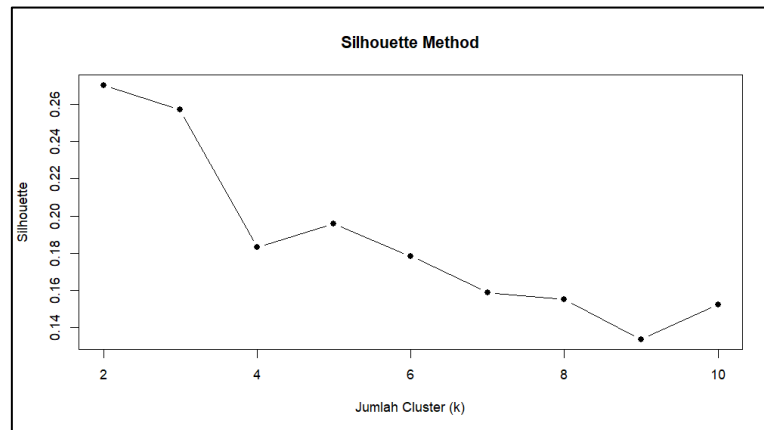
Uji Validitas

Dalam *K-Means*, diperlukan jumlah kluster yang diinginkan lalu selanjutnya obyek digabungkan sesuai keinginan jumlah kluster tersebut. Pada penelitian ini, jumlah kluster mulai dari $k = 2$ hingga $k = 4$. Pembatasan jumlah kluster hingga maksimum empat dipilih karena penggunaan kluster yang terlalu banyak dapat menyulitkan interpretasi hasil klusterisasi. Dengan jumlah kluster yang lebih sedikit dan terfokus, proses penarikan kesimpulan mengenai karakteristik atau pola pada tiap kluster menjadi lebih sederhana, sekaligus menghasilkan informasi yang lebih relevan dan mudah dipahami. Untuk memvalidasi jumlah kluster terbaik, dapat dilakukan dengan metode Elbow dan Silhouette.



Gambar 2. K Optimal Menggunakan Metode *Elbow*

Sumber: BPS Provinsi Kalimantan Barat, diolah.



Gambar 3. K Optimal Menggunakan Metode *Silhouette*

Sumber: BPS Provinsi Kalimantan Barat, diolah.

Dari Gambar 2 dan Gambar 3, terlihat bahwa jumlah kluster optimal $k = 2$. Hal ini dapat dilihat pada metode *Elbow* menunjukkan titik siku pada $k = 2$ yang menandakan penurunan variasi dalam kluster mulai melambat, dan diperkuat pula oleh metode *Silhouette* yang memberikan nilai tertinggi pada $k = 2$. Namun, untuk memastikan pemilihan jumlah kluster sudah optimal, dilakukan validasi kluster. Dari validasi menggunakan *Silhouette Index*, *Davies Bouldin Index*, dan *Dunn Index*, didapat nilai sebagai berikut.

Tabel 6. Validasi k Optimal *K-Means* Menggunakan 3 Index

k	<i>Silhouette Index</i>	<i>Davies Bouldin Index</i>	<i>Dunn Index</i>
(1)	(2)	(3)	(4)
k = 2	0,2703791	1,344479	0,6484975
k = 3	0,2569720	1,219100	0,5812994
k = 4	0,1831029	1,042349	0,2582777

Sumber: BPS Provinsi Kalimantan Barat, diolah.

Meskipun *Davies Bouldin Index* menunjukkan jumlah kluster optimal yakni $k = 4$, nilai *Silhouette* yang menurun menunjukkan bahwa kualitas pemisahan antarkluster justru memburuk. Oleh karena itu, $k = 2$ tetap menjadi pilihan optimal, karena memberikan keseimbangan terbaik antara pemisahan yang jelas antarkluster dan konsistensi internal.

