



SEGMENTASI PROVINSI PENGHASIL PADI DI INDONESIA BERDASARKAN KARAKTERISTIK PRODUKSI TAHUN 2024 MENGGUNAKAN CLUSTERING K-MEANS

Risbuwono Heru Cokro¹, Davin Anezta Ahmad², Reyner Archi Sanchia³, Trimono⁴,
Muhammad Nasrudin⁵^{1,2,3,4,5}Program Studi Sains Data , Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jawa TimurE-mail: 23083010104@upnjatim.ac.id¹, 23083010097@student.upnjatim.ac.id²,
23083010085@student.upnjatim.ac.id³, trimono.stat@upnjatim.ac.id⁴,
nasrudin.fasilkom@upnjatim.ac.id⁵

ABSTRACT

This study aims to cluster Indonesian provinces based on rice production characteristics in 2024 using the K-Means Clustering method. The three key variables analyzed include harvested area (ha), productivity (quintal/ha), and total production (tons). The data were obtained from official publications of Statistics Indonesia (BPS) and processed using Python on Google Colab. The Elbow Method was used to determine the optimal number of clusters, resulting in three main groups: high, medium, and low production provinces. The clustering results indicate that West Java, Central Java, and East Java belong to the high-production cluster, while most eastern provinces fall into the low-production group. Meanwhile, several provinces such as North Sumatra and West Nusa Tenggara are included in the medium cluster. This segmentation provides sharper insights into disparities in rice production capacity and can serve as a reliable basis for designing more targeted, region-specific, and data-driven agricultural policies.

Keywords: Rice production, K-Means, Clustering, Regional segmentation, Agricultural statistics

ABSTRAK

Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan karakteristik produksi padi tahun 2024 menggunakan metode K-Means Clustering. Tiga variabel utama yang dianalisis meliputi luas panen (ha), produktivitas (kuintal/ha), dan total produksi (ton). Data diperoleh dari publikasi resmi Badan Pusat Statistik (BPS) dan diolah menggunakan Python melalui platform Google Colab. Metode Elbow digunakan untuk menentukan jumlah kluster optimal, yang menghasilkan tiga kelompok utama: provinsi dengan produksi tinggi, sedang, dan rendah. Hasil klasterisasi menunjukkan bahwa Jawa Barat, Jawa Tengah, dan Jawa Timur tergolong dalam kluster produksi tinggi, sedangkan mayoritas provinsi di wilayah timur termasuk dalam kluster produksi rendah. Sementara itu, beberapa provinsi seperti Sumatera Utara dan NTB masuk dalam kluster menengah. Segmentasi ini memberikan gambaran yang

Article History

Received: Juni 2025
Reviewed: Juni 2025
Published: Juni 2025Plagiarism Checker No
235Prefix DOI :
[10.8734/Kohesi.v1i2.365](https://doi.org/10.8734/Kohesi.v1i2.365)Copyright : Author
Publish by : Kohesi

This work is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



lebih rinci mengenai ketimpangan kapasitas produksi padi antarwilayah, sehingga dapat menjadi dasar yang kuat dalam perumusan kebijakan pertanian yang lebih spesifik, terarah, dan berbasis data.

Kata Kunci: *Produksi padi, K-Means, Klasterisasi, Segmentasi wilayah, Statistik pertanian*

1. PENDAHULUAN

Indonesia dikenal sebagai negara agraris dengan ketergantungan tinggi terhadap sektor pertanian, khususnya komoditas padi yang menjadi makanan pokok utama masyarakat. Padi tidak hanya penting dari sisi ketahanan pangan, tetapi juga memiliki peran strategis dalam mendukung stabilitas ekonomi pedesaan dan kesejahteraan petani. Oleh karena itu, pemahaman mendalam terhadap kondisi produksi padi di berbagai wilayah sangat penting dalam rangka merumuskan kebijakan pertanian yang efektif dan responsif terhadap kebutuhan tiap daerah.

