

Application of K-Means Clustering for Grouping Plantation Production in West Pasaman Regency in 2024

Dini Andita Putri, Fitri Mudia Sari*, dan Chairina Wirdiastuti

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: fitrimudiasari@fmipa.unp.ac.id

Submitted : 08 Agustus 2025

Revised : 17 November 2025

Accepted : 01 Desember 2025

ABSTRACT

The plantation sector plays a strategic role in supporting the economy of West Pasaman Regency, but the uneven distribution of production across subdistricts has not been systematically grouped to identify areas with similar characteristics. This study analyzes the main plantation commodities, namely oil palm, coconut, rubber, cocoa, and patchouli, using the K-Means Clustering method, with the optimal number of clusters determined using the Elbow Method. The analysis produced three clusters: the first cluster consists of subdistricts with moderate and relatively varied production levels, the second cluster is dominated by rubber and cocoa, and the third cluster consists only of Kinali Subdistrict, which shows much higher production of oil palm, coconut, and patchouli. These findings indicate that K-Means Clustering is effective for mapping regional plantation profiles and addressing production imbalances. These results provide a practical basis for targeted policy formulation, more efficient resource allocation, and support for sustainable agricultural development, while contributing to empirical evidence on the application of clustering techniques in regional commodity analysis.

Keywords: Clustering, K-Means, Plantation, Production, West Pasaman



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Sektor perkebunan berperan strategis dalam pembangunan ekonomi Kabupaten Pasaman Barat karena menjadi penyumbang utama Produk Domestik Regional Bruto (PDRB) sekaligus sumber penghidupan masyarakat. Kontribusinya mencapai 38,88% terhadap PDRB, jauh melampaui sektor industri pengolahan maupun perdagangan yang masing-masing hanya 13,23% (BPS Kabupaten Pasaman Barat, 2024). Potensi ini ditopang oleh pemanfaatan 164.519 hektare lahan perkebunan atau 42,32% dari total wilayah, menjadikannya penggunaan lahan terbesar. Komoditas utama meliputi kelapa sawit, kelapa, karet, kakao, dan nilam, dengan kontribusi signifikan pada tingkat provinsi, yaitu 52,27% produksi kelapa sawit, 10,45% kakao, dan 85,12% nilam dari total produksi Sumatera Barat (BPS Kabupaten Pasaman Barat, 2025). Kondisi tersebut menegaskan bahwa Pasaman Barat tidak hanya vital di tingkat kabupaten, tetapi juga strategis dalam konteks pembangunan perkebunan provinsi.

Meskipun demikian, ketimpangan produksi antar kecamatan masih menjadi tantangan utama. Sebagian wilayah mencatat hasil sangat tinggi pada komoditas tertentu, sementara wilayah lain rendah atau bahkan tidak berproduksi. Kecamatan Kinali misalnya, menghasilkan kelapa sawit terbesar (72.473,15 ton), sementara Kecamatan Sasak Ranah Pasisie menonjol pada kelapa (269,80 ton) tetapi tidak menghasilkan karet (BPS Kabupaten Pasaman Barat, 2025). Perbedaan potensi tersebut memerlukan analisis berbasis data yang mampu mengelompokkan wilayah secara objektif berdasarkan karakteristik produksinya, karena tanpa pemetaan yang tepat pemerintah daerah berisiko melakukan alokasi sumber daya yang tidak optimal dan merumuskan kebijakan yang tidak sesuai dengan kondisi faktual serta perbedaan kapasitas setiap kecamatan.

Pemetaan yang akurat akan membantu pemerintah dalam menentukan strategi pengembangan wilayah sesuai potensi yang dimiliki masing-masing kecamatan. Salah satu metode yang relevan adalah *K-Means Clustering*, yaitu analisis multivariat non-hirarki yang membagi objek ke dalam sejumlah klaster berdasarkan tingkat kemiripan (Hair et al., 2014). Sejumlah penelitian terdahulu juga membuktikan efektivitas metode ini. Eliyani dkk. (2024) menemukan bahwa provinsi penghasil jagung yang memiliki produksi tinggi membentuk klaster tersendiri, sehingga strategi peningkatan produksi dapat disesuaikan tiap klaster. Sanusi dkk. (2022) menunjukkan bahwa penerapan *K-Means* pada

produksi kelapa sawit di Kabupaten Aceh Barat menghasilkan kluster dengan kebutuhan pengelolaan dan teknologi pertanian yang berbeda.

