

Analisis Kluster Produksi Cabai Menggunakan K-Means dan DBSCAN Berdasarkan Evaluasi Silhouette dan DBI

Muhammad Ridwanansyah Rahman¹, Jovan Bagas Pangestu², Masna Wati³, Joan Angelina Widians⁴

^{1,2,3,4}Program Studi Informatika, Universitas Mulawarman, Samarinda

e-mail: ¹mridwanansyahrahman@student.unmul.ac.id, ²2409106071@student.unmul.ac.id,

³masnawati@fkti.unmul.ac.id, ⁴angelwidians@unmul.ac.id

Abstrak

Cabai rawit (*Capsicum frutescens*) dan cabai keriting (*Capsicum annuum*) merupakan komoditas hortikultura strategis di Indonesia yang rentan terhadap fluktuasi harga akibat ketimpangan distribusi produksi antar provinsi. Penelitian ini bertujuan mengelompokkan 37 provinsi di Indonesia berdasarkan variabel produksi (kuintal) dan luas panen (hektare) kedua jenis cabai menggunakan algoritma K-Means dan Density-Based Spatial Clustering of Applications with Noise (DBSCAN). Data bersumber dari dua dataset Badan Pusat Statistik (BPS) tahun 2025 yang diintegrasikan menjadi satu dataset terpadu. Pra-pemrosesan meliputi integrasi data, penghapusan data tidak lengkap (Papua Selatan), dan normalisasi menggunakan Z-Score Standardization. Penentuan jumlah kluster optimal K-Means dilakukan melalui Elbow Method dan divalidasi dengan Silhouette Score serta Davies-Bouldin Index (DBI), menghasilkan $k=3$ sebagai konfigurasi terbaik (Silhouette=0,802; DBI=0,387). Parameter DBSCAN ditetapkan melalui analisis grafik k-distance dan trial and error ($\epsilon=1,6$; MinPts=3), menghasilkan tiga kluster dengan Jawa Timur teridentifikasi sebagai noise (Silhouette=0,811; DBI=0,573). K-Means unggul dalam separabilitas kluster (DBI lebih rendah), sedangkan DBSCAN unggul dalam kekompakan internal (Silhouette lebih tinggi). Tiga profil kluster K-Means terbentuk: Kluster Rendah (30 provinsi), Kluster Sedang (6 provinsi), dan Kluster Tinggi (1 provinsi: Jawa Timur). Hasil penelitian diharapkan menjadi acuan perumusan kebijakan pangan berbasis data di Indonesia.

Kata kunci—Cabai Rawit, Cabai Keriting, Clustering, K-Means, DBSCAN

1. PENDAHULUAN

Cabai rawit (*Capsicum frutescens*) dan cabai keriting (*Capsicum annuum*) termasuk komoditas hortikultura yang memiliki nilai ekonomi cukup tinggi di Indonesia. Selain menjadi kebutuhan utama masyarakat, cabai juga sering memengaruhi kondisi inflasi nasional [1]. Ketika produksi menurun atau distribusi terganggu, harga cabai biasanya mengalami kenaikan yang cukup signifikan.

Salah satu penyebab kondisi tersebut adalah ketidakmerataan produksi antar provinsi. Produksi cabai nasional masih banyak terpusat di Pulau Jawa, khususnya Jawa Timur dan Jawa Tengah. Sementara itu, beberapa wilayah lain seperti Kalimantan, Sulawesi, Maluku, dan Papua masih bergantung pada pasokan dari luar daerah [2]. Perbedaan kapasitas produksi ini menyebabkan selisih harga antar wilayah menjadi cukup tinggi, terutama saat terjadi gagal panen atau hambatan distribusi [3].

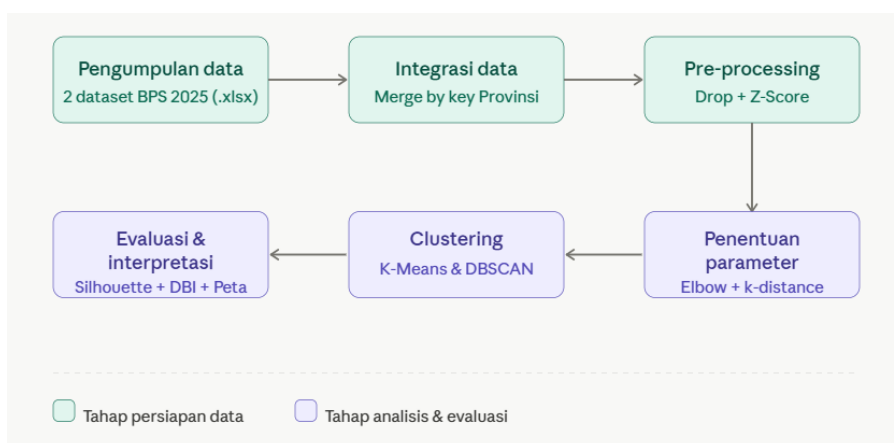
Untuk melihat pola produksi antar provinsi, salah satu metode yang dapat digunakan adalah *clustering*. Teknik ini termasuk bagian dari *unsupervised machine learning* yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik tanpa memerlukan label kelas [4]. Beberapa penelitian sebelumnya menunjukkan bahwa metode *clustering* cukup efektif diterapkan pada data pertanian. Penelitian [5] menggunakan K-Means untuk mengelompokkan produktivitas komoditas

