

Analisis Klaster Produksi Cabai Besar dan Cabai Rawit Antar Provinsi di Indonesia Menggunakan Algoritma K-Means.

M. Hafiz Budiman, Ferdy Ardiansyah, Eriski Aulia Rahmi, Mauludimas Nasution, Wulan Inda Sari, Siti Sarah

Universitas Labuhanbatu, Indonesia

*Email: nurhafizahisnaini2002@gmail.com, ferdyardiansyah506@gmail.com, eriskiauliarahmi@gmail.com, mauludimasnasution@gmail.com, wulaninda12345@gmail.com, sitisarah01519@gmail.com

ABSTRAK

Produksi cabai merupakan salah satu komponen penting dalam stabilitas pasokan komoditas hortikultura di Indonesia. Ketimpangan produksi cabai antarprovinsi menyebabkan fluktuasi harga dan ketidakstabilan pasokan, sehingga pemetaan wilayah berdasarkan kapasitas produksi menjadi sangat relevan. Penelitian ini bertujuan mengklusterkan provinsi di Indonesia berdasarkan produksi cabai besar dan cabai rawit menggunakan algoritma K-Means. Data diperoleh dari Badan Pusat Statistik (BPS) tahun 2023 dan dianalisis menggunakan pendekatan unsupervised learning. Proses penelitian meliputi pembersihan data, normalisasi, pemilihan parameter jumlah klaster, penerapan algoritma K-Means, dan evaluasi menggunakan Silhouette Coefficient. Hasil penelitian menunjukkan terbentuknya tiga klaster, namun hanya dua yang stabil yaitu klaster produksi rendah dan klaster produksi tinggi. Klaster produksi tinggi dihuni oleh provinsi Jawa Timur, Jawa Tengah, Jawa Barat, serta Sumatera Utara sebagai sentra utama. Nilai silhouette untuk klaster produksi rendah mencapai 0.50–0.75, menunjukkan pemisahan klaster yang kuat. Temuan ini dapat menjadi dasar perencanaan distribusi dan pengembangan wilayah produksi cabai nasional.

Kata kunci: cabai, clustering, k-means, produksi, provinsi.

ABSTRACT

Chili production plays a critical role in stabilizing Indonesia's horticultural supply chain. Regional disparities in chili production often lead to price fluctuations and supply instability, making it necessary to map provinces based on their production capacity. This study aims to cluster Indonesian provinces based on red chili and cayenne pepper production using the K-Means algorithm. Data were obtained from Statistics Indonesia (BPS) for 2023 and analyzed using an unsupervised learning approach. The workflow includes data preprocessing, normalization, cluster number selection, K-Means implementation, and evaluation using the Silhouette Coefficient. Results reveal three clusters, with only two considered stable: low-production and high-production clusters. High-production provinces include East Java, Central Java, West Java, and North Sumatra as primary production centers. The silhouette scores for the low-production cluster range from 0.50 to 0.75, indicating strong cohesiveness. These findings may guide policymakers in distribution planning and regional agricultural development.

Keywords: chili, clustering, k-means, production, province

PENDAHULUAN

Beberapa penelitian telah memanfaatkan algoritma *K-Means clustering* dalam konteks pertanian dan komoditas pangan di Indonesia. Laia dkk. (2022) menerapkan K-Means untuk mengelompokkan kawasan potensi pertanian karet produktif di Provinsi Sumatera Utara dan berhasil mengidentifikasi tiga klaster berdasarkan tingkat produktivitas lahan. Ginting & Riandari (2020) melakukan klasterisasi bibit tanaman kopi arabika berdasarkan karakteristik pertumbuhan untuk mendukung seleksi bibit unggul. Dalam konteks cabai, Abdullah & Fatah (2024) melakukan analisis produksi cabai rawit menggunakan K-Means dan menghasilkan pemetaan provinsi berdasarkan kategori produksi rendah, sedang, dan tinggi. Widiyanto dkk.

. Hafiz Budiman, Ferdy Ardiansyah, Eriski Aulia Rahmi, Mauludimas Nasution, Wulan Inda Sari, Siti Sarah

Analisis Klaster Produksi Cabai Besar dan Cabai Rawit Antar Provinsi di Indonesia Menggunakan Algoritma K-Means.