Temuan-temuan tersebut menunjukkan bahwa metode *K-Means Clustering* efektif digunakan untuk mengidentifikasi pola produksi dan karakteristik wilayah pertanian. Namun, masih berfokus pada satu jenis komoditas tertentu, sehingga belum mampu menggambarkan keterkaitan antarkomoditas dalam suatu wilayah secara menyeluruh. Kebaruan penelitian ini terletak pada penerapan analisis kluster untuk beberapa komoditas perkebunan sekaligus di Kabupaten Pasaman Barat. Melalui pendekatan multikomoditas, penelitian ini memberikan gambaran yang lebih lengkap tentang pola dan sebaran produksi perkebunan di setiap kecamatan, sehingga dapat menjadi dasar dalam menyusun strategi pengembangan wilayah yang lebih tepat dan sesuai dengan potensi daerah.

Tujuan penelitian ini adalah untuk mengelompokkan kecamatan di Kabupaten Pasaman Barat berdasarkan produksi komoditas perkebunan utama menggunakan metode *K-Means Clustering* guna mengidentifikasi pola karakteristik setiap kluster sebagai dasar perumusan strategi pembangunan yang lebih efektif dan tepat sasaran.

II. METODE PENELITIAN

A. Data dan Variabel Penelitian

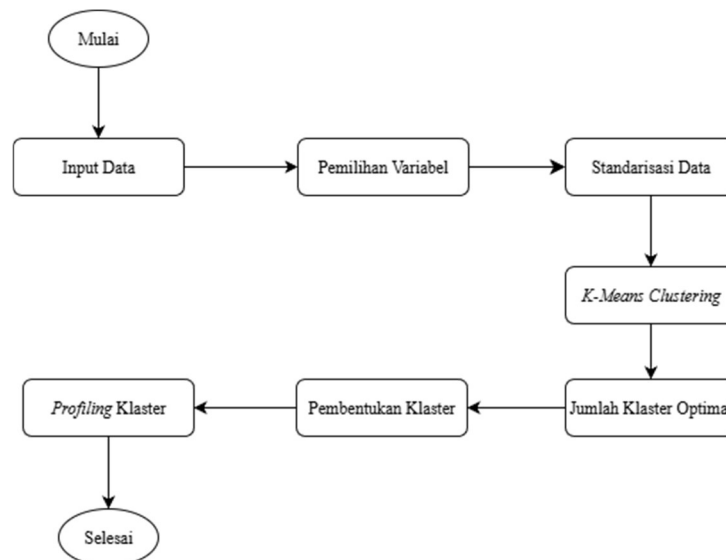
Penelitian ini menggunakan data sekunder yang diperoleh dari BPS Kabupaten Pasaman Barat yaitu dari publikasi Kabupaten Pasaman Barat Dalam Angka 2025, adapun data yang digunakan adalah data produksi perkebunan menurut kecamatan dan jenis tanaman (ton) tahun 2024. Objek amatan pada data ini adalah 11 kecamatan yang ada di Kabupaten Pasaman Barat, dan variabelnya yaitu hasil produksi perkebunan dalam satuan ton. Terdapat 5 komoditas hasil produksi perkebunan yang menjadi variabel dalam analisis ini, yaitu Kelapa Sawit (X_1), Kelapa (X_2), Karet (X_3), Kakao (X_4), dan Nilam (X_5).