pangan antar provinsi. Selanjutnya, hasil penelitian [6] menunjukkan bahwa DBSCAN mampu menghasilkan nilai *Silhouette Score* yang lebih baik pada pola data yang tidak beraturan. Selain itu, penelitian [7] serta [8] juga membahas penggunaan K-Means dan DBI pada kasus produktivitas bawang merah.

Walaupun demikian, penelitian yang secara khusus membandingkan K-Means dan DBSCAN pada data produksi cabai nasional masih cukup sedikit. Oleh karena itu, penelitian ini dilakukan untuk mengelompokkan provinsi berdasarkan produksi dan luas panen cabai menggunakan data BPS tahun 2025 sekaligus membandingkan performa kedua algoritma melalui *Silhouette Coefficient* dan DBI.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen menggunakan data sekunder. Tahapan penelitian dimulai dari pengumpulan dan integrasi data, dilanjutkan dengan prapemrosesan, proses klusterisasi menggunakan K-Means dan DBSCAN, hingga evaluasi kualitas kluster seperti yang ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

2.1 Sumber dan Deskripsi Data

Data penelitian berasal dari dua dataset Badan Pusat Statistik (BPS) tahun 2025, yaitu *Dataset Produksi Tanaman Sayuran dan Buah-Buahan Semusim Menurut Provinsi dan Jenis Tanaman*, dan *Dataset Luas Panen Tanaman Sayuran dan Buah-Buahan Semusim Menurut Provinsi dan Jenis Tanaman*. Fokus penelitian diarahkan pada komoditas cabai rawit dan cabai keriting untuk seluruh provinsi di Indonesia.

2.2 Integrasi Data

Karena data berasal dari dua file berbeda, proses integrasi dilakukan menggunakan Python dengan menggabungkan data berdasarkan provinsi sebagai primary key. Dari proses tersebut diperoleh empat variabel numerik utama, yaitu produksi cabai rawit, produksi cabai keriting, luas panen cabai rawit, dan luas panen cabai keriting.

2.3 Preprocessing Data

Terdapat dua tahapan preprocessing yang diterapkan. Cleaning data, penanganan data tidak lengkap: satu provinsi, yaitu Papua Selatan, dihapus dari dataset karena tidak memiliki data produksi dan luas panen cabai keriting (bernilai nol atau kosong), sehingga tidak memenuhi syarat kelengkapan fitur untuk proses clustering. Selain itu, baris agregat nasional ("Indonesia") juga dihapus karena bukan merupakan unit analisis individual. Setelah proses ini, jumlah observasi final adalah 37 provinsi. Kedua,

normalisasi data: seluruh fitur numerik dinormalisasi menggunakan *Z-Score Standardization* melalui fungsi *StandardScaler* dari pustaka *Scikit-learn*, sesuai persamaan (1):

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

di mana x adalah nilai asli fitur, μ adalah rata-rata kolom, dan σ adalah standar deviasi kolom. Normalisasi ini bertujuan untuk menyamakan skala antar variabel sebelum proses *clustering*, sehingga tidak ada satu variabel pun yang mendominasi perhitungan jarak secara tidak proporsional.

2.4 Algoritma K-Means

K-Means merupakan algoritma klusterisasi berbasis sentroid yang membagi data ke dalam beberapa kelompok berdasarkan kedekatan jarak data terhadap pusat klaster [9]. Penentuan jumlah klaster optimal dilakukan menggunakan *Elbow Method* dengan melihat penurunan nilai *Within-Cluster Sum of Squares* (WCSS). Setelah itu, hasil klaster dievaluasi menggunakan *Silhouette Score* dan DBI.