Setiap provinsi di Indonesia menunjukkan karakteristik produksi padi yang berbeda, tergantung pada faktor geografis, iklim, ketersediaan sarana produksi, serta praktik budidaya yang diterapkan. Tiga indikator utama yang sering digunakan untuk menggambarkan kondisi produksi padi adalah luas panen (hektar), produktivitas (kuintal per hektar), dan total produksi (ton). Ketiga indikator ini mencerminkan kapasitas dan efisiensi sektor pertanian padi di suatu wilayah. Perbedaan-perbedaan inilah yang seharusnya menjadi dasar dalam penyusunan kebijakan pertanian berbasis wilayah, agar intervensi pemerintah lebih tepat sasaran.

Sayangnya, kebijakan pertanian di Indonesia masih banyak yang bersifat umum dan belum mempertimbangkan secara detail perbedaan karakteristik antar provinsi. Sebagian besar program intervensi dilakukan dengan pendekatan agregatif yang menyamaratakan kebutuhan dan kapasitas wilayah. Padahal, pendekatan tersebut berisiko menciptakan ketidakefisienan dalam alokasi sumber daya dan berpotensi menimbulkan ketimpangan pembangunan antar daerah. Oleh karena itu, diperlukan pendekatan analisis data yang mampu mengelompokkan wilayah berdasarkan kemiripan karakteristik produksi, sebagai dasar penyusunan kebijakan yang lebih adaptif dan spesifik.

Penelitian ini bertujuan untuk melakukan segmentasi provinsi-provinsi di Indonesia berdasarkan data luas panen, produktivitas, dan total produksi padi pada tahun 2024 menggunakan metode K-Means Clustering. Metode ini memungkinkan identifikasi pola dan pengelompokan wilayah yang memiliki karakteristik produksi padi yang serupa. Dengan



pendekatan ini, diharapkan hasil segmentasi dapat memberikan informasi yang lebih tajam dan aplikatif untuk mendukung perencanaan kebijakan pangan dan pertanian di Indonesia, serta mendorong pembangunan pertanian yang lebih inklusif, efisien, dan berbasis data.

2. METODE PENELITIAN

2.1 Dataset dan Variabel

Data yang digunakan dalam penelitian ini merupakan data sekunder bersifat kuantitatif. Data sekunder adalah data yang diperoleh secara tidak langsung dari pihak lain yang telah mengumpulkannya sebelumnya untuk keperluan tertentu. Dalam konteks penelitian ini, data diperoleh dari publikasi resmi instansi pemerintah yang memuat statistik pertanian, khususnya terkait produksi padi di Indonesia tahun 2024. Adapun variabel yang digunakan dalam penelitian meliputi:

1. Luas Panen (ha): total luas lahan yang digunakan untuk panen padi per provinsi
2. Produktivitas (ku/ha): hasil panen per hektar (kuintal/hektar) sebagai indikator efisiensi produksi
3. Produksi (ton): total produksi padi dalam satuan ton per provinsi
4. Wilayah: seluruh provinsi di Indonesia
5. Periode: Tahun 2024

Sumber utama data adalah Badan Pusat Statistik (BPS), lembaga resmi pemerintah yang menyediakan data statistik nasional, termasuk statistik pertanian tahunan. Data yang digunakan telah melalui tahap verifikasi dan validasi untuk memastikan keandalan dan representasi kondisi riil pertanian padi nasional.

2.2 Alur Analisis

Penelitian ini menggunakan pendekatan kuantitatif dengan metode unsupervised learning, yaitu analisis klaster K-Means, yang digunakan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan karakteristik produksinya. Metode K-Means dipilih karena kemampuannya dalam menyegmentasi data numerik ke dalam klaster yang memiliki karakteristik serupa berdasarkan jarak Euclidean antar observasi.