Tabel 1. Variabel Penelitian

Variabel	Keterangan
X_1	Kelapa Sawit
X_2	Kelapa
X_3	Karet
X_4	Kakao
X_5	Nilam

B. Tahapan Analisis Data

Penelitian ini menggunakan metode *K-Means Clustering* untuk mengelompokkan kecamatan berdasarkan hasil produksi komoditas perkebunan. Analisis dilakukan dengan bantuan *software RStudio* melalui beberapa tahapan, yaitu:



Gambar 1. Flowchart Penelitian

1. Standarisasi Data

Standarisasi data dilakukan dengan metode *Z-score* agar seluruh variabel produksi berada pada skala yang sama (Mohamad & Usman, 2013). Berikut adalah rumus standarisasi *Z-score*:

$$Z = \frac{x - \mu}{\sigma} \quad (1)$$

dimana,

μ = rata-rata sample

σ = standar deviasi

2. Menentukan Jumlah Kluster Optimal

Jumlah kluster optimal ditentukan menggunakan metode *Elbow*, yaitu teknik yang mengevaluasi nilai *Within-Cluster Sum of Squares* (WCSS) terhadap jumlah kluster (k). WCSS merupakan ukuran sebaran data dalam satu kluster, yang semakin kecil nilainya menunjukkan kluster semakin homogen. Rumus WCSS dapat dituliskan sebagai berikut:

$$WCSS = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (2)$$

Dimana k adalah jumlah kluster, C_i adalah kluster ke- i , x adalah data dalam kluster C_i , dan μ_i adalah *centroid* kluster C_i . Selanjutnya pada algoritma *K-Means*, pengelompokan data dilakukan berdasarkan kedekatan data dengan *centroid*. Ukuran kedekatan tersebut umumnya dihitung menggunakan jarak *Euclidean*, dengan rumus:

$$d(x_i, x_j) = \sqrt{\sum_{p=1}^m (x_{ip} - x_{jp})^2} \quad (3)$$

Dengan m adalah jumlah variabel, x_i adalah amatan ke- i , dan x_j adalah amatan ke- j .

Proses metode *Elbow* dilakukan dengan menghitung WCSS untuk berbagai nilai k . Titik ketika penurunan WCSS mulai melambat dan membentuk sudut (*elbow*) dianggap sebagai jumlah kluster terbaik (Umargono et al., 2020). Penggunaan metode ini dipilih karena telah terbukti efektif dalam penelitian sebelumnya. Irawan dkk. (2025) berhasil mengintegrasikan *K-Means* dan *Elbow Method* untuk pemetaan komoditas perkebunan (*Baccaurea* spp.) sehingga menghasilkan jumlah kluster yang lebih representatif.

3. Hasil Klasterisasi

Pada hasil klasterisasi memuat tahap pembentukan kluster dan *profiling* kluster. Objek akan dikelompokkan ke dalam kluster yang sesuai dengan karakteristik masing-masing. Objek digabungkan berdasarkan pada tingkat kemiripan karakteristik yang diukur menggunakan ukuran kemiripan. Selanjutnya pada tahap *profiling* akan dilakukan pendeskripsian karakteristik setiap kluster untuk menjelaskan perbedaan antar kluster, karena setiap kluster yang terbentuk akan memiliki karakteristiknya masing-masing sehingga berbeda antara satu kluster dengan lainnya dan juga akan diketahui pola distribusi dari datanya.

Analisis ini dilakukan untuk mengidentifikasi karakteristik kluster produksi perkebunan di Kabupaten Pasaman Barat, sehingga dapat menjadi dasar dalam pengambilan kebijakan pengembangan sektor tersebut

III. HASIL DAN PEMBAHASAN

A. Standarisasi Data

Perbedaan skala antar variabel pada data produksi perkebunan, seperti kelapa sawit yang mencapai puluhan ribu ton dibanding nilam yang hanya puluhan, menjadikan standarisasi sangat penting dalam analisis *K-Means*. Tanpa standarisasi, variabel berskala besar akan mendominasi perhitungan jarak sehingga hasil kluster menjadi bias.

Misalnya, produksi kelapa sawit yang mencapai 72.473,15 ton, sedangkan produksi nilam hanya berkisar antara 3,57 hingga 65,70 ton. Perbedaan skala ini menyebabkan variabel dengan nilai besar, seperti kelapa sawit, mendominasi perhitungan jarak pada metode *K-Means* sehingga pola dari produksi perkebunan lain yang nilainya kecil menjadi kurang berpengaruh.