2.5 Algoritma DBSCAN

DBSCAN merupakan metode klusterisasi berbasis kepadatan yang dapat membentuk klaster tanpa menentukan jumlah kelompok di awal [10]. Algoritma ini menggunakan parameter ϵ sebagai radius kedekatan dan *MinPts* sebagai jumlah minimum titik dalam radius tertentu. Nilai ϵ ditentukan menggunakan grafik *k-distance*, sedangkan *MinPts* diperoleh melalui beberapa pengujian sesuai karakteristik data

2.6 Metrik Evaluasi

Evaluasi kualitas klaster dilakukan secara internal menggunakan dua metrik tanpa ground truth. Pertama, *Silhouette Coefficient* (SC) mengukur kemiripan setiap titik data dengan klasternya sendiri dibandingkan klaster terdekat lainnya, dengan rentang nilai $[-1, +1]$ [11]. Nilai SC mendekati +1 menunjukkan klaster yang kompak dan terpisah dengan baik. SC dihitung melalui persamaan (2):

$$SC(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}} \quad (2)$$

di mana $a(i)$ adalah rata-rata jarak intra-klaster dan $b(i)$ adalah rata-rata jarak terkecil ke klaster lain untuk titik i .

Kedua, *Davies-Bouldin Index* (DBI) mengukur rasio antara dispersi intra-klaster dan jarak antar-sentroid [12]. Nilai DBI yang lebih rendah (mendekati nol) menunjukkan klaster yang lebih baik. DBI dihitung melalui persamaan (3):

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{S_i + S_j}{M_{ij}} \right) \quad (3)$$

di mana k adalah jumlah klaster, s_i adalah dispersi rata-rata klaster ke- i , dan $d(c_i, c_j)$ adalah jarak antara sentroid klaster i dan j .

3. HASIL DAN PEMBAHASAN

3.1 Hasil Integrasi dan Preprocessing Data

Proses integrasi dari dua dataset BPS berhasil menghasilkan satu dataset terpadu dengan empat variabel fitur numerik per provinsi. Ditemukan ketidaklengkapan data pada satu provinsi, yaitu Papua Selatan, yang tidak memiliki data produksi dan luas panen cabai keriting. Mengingat ketidaklengkapan

ini mencakup dua dari empat fitur utama, diputuskan untuk menghapus provinsi tersebut dari dataset guna menjaga integritas analisis. Selain itu, baris agregat "Indonesia" yang merepresentasikan nilai rata-rata nasional juga dihapus karena tidak merepresentasikan unit wilayah individual. Dataset final hasil integrasi terdiri dari 37 provinsi seperti yang ditunjukkan pada Tabel 1.

Table 1. Hasil Integrasi Data

Provinsi	Produksi Rawit (Kw)	Produksi Keriting (Kw)	Luas Panen Rawit (Ha)	Luas Panen Keriting (Ha)
Aceh	859.676,66	1.062.152,06	6.852,88	9631,62
Sumatera Utara	1.176.582,82	2.455.138,08	9.886,21	16.169,10
Sumatera Barat	314.904,30	1.217.582,91	3.526,20	9.411,40
Riau	66.824,69	67.879,40	1.039,26	895,90
Jambi	421.740,79	2.139.753,27	1.363,49	6.195,30
Papua Barat	3.720,90	310,94	249,84	108,40
Papua Barat Daya	30.213,61	3.963,20	276,32	48,30
Papua	1.368,00	504,00	86,60	50,00
Papua Tengah	26.407,00	663,00	456,50	14,00
Papua Pegunungan	253,60	80,70	91,00	29,70

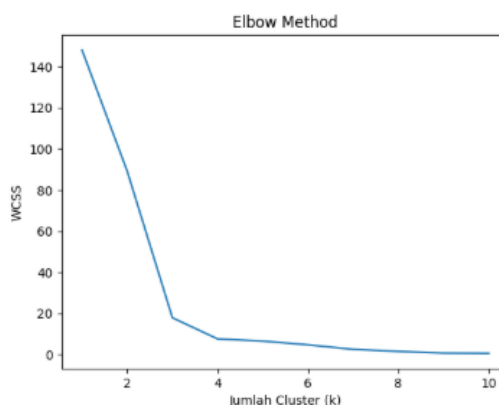
Selanjutnya, seluruh fitur numerik dinormalisasi menggunakan Z-Score Standardization melalui fungsi StandardScaler. Normalisasi bertujuan menyamakan skala antar variabel agar tidak ada satu variabel yang mendominasi perhitungan jarak secara tidak proporsional. Hasil preprocessing ditunjukkan pada Tabel 2.

Tabel 2. Hasil Preprocessing Data

Provinsi	Produksi Rawit (Kw)	Produksi Keriting (Kw)	Luas Panen Rawit (Ha)	Luas Panen Keriting (Ha)
Aceh	0,325173	1,046398	0,071710	1,465091
Sumatera Utara	0,597935	3,088804	0,277388	2,836303
Sumatera Barat	-0,143715	1,274291	-0,153858	1,418900
Riau	-0,357238	-0,411413	-0,322487	-0,367194
Jambi	-0,051760	2,626385	-0,300502	0,744336
Papua Barat	-0,411551	-0,510482	-0,376014	-0,532369
Papua Barat Daya	-0,388749	-0,505127	-0,374219	-0,544975
Papua	-0,413576	-0,510199	-0,387083	-0,544618
Papua Tengah	-0,392025	-0,509966	-0,362002	-0,552169
Papua Pegunungan	-0,414535	-0,510820	-0,386785	-0,548876

