JURNAL STATISTIKA DAN SAINS DATA
Volume 3 Edisi 1, Oktober, 2025, 82-93
https://journal.student.uny.ac.id/index.php/jssd
https://dx.doi.org/10.21831/jssd.v3i1.25385

# Analisis perbandingan K-Means, K-Medoids, dan Fuzzy C-Means untuk pengelompokan provinsi di Indonesia berdasarkan produksi padi tahun 2024

Ilham Mujahidin*, Universitas Terbuka
Siti Hadijah Hasanah, Uniersitas Terbuka
*e-mail: ilhammujahidin17@email.com

**Abstrak.** Padi menjadi salah satu komoditas penting dalam sektor pertanian Indonesia yang berperan krusial dalam menjaga ketahanan pangan nasional. Tujuan dari penelitian ini adalah untuk melakukan pengelompokan terhadap 38 provinsi di Indonesia berdasarkan kemiripan produksi padi guna memahami variasi spasial kinerja pertanian di berbagai provinsi. Dalam analisis ini digunakan tiga metode *clustering* yaitu K-Means, K-Medoids, dan Fuzzy C-Means. Analisis dilakukan dengan mempertimbangkan dua variabel utama yakni luas panen dan jumlah produksi padi tahun 2024 yang datanya bersumber dari Badan Pusat Statistik. Evaluasi kualitas klaster dilakukan menggunakan *Davies-Bouldin Index* (DBI). Berdasarkan hasil analisis, diketahui bahwa metode K-Means menghasilkan *clustering* paling optimal dengan memperoleh nilai DBI terendah sebesar 0,276 saat jumlah klaster $K = 8$, dibandingkan dengan K-Medoids (0,279 pada $K = 8$) dan Fuzzy C-Means (0,285 pada $K = 2$). Klaster yang terbentuk menunjukkan adanya pemisahan yang jelas antara provinsi dengan tingkat produksi tinggi sampai rendah. Provinsi dengan intensifikasi pertanian tinggi dan kontribusi besar terhadap produksi tergabung dalam klaster utama, sedangkan wilayah dengan keterbatasan sumber daya dan produksi rendah membentuk klaster tersendiri. Beberapa klaster lainnya mencerminkan karakteristik produksi sedang sampai tinggi dengan potensi pengembangan yang bervariasi. Temuan ini mencerminkan adanya keragaman kondisi agrikultur yang dipengaruhi oleh infrastruktur, intensifikasi, serta faktor geografis dan iklim.

***Kata kunci: K-means, K-medoids, fuzzy C-means, davies-bouldin index, produksi padi.***

***Abstract.*** *Rice is one of the important commodities in Indonesia's agricultural sector that plays a crucial role in maintaining national food security. The purpose of this study is to cluster 38 provinces in Indonesia based on similarities in rice production in order to understand the spatial variation of agricultural performance in different provinces In this analysis, three clustering methods were used, namely K-Means, K-Medoids, and Fuzzy C-Means. The analysis was conducted by considering two main variables, namely the area of harvest and the amount of rice production in 2024, whose data were sourced from the Central Bureau of Statistics. Evaluation of cluster quality was conducted using the Davies-Bouldin Index (DBI). Based on the analysis, it is known that the K-Means method produces the most optimal clustering by obtaining the lowest DBI value of 0.276 when the number of clusters $K = 8$, compared to K-Medoids (0.279 at $K = 8$) and Fuzzy C-Means (0.285 at $K = 2$). The clusters formed show a clear separation between provinces with high and low production levels. Provinces with high agricultural intensification and a large contribution to production belong to the main cluster,*

*while areas with limited resources and low production form a separate cluster. Several other clusters reflect medium to high production characteristics with varying development potential. This finding reflects the diversity of agricultural conditions influenced by infrastructure, intensification, and geographical and climatic factors.*

***Keywords: K-means, K-medoids, fuzzy C-means, davies-bouldin indeks, rice production.***

## PENDAHULUAN

Sektor pertanian, khususnya komoditas padi memiliki peranan krusial dalam menjaga ketahanan pangan nasional di Indonesia (Quirinno et al., 2024). Luas panen dan produksi padi menjadi indikator utama dalam menilai kinerja sektor ini. Menurut data dari Badan Pusat Statistik (BPS) tahun 2024, luas panen padi diprediksikan memperoleh sekitar 10,05 juta hektare, turun 167,25 ribu hektare atau 1,64 persen dibandingkan dengan tahun sebelumnya yang mencatatkan luas panen 10,21 juta hektare. Seiring dengan penurunan luas panen tersebut, produksi padi juga diperkirakan menurun menjadi sekitar 53,14 juta ton GKG, turun 838,27 ribu ton GKG (1,55 persen) dibanding produksi tahun 2023 yang mencapai 53,98 juta ton GKG. Penurunan ini kemungkinan besar dipengaruhi oleh berbagai faktor, seperti perubahan iklim, kebijakan pertanian, serta penggunaan teknologi dan infrastruktur yang berbeda di setiap provinsi (Malau et al., 2025).