Menurut Siregar & Nursal (2020), algoritma K-Means bekerja secara iteratif untuk meminimalkan within-cluster sum of squares (WCSS) dan membagi data ke dalam k kelompok yang optimal. Selain itu, Wardhani et al. (2021) menjelaskan bahwa K-Means cocok diterapkan pada data statistik pertanian yang bersifat multivariat dan dapat divisualisasikan dengan baik.

Dalam penelitian ini, proses analisis dilakukan menggunakan Google Colab dan bahasa pemrograman Python. Tahapan analisis yang dilakukan antara lain:

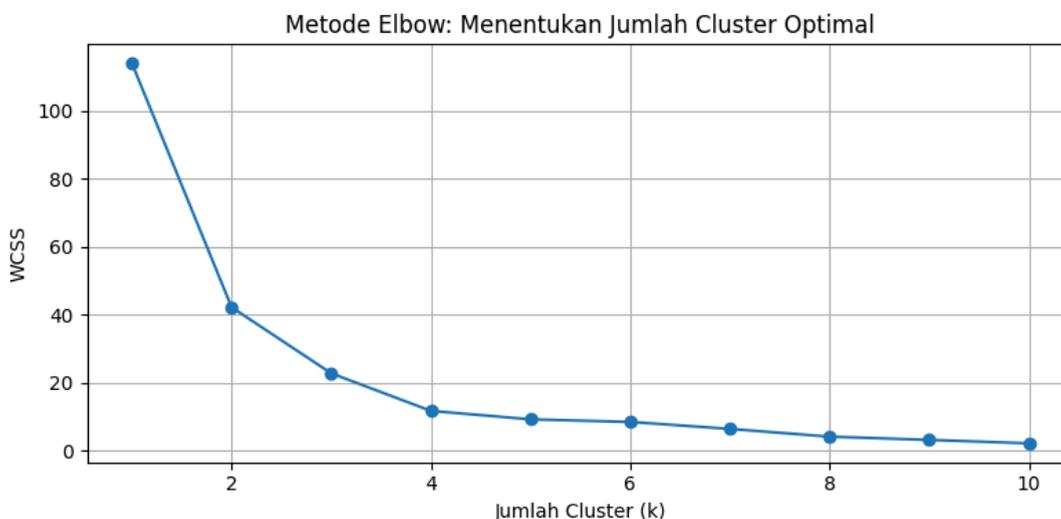
1. Pra-pemrosesan data, termasuk transformasi nama kolom, pengecekan missing values, dan normalisasi data numerik menggunakan metode Min-Max Scaling
2. Penentuan jumlah kluster optimal menggunakan metode Elbow Method, dengan cara mengevaluasi perubahan nilai WCSS untuk berbagai nilai k
3. Penerapan algoritma K-Means untuk membentuk kluster berdasarkan variabel luas panen, produktivitas, dan produksi
4. Visualisasi hasil klusterisasi melalui scatter plot dan peta tematik untuk menampilkan distribusi kluster antar provinsi

Menurut Herdiana et al. (2025), metode Elbow efektif untuk menentukan nilai k optimal dengan mengamati titik tekuk ("elbow") pada grafik WCSS. Berdasarkan hasil evaluasi internal tersebut, jumlah kluster optimal yang digunakan dalam penelitian ini adalah tiga kluster ($k = 3$).

