

ALGORITMA DBSCAN DAN SHARED NEAREST NEIGHBOR DALAM PENGELOMPOKKAN SPASIAL PRODUKTIVITAS JERUK SIAM DI INDONESIA

Nur Sofi Sely Oktavia

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya
nursofiselyo05@student.ub.ac.id

Atiek Iriany

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya
atiekirianny@ub.ac.id*

Ani Budi Astuti

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Brawijaya
ani_budi@ub.ac.id

Abstrak

Penerapan dua algoritma analisis *density-based cluster*, yaitu DBSCAN dan *Shared Nearest Neighbor* (SNN), dengan integrasi *Principal Component Analysis* (PCA) pada Produktivitas Jeruk Siam di Indonesia tahun 2023. Variabel yang digunakan terdiri dari 7 indikator jeruk siam, yaitu produktivitas (Ton/Pohon), produksi (Ton), pertumbuhan produksi (%), luas panen (Hektar), rata-rata suhu bulanan (°C), rata-rata kelembaban bulanan (%), dan rata-rata curah hujan bulanan (mm). Data yang digunakan merupakan data sekunder tahun 2023 yang diperoleh melalui publikasi BPS dan Kementerian Pertanian RI. Hasil penelitian menunjukkan bahwa SNN memiliki stabilitas pengelompokan yang lebih baik dibandingkan DBSCAN, dan penerapan PCA meningkatkan kinerja DBSCAN dan SNN. Model terbaik diperoleh dari SNN pada data PCA dengan tiga komponen utama (PC3), dengan *Silhouette Coefficient* sebesar 0,872. Algoritma ini menghasilkan 3 cluster, yaitu Cluster 0 mencakup 32 provinsi dengan skala produksi besar dan kondisi agroklimat yang beragam, Cluster 1 terdiri dari 3 provinsi dengan produksi kecil namun pertumbuhan sangat tinggi, sehingga terpisah dari sentra produksi utama, dan Cluster 2 mencakup 3 provinsi yang memiliki karakteristik lokal unik dengan skala produksi rendah hingga nol.

Kata Kunci: DBSCAN, SNN, Produktivitas Jeruk

Abstract

The study applies two density-based clustering algorithms, DBSCAN and Shared Nearest Neighbor (SNN), integrated with Principal Component Analysis (PCA), to analyze the productivity of Siam citrus in Indonesia in 2023. The analysis utilizes seven indicators: productivity (tons/tree), production (tons), production growth (%), harvested area (ha), average monthly temperature (°C), average monthly humidity (%), and average monthly rainfall (mm). The dataset consists of 2023 secondary data obtained from the Indonesian Central Bureau of Statistics (BPS) and the Ministry of Agriculture. The results indicate that SNN provides more stable clustering performance than DBSCAN, and the application of PCA further enhances both methods. The best model is achieved using SNN on PCA-reduced data with three principal components (PC3), yielding a *Silhouette Coefficient* of 0.872. This model produces three clusters: Cluster 0 includes 32 provinces characterized by large-scale production and diverse agroclimatic conditions; Cluster 1 consists of 3 provinces with small production but exceptionally high growth, separating them from the major production centers; and Cluster 2 comprises 3 provinces with unique local characteristics and low to near-zero production levels.

Keywords: content, formatting, article.

PENDAHULUAN

Analisis *cluster* merupakan metode statistika yang digunakan untuk mengidentifikasi kelompok objek berdasarkan kemiripan karakteristik tertentu (Han, Kamber, & Pei, 2022). Pada data spasial, analisis cluster berperan untuk mengungkap pola persebaran

dan hubungan kedekatan antarwilayah, terutama ketika data menunjukkan distribusi yang tidak beraturan dan mengandung *noise* (Shah & Kusiak, 2021). Metode *density-based clustering* menawarkan solusi dengan membentuk cluster berdasarkan tingkat kepadatan titik.

DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) merupakan metode clustering berbasis kepadatan yang mengelompokkan titik pada area padat dan memisahkan titik jarang sebagai noise (Misra et al., 2020). DBSCAN mampu mengidentifikasi cluster dengan bentuk bebas dan tidak bergantung pada asumsi jumlah cluster awal. Performa DBSCAN cukup sensitif terhadap pemilihan parameter ϵ dan MinPts, sehingga parameter yang kurang tepat dapat menghasilkan struktur cluster yang tidak stabil atau tidak akurat (Schubert & Zimek, 2019).