Oleh karena itu, perlu dilakukan standarisasi agar semua variabel berada pada skala yang sebanding sehingga kontribusi setiap komoditas dalam pembentukan kluster menjadi setara dan pola sebenarnya dapat terlihat. Standarisasi dilakukan dengan menggunakan *software Rstudio* dengan fungsi *scale()*. Adapun hasil standarisasi ditampilkan pada Tabel 2.

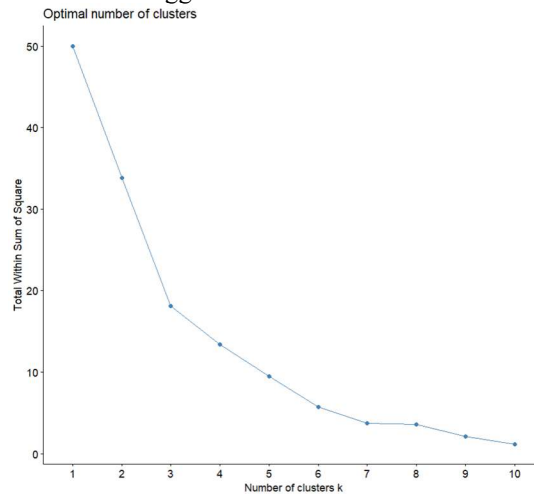
Tabel 2. Nilai Standarisasi Data

Amatan	X_1	X_2	X_3	X_4	X_5
1	0.01513949	-0.3895036	-0.9499116	0.32143819	-1.3416994
2	-0.32945906	-0.3830717	0.5966590	1.33931955	-0.4615832
3	-0.02070332	-0.3188667	1.0522064	1.41064655	-0.3862767
...
10	-1.24590330	0.3908715	-1.1525199	-0.94820967	-1.5066345
11	2.26986576	2.9295888	-1.1254979	-1.64683087	0.9202684

Pada Tabel 2 disajikan nilai standarisasi data sebelumnya dimana telah didapatkan satuan yang seragam untuk setiap variabel sehingga analisis kluster dapat dilakukan dengan lebih optimal.

B. Menentukan Jumlah Kluster Optimal

Sebelum dilakukan analisis, peneliti menentukan jumlah kluster yang akan dibentuk terlebih dahulu. Penentuan jumlah kluster pada penelitian ini dilakukan menggunakan metode *Elbow*. Berikut ditampilkan pada Gambar 1.



Gambar 2. Jumlah Kluster Optimal

Berdasarkan Gambar 2 diketahui jumlah kluster optimal adalah tiga, karena setelah $k = 3$ penurunan WCSS tidak lagi signifikan sehingga tiga kluster dinilai cukup merepresentasikan pola data secara keseluruhan. Secara praktis, tiga kluster ini dapat merepresentasikan perbedaan karakteristik antar kecamatan dalam hal produksi perkebunan, misalnya kelompok kecamatan dengan produksi tinggi, sedang, dan rendah. Dengan demikian, penggunaan tiga kluster dinilai cukup untuk memberikan pemetaan yang jelas terhadap variasi potensi perkebunan di wilayah penelitian.

C. Hasil Klasterisasi

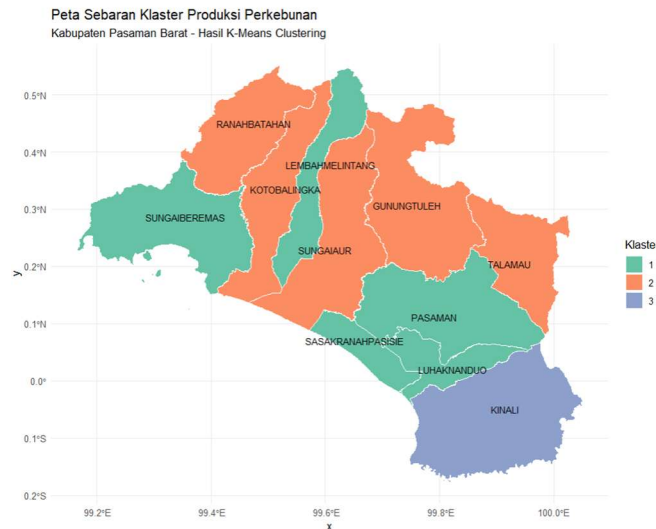
Untuk memberikan pemahaman yang lebih jelas terkait karakteristik tiap kluster, disajikan Tabel 3 yang memuat nilai rata-rata produksi masing-masing komoditas. Tabel ini bertujuan memperlihatkan kecenderungan komoditas unggulan yang membedakan antar kluster.