3. HASIL DAN DISKUSI

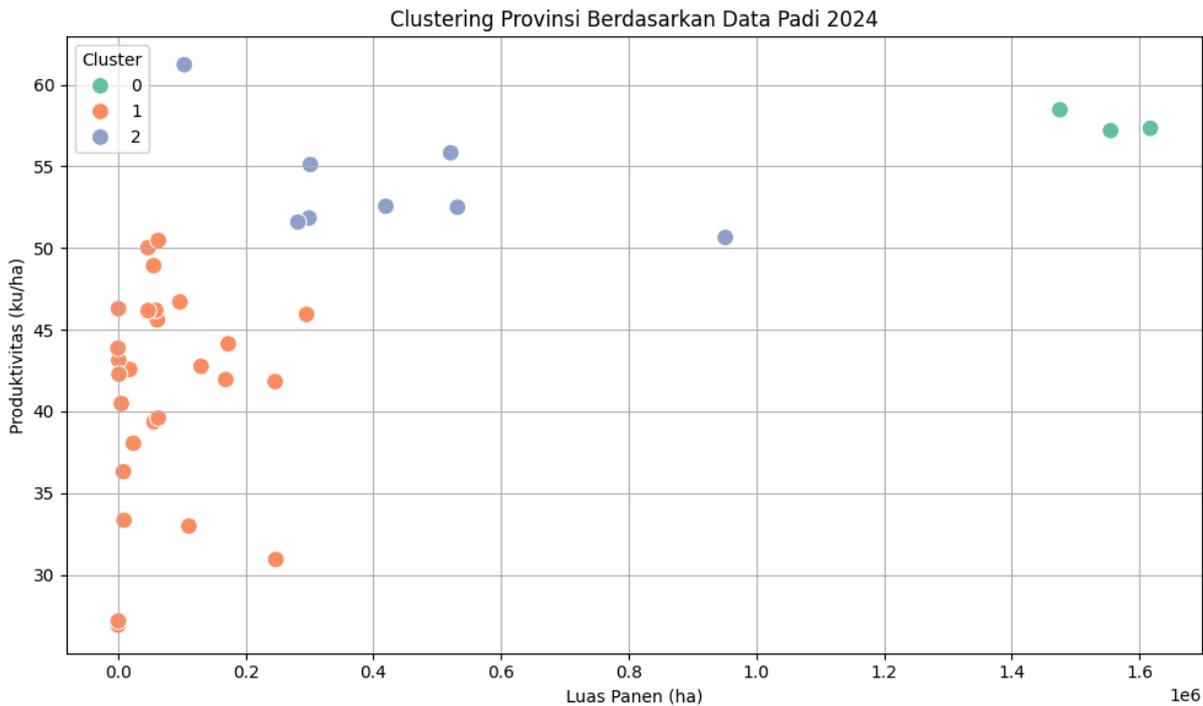
3.1 Penentuan Jumlah Kluster

Untuk menentukan jumlah kluster yang optimal, digunakan metode Elbow, yaitu dengan memplot nilai WCSS (Within-Cluster Sum of Squares) terhadap jumlah kluster. Titik "siku" pada kurva menunjukkan jumlah kluster terbaik, yang dalam kasus ini diperoleh $k = 3$. Berdasarkan metode Elbow, jumlah kluster optimal ditentukan sebanyak 3 kluster. Provinsi-provinsi di Indonesia kemudian dikelompokkan ke dalam tiga kategori berdasarkan kemiripan dalam hal luas panen, produktivitas, dan total produksi padi. Visualisasi kluster ditampilkan pada Gambar 1 berikut:





3.2 Penerapan K-Means



Data yang telah distandarisasi kemudian dikelompokkan menggunakan algoritma KMeans dari pustaka scikit-learn. Masing-masing provinsi kemudian mendapatkan label kluster berdasarkan kedekatan statistik dengan pusat kluster. Setelah proses clustering, dilakukan visualisasi 2D menggunakan scatter plot antara variabel seperti luas panen dan produktivitas untuk menunjukkan sebaran kluster. Selain itu, setiap kluster dianalisis secara deskriptif guna memahami karakteristik rata-rata dari provinsi-provinsi di dalamnya.

3.3 Statistika Deskriptif

Setelah dilakukan proses klusterisasi menggunakan algoritma K-Means, provinsi-provinsi di Indonesia terbagi ke dalam tiga kluster berdasarkan kemiripan karakteristik dalam indikator luas panen, produktivitas, dan total produksi padi tahun 2024. Untuk memahami lebih lanjut profil dari masing-masing kluster, dilakukan analisis statistik deskriptif terhadap nilai rata-rata dari ketiga variabel tersebut pada tiap kluster. Tabel berikut menyajikan ringkasan statistik deskriptif dari masing-masing kluster.