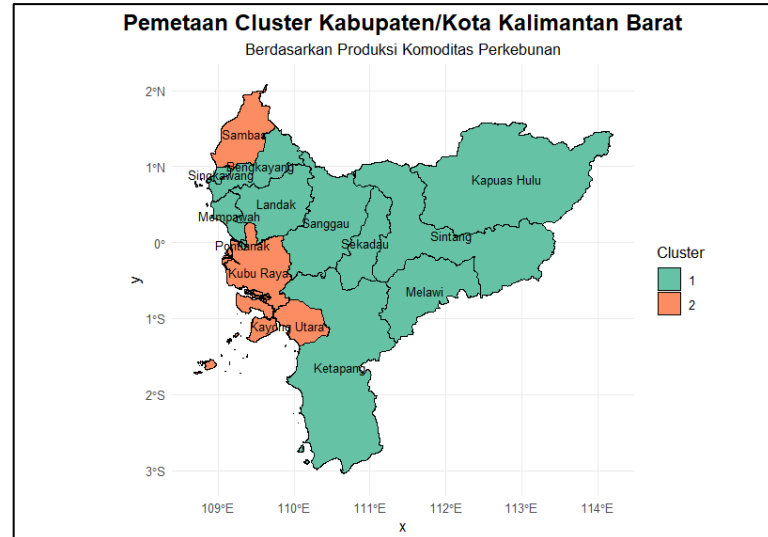
Tabel 7. Hasil Klasterisasi Algoritma *K-Means*

Jumlah k	Klaster	
	1	2
(1)	(2)	(3)
k = 2	3	11

Sumber: BPS Provinsi Kalimantan Barat, diolah.

Profil Kluster

Gambar dan tabel di bawah ini menunjukkan pemetaan serta ringkasan statistik jumlah usaha pertanian perorangan (UTP) tanaman perkebunan di Provinsi Kalimantan Barat berdasarkan hasil Sensus Pertanian 2023.



Gambar 4. Pemetaan Kluster Berdasarkan Kabupaten/Kota Provinsi Kalimantan Barat
Sumber: BPS Provinsi Kalimantan Barat, diolah

Tabel 8. Rata-rata Variabel Berdasarkan Kluster Pada Algoritma *K-Means*

Variabel	Kluster 1	Kluster 2
(1)	(2)	(3)
Cengkeh	33,5	78
Kakao	27	7
Karet	19.228	15.192
Kelapa Sawit	20.933	18.637
Kelapa	1.311	12.998
Kopi	33,5	472
Lada	753	1.071
Teh	0,0909	1

Sumber: BPS Provinsi Kalimantan Barat, diolah.

Dari Tabel 8, didapat bahwa masing-masing kluster memiliki pola komoditas yang berbeda berdasarkan jumlah Usaha Pertanian Perorangan (UTP). Kluster 1 terdiri atas 11 kabupaten/kota, yaitu Bengkayang, Landak, Mempawah, Sanggau, Ketapang, Sintang, Kapuas Hulu, Sekadau, Melawi, Pontianak, dan Singkawang, memiliki karakteristik utama berupa jumlah UTP tanaman perkebunan yang relatif tinggi pada komoditas karet dan kelapa sawit, dengan rata-rata masing-masing sebesar 19.228 dan 20.933 UTP. Selain itu, komoditas kelapa sawit menjadi komoditas yang paling dominan dibandingkan komoditas lainnya. Kondisi ini mencerminkan karakteristik daerah dengan luas lahan perkebunan sawit yang signifikan, banyaknya petani maupun perusahaan yang terlibat, serta rantai pasok sawit yang sudah terbentuk. Kluster ini cenderung lebih monokultur dengan ketergantungan yang kuat pada satu komoditas, sehingga perkembangan ekonomi daerahnya sangat dipengaruhi oleh pasar minyak sawit.

Sementara itu, jumlah UTP pada komoditas cengkeh, kakao, kopi, lada, dan teh cenderung lebih rendah. Oleh karena itu, kluster ini dapat dikategorikan sebagai wilayah dengan dominasi UTP pada

komoditas karet dan kelapa sawit. Sedangkan klaster 2 terdiri atas 3 kabupaten/kota, yaitu Sambas, Kayong Utara, dan Kubu Raya. Klaster ini ditandai oleh jumlah UTP yang relatif tinggi pada komoditas kelapa, dengan rata-rata sebesar 12.998 UTP, serta memiliki jumlah UTP yang lebih tinggi pada komoditas cengkeh, kakao, kopi, lada, dan teh dibandingkan Klaster 1. Meskipun rata-rata UTP kelapa sawit masih cukup besar, nilai tersebut lebih rendah dibandingkan Klaster 1. Dengan demikian, klaster ini dapat diidentifikasi sebagai wilayah yang memiliki keragaman komoditas perkebunan yang lebih tinggi dengan dominasi UTP pada komoditas kelapa.