(2024) memperluas cakupan dengan membandingkan algoritma K-Means dan K-Medoids untuk mengelompokkan distribusi perdagangan komoditas cabai di Indonesia. Penelitian lain oleh Yasmin dkk. (2023) menggunakan K-Means untuk klasifikasi citra cabai keriting berbasis ekstraksi warna dan tekstur, sementara Kholila dkk. (2023) memetakan kondisi lingkungan tanam berdasarkan parameter agronomis. Secara umum, penelitian-penelitian tersebut berfokus pada aspek produktivitas, citra digital, atau distribusi perdagangan, namun belum secara spesifik melakukan klusterisasi berbasis kapasitas produksi simultan antara cabai besar dan cabai rawit di tingkat provinsi dengan evaluasi kualitas klaster menggunakan Silhouette Coefficient.

Cabai merupakan komoditas hortikultura yang memiliki nilai ekonomi tinggi di Indonesia. Komoditas ini menjadi bagian penting dalam struktur konsumsi rumah tangga dan sangat berpengaruh terhadap laju inflasi pangan. Tingginya elastisitas permintaan cabai, ditambah dengan pola produksi yang tidak merata antarprovinsi, membuat harga cabai rentan berfluktuasi. Ketidakstabilan harga sering kali bukan hanya disebabkan oleh perubahan musiman, tetapi juga ketimpangan produksi dan ketergantungan terhadap wilayah sentra. Produksi cabai di Indonesia pada umumnya terkonsentrasi pada beberapa provinsi seperti Jawa Timur, Jawa Tengah, dan Sumatera Utara. Sementara provinsi lainnya cenderung memiliki kapasitas produksi yang lebih rendah. Ketimpangan ini menimbulkan tantangan dalam perencanaan distribusi, terutama untuk menjaga kelancaran suplai di luar wilayah sentra produksi. Pemerintah membutuhkan informasi berbasis data untuk memahami pola distribusi produksi dan menentukan strategi pengembangan yang tepat (Nuraeni & Maulana Firdaus, 2025; Septianto et al., 2025) (Septianto et al., 2025).

Penelitian sebelumnya banyak membahas analisis regresi, prediksi harga, produktivitas berdasarkan kondisi cuaca, serta distribusi pasokan. Akan tetapi, kajian yang memetakan provinsi berdasarkan kapasitas produksi cabai menggunakan teknik machine learning, khususnya algoritma K-Means, masih sangat terbatas. Padahal, K-Means memiliki kemampuan mengidentifikasi struktur kelompok berdasarkan kemiripan karakteristik produksi.

Kesenjangan penelitian yang diidentifikasi dalam kajian ini adalah minimnya studi yang secara simultan mengelompokkan provinsi berdasarkan dua variabel produksi utama cabai (cabai besar dan cabai rawit) dalam satu kerangka analisis klaster berbasis *machine learning*. Sebagian besar penelitian terdahulu cenderung memisahkan analisis antarjenis cabai atau hanya menggunakan satu indikator produksi. Selain itu, evaluasi kualitas klaster menggunakan Silhouette Coefficient masih jarang diterapkan dalam konteks pemetaan produksi hortikultura nasional, padahal indikator ini penting untuk menilai validitas dan stabilitas kelompok yang terbentuk. Penelitian ini menutup gap tersebut dengan melakukan klusterisasi provinsi berdasarkan produksi cabai besar dan cabai rawit. Analisis dilakukan secara objektif menggunakan data BPS, sehingga hasilnya dapat memberikan gambaran empiris mengenai perbedaan kapasitas produksi antarprovinsi. Selain itu, penelitian ini menambahkan evaluasi kualitas klaster menggunakan Silhouette Coefficient untuk memastikan stabilitas hasil klaster

Kebaruan (*novelty*) penelitian ini terletak pada tiga aspek utama. Penelitian ini mengintegrasikan dua variabel produksi cabai sekaligus (cabai besar dan cabai rawit) dalam