Tabel 3. Karakteristik Kluster

Kluster	X_1	X_2	X_3	X_4	X_5
1	31134,77	116,684	150,926	290,75	24,058
2	29432,35	65,690	734,836	376,49	37,180
3	72473,15	932,900	9,460	201,30	52,530

Berdasarkan Tabel 3, didapatkan hasil bahwa kluster 1 mempunyai produksi yang relatif merata di semua komoditas. Kluster 2 menunjukkan keunggulan pada komoditas karet dan kakao. Sementara itu, kluster 3 dominan pada kelapa sawit, kelapa, dan nilam, tetapi rendah pada komoditas lainnya.

Berikut ini ditampilkan Gambar 3 yang menampilkan hasil pengelompokan kecamatan yang ada di Kabupaten Pasaman Barat menggunakan *K-Means Clustering* yang diolah menggunakan *software Rstudio*.



Gambar 3. Visualisasi *K-Means Clustering*

Berdasarkan Gambar 3 terlihat bahwa terdapat tiga kluster kecamatan yang terbentuk. Kluster pertama ditandai dengan warna hijau, kluster kedua dengan warna oranye, dan kluster ketiga dengan warna biru.

Kluster 1 berisi lima kecamatan dengan produksi yang relatif merata di semua jenis komoditas. Hal ini menunjukkan bahwa wilayah-wilayah tersebut memiliki potensi diversifikasi yang baik karena tidak bergantung pada satu komoditas saja. Pola ini dipengaruhi oleh luas lahan yang sedang dan pemanfaatan lahan yang beragam, sehingga setiap komoditas bisa dikembangkan secara bersamaan. Selain itu, sebagian besar wilayah di kluster ini sudah memiliki akses terhadap teknologi pertanian sederhana dan sistem irigasi dasar, yang membantu petani mengelola lebih dari satu jenis tanaman (BPS Kabupaten Pasaman Barat, 2025). Karena itu, strategi pengembangan yang tepat adalah meningkatkan produktivitas beberapa komoditas sekaligus, misalnya melalui pelatihan terpadu bagi petani dan penyediaan sarana produksi pertanian yang merata.

Kluster 2 juga beranggotakan lima kecamatan, dengan karakteristik produksi yang menunjukkan kluster ini lebih unggul dalam produksi karet dan kakao. Keunggulan ini berkaitan dengan kondisi alam dan jenis tanah yang cocok untuk tanaman tahunan seperti karet dan kakao, yaitu daerah dataran menengah dengan curah hujan tinggi dan tanah liat lempung. Selain itu, wilayah ini biasanya jauh dari pusat perdagangan besar, sehingga petani lebih memilih menanam komoditas yang bisa disimpan lama dan memiliki harga stabil. Hal ini membuka peluang bagi pemerintah daerah untuk mengembangkan sentra industri hilir berbasis karet dan kakao, misalnya pengolahan lateks dan fermentasi biji kakao, serta memperkuat rantai nilai lokal melalui pelatihan teknis dan insentif investasi sektor agroindustri.

Sementara itu, kluster 3 hanya terdiri dari Kecamatan Kinali, yang menunjukkan dominasi tinggi pada kelapa sawit, kelapa, dan nilam. Hal ini disebabkan oleh lahan yang luas, infrastruktur transportasi yang baik, serta lokasi yang dekat dengan jalur distribusi utama ke luar daerah. Kondisi alamnya berupa dataran rendah dengan suhu yang cukup panas juga sangat mendukung pertumbuhan nilam. Dengan potensi tersebut, wilayah ini bisa menjadi pusat produksi utama dan daerah ekspor unggulan. Strategi pengembangannya dapat difokuskan pada penguatan rantai pasok, pembangunan industri pengolahan, dan perluasan akses pasar ekspor secara berkelanjutan.