Shared Nearest Neighbor (SNN) clustering sebagai alternatif DBSCAN menawarkan pendekatan yang lebih robust dengan menilai kedekatan berdasarkan jumlah tetangga terdekat yang sama, sehingga lebih stabil pada data berdensitas tidak seragam (Campello dkk., 2021). Metode ini lebih tahan terhadap variasi kepadatan dan lebih sesuai untuk data dengan struktur yang kompleks. Kedua metode ini digunakan untuk memperoleh pola pengelompokan yang lebih akurat pada data spasial.

Pendekatan *clustering* relevan ketika diterapkan pada bidang agrokomples yang memiliki variasi kondisi agroklimat, topografi, serta praktik budidaya antarwilayah. Variasi tersebut menghasilkan perbedaan produktivitas komoditas pertanian yang cukup signifikan (Wahyuni & Yuliana, 2022). Analisis spasial diperlukan untuk untuk memetakan pola produktivitas secara ilmiah dan memahami keterkaitannya dengan faktor geografis (Putra & Wulandari, 2021). Melalui *clustering*, wilayah pertanian dapat dikelompokkan berdasarkan kesamaan karakteristik produksi dan kondisi lingkungan sehingga pola tersebut lebih mudah diinterpretasikan.

Jeruk siam merupakan salah satu komoditas hortikultura unggulan di Indonesia dengan kontribusi produksi terbesar dibandingkan jenis lain. Berdasarkan data BPS (2025), jeruk siam menyumbang lebih dari 60% produksi jeruk nasional dengan volume mencapai lebih dari 1 juta ton per tahun. Konsumsi domestik yang tinggi serta persebaran produksi yang luas menunjukkan komoditas ini penting bagi sektor pertanian. Variasi produktivitas antarwilayah mencerminkan perbedaan kondisi agroklimat dan teknik budidaya, sehingga analisis spasial diperlukan untuk mengidentifikasi potensi dan tantangan produksi di

setiap wilayah (FAO, 2021). Penelitian ini bertujuan mengelompokkan wilayah produksi jeruk siam menggunakan metode DBSCAN dan SNN untuk memperoleh pola spasial yang representatif sebagai dasar perencanaan dan strategi peningkatan produktivitas pertanian secara berkelanjutan (Purwanto, Sari, & Nugraha, 2021).

KAJIAN TEORI

ANALISIS CLUSTER

Analisis *cluster* merupakan teknik statistika multivariat yang digunakan untuk mengelompokkan objek berdasarkan tingkat kemiripan karakteristiknya. Hasil pengelompokan yang baik ditandai oleh tingginya homogenitas objek di dalam cluster dan heterogenitas yang jelas antar-cluster, sehingga struktur kelompok dapat diinterpretasikan secara konsisten (Xu, 2020). Analisis *cluster* juga dapat digunakan untuk mendeteksi keberadaan outlier. Secara umum, *clustering* dapat dikelompokkan ke dalam beberapa pendekatan utama yaitu metode partisi, hierarki, *density-based*, dan *grid-based* (Han dkk., 2022).

DENSITY-BASED CLUSTER

Menurut Han dkk. (2022), metode partisi dan hierarki kurang efektif ketika pola data tidak berbentuk bulat atau teratur. Sebaliknya, metode *density-based clustering* mampu menangani pola kompleks dengan mendefinisikan *cluster* sebagai area berkepadatan tinggi yang dipisahkan oleh wilayah berkepadatan rendah, sementara titik yang jarang dianggap sebagai *noise*. Pendekatan ini bersifat non-parametrik karena tidak memerlukan asumsi distribusi maupun penentuan jumlah cluster di awal (Zhao & Qin, 2021).

DENSITY-BASED SPATIAL CLUSTERING OF APPLICATIONS WITH NOISE

DBSCAN merupakan algoritma *density-based* yang membentuk cluster berdasarkan kepadatan titik dalam suatu radius (ϵ) dan jumlah minimum titik tetangga (MinPts) (Ester et al., 1996). Suatu titik p memiliki himpunan tetangga dalam radius ϵ yang disebut *epsilon-neighborhood*, dirumuskan sebagai berikut:

$$N_{Eps}(p) = \{q \in D | dist(p, q) \leq \epsilon\} \quad (1)$$

dengan $dist(p, q)$ biasanya berupa jarak *Euclidean*:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ki} - x_{kj})^2} \quad (2)$$

Berdasarkan kepadatan, titik diklasifikasikan menjadi *core point* (memiliki $\geq \text{MinPts}$ dalam N_ϵ), *border point* (tetangga *core* namun jumlah titik $< \text{MinPts}$), dan *noise* (tidak termasuk keduanya). Struktur cluster dibangun melalui tiga hubungan utama: *direct density-reachable* (titik p berada dalam $N_\epsilon(q)$ dan q memenuhi syarat *core*), *density-reachable* (rantai *direct density-reachable* secara transitif), dan *density-connected* (dua titik terhubung melalui titik ketiga). DBSCAN membentuk cluster jika seluruh titik dalam cluster saling *density-connected*, sedangkan titik yang tidak memenuhi aturan tersebut digolongkan sebagai *noise*. Dengan pendekatan ini, DBSCAN efektif mengelompokkan pola bentuk arbitrer, tidak memerlukan jumlah *cluster* di awal, serta robust terhadap *outlier*.