. Hafiz Budiman, Ferdy Ardiansyah, Eriski Aulia Rahmi, Mauludimas Nasution, Wulan Inda Sari, Siti Sarah

Analisis Klaster Produksi Cabai Besar dan Cabai Rawit Antar Provinsi di Indonesia Menggunakan Algoritma K-Means.

satu proses klasterisasi, sehingga menghasilkan pemetaan yang lebih komprehensif dibandingkan studi sebelumnya yang hanya berfokus pada satu jenis cabai. Penelitian ini mengevaluasi stabilitas klaster secara kuantitatif melalui Silhouette Coefficient, yang tidak hanya mengidentifikasi jumlah klaster optimal tetapi juga mengukur kohesi internal dan separasi antarklaster. Penelitian ini secara eksplisit mengidentifikasi ketidakstabilan klaster menengah sebagai temuan kritis yang merefleksikan karakter produksi cabai yang fluktuatif dan tidak merata. Dengan demikian, penelitian ini tidak sekadar mereplikasi metode klasterisasi, tetapi memperluas pemahaman tentang struktur produksi cabai nasional serta menyediakan dasar evaluatif bagi perumusan kebijakan berbasis bukti.

Tujuan penelitian ini adalah mengelompokkan provinsi di Indonesia berdasarkan tingkat produksi cabai besar dan cabai rawit tahun 2023 menggunakan algoritma K-Means, mengevaluasi kualitas klaster yang terbentuk melalui Silhouette Coefficient, mengidentifikasi provinsi-provinsi yang berperan sebagai sentra produksi utama, serta menyediakan dasar analitis yang dapat digunakan untuk menyusun rekomendasi kebijakan produksi dan distribusi cabai nasional yang lebih tepat sasaran dan berbasis data. Penelitian ini memberikan manfaat praktis, teoretis, dan kebijakan. Secara praktis, hasil klasterisasi dapat digunakan oleh Kementerian Pertanian, Badan Pangan Nasional, dan pemerintah daerah untuk mengidentifikasi provinsi prioritas dalam program peningkatan produksi cabai, serta merancang skema distribusi antardaerah yang lebih efisien guna menekan fluktuasi harga. Secara teoretis, penelitian ini memperkaya literatur klasterisasi dalam konteks agrikultur tropis dan mendemonstrasikan penerapan evaluasi silhouette sebagai instrumen diagnostik kualitas klaster. Secara kebijakan, penelitian ini menyediakan kerangka analitis berbasis data yang dapat diintegrasikan ke dalam sistem informasi pertanian nasional untuk pemantauan produksi secara periodik dan adaptif.

Hasil penelitian ini memperkuat argumentasi bahwa metode klasterisasi unsupervised learning seperti K-Means efektif digunakan untuk pemetaan sektor pertanian berbasis indikator produksi, namun juga mengungkap keterbatasan pendekatan univariat dalam menangkap kompleksitas agronomis. Rendahnya stabilitas klaster menengah mengindikasikan bahwa variabel produksi saja tidak cukup untuk membedakan provinsi dengan karakteristik transisional. Oleh karena itu, diperlukan pengembangan model klasterisasi multivariat yang mengintegrasikan variabel kontekstual seperti luas panen, produktivitas, iklim, dan infrastruktur pertanian. Temuan ini membuka peluang riset lanjutan dalam pengembangan hybrid clustering atau ensemble methods untuk meningkatkan akurasi pemetaan.

METODE

Data Penelitian

Data penelitian bersumber dari Badan Pusat Statistik (BPS) melalui Query Builder Hortikultura tahun 2023. Dua variabel utama digunakan sebagai indikator produksi, yaitu produksi cabai besar (ton) dan produksi cabai rawit (ton) pada tingkat provinsi. Dataset mencakup seluruh provinsi di Indonesia sesuai pembagian administrasi terbaru. Data disusun

. Hafiz Budiman, Ferdy Ardiansyah, Eriski Aulia Rahmi, Mauludimas Nasution, Wulan Inda Sari, Siti Sarah

Analisis Klaster Produksi Cabai Besar dan Cabai Rawit Antar Provinsi di Indonesia Menggunakan Algoritma K-Means.

dalam bentuk tabular dengan provinsi sebagai atribut identitas (metadata), sementara dua variabel produksi digunakan sebagai fitur utama dalam proses klasterisasi.