Secara keseluruhan, hasil ini menunjukkan bahwa perbedaan potensi antarwilayah tidak hanya disebabkan oleh jumlah produksi, tetapi juga oleh faktor lingkungan seperti luas lahan, kondisi alam (agroklimat), akses teknologi, serta infrastruktur dan pasar. Oleh karena itu, penerapan metode *K-Means Clustering* tidak hanya membantu memetakan potensi setiap wilayah, tetapi juga dapat digunakan sebagai dasar untuk perencanaan pembangunan perkebunan yang

lebih tepat dan sesuai karakteristik daerah. Temuan ini sejalan dengan studi-studi terdahulu yang memanfaatkan *K-Means* untuk pemetaan potensi komoditas dan merekomendasikan intervensi terfokus berdasarkan klaster. Oleh karena itu, pemerintah daerah disarankan menggunakan hasil analisis ini sebagai basis zonasi kebijakan mengalokasikan dukungan teknis, investasi hilirisasi, serta program-program konservasi lahan sesuai karakteristik tiap klaster

IV. KESIMPULAN

Berdasarkan hasil penelitian, diperoleh bahwa jumlah klaster optimal pada produksi perkebunan di Kabupaten Pasaman Barat adalah tiga klaster, yang ditentukan melalui metode *Elbow*. Klaster 1 terdiri dari kecamatan dengan produksi yang relatif merata di semua komoditas dan cocok untuk strategi *diversifikasi*. Klaster 2 menunjukkan spesialisasi pada karet dan kakao, sehingga berpotensi dikembangkan sebagai sentra industri hilir berbasis komoditas tersebut. Sementara itu, Klaster 3 yang hanya mencakup Kecamatan Kinali, unggul dalam kelapa sawit, kelapa, dan nilam, menjadikannya pusat produksi utama yang strategis untuk penguatan rantai pasok dan ekspor. Hasil ini menegaskan bahwa metode *K-Means Clustering* dapat mengelompokkan wilayah sesuai potensi unggulannya dan menjadi dasar strategi pengembangan perkebunan yang lebih tepat sasaran. Penelitian selanjutnya disarankan untuk menambahkan variabel pendukung seperti luas lahan, produktivitas per hektare, tenaga kerja, atau faktor sosial-ekonomi, agar analisis klaster yang dihasilkan lebih komprehensif dan dapat digunakan dalam perencanaan kebijakan pembangunan pertanian berkelanjutan.

DAFTAR PUSTAKA

- BPS Kabupaten Pasaman Barat. (2024). *Top Five Komoditas Usaha Pertanian di Kabupaten Pasaman Barat*.
BPS Kabupaten Pasaman Barat. (2025). *Kabupaten Pasaman Barat Dalam Angka 2025*.
Eliyani, Dwiasnati, S., Arif, S. M., Avrizal, R., & Fatimah, N. (2024). Application of Machine Learning in Clustering Maize Producing Regions in Indonesia. *Computer Engineering and Applications*, 13(2).
Hair, J. F. ., Black, W. C. ., Babin, B. J. ., & Anderson, R. E. . (2014). *Multivariate Data Analysis*. Pearson Education Limited.
Irawan, H., Fawzi, M. I., & Hidayat, S. (2025). Integration of K-Means Clustering and Elbow Method for Mapping *Baccaurea* spp. Distribution to Support Agroindustrial Development in West Sulawesi. *Jurnal Teknik Informatika (JUTIF)*, 6(2). <https://doi.org/10.52436/1.jutif.6.2.4671>
Mohamad, I. Bin, & Usman, D. (2013). Standardization and Its Effects on K-Means Clustering Algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, 6(17), 3299–3303. <https://doi.org/10.19026/rjaset.6.3638>
Sanusi, Juliwardi, I., Ardiansyah, M., Astrianda, N., & Elviajakfar, A. (2022). Clustering of Palm Oil Production Results with K-Means Algorithm in West Aceh Regency. *Jurnal Inotera*, 7(1), 34–41. <https://doi.org/10.31572/inotera.vol7.iss1.2022.id168>
Umargono, E., Suseno, J. E., & Gunawan, V. (2020). *K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula*.