Cluster	Luas Panen (ha)	Produktivitas (ku/ha)	Produksi (ton)
0	1549041.43	57.66	8929537.42



1	73689.73	41.26	308539.96
2	426173.54	53.91	2252941.94

1. Klaster 0 merupakan kelompok provinsi dengan produksi padi sangat tinggi, ditandai oleh luas panen terbesar, produktivitas tinggi, dan total produksi mencapai hampir 9 juta ton.
2. Klaster 1 mencerminkan provinsi dengan produksi padi sangat rendah, memiliki luas panen dan produktivitas paling kecil, serta total produksi di bawah 500 ribu ton.
3. Klaster 2 menggambarkan provinsi dengan produksi padi kategori sedang, memiliki luas panen dan produktivitas yang cukup baik dengan total produksi sekitar 2,25 juta ton.

Hasil segmentasi ini memberikan gambaran yang lebih tajam terhadap heterogenitas wilayah dalam produksi padi nasional. Dengan adanya klasterisasi ini, pengambilan keputusan oleh pemerintah pusat maupun daerah dapat dilakukan secara lebih tepat sasaran, efisien, dan sesuai kebutuhan masing-masing wilayah.

3.4 Menampilkan Kelompok Statistika Deskriptif

Setelah dilakukan klasterisasi terhadap 38 provinsi di Indonesia berdasarkan data produksi padi tahun 2024, provinsi-provinsi tersebut terbagi ke dalam tiga klaster utama. Pengelompokan dilakukan berdasarkan tiga indikator utama, yaitu luas panen (hektar), produktivitas (kuintal per hektar), dan total produksi (ton). Setiap klaster menunjukkan pola karakteristik yang berbeda-beda, yang dapat dijadikan dasar dalam penyusunan kebijakan pertanian berbasis wilayah. Untuk memahami hasil segmentasi ini lebih lanjut, berikut disajikan daftar provinsi pada masing-masing klaster hasil algoritma K-Means.

Cluster	Provinsi
0	Jawa Tengah, Jawa Timur, Jawa Barat.
1	Riau, Kepulauan Bangka Belitung, Sumatera Barat, Bengkulu, Jambi, DKI Jakarta, Kepulauan Riau, DI Yogyakarta, Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Nusa Tenggara Timur, Kalimantan Barat, Maluku, Sulawesi Barat, Sulawesi Tenggara, Gorontalo, Maluku Utara, Kalimantan Utara, Sulawesi Tengah, Sulawesi Utara, Papua Selatan, Papua, Papua Pegunungan, Papua Tengah, Papua Barat Daya, Papua Barat.
2	Banten, Sumatera Utara, Sulawesi Selatan, Nusa Tenggara Barat, Bali, Aceh, Lampung, Sumatera Selatan.

Hasil klasterisasi menunjukkan bahwa Klaster 0 terdiri dari tiga provinsi sentra produksi padi nasional (Jawa Tengah, Jawa Timur, dan Jawa Barat) dengan produksi sangat tinggi. Klaster 1 mencakup mayoritas provinsi di Kalimantan, Papua, Maluku, dan beberapa wilayah lain dengan produksi padi rendah akibat keterbatasan lahan atau produktivitas. Sementara itu, Klaster 2 berisi provinsi dengan produksi menengah seperti Sumatera Utara, Sulawesi Selatan, dan NTB yang memiliki potensi untuk ditingkatkan.