Proses memahami pola dan karakteristik sektor pertanian di berbagai provinsi membutuhkan pendekatan analisis berbasis data, salah satunya melalui metode *clustering* atau pengelompokan. *Clustering* merupakan salah satu metode dalam data mining yang berfungsi untuk mengelompokkan data ke dalam beberapa kelompok berdasarkan kesamaan atribut atau karakteristik tertentu (Zheng et al., 2020). Algoritma *clustering* melakukan pembagian data ke dalam beberapa klaster yang di dalamnya terdapat kesamaan yang tinggi antar anggota, sementara antar klaster menunjukkan perbedaan yang signifikan (Ling et al., 2025). Tujuan utamanya adalah mengeksplorasi data guna menemukan struktur atau relasi yang tidak langsung terlihat, sehingga mampu memberikan gambaran yang lebih jelas mengenai kategori atau kelompok yang terdapat di dalamnya (Hendrastuty, 2024). *Clustering* dapat dilakukan dengan berbagai metode, di antaranya: K-Means, K-Medoids, dan Fuzzy C-Means. Keunggulan metode K-Means terletak pada kesederhanaan prosesnya serta efisiensinya saat mengklaster data berjumlah besar dengan waktu pemrosesan yang relatif cepat (Kousis & Tjortjis, 2021). K-Medoids merupakan metode cluster yang memiliki ketahanan yang lebih baik terhadap outlier, karena menggunakan medoids sebagai pusat cluster (Soesmono et al., 2025). Sementara itu, metode Fuzzy C-Means menjalankan proses klasterisasi dengan cara mendistribusikan data ke setiap klaster berdasarkan prinsip teori fuzzy. Metode ini mengukur sejauh mana tingkat kemungkinan suatu data termasuk sebagai anggota suatu klister (Turrahma et al., 2023).

Berdasarkan temuan dari Andini dkk. (2024), K-Means menjadi metode yang paling efektif digunakan dalam pengelompokan telur ayam ras di Provinsi Sumatera Selatan. Hal ini dibuktikan oleh nilai *Davies Bouldin Index* (DBI) sebesar 0,193 pada lima klaster yang lebih rendah dibandingkan DBI sebesar 0,268 dari algoritma K-Medoids pada tiga cluster. Sebaliknya, Rudianto dan Wijayanto (2023) menemukan bahwa metode K-Medoids memberikan hasil pengelompokan terbaik dengan lima klaster, berdasarkan evaluasi rasio Sw/Sb yang mana K-Medoids mempunyai rasio yang lebih kecil dibandingkan dengan K-Means. Di sisi lain Firdaus dkk., (2021) menemukan metode Fuzzy C-Means memperoleh nilai *Silhouette Index* (SI) sebesar 0,818, yang lebih tinggi dibandingkan nilai 0,569 dari algoritma

K-Means. Temuan ini mengindikasikan bahwa performa Fuzzy C-Means lebih optimal dibandingkan K-Means.

Penelitian ini bertujuan untuk melakukan perbandingan antara metode K-Means, K-Medoids, dan Fuzzy C-Means untuk mengelompokkan provinsi-provinsi di Indonesia menggunakan data produksi padi tahun 2024. Dari analisis ini, diharapkan memperoleh gambaran yang lebih jelas terkait kondisi pertanian dari setiap provinsi, sehingga dapat mendukung penyusunan kebijakan pertanian yang lebih efektif dan terstruktur.

## METODE

### Pengumpulan Data

Penelitian ini memanfaatkan data sekunder produksi padi di Indonesia tahun 2024 yang didapatkan melalui publikasi resmi dari Badan Pusat Statistik (BPS). Variabel yang dianalisis meliputi luas panen dan jumlah produksi padi. Informasi mengenai data tersebut dapat dilihat pada Tabel 1.

**Tabel 1. Tabel 1.** Data luas panen dan jumlah produksi padi tahun 2024

| No | Provinsi | Luas Panen | Produksi Padi |
|---|---|---|---|
| 1 | Aceh | 301196,4 | 1659966 |
| 2 | Sumatera Utara | 419463,5 | 2204876 |
| 3 | Sumatera Barat | 295279 | 1356468 |
| 4 | Riau | 56421,96 | 222055,7 |
| 5 | Jambi | 61625,68 | 281022,1 |
| 6 | Sumatera Selatan | 521092,2 | 2909412 |
| 7 | Bengkulu | 55775,09 | 272848,6 |
| 8 | Lampung | 531715,1 | 2791348 |
| 9 | Kep. Bangka Belitung | 18202,56 | 77489,79 |
| 10 | Kep. Riau | 113,33 | 305,09 |
| 11 | Dki Jakarta | 498,31 | 2306,54 |
| -- | -- | -- | -- |
| 38 | Papua Pegunungan | 9,66 | 42,38 |