Tahap Pra-pemrosesan

Tahap pra-pemrosesan dilakukan untuk memastikan kualitas data sebelum dianalisis lebih lanjut. Nilai kosong atau tidak valid dibersihkan dan dikonversi; seluruh atribut numerik dipastikan berada dalam tipe bilangan yang sesuai. Normalisasi Z-Score diterapkan pada kedua variabel produksi guna menyetarakan skala dan mengurangi dominasi nilai ekstrem. Hanya dua fitur produksi yang digunakan dalam model, sesuai fokus penelitian pada kapasitas produksi cabai antarprovinsi.

Klasterisasi dengan K-Means

Algoritma K-Means digunakan untuk mengelompokkan provinsi berdasarkan kemiripan nilai produksi. Proses klasterisasi dilakukan dengan jarak Euclidean dan inisialisasi centroid menggunakan metode k-means. Jumlah klaster ditetapkan sebanyak tiga berdasarkan eksplorasi awal distribusi data dan kebutuhan untuk membedakan kelompok produksi rendah, sedang, dan tinggi. Model dijalankan hingga konvergen atau mencapai batas maksimum iterasi. (Ginting & Riandari, 2020; Laia et al., 2022), (Kurniawan et al., 2020)

Evaluasi Klaster

Silhouette Coefficient digunakan untuk mengevaluasi kualitas klaster dengan rumus :

$$s = (b - a) / \max(a, b)$$

dengan:

a = jarak rata-rata antaranggota dalam klaster,

b = jarak rata-rata ke klaster terdekat.

Nilai silhouette berkisar -1 hingga 1.

Tools

Seluruh proses analisis dilakukan menggunakan perangkat Orange Data Mining yang memfasilitasi pemodelan K-Means, visualisasi scatter plot, dan penilaian silhouette. Pipeline analisis mencakup impor data, seleksi atribut, normalisasi, klasterisasi, dan visualisasi hasil.

Workflow Penelitian

Workflow penelitian dimulai dari pengambilan data produksi cabai besar dan cabai rawit tahun 2023 yang diperoleh melalui Badan Pusat Statistik. Data kemudian dipilih berdasarkan variabel yang relevan, yaitu dua fitur produksi dan satu variabel provinsi sebagai metadata. Setelah itu dilakukan proses pembersihan data untuk mengatasi nilai kosong dan memastikan kualitas dataset tetap terjaga. (Ohana-Levi et al., 2019; Zhu et al., 2025)

Tahap berikutnya adalah normalisasi menggunakan skala Z-Score agar kedua fitur memiliki rentang nilai yang sebanding. Data yang telah dinormalisasi selanjutnya diproses menggunakan algoritma K-Means untuk membentuk tiga klaster berdasarkan pola produksi cabai. Hasil klaster kemudian dievaluasi menggunakan Silhouette Coefficient untuk melihat tingkat stabilitas dan pemisahan antar klaster.

Seluruh keluaran divisualisasikan melalui Scatter Plot, Bar Plot, dan Box Plot untuk memberikan gambaran yang jelas mengenai pola dan perbedaan tiap klaster. Tahapan akhir adalah menyimpan hasil pemrosesan dan melakukan interpretasi untuk digunakan pada bagian hasil dan pembahasan.

. Hafiz Budiman, Ferdy Ardiansyah, Eriski Aulia Rahmi, Mauludimas Nasution, Wulan Inda Sari, Siti Sarah

Analisis Klaster Produksi Cabai Besar dan Cabai Rawit Antar Provinsi di Indonesia Menggunakan Algoritma K-Means.