3.5 Pemetaan Hasil Cluster

Untuk memberikan gambaran visual terhadap hasil klasterisasi produksi padi tahun 2024, dilakukan pemetaan provinsi berdasarkan klaster yang telah ditentukan. Setiap provinsi ditandai dengan warna berbeda sesuai klaster masing-masing: biru untuk Klaster 0 (produksi tinggi), hijau untuk Klaster 1 (produksi rendah), dan oranye untuk Klaster 2 (produksi sedang). Pemetaan ini bertujuan mempermudah identifikasi spasial wilayah produksi padi serta membantu perumusan kebijakan berbasis wilayah secara lebih efektif.



Dari peta tersebut terlihat bahwa Klaster 0 (biru) hanya mencakup provinsi di Pulau Jawa bagian tengah dan timur, yaitu Jawa Tengah, Jawa Timur, dan Jawa Barat, yang menjadi pusat utama produksi padi nasional. Klaster 1 (hijau) tersebar luas di wilayah Papua, Maluku, Kalimantan, sebagian Sulawesi, dan beberapa provinsi di Sumatera, menunjukkan tantangan dalam hal produksi padi yang relatif rendah. Sementara itu, Klaster 2 (oranye) terkonsentrasi di provinsi-provinsi seperti Sumatera Utara, Sumatera Selatan, Sulawesi Selatan, dan NTB, yang berada di level produksi menengah dan berpotensi untuk dikembangkan lebih lanjut.



KESIMPULAN

Penelitian ini bertujuan untuk mengelompokkan provinsi-provinsi di Indonesia berdasarkan karakteristik produksi padi tahun 2024 menggunakan metode K-Means Clustering. Dengan mempertimbangkan tiga variabel utama—luas panen, produktivitas, dan total produksi—penelitian ini berhasil mengidentifikasi tiga kluster utama yang mencerminkan tingkat produksi padi yang berbeda di masing-masing wilayah. Penentuan jumlah kluster dilakukan melalui Elbow Method yang menunjukkan bahwa segmentasi optimal terjadi pada tiga kluster.

Kluster pertama terdiri dari provinsi dengan tingkat produksi padi sangat tinggi, yaitu Jawa Barat, Jawa Tengah, dan Jawa Timur. Kluster ketiga berisi provinsi-provinsi dengan produksi sedang seperti Sumatera Utara, Sulawesi Selatan, NTB, dan lainnya yang menunjukkan potensi pengembangan lebih lanjut. Sementara itu, kluster kedua mencakup mayoritas provinsi di wilayah timur Indonesia seperti Papua, Maluku, dan sebagian Kalimantan yang memiliki produksi relatif rendah akibat keterbatasan lahan dan efisiensi produksi.

Hasil segmentasi ini memberikan wawasan penting bagi pembuat kebijakan untuk merancang strategi pertanian yang lebih tepat sasaran dan berbasis wilayah. Dengan memahami perbedaan karakteristik produksi antarprovinsi, intervensi pemerintah dapat diarahkan secara lebih efisien, baik dalam alokasi sumber daya, penyuluhan, maupun pembangunan infrastruktur pertanian. Oleh karena itu, pendekatan analisis data seperti ini sangat relevan untuk mendukung ketahanan pangan nasional yang inklusif dan berkelanjutan.