(Sumber: Badan Pusat Statistik, 2024)

### Metode Penelitian

Penelitian ini menggunakan tiga metode *clustering* atau pengelompokan yaitu K-Means, K-Medoids, dan Fuzzy C-Means. Perbandingan nilai *Davies-Bouldin Index* (DBI) digunakan dalam memilih metode yang paling optimal.

a. K-Means

K-Means adalah metode pengelompokan yang mengklasifikasikan data dengan mempertimbangkan kedekatannya terhadap pusat klaster atau *centroid* (Triayudi et al., 2024). K-Means bekerja dengan mengelompokkan data ke dalam $c$ klaster dengan meminimalkan variasi intra-cluster. Proses iteratif dimulai dengan inisialisasi pusat cluster $v_k$, lalu data dikelompokkan berdasarkan jarak Euclidean terdekat, kemudian pusat cluster diperbarui menggunakan:

$$v_k = \frac{1}{|C_K|} \sum_{x_i \epsilon C_k} x_i \quad (1)$$

Di mana $C_K$ adalah himpunan data pada cluster ke-$k$, dan $|C_K|$ adalah jumlah data dalam cluster tersebut.

Fungsi objektif yang diminimalkan adalah:

$$J = \sum_{k=1}^{c} \sum_{x_i \epsilon C_k} x_i \, \|x_i - v_k\|^2 \quad (2)$$

Jarak antara dua titik $x_i$ dan $x_j$ menggunakan jarak Euclidean:

$$d(x_i, x_j) = \sqrt{\sum_{j=1}^{p} (X_{il} - C_{jl})^2} \quad (3)$$

b. K-Medoids

K-Medoids serupa dengan K-Means, tetapi metode ini menggunakan medoid sebagai pusat klister, yaitu data aktual yang meminimalkan total jarak ke semua titik dalam klaster (Supriyadi et al., 2021). Fungsi objektif:

$$J = \sum_{k=1}^{c} \sum_{x_i \epsilon C_k} d(x_i, m_k) \quad (4)$$

di mana $m_k$ adalah medoid (data representatif) dari cluster ke-$k$, dan $d(x_i, m_k)$ dapat berupa:

- Euclidean: $d(x_i, m_k) = \sqrt{\sum_{j=1}^{p} (X_{il} - C_{jl})^2}$
- Manhattan: $d(x_i, m_k) = \sum_{l=1}^{p} |x_{il} - m_{kl}|$

Pemilihan medoid dilakukan dengan mencari titik $x_j$ dalam cluster yang meminimalkan:

$$\sum_{x_i \epsilon C_k} d(x_i, x_j) \quad (5)$$

c. Fuzzy C-Means

Fuzzy C-Means adalah metode pengelompokan yang menentukan klaster optimal dengan menggunakan derajat keanggotaan sebagai dasar dalam menentukan sejauh mana suatu data termasuk dalam klaster tertentu. Selain itu, Fuzzy C-Means mampu membentuk klaster yang lebih bervariasi dan terukur (Nur et al., 2023). Fuzzy C-Means menggunakan pendekatan *soft clustering*, di mana data dapat tergabung ke dalam beberapa klaster dengan tingkat keanggotaan $\mu_{ik} \epsilon [0, 1]$ (Herijanto dan Riti, 2024). Fungsi objektif yang diminimalkan:

$$J_m = \sum_{i=1}^{n} \sum_{k=1}^{c} (\mu_{ik})^m \, \|x_i - v_k\|^2 \quad (6)$$

di mana:

- $\mu_{ik}$ : derajat keanggotaan data $x_i$ terhadap klaster $k$
- $m$ : parameter fuzziness $(m > 1)$

Rumus pembaruan pusat klaster:

$$v_k = \frac{\sum_{i=1}^{n} (\mu_{ik})^m \, x_i}{\sum_{i=1}^{n} (\mu_{ik})^m} \quad (7)$$

Rumus pembaruan derajat keanggotaan:

$$\mu_{ik} = \left( \sum_{j=1}^{c} \left( \frac{\|x_i - v_k\|}{\|x_i - v_j\|} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (8)$$

d. Evaluasi dengan *Davies-Bouldin Index* (DBI)

Metode DBI berfungsi dalam mengukur kualitas hasil pengelompokan dengan mempertimbangkan seberapa kompak dan terpisah suatu klaster. Nilai DBI dihitung dengan:

$$DBI = \frac{1}{c} \sum_{i=1}^{c} R_i \quad (9)$$

dengan:

$$R_i = \max_{j \neq 1} \left( \frac{S_i + S_j}{M_{ij}} \right) \quad (10)$$

Dimana

- $S_i = \frac{1}{|C_i|} \sum_{x \epsilon C_i} ||x - v_i||$: rata-rata jarak dalam klaster ke-$i$
- $M_{ij} = ||v_i - v_j||$ : jarak antar pusat klaster $i$ dan $j$

Nilai DBI yang lebih kecil menandakan hasil pengelompokan yang lebih optimal karena klaster lebih kompak dan terpisah dengan baik (Singh et al., 2020).