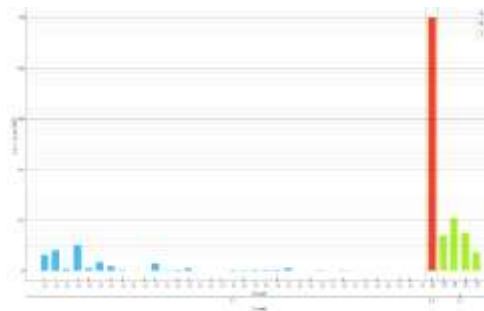
Gambar 1. Workflow Penelitian

Sumber: Dokumentasi Penelitian, 2026

HASIL DAN PEMBAHASAN

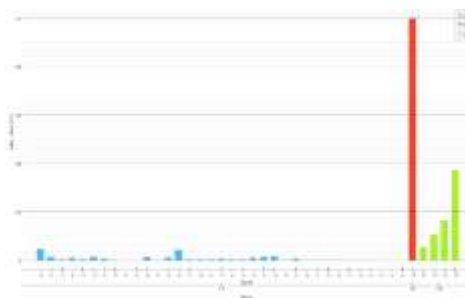
Gambaran Umum Data

Data penelitian terdiri dari produksi cabai besar dan cabai rawit pada tingkat provinsi tahun 2023 yang diperoleh dari Badan Pusat Statistik. Variabel provinsi digunakan sebagai identitas, sementara dua variabel produksi dijadikan fitur utama untuk proses klasterisasi. Distribusi awal data menunjukkan adanya perbedaan produksi yang cukup mencolok antarwilayah. Beberapa provinsi memiliki kapasitas produksi yang besar dan menjadi sentra hortikultura, sementara sebagian wilayah lainnya memiliki produksi yang relatif rendah. Penyebaran nilai yang tidak merata ini membentuk dasar yang kuat untuk dilakukan klasterisasi dalam rangka mengidentifikasi pola pengelompokan antar provinsi.



Gambar 2. Distribusi Produksi Cabai Besar Tahun 2023

Sumber: Hasil Olah Data Menggunakan Orange Data Mining, 2026



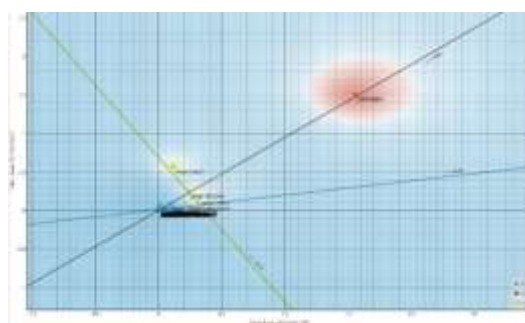
Gambar 3. Distribusi Produksi Cabai Rawit Tahun 2023

Sumber: Hasil Olah Data Menggunakan Orange Data Mining, 2026

Statistik deskriptif menunjukkan bahwa variabel produksi cabai besar memiliki sebaran yang sangat lebar antarprovinsi, mulai dari wilayah dengan kapasitas produksi rendah hingga provinsi yang mencatatkan produksi dalam skala besar. Variabel cabai rawit juga menunjukkan pola variasi yang serupa, sehingga kedua fitur ini memiliki pengaruh yang cukup besar dalam proses pengelompokan. Rentang nilai yang berbeda jauh antarprovinsi menegaskan perlunya normalisasi agar proses klasterisasi dapat berjalan lebih stabil dan tidak dipengaruhi dominasi nilai ekstrem.

Proses klasterisasi menggunakan K-Means dengan nilai $k=3$ menghasilkan tiga kelompok utama. Klaster pertama berisi provinsi-provinsi dengan kapasitas produksi cabai besar dan cabai rawit yang rendah. Klaster ketiga memuat provinsi dengan tingkat produksi yang konsisten tinggi dan sebagian besar merupakan sentra hortikultura nasional seperti Jawa Timur, Jawa Tengah, Jawa Barat, dan Sumatera Utara. Sementara itu, klaster kedua menunjukkan karakteristik campuran dan mencerminkan wilayah dengan pola produksi yang tidak stabil atau berada pada rentang menengah. Hasil visualisasi scatter plot memperlihatkan pemisahan antarklaster yang cukup jelas antara kelompok berproduksi tinggi dan kelompok berproduksi rendah, meskipun klaster menengah tampak kurang terdefinisi dengan baik. (Thongnim et al., 2025; Van Arkel & Kaleita, 2014; Zhang et al., 2024)