Perbedaan tajam antara komoditas utama dan komoditas minor mencerminkan struktur perkebunan yang tidak merata antarkomoditas, sehingga klaster terbentuk terutama berdasarkan besarnya skala produksi komoditas unggulan. Hal ini memperlihatkan bahwa perbedaan intensitas usaha pada komoditas tertentu menjadi faktor dominan dalam pemisahan kelompok oleh algoritma *K-Means*.

KESIMPULAN DAN SARAN

Berdasarkan hasil analisis *K-Means*, diperoleh jumlah klaster optimal sebanyak 2 klaster berdasarkan evaluasi menggunakan *Silhouette Index*, *Davies-Bouldin Index*, dan *Dunn Index*. Klaster 1 terdiri atas 11 kabupaten/kota yang didominasi oleh UTP komoditas karet dan kelapa sawit, sedangkan Klaster 2 terdiri atas 3 kabupaten/kota yang didominasi oleh UTP komoditas kelapa dan memiliki keragaman komoditas perkebunan yang lebih tinggi. Hasil pengelompokan menunjukkan adanya perbedaan karakteristik UTP tanaman perkebunan antarwilayah di Provinsi Kalimantan Barat. Informasi tersebut dapat menjadi dasar dalam penyusunan strategi pengembangan subsektor perkebunan yang lebih sesuai dengan potensi dan karakteristik masing-masing wilayah.

Namun demikian, penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Variabel yang digunakan dalam proses klasterisasi masih terbatas pada kuantitas Usaha Pertanian Perorangan (UTP) per komoditas, sehingga belum sepenuhnya merepresentasikan kinerja dan kondisi usaha perkebunan secara menyeluruh. Penelitian ini juga belum mempertimbangkan variabel penting lainnya, seperti produktivitas, luas lahan, dan nilai produksi, yang berpotensi memberikan gambaran yang lebih akurat mengenai struktur dan tingkat perkembangan subsektor perkebunan antarwilayah. Sehingga, penelitian selanjutnya disarankan untuk menggunakan variabel yang lebih luas, agar dapat menganalisis dinamika perkembangan usaha perkebunan secara lebih mendalam. Selain itu, penerapan dan perbandingan metode klasterisasi lain juga dapat dilakukan untuk memperoleh hasil pengelompokan yang lebih *robust* dan memperkaya perspektif analisis dalam mendukung perumusan kebijakan pengembangan subsektor perkebunan di tingkat daerah.