**Analisis Data**

Analisis data dilaksanakan menggunakan bahasa pemrograman Python melalui *Google Colaboratory*. Proses analisis terdiri atas beberapa tahapan sebagai berikut:

1. Persiapan Data

   Data diimpor ke Google Colaboratory menggunakan pustaka Python yang relevan. Setelah data berhasil dimuat, dilakukan peninjauan awal terhadap struktur dan ringkasan statistik untuk memahami karakteristik umum data.
2. Standarisasi Data

   Data numerik distandarisasi agar seluruh variabel berada pada skala yang sebanding. Proses ini bertujuan untuk mencegah dominasi variabel tertentu dalam perhitungan jarak selama proses *clustering*.
3. Penentuan Jumlah Klaster

   Penentuan jumlah klaster dilakukan secara eksploratif dengan menguji beberapa nilai K, yaitu K=2 hingga K=9. Tujuan dari langkah ini adalah untuk membandingkan kualitas hasil klaster berdasarkan jumlah klaster yang berbeda.
4. Proses *Clustering*

   Proses *clustering* dilakukan menggunakan tiga metode, yaitu:
   a. K-Means yang mengelompokkan data berdasarkan pusat rata-rata (*centroid*).
   b. K-Medoids yang membentuk klaster berdasarkan titik median sehingga lebih tahan terhadap outlier.
   c. Fuzzy C-Means memungkinkan sebuah data tergabung dalam beberapa klaster dengan derajat keanggotaan yang berbeda secara simultan.
5. Evaluasi Klaster

   Hasil pengelompokan dari setiap metode dan jumlah klaster dievaluasi menggunakan *Davies-Bouldin Index* (DBI). Nilai DBI yang semakin kecil

menghasilkan kualitas klaster yang lebih optimal, karena menandakan adanya pemisahan yang jelas antar klaster serta kekompakan yang tinggi dalam setiap klaster.

6. Visualisasi Klaster

   Visualisasi hasil klaster dilakukan untuk memperoleh gambaran mengenai pola distribusi dan pemisahan antar klaster, sehingga dapat mendukung interpretasi hasil analisis secara visual.

## HASIL DAN PEMBAHASAN

Tiga metode pengelompokan yang digunakan adalah K-Means, K-Medoids, dan Fuzzy C-Means. Data dari 38 provinsi dianalisis dan dikelompokkan menggunakan ketiga metode tersebut.

### Proses Clustering

Proses pengelompokan diawali dengan menetapkan jumlah klaster. Selanjutnya untuk memperoleh hasil yang optimal, diperlukan percobaan dalam menentukan jumlah klaster. Jumlah klaster yang dipilih yaitu $K = 2, K = 3, K = 4, K = 5, K = 6, K = 7, K = 8$, dan $K = 9$. Setiap nilai $K$ dievaluasi menggunakan DBI guna memilih jumlah klaster yang paling baik atau optimal.

**Tabel 2.** Hasil percobaan dengan metode K-Means

| Percobaan | $K1$ | $K2$ | $K3$ | $K4$ | $K5$ | $K6$ | $K7$ | $K8$ | $K9$ |
|---|---|---|---|---|---|---|---|---|---|
| $K = 2$ | 4 | 34 | - | - | - | - | - | - | - |
| $K = 3$ | 3 | 27 | 8 | - | - | - | - | - | - |
| $K = 4$ | 9 | 3 | 25 | 1 | - | - | - | - | - |
| $K = 5$ | 3 | 23 | 3 | 1 | 8 | - | - | - | - |
| $K = 6$ | 5 | 3 | 19 | 1 | 3 | 7 | - | - | - |
| $K = 7$ | 3 | 12 | 3 | 4 | 1 | 11 | 4 | - | - |
| $K = 8$ | 1 | 3 | 12 | 1 | 11 | 4 | 2 | 4 | - |
| $K = 9$ | 11 | 3 | 2 | 4 | 1 | 4 | 4 | 1 | 8 |

Berdasarkan hasil pada Tabel 2 menunjukkan jumlah anggota masing-masing klaster dari hasil metode K-Means untuk berbagai jumlah klaster ($K = 2$ hingga $K = 9$). Dari tabel ini dapat dilihat bagaimana distribusi data ke dalam masing-masing klaster berubah sesuai dengan nilai $K$. Pemilihan jumlah klaster yang optimal akan ditentukan pada bagian evaluasi menggunakan indeks DBI.