Evaluasi menggunakan Silhouette Coefficient memberikan gambaran kualitas struktur klaster yang terbentuk. Klaster berproduksi tinggi dan rendah menunjukkan nilai silhouette yang positif, menandakan bahwa kedua kelompok tersebut memiliki kedekatan internal yang kuat serta jarak yang cukup besar dengan klaster lainnya. Sebaliknya, klaster kedua memiliki nilai silhouette yang rendah hingga mendekati nol. Hal ini menunjukkan bahwa sebagian provinsi dalam klaster tersebut memiliki kedekatan yang serupa dengan dua klaster lainnya sehingga pemisahannya kurang stabil.



Gambar 3. Scatter Plot Klaster Produksi Cabai Tahun 2023

Sumber: Hasil Analisis Menggunakan Orange Data Mining, 2026

Hasil klasterisasi menggambarkan bahwa pola produksi cabai di Indonesia tidak terdistribusi secara merata dan memiliki kecenderungan untuk berkumpul berdasarkan kemampuan produksi. Kelompok provinsi dengan produksi tinggi mencerminkan wilayah-wilayah yang memiliki infrastruktur pertanian yang baik, luas lahan yang memadai, serta praktik budidaya yang intensif. Sementara itu, provinsi yang masuk dalam klaster produksi

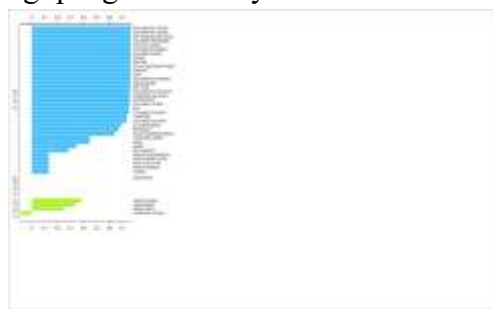
. Hafiz Budiman, Ferdy Ardiansyah, Eriski Aulia Rahmi, Mauludimas Nasution, Wulan Inda Sari, Siti Sarah

Analisis Klaster Produksi Cabai Besar dan Cabai Rawit Antar Provinsi di Indonesia Menggunakan Algoritma K-Means.

rendah umumnya memiliki keterbatasan pada faktor teknis, geografis, atau sumber daya yang memengaruhi produktivitas cabai.

Hasil evaluasi silhouette memperkuat pemahaman bahwa hanya dua klaster yang benar-benar stabil. Klaster menengah atau klaster kedua cenderung memuat provinsi dengan pola produksi yang tumpang tindih, sehingga sulit dibedakan secara tegas dari dua klaster lainnya. Kondisi ini sejalan dengan fenomena produksi hortikultura yang sering kali dipengaruhi faktor musiman, ketergantungan pada kondisi iklim, hingga fluktuasi luas panen yang menyebabkan karakteristik produksi tidak selalu konsisten.

Temuan penelitian ini menunjukkan bahwa proses klasterisasi dapat menjadi alat yang efektif untuk memetakan potensi dan pola produksi antarprovinsi. Namun demikian, hasil yang kurang stabil pada klaster tertentu mengindikasikan bahwa variabel produksi saja belum sepenuhnya mampu menggambarkan kompleksitas sektor cabai nasional. Penambahan variabel iklim, produktivitas, penggunaan pupuk, serta distribusi pasar diperkirakan mampu menghasilkan klaster yang lebih representatif. Meskipun demikian, hasil penelitian ini sudah cukup memberikan gambaran umum untuk mendukung pengambilan kebijakan terkait distribusi produksi dan strategi penguatan wilayah sentra cabai di Indonesia.