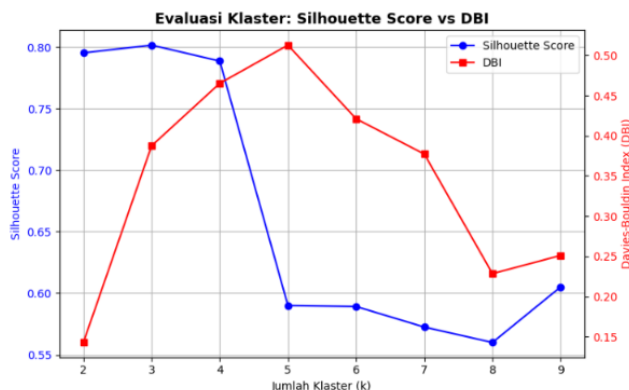
3.2 Penentuan Parameter Optimal K-Means

Pengujian jumlah kluster menggunakan *Elbow Method* dilakukan pada rentang $k=2$ hingga $k=5$. Dari grafik yang diperoleh, penurunan nilai WCSS mulai melandai pada $k=3$ sehingga titik tersebut dipilih sebagai kandidat jumlah kluster terbaik seperti yang terlihat pada Gambar 2.



Gambar 2. Grafik Elbow Method (WCSS vs Jumlah Kluster k)

Pada tahap kedua, konfirmasi kuantitatif dilakukan melalui *Silhouette Score* dan *Davies-Bouldin Index* (DBI), hasilnya disajikan pada Gambar 3. Kedua metrik memperlihatkan pola yang berlawanan: *Silhouette Score* mencapai nilai tertinggi pada $k=3$, sedangkan DBI mencapai nilai terendah pada $k=2$.



Gambar 3. Perbandingan Silhouette Score dan DBI per Nilai k

Rangkuman ketiga metrik evaluasi disajikan pada Tabel 3, yang menjadi dasar penetapan $k=3$ sebagai konfigurasi optimal.

Tabel 3. Perbandingan Nilai Evaluasi K-Means pada Berbagai Nilai k

Jumlah Kluster (k)	WCSS	Silhouette Score	Davies-Bouldin Index
2	89,1302	0,795	0,143
3	17,8936	0,802	0,387
4	7,6131	0,789	0,465
5	6,5520	0,590	0,513

Berdasarkan Tabel 3, konfigurasi $k = 2$ unggul pada DBI (0,143), namun hanya didukung satu metrik. Sebaliknya, $k=3$ didukung oleh dua metrik secara bersamaan: *Silhouette Score* tertinggi (0,802 dibandingkan 0,795) dan konfirmasi *Elbow Method* yang menunjukkan penurunan WCSS paling signifikan tepat pada transisi menuju $k=3$. Selain itu, dua kluster dinilai terlalu kasar secara interpretatif, sebab provinsi dengan kapasitas produksi sedang seperti Sumatera Utara dan Jawa Barat akan tersatukan dengan kelompok berkapasitas rendah, sehingga diferensiasi yang diperlukan untuk rekomendasi kebijakan pangan tidak dapat dilakukan. Nilai DBI $k=3$ sebesar 0,387 masih tergolong valid dan menunjukkan separasi antara kluster yang memadai. Dengan demikian, $k=3$ ditetapkan sebagai konfigurasi yang paling optimal.

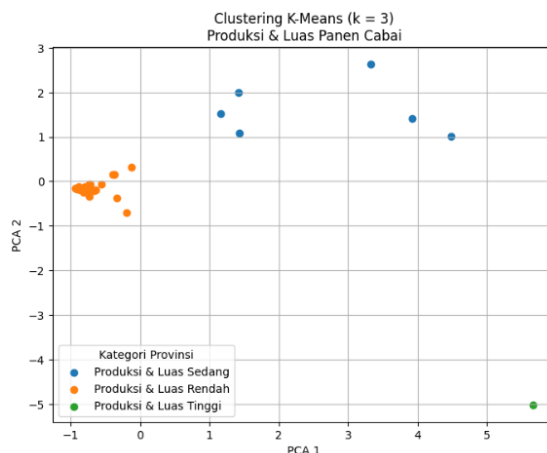
3.3 Hasil Clustering K-Means ($k=3$)

Dengan konfigurasi $k=3$, algoritma K-Means menghasilkan tiga kluster yang terdistribusi sebagaimana ditunjukkan pada Tabel 4.