**Tabel 3.** Hasil percobaan dengan metode K-medoids

| Percobaan | $K1$ | $K2$ | $K3$ | $K4$ | $K5$ | $K6$ | $K7$ | $K8$ | $K9$ |
|---|---|---|---|---|---|---|---|---|---|
| $K = 2$ | 4 | 34 | - | - | - | - | - | - | - |
| $K = 3$ | 3 | 25 | 10 | - | - | - | - | - | - |
| $K = 4$ | 6 | 3 | 4 | 25 | - | - | - | - | - |
| $K = 5$ | 4 | 3 | 11 | 14 | 6 | - | - | - | - |
| $K = 6$ | 11 | 3 | 1 | 6 | 3 | 14 | - | - | - |
| $K = 7$ | 11 | 3 | 5 | 5 | 3 | 10 | 1 | - | - |
| $K = 8$ | 11 | 3 | 1 | 10 | 2 | 5 | 5 | 1 | - |
| $K = 9$ | 1 | 3 | 4 | 8 | 2 | 4 | 4 | 11 | 1 |

Tabel 3 menyajikan hasil jumlah anggota klaster dari metode K-Medoids untuk setiap percobaan jumlah klaster ($K$). Seperti pada K-Means, informasi ini penting untuk melihat sebaran jumlah anggota klaster dan menganalisis kestabilan hasil pengelompokan.

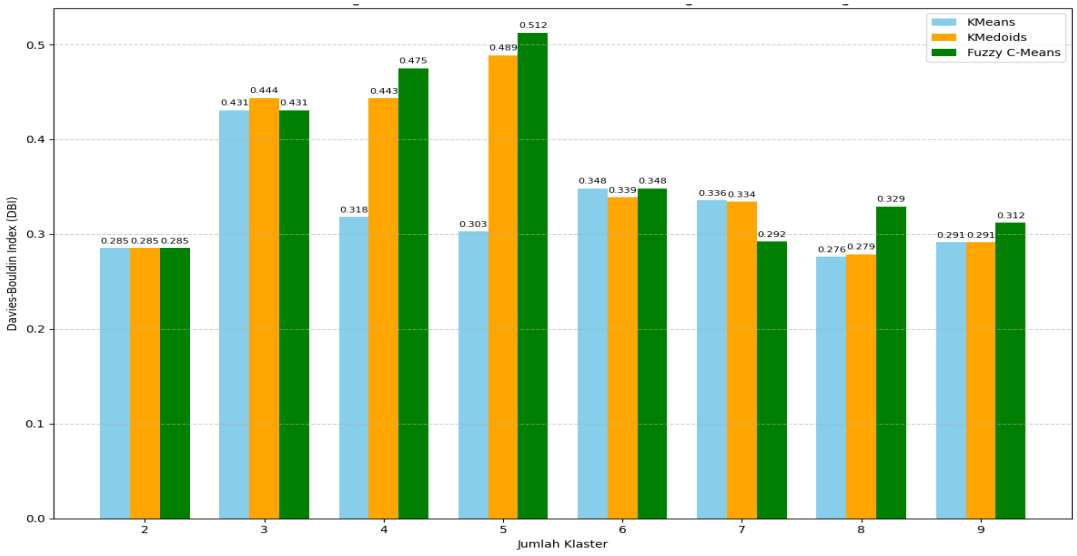
**Tabel 4.** Hasil percobaan dengan metode Fuzzy C-means

| Percobaan | $K1$ | $K2$ | $K3$ | $K4$ | $K5$ | $K6$ | $K7$ | $K8$ | $K9$ |
|---|---|---|---|---|---|---|---|---|---|
| $K = 2$ | 34 | 4 | - | - | - | - | - | - | - |
| $K = 3$ | 8 | 3 | 27 | - | - | - | - | - | - |
| $K = 4$ | 9 | 3 | 23 | 3 | - | - | - | - | - |
| $K = 5$ | 3 | 3 | 19 | 6 | 7 | - | - | - | - |
| $K = 6$ | 1 | 3 | 5 | 3 | 7 | 19 | - | - | - |
| $K = 7$ | 1 | 1 | 6 | 2 | 6 | 19 | 3 | - | - |
| $K = 8$ | 1 | 1 | 6 | 1 | 6 | 19 | 2 | 2 | - |
| $K = 9$ | 2 | 2 | 5 | 1 | 5 | 11 | 1 | 1 | 10 |

Tabel 4 memperlihatkan hasil pengelompokan menggunakan Fuzzy C-Means yaitu jumlah provinsi yang memiliki keanggotaan dominan pada masing-masing klaster. Fuzzy C-Means memungkinkan satu objek menjadi bagian dari beberapa klaster, namun ditampilkan klaster dominannya. Distribusi anggota untuk masing-masing $K$ juga diamati di sini.

**Evaluasi Clustering dengan Metode DBI**

Setelah pengelompokan menggunakan metode K-Means, K-Medoids, dan Fuzzy C-Means. Selanjutnya adalah mengevaluasi klaster dengan metode DBI (*Davies-Bouldin Indeks*).