SHARED NEAREST-NEIGHBOR CLUSTERING

Shared Nearest Neighbor (SNN) merupakan pendekatan kesamaan tidak langsung yang dirancang untuk mengatasi kelemahan jarak *Euclidean* pada data berdimensi tinggi, di mana jarak sering kehilangan makna. Prinsip dasarnya adalah bahwa dua titik dianggap mirip apabila mereka berbagi tetangga terdekat yang sama, meskipun jarak langsung antar titik tidak menunjukkan kedekatan tersebut (Schubert & Zimek, 2019). Kesamaan SNN dihitung berdasarkan jumlah tetangga yang sama dalam daftar *k-nearest neighbors* (k-NN) untuk tiap titik. Secara matematis konsep ini dirumuskan sebagai berikut (Schubert & Zimek, 2019):

$$\text{SNN}_{\text{sim}}(p, q) = |\text{NN}(p) \cap \text{NN}(q)| \quad (3)$$

dengan $\text{NN}(p)$ dan $\text{NN}(q)$ adalah himpunan tetangga terdekat dari titik p dan q . Pendekatan ini mampu menangani variasi kepadatan data, karena penilaian kesamaan tidak bergantung pada jarak absolut, melainkan pada struktur tetangga lokal. SNN lebih stabil terhadap perubahan densitas maupun dimensi ruang, dan memberikan penskalaan otomatis pada area berpenduduk jarang maupun padat. SNN *density-based clustering*, kesamaan digunakan untuk membentuk *SNN graph* yang kemudian diproses menggunakan prinsip DBSCAN. Algoritma SNN meliputi penghitungan kesamaan antar objek, pembentukan daftar *k-nearest neighbors*, konstruksi *SNN graph*, perhitungan kepadatan lokal, identifikasi titik representatif,

pembentukan *cluster*, serta penandaan *noise* (Ertoz dkk., 2003).

SILHOUETTE COEFFICIENT

Silhouette coefficient menilai kualitas hasil *clustering* melalui perbandingan *cohesion* (kedekatan objek dengan *cluster*-nya) dan *separation* (kedekatan objek dengan *cluster* lain). Nilainya dihitung dengan rumus sebagai berikut (Tan dkk., 2012):

$$SC_i = \frac{b_i - a_i}{\max\{b_i, a_i\}} \quad (4)$$

dengan a_i sebagai jarak rata-rata objek ke- i terhadap anggota *cluster*-nya, dan b_i sebagai jarak rata-rata objek ke- i terhadap *cluster* terdekat lainnya. Nilai SC berada pada rentang -1 hingga 1, di mana nilai positif menunjukkan pengelompokan yang baik. Rata-rata SC seluruh objek digunakan sebagai indikator kualitas *cluster* secara keseluruhan. Kriteria nilai *silhouette coefficient* adalah sebagai berikut (Kaufman dan Rousseeuw, 1990) : $1 \geq SC > 0,7$ struktur kuat, $0,7 \geq SC > 0,5$ struktur baik, $0,5 \geq SC > 0,25$ struktur lemah, dan $SC \leq 0,25$ struktur buruk.

PRINCIPAL COMPONENT ANALYSIS

PCA adalah metode reduksi dimensi yang membentuk variabel baru (*principal component*) berupa kombinasi linier dari variabel asli, di mana komponen-komponen tersebut saling tidak berkorelasi (Sharma, 1996). Misalkan vektor data $X' = [X_1, X_2, \dots, X_p]$ dengan matriks kovarian Σ , dan nilai *eigen* $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$. Persamaan kombinasi linier pada PCA sebagai berikut (Johnson dan Wichern, 2007).

$$Y_i = a'_i X = a'_{i1} X_1 + a'_{i2} X_2 + \dots + a'_{ip} X_p \quad (5)$$

Ragam dan peragam dari variabel Y sebagai berikut.

$$\text{var}(Y_i) = a'_i \Sigma a_i \quad i = 1, 2, \dots, p \quad (6)$$

$$\text{cov}(Y_i, Y_j) = a'_i \Sigma a_j \quad i, j = 1, 2, \dots, p; i \neq j \quad (7)$$