Gambar 4. Silhouette Plot Evaluasi Klaster K-Means

Sumber: Hasil Analisis Menggunakan Orange Data Mining, 2026

KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan provinsi di Indonesia ke dalam tiga klaster berdasarkan produksi cabai besar dan cabai rawit tahun 2023, di mana klaster berproduksi tinggi dan rendah memiliki struktur yang stabil, sedangkan klaster menengah menunjukkan nilai silhouette yang rendah dan mencerminkan karakter produksi yang tidak konsisten. Hasil ini menggambarkan bahwa variasi produksi cabai antarwilayah sangat dipengaruhi perbedaan kapasitas dan kondisi agronomis, sehingga pemetaan menggunakan klasterisasi dapat membantu mengidentifikasi wilayah prioritas untuk penguatan produksi dan perencanaan distribusi. Meskipun demikian, hasil klaster belum sepenuhnya merepresentasikan kompleksitas sektor cabai karena penelitian ini hanya menggunakan dua variabel produksi, sehingga diperlukan integrasi variabel tambahan serta penggunaan metode klasterisasi lain pada penelitian selanjutnya untuk memperoleh pemetaan yang lebih akurat dan komprehensif.

. Hafiz Budiman, Ferdy Ardiansyah, Eriski Aulia Rahmi, Mauludimas Nasution, Wulan Inda Sari, Siti Sarah

Analisis Klaster Produksi Cabai Besar dan Cabai Rawit Antar Provinsi di Indonesia Menggunakan Algoritma K-Means.

DAFTAR PUSTAKA

- Ginting, B., & Riandari, F. (2020). Implementasi metode K-Means clustering dalam pengelompokan bibit tanaman kopi arabika. *Jurnal Nasional Komputasi Dan Teknologi Informasi*, 3(2), 151–157. <https://doi.org/10.32672/jnkti.v3i2.2381>
- Kurniawan, Y. B., Winarno, & Wiranto. (2020). Soil moisture clustering using the K-Means clustering method in the UNS's agricultural laboratory at Jumantono. *AIP Conference Proceedings*, 2217, 30107. <https://doi.org/10.1063/5.0000861>
- Laia, S., Pane, D. H., & Affandi, E. (2022). Implementasi metode K-Means untuk mengelompokkan kawasan potensi pertanian karet produktif. *Jurnal Sistem Informasi Triguna Dharma*, 1(4), 282. <https://doi.org/10.53513/jursi.v1i4.5224>
- Nuraeni, F., & Maulana Firdaus, M. H. (2025). Model Klasterisasi Data Pertanian Menggunakan Algoritma K-Means. *Jurnal Algoritma*, 22(2), 890–898. <https://doi.org/10.33364/algoritma/v.22-2.2054>
- Ohana-Levi, N., Paz-Kagan, T., Peeters, A., Bar, P., & Karnieli, A. (2019). A weighted multivariate spatial clustering model to determine irrigation management zones. *Computers and Electronics in Agriculture*, 166, 104979. <https://doi.org/10.1016/j.compag.2019.05.012>
- Septianto, M. A., Faqih, A., & Rinaldi, A. R. (2025). Klasterisasi data produksi pertanian di Kabupaten Cirebon dengan algoritma K-Means. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(2), 260–269. <https://doi.org/10.23960/jitet.v13i2.6174>
- Thongnim, P., Srinil, P., & Phukseng, T. (2025). Data-driven clustering of smart farming to optimize agricultural practices through machine learning. *Bulletin of Electrical Engineering and Informatics*, 14(2), 1343–1354. <https://doi.org/10.11591/eei.v14i2.9343>
- Van Arkel, Z., & Kaleita, A. L. (2014). Identifying sampling locations for field-scale soil moisture estimation using K-Means clustering. *Water Resources Research*, 50(8), 7050–7057. <https://doi.org/10.1002/2013WR015015>
- Zhang, B., Zhang, H., Wang, Y., & others. (2024). Research on automatic alignment for corn harvesting based on Euclidean clustering and K-Means clustering. *Agriculture*, 14(11), 2071. <https://doi.org/10.3390/agriculture14112071>
- Zhu, F., Liu, Y., Zhou, J., & Wang, H. (2025). Using Constrained K-Means Clustering for Soil Texture Mapping with Limited Soil Samples. *Agronomy*, 15(5), 1220. <https://doi.org/10.3390/agronomy15051220>



© 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>).