**Gambar 1.** Perbandingan nilai DBI untuk berbagai metode clustering

Berdasarkan perbandingan nilai DBI (Gambar 1) bahwa K-Means memiliki nilai indeks terkecil saat $K = 8$ dengan nilai 0,276. Metode K-Medoids juga memperoleh nilai indeks terkecil saat $K = 8$ dengan nilai 0,279. Sedangkan Fuzzy C-Means mempunyai nilai indeks terkecil saat $K = 2$ dengan nilai 0,285. DBI dengan menggunakan metode K-Means sebesar 0,276 mempunyai nilai paling kecil dari metode lainnya. Nilai yang mendekati nol ini menunjukkan bahwa proses pengelompokan data berjalan dengan baik. Menurut Fathurrahman

et al., (2023) nilai DBI yang mendekati nol menunjukkan bahwa hasil pengelompokan data semakin optimal. Dengan demikian, K-Means merupakan metode yang paling optimal dalam pengelompokan data produksi padi jika dibandingkan dengan K-Medoids dan Fuzzy C-Means. Hasil ini sejalan dengan penelitian yang dilakukan oleh Syukron et al., (2022), yang menunjukkan bahwa K-Means menghasilkan struktur klaster yang lebih baik dan kinerja yang lebih optimal dibandingkan K-Medoids dan Fuzzy C-Means.

**Hasil Clustering**

Hasil menunjukkan bahwa metode K-Means mempunyai nilai DBI terkecil (0,276) dibandingkan metode lainnya, sehingga menghasilkan klaster dengan validitas terbaik.

**Tabel 5.** Hasil pengelompokan dengan metode K-Means saat $K = 8$

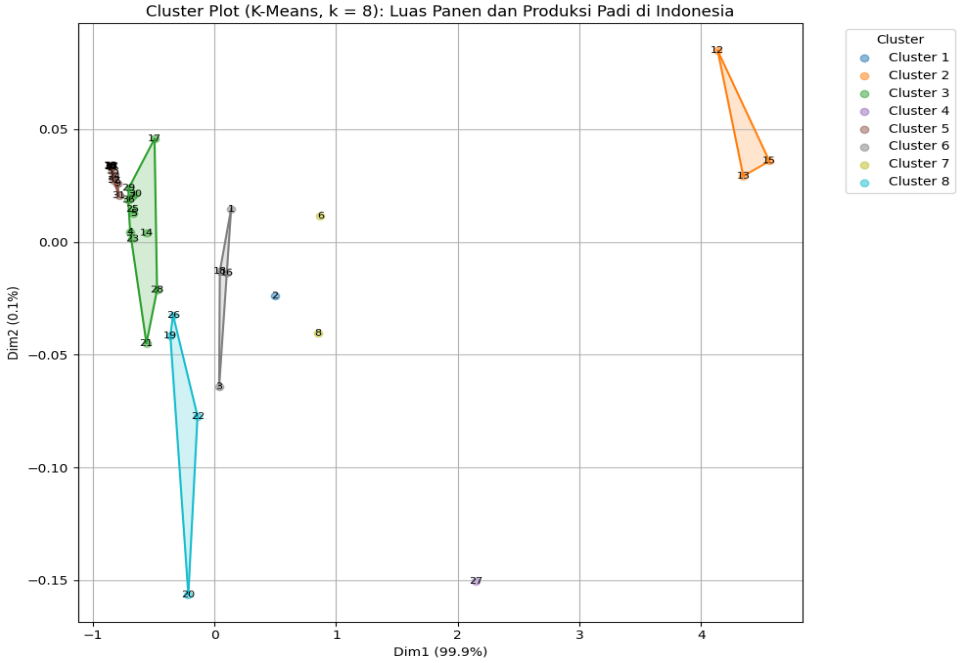
| Klaster | Provinsi |
|---|---|
| Klaster 1 | Sumatera Utara |
| Klaster 2 | Jawa Tengah, Jawa Barat, Jawa Timur |
| Klaster 3 | DI Yogyakarta, Gorontalo, Sulawesi Barat, Kalimantan Timur, Bali, Sulawesi Utara, Kalimantan Tengah, Bengkulu, Jambi, Riau, Sulawesi Tenggara, Papua Selatan |
| Klaster 4 | Sulawesi Selatan |
| Klaster 5 | Kep. Riau, Kep. Bangka Belitung, Papua Barat, Maluku Utara, Papua Barat Daya, Maluku, Papua Tengah, Kalimantan Utara, Papua, DKI Jakarta, Papua Pegunungan |
| Klaster 6 | Aceh, Nusa Tenggara Barat, Banten, Sumatera Barat |
| Klaster 7 | Lampung, Sumatera Selatan |
| Klaster 8 | Kalimantan Selatan, Sulawesi Tengah, Kalimantan Barat, Nusa Tenggara Timur |

Hasil pengelompokan provinsi di Indonesia menunjukkan bahwa metode K-Means dengan K=8 (Tabel 5) menghasilkan pola yang menggambarkan perbedaan karakteristik produksi padi antar wilayah. Provinsi dengan produksi padi yang tinggi cenderung tergabung dalam klaster yang sama, sedangkan provinsi dengan produksi rendah dikelompokkan secara terpisah.