Tabel 4. Komposisi Kluster Hasil K-Means ($k=3$)

Kluster	Label	Provinsi	Jumlah
Kluster 0	Produksi & Luas Rendah	Riau, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, NTB, NTT, serta seluruh provinsi Kalimantan, Sulawesi, Maluku, dan Papua (kecuali yang terklasifikasi lain)	30
Kluster 1	Produksi & Luas Sedang	Aceh, Sumatera Utara, Sumatera Barat, Jambi, Jawa Barat, Jawa Tengah	6
Kluster 2	Produksi & Luas Tinggi	Jawa Timur	1

Visualisasi hasil clustering K-Means menggunakan reduksi dimensi *Principal Component Analysis* (PCA) menjadi dua dimensi ditunjukkan pada Gambar 4. *Scatter plot* tersebut memperlihatkan bahwa ketiga kluster memiliki separabilitas yang jelas dan tidak terdapat tumpang tindih yang signifikan antar kluster.



Gambar 4. Scatter Plot PCA Hasil Clustering K-Means (k=3)

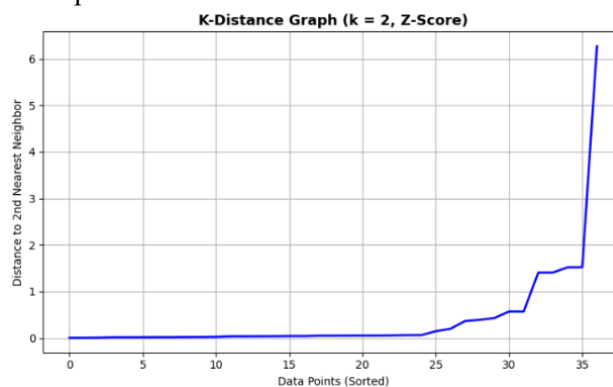
3.4 Penentuan Parameter DBSCAN dan Hasil Clustering

Pada DBSCAN, penggunaan MinPts sebesar 4 atau 5 membuat seluruh data cenderung tergabung menjadi satu kelompok besar. Karena jumlah data relatif sedikit, nilai MinPts kemudian diturunkan menjadi 3 agar pola kepadatan lokal lebih mudah terdeteksi [13]. Ringkasan pengujian awal tersebut ditunjukkan pada Tabel 5

Tabel 5. Ringkasan Observasi Awal Penentuan MinPts

MinPts	Range ϵ (Epsilon)	Jumlah Kluster	Kesimpulan
5	1,0 – 2,0	1	<i>Tidak Optimal (Single Cluster)</i>
4	1,0 – 2,0	1	<i>Tidak Optimal (Single Cluster)</i>
3	1,0 – 2,0	> 1	Dilanjutkan ke Tahap Optimasi

Setelah menetapkan MinPts=3, tahap selanjutnya adalah menentukan nilai ϵ (epsilon) yang optimal. Analisis grafik k-distance dengan $k=2$ ($k=\text{MinPts} - 1$) sebagaimana ditunjukkan pada Gambar 5 mengidentifikasi adanya titik siku (elbow point) yang jelas pada rentang jarak 0,7 hingga 1,8. Rentang ini menjadi panduan untuk eksplorasi nilai ϵ secara lebih terfokus.



Gambar 5. Grafik K-Distance ($k=2$) untuk Penentuan Parameter ϵ DBSCAN

Tabel 6. Evaluasi Metrik Silhouette dan DBI pada Berbagai Nilai ϵ (MinPts=3)

ϵ (Epsilon)	Jumlah Klaster	Silhouette Score	DBI ↓	Keterangan
1,0 - 1,5	1	-	-	Single Cluster
1,6	3	0,811	0,573	Optimal
1,7	3	0,811	0,573	Optimal
1,8	2	0,803	0,325	Merging Cluster

Hasil evaluasi pada Tabel 6 menunjukkan bahwa $\epsilon=1,0 - 1,5$ tidak berhasil memecah data menjadi lebih dari satu klaster, sehingga tidak dapat dievaluasi. Nilai $\epsilon=1,6$ dan $\epsilon=1,7$ keduanya menghasilkan tiga klaster dengan nilai Silhouette Score dan DBI yang identik, sehingga keduanya dikategorikan sebagai parameter optimal. Sementara $\epsilon=1,8$ menyebabkan penggabungan klaster (*merging cluster*) menjadi hanya dua kelompok dengan Silhouette yang lebih rendah. Meskipun DBI pada $\epsilon=1,8$ lebih baik dibandingkan $\epsilon=1,6$, konfigurasi ini tidak dipilih karena adanya penurunan homogenitas internal kelompok akibat penggabungan paksa tersebut, yang membuat struktur klasifikasinya kurang merepresentasikan pola sebaran asli untuk interpretasi kebijakan pangan.