DAFTAR PUSTAKA

- Afiantoro, F., Suhartati, T., & Kifli, F. W. (2025). Strategi Manajemen Lanskap Berkelanjutan dalam Perkebunan Kelapa Sawit untuk Meningkatkan Kesejahteraan Masyarakat Lokal. *Flora: Jurnal Kajian Ilmu Pertanian dan Perkebunan*, 2(3), 01-17.
- Awaliyah, L., Rahaningsih, N., & Dana, R. D. (2024). Implementasi Algoritma K-Means Dalam Analisis Cluster Korban Kekerasan Di Provinsi Jawa Barat. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 188-195.
- BPS Provinsi Kalimantan Barat. (2025). *Provinsi Kalimantan Barat Dalam Angka 2025*.
- BPS Provinsi Kalimantan Barat. (2024). *Hasil Pencacahan Lengkap Sensus Pertanian 2023 -Tahap II:Usaha Pertanian Perorangan (UTP) Tanaman Perkebunan Provinsi Kalimantan Barat*.
- BPS Provinsi Kalimantan Barat. (2024). *Hasil Survei Ekonomi Pertanian (SEP) 2024 Provinsi Kalimantan Barat*.
- Farahdinna, F., Nurdiansyah, I., Suryani, A., & Wibowo, A. (2019). Perbandingan algoritma k-means dan k-medoids dalam klasterisasi produk asuransi perusahaan nasional. *Jurnal Ilmiah Fifi*, 11(2), 208-214.
- Ferdiansyah, F. R., Nugraha, R. W., Sofian, R., Purwanto, H., Saepudin, D., & Andriansyah, E. (2024). Implementation of K-Means and DBSCAN algorithms: A bibliometric review. Proceedings of the Widyatama International Conference on Engineering 2024 (WICOENG 2024), *Advances in Engineering Research*, 192–202. https://doi.org/10.2991/978-94-6463-618-5_21.
- Hilyawan, F. R., Syafrullah, M., & Sudija, I. (2025). A Systematic Literature Review on the Optimization of K-Means and Agglomerative Clustering for Student Performance Segmentation: A Comparative Analysis of Elbow and Silhouette Methods. *Edcomtech: Jurnal Kajian Teknologi Pendidikan*, 10(1), 73–88. <https://doi.org/10.17977/um039v10i12025p73-88>.
- Ikotun, A. M., Almutari, M. S., & Ezugwu, A. E. (2021). K-means-based nature-inspired metaheuristic algorithms for automatic data clustering problems: Recent advances and future directions. *Applied Sciences*, 11(23), 11246.
- Ishak, R., & Bengnga, A. (2024). Clustering Prestasi Akademik Lulusan Menggunakan Metode K-Means. *Jambura Journal of Electrical and Electronics Engineering*, 6(1), 76-81.
- Margareta, T., Satyahadewi, N., & Pertiwi, R. (2025). Perbandingan Kinerja Algoritma K-Means dan K-Medoids Pada Pengelompokan Usaha Pertanian Perorangan Tanaman Pangan di Provinsi Kalimantan Barat. In *Jurnal Forum Analisis Statistik (FORMASI)* (Vol. 5, No. 1, pp. 35-46).
- Martarelli, N. J., & Nagano, M. S. (2019). Agrupamento de dados mistos baseados no algoritmo k-means: Uma revisão sistemática da literatura. In P. Miranda, C. Costa, P. Isaías, & F. M. Santoro (Eds.), *IADIS Conferencia Ibero Americana Computação Aplicada 2019* (pp. 199–206). IADIS. *IADIS Digital Library*.
- Nafri, T. C., Pinasty, S., Dini, S. K., & Nurcahayani, H. (2025). Studi Klasterisasi Usaha Pertanian Perorangan di Kabupaten Bantul Tahun 2023 dengan Pendekatan Hirarki: Studi Klasterisasi Usaha Pertanian Perorangan di Kabupaten Bantul Tahun 2023 dengan Pendekatan Hirarki. *Emerging Statistics and Data Science Journal*, 3(1), 462-471.
- Nasari, F., Tanjung, D. H., & Handayani, F. (2023). Optimasi Metode K-Means dan K-Medoids Berdasarkan Jumlah Cluster dan Nilai DBI Dalam Pengelompokan Produksi Kelapa Sawit Di Provinsi Riau. *INFOSYS (Information System) Journal*, 7(2), 129-141.
- Natalia, N., Kusriani, N., & Oktoriana, S. (2025). ANALISIS PENENTUAN KOMODITAS UNGGULAN SUB SEKTOR PERTANIAN TANAMAN PANGAN DI KABUPATEN BENGKAYANG. *Jurnal Sains Pertanian Equator*, 14(3), 607-615.

- Putra, A., Abdullah, D., & Daud, M. (2024). Klasterisasi kualitas biji kopi berdasarkan taraf penyusutan menggunakan metode k-harmonic means dengan validasi *silhouette index* dan c-index. *Jurnal Janitra Informatika dan Sistem Informasi*, 4(2), 74-86.
- Rhomaningtias, L., Kusharyadi, M. N., Jatindra, R. W. S., & Nasrudin, M. (2025). ANALISIS PERBANDINGAN METODE K-MEDOID DAN AGGLOMERATIVE HIERARCHICAL CLUSTERING PADA DATA KONSUMSI REMPAH-REMPAH DI KABUPATEN/KOTA. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3).
- Wahidah, N., Juwita, O., & Arifin, F. N. (2023). Pengelompokan Daerah Rawan Bencana di Kabupaten Jember Menggunakan Metode K-Means Clustering. *INFORMAL*, 8(1), 22-29.