- Klaster 2 merupakan kelompok utama yang terdiri dari wilayah dengan intensifikasi pertanian tinggi, infrastruktur yang memadai, dan produktivitas lahan yang tinggi. Klaster ini berperan besar dalam menyokong produksi padi di Indonesia.
- Klaster 5 mencakup wilayah-wilayah dengan tingkat produksi yang sangat rendah. Wilayah dalam klaster ini umumnya berada di kawasan timur Indonesia, dengan tantangan geografis dan keterbatasan sarana produksi yang signifikan.
- Klaster 3 dan Klaster 8 mencerminkan kelompok provinsi dengan karakteristik produksi rendah hingga menengah, tersebar di berbagai wilayah Indonesia. Meskipun kontribusinya terhadap produksi nasional tidak dominan, provinsi dalam kedua klaster ini menunjukkan kemiripan dalam pola produksi yang relatif terbatas namun konsisten.
- Klaster 4 terdiri dari satu provinsi dengan produksi tinggi di luar Pulau Jawa. Provinsi ini menempati posisi tersendiri karena perannya yang signifikan sebagai lumbung padi regional, meskipun skalanya tidak sebesar klaster utama.
- Klaster 1 dan Klaster 7 mencerminkan kelompok provinsi dengan tingkat produksi yang cukup tinggi, meskipun tidak sebanding dengan wilayah utama. Provinsi-provinsi dalam kedua klaster ini memiliki kontribusi yang penting, khususnya di luar Pulau Jawa.

- Klaster 6 mencakup wilayah dengan kategori produksi sedang. Provinsi-provinsi dalam klaster ini memiliki potensi pengembangan yang cukup baik dan dapat meningkat melalui intensifikasi dan perluasan lahan pertanian.

Secara keseluruhan, hasil pengelompokan ini mencerminkan adanya variasi spasial dalam produksi padi di Indonesia. Perbedaan ini dipengaruhi oleh potensi sumber daya, tingkat intensifikasi pertanian, infrastruktur pertanian, serta kondisi geografis dan iklim yang beragam di tiap wilayah.




**Gambar 2.** Visualisasi hasil pengelompokan dengan metode K-Means saat K=8

Visualisasi klaster yang terbentuk dari penerapan metode K-Means (Gambar 2) menunjukkan hasil pengelompokan provinsi berdasarkan produksi padi ke dalam delapan klaster. Terlihat bahwa provinsi dengan karakteristik produksi yang tinggi cenderung membentuk kelompok tersendiri yang terpisah secara jelas dari kelompok lainnya, mencerminkan kontribusi dominan terhadap produksi padi di Indonesia. Sementara itu, provinsi dengan tingkat produksi yang rendah tersebar dalam beberapa kelompok berbeda, menunjukkan kemiripan dalam pola produksi meskipun berasal dari wilayah geografis yang beragam. Pola penyebaran klaster ini mengindikasikan adanya perbedaan potensi sumber daya, tingkat intensifikasi pertanian, serta kondisi geografis dan infrastruktur yang memengaruhi kapasitas produksi padi antarwilayah di Indonesia.

## SIMPULAN

Berdasarkan hasil analisis, provinsi-provinsi di Indonesia berhasil dikelompokkan berdasarkan data produksi padi tahun 2024. Pengelompokan menggunakan metode K-Means memberikan hasil paling optimal dengan nilai DBI terendah sebesar 0,276 pada $K = 8$, lebih baik dibandingkan dengan K-Medoids (0,279 pada $K = 8$) dan Fuzzy C-Means (0,285 pada $K = 2$). Temuan ini mengindikasikan bahwa metode K-Means dapat membentuk klaster yang lebih terpisah dan kompak, sesuai prinsip bahwa semakin kecil nilai DBI semakin optimal kualitas pengelompokan. Visualisasi dan analisis lanjutan terhadap hasil pengelompokan

mengungkapkan adanya pola spasial yang mencerminkan perbedaan karakteristik agrikultur antar wilayah. Provinsi dengan produksi tinggi cenderung tergabung dalam klaster yang sama, sedangkan wilayah dengan tingkat produksi lebih rendah tersebar dalam klaster tersendiri. Secara keseluruhan, hasil pengelompokan ini mencerminkan keragaman potensi, tingkat intensifikasi, infrastruktur pertanian, serta kondisi geografis dan iklim yang memengaruhi pola produksi padi di Indonesia.

## DAFTAR PUSTAKA

Andini, R., Aminisari, S. T., Husin, V. A. F., & Jambak, M. I. (2024). Perbandingan algoritma K-Means dan K-Medoids untuk klasterisasi produksi telur ras ayam kampung di Provinsi Sumatera Selatan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(3), 3910–3915.