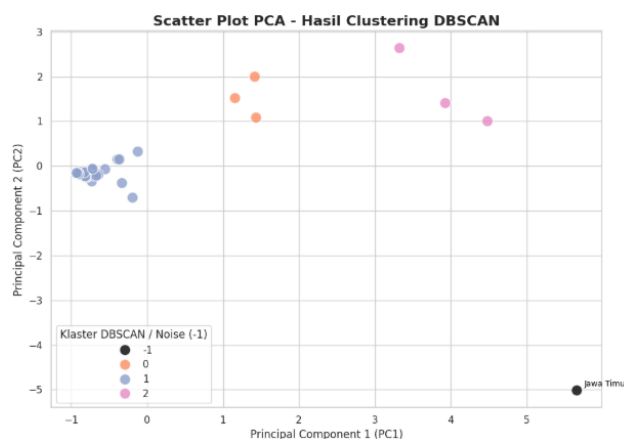
Dari dua nilai optimal, dipilih $\epsilon=1,6$ sebagai parameter final berdasarkan prinsip parsimoni dalam pemodelan statistik: ketika dua konfigurasi menghasilkan performa yang setara, dipilih parameter yang lebih konservatif (nilai lebih kecil). Radius lingkungan yang lebih ketat ($\epsilon=1,6$) berarti algoritma menggunakan kriteria kepadatan yang lebih presisi, menghasilkan klaster yang lebih kompak secara geometris. Pendekatan ini sejalan dengan prinsip pemilihan parameter yang dikemukakan Driyandita et al. [13].

DBSCAN dengan $\epsilon = 1,6$ dan MinPts=3 menghasilkan tiga klaster utama dan menetapkan Jawa Timur sebagai noise seperti yang ditunjukkan pada Tabel 7. Jawa Timur teridentifikasi sebagai noise karena karakteristik produksi dan luas panennya yang ekstrem, sehingga gagal memenuhi kriteria *core point* dalam radius ϵ yang ditentukan. Fenomena ini menegaskan keunggulan inheren DBSCAN dalam mengidentifikasi *outlier* secara otomatis tanpa asumsi distribusi data [14]. Secara kontekstual, temuan ini sangat valid mengingat posisi Jawa Timur sebagai sentra produksi cabai nasional dengan skala yang jauh melampaui rata-rata provinsi lainnya.

Tabel 7. Komposisi Klaster Hasil DBSCAN ($\epsilon=1,6$; MinPts=3)

Klaster	Label	Provinsi	Jumlah
Klaster 1	Produksi & Luas Rendah	Riau, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, NTB, NTT, seluruh provinsi Kalimantan, Sulawesi, Maluku, dan Papua (kecuali yang terklasifikasi lain)	30
Klaster 0	Produksi & Luas Sedang	Aceh, Sumatera Barat, Jambi	3
Klaster 2	Produksi & Luas Tinggi	Sumatera Utara, Jawa Barat, Jawa Tengah	3
Noise (-1)	<i>Outlier</i>	Jawa Timur	1

Visualisasi hasil clustering DBSCAN menggunakan reduksi dimensi PCA yang ditunjukkan pada Gambar 6 yang memperlihatkan separabilitas jelas antar-kelompok serta posisi Jawa Timur sebagai pencilan (*noise*).



Gambar 6. Scatter Plot PCA Hasil Clustering DBSCAN ($\epsilon=1,6$; MinPts=3)

3.5 Perbandingan Metrik Evaluasi K-Means dengan DBSCAN

Analisis komparatif pada bagian ini bertujuan mengevaluasi perbedaan mendasar antara K-Means dan DBSCAN dalam memproses data produksi cabai di Indonesia. Pengujian dilakukan dengan membandingkan metrik evaluasi internal (Silhouette Score dan Davies-Bouldin Index) terhadap karakteristik geometris dan fungsional hasil partisi masing-masing algoritma. Ringkasan perbandingan performa kedua metode tersebut disajikan pada Tabel 8.

Tabel 8. Perbandingan Metrik Evaluasi K-Means dan DBSCAN

Algoritma	Parameter	Jml. Kluster	Silhouette	DBI
K-Means	$k = 3$	3	0,802	0,387
DBSCAN	$\epsilon=1,6$; MinPts=3	3 + 1 noise	0,811	0,573

Berdasarkan Tabel 8, baik model kluster optimal DBSCAN yang menggunakan nilai $\epsilon=1,6$ dan MinPts=3, maupun model kluster optimal K-Means yang menggunakan $k=3$, menunjukkan bahwa data produksi dan luas panen cabai per provinsi ini paling cocok dikelompokkan menjadi 3 kluster. Hal ini dibuktikan oleh konsistensi hasil pengelompokan kedua algoritma yang sama-sama membentuk tiga kelompok provinsi dengan karakteristik yang berbeda satu sama lain.