Penentuan komponen dilakukan dengan memaksimalkan ragam $a' \Sigma a$ di bawah kendala $a' a = 1$. Melalui metode *Lagrange*:

$$(\Sigma - \lambda I) a = 0 \quad (8)$$

sehingga λ adalah *eigenvalue* dan a adalah *eigenvector* dari Σ . Ragam PC ke- i adalah:

$$\text{var}(Y_i) = \lambda_i \quad (9)$$

Total keragaman data adalah jumlah seluruh *eigenvalue* dan proporsi keragaman yang dijelaskan sampai komponen ke- k dihitung dengan rumus berikut:

$$\text{tr}(\Sigma) = \lambda_1 + \lambda_2 + \dots + \lambda_p \quad (10)$$

$$\text{Proporsi} = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \quad k = 1, 2, \dots, p \quad (11)$$

Banyak *principal component* dengan keragaman total bernilai 80 persen hingga 90 persen dinilai baik untuk menjelaskan keragaman data (Johnson dan Wichern, 2007). Persamaan untuk menghitung *component score* dari setiap data adalah sebagai berikut:

$$SK_{ij} = \alpha_j X_i \quad (12)$$

KAISER MAYER OLKIN

Uji *Kaiser-Meyer-Olkin* (KMO) digunakan untuk menilai kecukupan sampel sebelum menerapkan *Principal Component Analysis*. Nilai KMO dihitung berdasarkan perbandingan antara jumlah kuadrat korelasi Pearson dan gabungan kuadrat korelasi Pearson serta korelasi parsial antar variabel, dengan rumus sebagai berikut (Astuti, dkk, 2023):

$$KMO = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p \alpha_{ij}^2}, i \neq j \quad (13)$$

Kriteria kecukupan sampel menurut Sharma adalah nilai KMO $\geq 0,90$ dianggap baik sekali, nilai KMO $\geq 0,80$ baik, nilai KMO $\geq 0,70$ sedang, nilai KMO $\geq 0,60$ cukup, nilai KMO $\geq 0,50$ kurang, sedangkan nilai $< 0,50$ menunjukkan bahwa data tidak cocok untuk PCA.

BARLETT'S TEST OF SPHERICITY

Bartlett's Test of Sphericity digunakan untuk menilai apakah variabel-variabel dalam data saling berkorelasi sehingga layak dianalisis menggunakan PCA. Pengujian dilakukan dengan hipotesis sebagai berikut:

$H_0 : \mathbf{R} = \mathbf{I}$ (Tidak terdapat korelasi antar variabel)

$H_1 : \mathbf{R} \neq \mathbf{I}$ (Terdapat korelasi antar variabel)

Statistik Uji:

$$\chi_{hitung}^2 = \left((n-1) - \frac{2p+5}{6} \right) \ln(\det(\mathbf{R})) \quad (14)$$

H_0 ditolak jika $\chi_{hitung}^2 > \chi_{\frac{p(p-1)}{2}}^2$ atau $p - \text{value} < \alpha$, yang artinya terdapat korelasi antar variabel cukup kuat sehingga PCA dapat dilakukan.

JERUK SIAM

Jeruk Siam merupakan varietas jeruk lokal yang populer dan banyak dibudidayakan di Indonesia. Jeruk Siam mendominasi produksi jeruk nasional dan banyak ditemukan di pasar tradisional maupun pasar modern (BRMP, 2023). Varietas-varietas Jeruk Siam tersebar di berbagai provinsi, seperti Jawa

Timur, Sumatera Utara, Kalimantan Barat, Sumatera Barat, dan Kalimantan Selatan (BRMP, 2023).

Data dari BPS menunjukkan bahwa pada tahun 2017 produksi Jeruk Siam/keprok nasional mencapai sekitar 2.165.189 ton. Provinsi penyumbang produksi terbesar antara lain Jawa Timur (~41,5 %), Sumatera Utara (~20,1 %), Bali (~7,8 %), Kalimantan Barat (~5,8 %) dan Kalimantan Selatan (~5,7 %) (BRMP, 2020). Produksi nasional jeruk terus menunjukkan bahwa jeruk lokal tetap diminati konsumen Indonesia (Direktorat Jenderal Hortikultura, 2025).

Berbagai varietas Jeruk Siam seperti Siam Pontianak, Siam Madu, Siam Banjar telah dibudidayakan di lahan dataran rendah maupun dataran tinggi, menyesuaikan kondisi agroklimat setempat (BRMP, 2023). Varietas ini dikenal memiliki produktivitas per pohon yang baik, citarasa manis, kulit buah yang relatif tipis dan mudah dikupas, sehingga menarik bagi konsumen dan petani. Hal ini menjadikan Jeruk Siam sebagai komoditas hortikultura strategis dengan potensi luas untuk dikembangkan.