DAFTAR PUSTAKA

- [1] Wardhani, D. K., Astuti, D. M., & Huda, M. (2021). Implementasi Metode K-Means untuk Segmentasi Wilayah Berdasarkan Indikator Sosial Ekonomi. *Jurnal Sains dan Informatika*, 7(2), 55-62. <https://doi.org/10.21009/jsi.007.2.05>
- [2] Siregar, M. F., & Nursal, D. (2020). Segmentasi Wilayah Menggunakan K-Means Clustering untuk Perencanaan Pembangunan Daerah. *Jurnal Ilmiah Informatika KOMPUTA*, 9(2), 107-116. <https://doi.org/10.37676/komputa.v9i2.347>
- [3] Herdiana, I., Kamal, M. A., Triyani, M. N. E., & Renny. (2025). A More Precise Elbow Method for Optimum K-means Clustering. *arXiv preprint arXiv:2502.00851*.
- [4] Cokro, R. H., Jagaddatri, A., Sanchia, R. A., Pratama, A. R., & Damaliana, A. T. (2025). Analisis Kluster Provinsi Di Indonesia Berdasarkan Akses Dan Infrastruktur Listrik Menggunakan Algoritma K-Means. *Kohesi: Jurnal Multidisiplin Saintek*, 8(1).
- [5] Marwanti, M., Adi, S. H., Sosiawan, H., Sarwani, M., Irianto, G., & Wahab, M. I. (2023). Disrupsi sistem produksi padi nasional: Mampukah Indonesia memenuhi kebutuhan beras di tahun 2045? *Jurnal Triton*, 14(2), 403-421. <https://doi.org/10.47687/jt.v14i2.588>
- [6] Hakim, I., Rafid, M., & Anggraini, F. (2022). Pemanfaatan machine learning dengan algoritma X-means untuk pemetaan luas panen, produktivitas, dan produksi padi



- Indonesia. Building of Informatics, Technology and Science (BITS), 4(3), 1483-1494.
<https://doi.org/10.47065/bits.v4i3.2654>
- [7] Anisah, S., & Wachidah, L. (2023). Penerapan metode non-hierarchical clustering K-means untuk pengelompokan kabupaten/kota di Provinsi Jawa Barat berdasarkan faktor produksi padi tahun 2021. Bandung Conference Series: Statistics, 3(2), 323-332.
<https://doi.org/10.29313/bcss.v3i2.8148>
- [8] Wahyudin, E., Rudin, R. A., Kaslani, & Permana, S. E. (2024). Penerapan data mining pengelompokan produktivitas padi menggunakan algoritma K-means pada Provinsi Jawa Barat. JATI (Jurnal Mahasiswa Teknik Informatika), 8(1), 522-528.
- [9] Pribadi, T., Irsyada, R., Audytra, H., & Fatah, D. A. (2020). Implementasi algoritma K-means untuk klasterisasi potensi desa pada sektor produksi pertanian di Kabupaten Bojonegoro. Jurnal Teknik Informatika (JUTIN), 9(1), 20-23.
- [10] Hedyati, D., & Suartana, I. M. (2021). Penerapan principal component analysis (PCA) untuk reduksi dimensi pada proses clustering data produksi pertanian di Kabupaten Bojonegoro. Journal Information Engineering and Educational Technology (JIEET), 5(2), 49-54.
- [11] Dinesta, E. O, "Analisis peran sektor pertanian terhadap perekonomian provinsi Jawa Timur tahun 2010-2021", Institutional Repository UIN Syarif Hidayatullah Jakarta, 2024
- [12] Prasmita Dian Wijayati, "PERMINTAAN PANGAN SUMBER KARBOHIDRAT DI INDONESIA", *Jurnal Analisis Kebijakan Pertanian*, Vol. 17 No. 1, Juni 2019
- [13] I. W. Mulyana, "Dampak perubahan iklim terhadap produksi padi di Indonesia," *Jurnal Agroklimat*, vol. 10, no. 2, pp. 100-110, 2022..
- [14] Malau, Leo & Rambe, Khoiru & Ulya, Nur Arifatul & Purba, Arly. "Dampak perubahan iklim terhadap produksi tanaman pangan di indonesia". *Jurnal Penelitian Pertanian Terapan*. 23. 34-46.10.25181/jppt.v23i1.2418, 2023.
- [15] Sholeh, M., & Aeni, K. "Perbandingan Evaluasi Metode Davies Bouldin, Elbow dan Silhouette pada Model Clustering dengan Menggunakan Algoritma K-Means". *STRING (Satuan Tulisan Riset Dan Inovasi Teknologi)*, 8(1), 56.
<https://doi.org/10.30998/string.v8i1.16388>, 2023