Badan Pusat Statistik. (2025). *Ringkasan Eksekutif luas Panen, dan Produksi Padi di Indonesia 2024 (Angka Tetap)*. Badan Pusat Statistik.

Fathurrahman, Harini, S., & Kusumawati, R. (2023). Evaluasi Clustering K-Means dan K-Medoid Pada Persebaran Covid-19 di Indonesia dengan Metode Davies-Bouldin Index (DBI). In *Jurnal MNEMONIC, 6*(2), 117-128.

Firdaus, H. S., Nugraha, A. L., Sasmito, B., Awaluddin, M., & Nanda, C. A. (2021). Perbandingan metode Fuzzy C-Means dan K-Means untuk pemetaan daerah rawan kriminalitas di Kota Semarang. *Elipsoida*, *4*(1), 58–64.

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (Jima-Ilkom)*, *3*(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Kousis, A., & Tjortjis, C. (2021). Data Mining Algorithms for Smart Cities: A Bibliometric Analysis. In *Algorithms*, *14*(8), 1-35. https://doi.org/10.3390/a14080242

Ling, L. S., dan Weiling, C. T. (2025). Enhancing Segmentation: A Comparative Study of Clustering Methods. *Institute of Electrical and Electronics Engineers (IEEE)*, vol. 13, 47418-47439. https://doi.org/10.1109/ACCESS.2025.3550339

Malau, H., Saragih, J. R., & Purba, T. (2025). Strategi Perencanaan Wilayah dalam Menanggulangi Dampak Perubahan Iklim pada Kawasan Pertanian. *PESHUM: Jurnal Pendidikan, Sosial dan Humaniora, 4*(2), 2316-2323.

Nur, I. M., Syifa, A. N. L., Kharis, M., Permatasari, S. H. 2023. Implementasi Fuzzy C-Means dalam Pengelompokan Hasil Panen Padi di Provinsi Bali. *Variance: Journal of Statistics and its Applications, 5*(1), 13-24. https://doi.org/10.30598/variancevol5iss1page132-4

Quirinno, R. S., Murtiana, S., Asmoro, N. (2024). Peran Sektor Pertanian dalam Meningkatkan Ketahanan Pangan dan Ekonomi Nasional. NUSANTARA*: Jurnal Ilmu Pengetahuan Sosial, 11*(7), 2811-2822. https://doi.org/10.31604/jips.v11i7.2024

Rudianto, R. D., & Wijayanto, A. W. (2023). Analisis perbandingan K-Means dan K-Medoids dalam pengelompokan provinsi berdasarkan Indeks Demokrasi Indonesia 2021. *Komputika: Jurnal Sistem Komputer*, *13*(1), 19–27. https://doi.org/10.34010/komputika.v13i1.10812

Singh, A. K., Mittal, S., Malhotra, P., dan Srivastava, Y. V. (2020). Clustering Evaluation by Davies Bouldin Indeks (DBI) in Cereal data using K-Means. *Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020*, pp. 306–310. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00057

Supriyadi, A., Triayudi, A., & Sholihati, I. D. (2021). Perbandingan Algoritma K-Means dengan K-Medoids pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas. *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, *6*(2), 229-240.

Soesmono, S., Pertiwi, R., Saputri, B., Putri, N., & Widodo, E. (2025). Pengelompokan Provinsi di Indonesia Berdasarkan Tingkat Pengangguran Tahun 2023 Menggunakan K-Medoids. *Emerging Statistics and Data Science Journal*, *3*(1), 498-515.

Syukron, H., Fauzi Fayyad, M., Junita Fauzan, F., Ikhsani, Y., & Gurning, U. R. (2022). Perbandingan K-Means K-Medoids dan Fuzzy C-Means untuk Pengelompokan Data Pelanggan dengan Model LRFM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science 2*(2), 76–83.

Tendean, T., & Purba, W. (2020). Analisis Cluster Provinsi Indonesia Berdasarkan Produksi Bahan Pangan Menggunakan Algoritma K-Means. *Jurnal Sains dan Teknologi*, *1*(2), 5-11. https://doi.org/10.34013/saintek.v1i 2.31

Triayudi, A., Pinastawa, I. W. R., Pratama, J. D., Ningsih, S. (2024). *Clustering*. Yogyakarta: PT Penamudamedia.

Turrahma R. N., Putra, A. N. C., Alfajri, M. D., Gusmanto, R., dan Oktoeberza, W. K. Z. (2023). Implementasi *Fuzzy C-Means* untuk Clustering data Harga Saham Harian Pada PT. Astra International TBK. *Jurnal Rekursif, 11*(1), 64-69.

Zheng, A. Cai, J., Yang, H., and Zhao, X. (2020). CPGAN: Curve Clustering Architecture Based on Projected Latent Vector of Generative Adversarial Network. *Institute of Electrical and Electronics Engineers (IEEE).* https://doi.org /10.1109/access.2020.299