Ditinjau dari nilai DBI, K-Means menghasilkan nilai yang lebih rendah dibandingkan DBSCAN, yang menunjukkan bahwa pemisahan antar-sentroid kluster K-Means lebih tegas secara proporsional terhadap dispersi internalnya. Kondisi ini menjadikan seluruh 37 provinsi terklasifikasi ke dalam tiga kelompok tanpa pengecualian, sehingga memudahkan interpretasi secara menyeluruh. Di sisi lain, DBSCAN menghasilkan nilai Silhouette Score yang sedikit lebih tinggi dibandingkan K-Means, yang menunjukkan bahwa anggota dalam setiap kluster DBSCAN memiliki kemiripan karakteristik yang lebih kuat satu sama lain. Selain itu, kemampuan DBSCAN dalam mengidentifikasi Jawa Timur sebagai noise merupakan keunggulan algoritmik tersendiri. Noise dalam konteks DBSCAN bukan berarti data yang bermasalah, melainkan menandakan bahwa provinsi tersebut memiliki karakteristik yang sangat berbeda dari provinsi lainnya sehingga tidak dapat dimasukkan ke dalam kluster manapun, memberikan informasi tentang data yang terpencil yang tidak dapat diperoleh dari K-Means [15].

Temuan bahwa Jawa Timur menjadi outlier dalam DBSCAN sekaligus satu-satunya anggota Kluster Tinggi dalam K-Means secara konsisten mengonfirmasi bahwa provinsi ini memiliki karakteristik produksi dan luas panen cabai yang benar-benar outlier dalam skala nasional. Kondisi ini juga mengindikasikan bahwa ketergantungan pasokan nasional terhadap satu provinsi dominan merupakan kelemahan struktural dalam sistem ketahanan pangan cabai Indonesia yang perlu mendapat perhatian kebijakan.

4. KESIMPULAN

Berdasarkan seluruh rangkaian eksperimen dan analisis yang telah dilakukan, penelitian ini menghasilkan beberapa kesimpulan sebagai berikut:

1. K-Means dengan $k=3$ terbukti menjadi konfigurasi optimal berdasarkan nilai Silhouette Score tertinggi dari rentang k yang diuji, serta didukung oleh titik penurunan WCSS paling signifikan pada Elbow Method. Tiga kluster yang terbentuk memisahkan secara tegas 30 provinsi pada kluster rendah, 6 provinsi pada kluster sedang, dan Provinsi Jawa Timur secara tunggal pada kluster tinggi.
2. DBSCAN dengan $\epsilon=1,6$ dan $\text{MinPts}=3$ berhasil membentuk tiga kluster, serta secara unik mampu mendeteksi Jawa Timur sebagai noise, mengonfirmasi secara algoritmik status provinsi tersebut sebagai *outlier* produksi dalam skala nasional.
3. Analisis komparatif menunjukkan bahwa K-Means dan DBSCAN memiliki keunggulan pada dimensi yang berbeda: DBSCAN unggul pada Silhouette Score, sedangkan K-Means unggul pada DBI.
4. Berdasarkan hasil pengujian model kluster optimal K-Means dengan $k=3$ dan model kluster optimal DBSCAN dengan $\epsilon=1,6$ dan $\text{MinPts}=3$, data produksi dan luas panen cabai rawit serta cabai keriting per provinsi di Indonesia ini paling cocok dikelompokkan menjadi 3 kluster. Kedua algoritma secara konsisten membentuk tiga kelompok provinsi yang mencerminkan perbedaan karakteristik produksi secara signifikan, yaitu kelompok provinsi dengan produksi dan luas panen rendah, sedang, dan tinggi.
5. Dominasi Jawa Timur sebagai satu-satunya provinsi dalam kluster produktivitas tertinggi, sekaligus teridentifikasi sebagai data pencilan dalam DBSCAN, mengindikasikan ketimpangan struktural yang signifikan dalam distribusi produksi cabai nasional. Temuan ini menegaskan urgensi program pemberdayaan pertanian di provinsi-provinsi kluster rendah, khususnya di wilayah Kalimantan, Sulawesi, Maluku, dan Papua.

5. SARAN

Penelitian berikutnya dapat menambahkan variabel lain seperti curah hujan, kondisi cuaca, dan fluktuasi harga pasar agar analisis yang dihasilkan menjadi lebih lengkap. Selain itu, metode *soft clustering* seperti *Fuzzy C-Means* juga dapat dicoba untuk melihat kemungkinan suatu provinsi memiliki kedekatan dengan lebih dari satu kluster.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Badan Pusat Statistik (BPS) Republik Indonesia atas ketersediaan data publik yang menjadi fondasi penelitian ini, serta kepada seluruh pihak yang telah memberikan dukungan akademik dalam penyelesaian penelitian ini.

DAFTAR PUSTAKA

- [1] A. D. Lestari, E. Erlikasna, C. Ridho, I. Breta, and M. Daniyal, "Dampak Fluktuasi Harga Beras , Bawang Merah , Cabai Terhadap Inflasi Impact of Price Fluctuations of Rice , Shallots , Chilies on Inflation," *J. Sos. Ekon. Pertan.*, vol. 20, no. 2, pp. 219–226, 2024, doi: 10.20956/jsep.v20i2.35057.
- [2] I. Marina, D. Sukmawati, E. Juliana, and N. Safa, "Dinamika Pasar Komoditas Pangan Strategis : Analisis Fluktuasi Harga Dan Produksi," *J. Ilm. Pertan.*, vol. 12, no. 1, pp. 160–168, 2024, doi: 10.35138/paspalum.v12i1.700.

- [3] T. S. Wahyuni, R. Satriani, and A. N. Mandamdari, "Pengaruh Fluktuasi Harga Cabai Rawit Merah Terhadap Inflasi di Kabupaten Banyumas," *J. Pemikir. Masy. Ilm. Berwawasan Agribisnis*, vol. 10, no. 2, pp. 1866–1877, 2024, doi: 10.25157/ma.v10i2.13684.
- [4] P. D. Kusuma, *Machine Learning Teori, Program, dan Studi Kasus*. 2020. [Online]. Available: <https://openlibrary.telkomuniversity.ac.id/pustaka/160514/machine-learning-teori-program-dan-studi-kasus.html>
- [5] A. Novita, M. S. Iin Ernawati., and M. K. Nurul Chamidah, "Klasterisasi Provinsi Di Indonesia Berdasarkan Produktivitas Komoditas Pangan Menggunakan Algoritma K-Means," vol. 3, no. 2, pp. 462–471, 2022, [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/2194>
- [6] M. Putri, C. Dewi, E. P. Siam, G. A. Wijayanti, N. Aulia, and R. Nooraeni, "Comparison of DBSCAN and K-Means Clustering for Grouping the Village Status in Central Java 2020," *J. Mat. Stat. dan Komputasi*, vol. 17, no. 3, pp. 394–404, 2021, doi: 10.20956/j.v17i3.11704.
- [7] N. Nafiiyah, D. T. A. Erlianto, and R. Wardhani, "Perbandingan metode clustering," *J. Teknol. dan Sist. Inf. Univrab*, vol. 11, no. 1, pp. 469–475, 2026, doi: 10.36341/rabit.v11i1.6980.
- [8] S. N. Sofyan and Z. Sitorus, "Implementasi Data Mining Untuk Clustering Produktivitas Bawang Merah Menggunakan Metode K-Means," *J. Multimed. Dan Teknol. Inf.*, vol. 7, no. 2, pp. 109–121, 2025, doi: 10.54209/jatilima.v7i02.1442.
- [9] M. N. Sutoyo and I. Slamet, *Buku Ajar Data Mining*. 2024. [Online]. Available: https://www.researchgate.net/publication/377415198_BUKU_AJAR_DATA_MINING
- [10] S. Mutiah, Y. Hasnataeni, A. Fitrianto, and L. M. R. D. Jumansyah, "Perbandingan Metode Klastering K-Means dan DBSCAN dalam Identifikasi Kelompok Rumah Tangga Berdasarkan Fasilitas Sosial Ekonomi di Jawa Barat," *Teorema Teor. dan Ris. Mat.*, vol. 9, no. 2, pp. 247–260, 2024, doi: 10.25157/teorema.v9i2.16290.
- [11] D. Chicco, A. Campagner, A. Spagnolo, D. Ciucci, and G. Jurman, "The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters," 2025, doi: 10.7717/peerj-cs.3309.
- [12] N. T. Astuti, Sunardi, and A. Fadlil, "Optimalisasi Pemetaan Klaster Kependudukan Provinsi Di Indonesia Menggunakan Elbow Method Dan Davies- Bouldin Index," *JIP (Jurnal Inform. Polinema)*, vol. 12, no. 2, pp. 295–304, 2026, doi: 10.33795/jip.v12i2.9162.
- [13] B. A. H. Driyandita, I. P. E. N. Kencana, and I. G. N. L. Wijayakusuma, "Analisis Pemilihan Parameter pada Algoritma DBSCAN untuk Pengelompokan Titik Api di Indonesia," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 3, pp. 793–803, 2025, doi: 10.52436/1.jpti.703.
- [14] A. A. BUSHRA and G. YI, "Comparative Analysis Review of Pioneering DBSCAN and Successive Density-Based Clustering Algorithms," *IEEE Access*, vol. 9, pp. 87918–87935, 2021, doi: 10.1109/ACCESS.2021.3089036.
- [15] A. Jain, K. Rathi, and Y. Ganguly, "A Comparative Analysis of DBSCAN, K-Means and Agglomerative Clustering Algorithms for Geospatial Data," *Atl. Press*, pp. 212–221, 2025, doi: 10.2991/978-94-6463-716-8_18.