METODE

Data yang digunakan dalam penelitian merupakan data sekunder yang diperoleh melalui publikasi Badan Pusat Statistik (BPS) dan Kementerian Pertanian Republik Indonesia pada tahun 2023. Data terdiri dari tujuh variabel, yaitu produktivitas jeruk siam (X_1 , Ton/Hektar), produksi jeruk siam (X_2 , Ton), pertumbuhan produksi jeruk siam (X_3 , %), luas panen jeruk siam (X_4 , Hektar), rata-rata suhu bulanan (X_5 , °C), rata-rata kelembaban bulanan (X_6 , %), dan rata-rata curah hujan bulanan (X_7 , mm). Selain itu, data spasial berbentuk *shapefile* wilayah administrasi provinsi diperoleh dari Badan Informasi Geospasial (BIG) untuk keperluan pemetaan dan analisis spasial. Adapun langkah-langkah analisis yang dilakukan pada penelitian ini sebagai berikut:

1. Memilih variabel penelitian yang akan digunakan.
2. Memeriksa asumsi dan melakukan reduksi dimensi menggunakan PCA.
3. Melakukan *clustering* menggunakan algoritma DBSCAN dan SNN dengan parameter tertentu.
4. Menghitung indeks *Silhouette Coefficient* (tanpa *noise*) untuk mengevaluasi hasil *clustering*.

- Menentukan metode *clustering* terbaik berdasarkan nilai *Silhouette Coefficient*.
- Mengulangi langkah 2-5 pada data tanpa reduksi dimensi (tanpa PCA).
- Mendeskripsikan dan menganalisis hasil *clustering* yang diperoleh.

HASIL DAN PEMBAHASAN

STATISTIKA DESKRIPTIF

Hasil statistika deskriptif masing-masing variabel penelitian dapat dilihat pada Tabel 1.

Tabel 1. Statistika Deskriptif

Variabel	Mean	SD
Produktivitas Jeruk Siam (X_1 , Ton/Hektar)	0,1003	0,1078
Produksi Jeruk Siam (X_2 , Ton)	74.502,63	234.733,8
Pertumbuhan Produksi Jeruk Siam (X_3 , %)	4,61	51,04
Luas Panen (X_4 , Hektar)	577.282	1.233.576
Rata-rata Suhu Bulanan (X_5 , °C)	25,93	6,23
Rata-rata Kelembaban Bulanan (X_6 , %)	79,70	13,58
Rata-rata Curah Hujan Bulanan (X_7 , mm)	183,16	65,66

Tabel di atas menunjukkan bahwa variabel produksi dan luas panen memiliki rata-rata tinggi dengan variasi yang besar antar wilayah, menunjukkan ketimpangan kapasitas produksi. Produktivitas, suhu, dan kelembaban relatif stabil, mencerminkan kondisi lahan yang konsisten, sedangkan curah hujan bervariasi sedang sesuai perbedaan iklim regional.

UJI KECUKUPAN SAMPEL (KMO)

Uji KMO digunakan untuk menilai kecukupan data sebelum dilakukan PCA. Hasil perhitungan menunjukkan Overall MSA = 0,55, dengan MSA tiap variabel:

Tabel 2. Nilai MSA Variabel (7 Variabel)

Variabel	MSA
X_1	0,52
X_2	0,49
X_3	0,56
X_4	0,48
X_5	0,61

X_6	0,60
X_7	0,76
Rata-rata	0,55

Berdasarkan nilai rata-rata MSA, data cukup layak untuk PCA, meskipun terdapat 2 variabel yang memiliki nilai KMO <0,5, menandakan korelasi rendah dengan variabel lain.

PEMERIKSAAN KORELASI

Bartlett's Test of Sphericity digunakan untuk memeriksa apakah terdapat korelasi antar variabel dengan hipotesis uji sebagai berikut:

$H_0 : \mathbf{R} = \mathbf{I}$ (Tidak terdapat korelasi antar variabel)

$H_1 : \mathbf{R} \neq \mathbf{I}$ (Terdapat korelasi antar variabel)

Hasil *Bartlett's Test of Sphericity* dapat dilihat pada Tabel 3.

Tabel 3. Pemeriksaan Korelasi

Chi-Square	df	p-value
107,192	21	$<2,22 \times 10^{-16}$

Berdasarkan Tabel 3, diperoleh nilai *chi-square* sebesar 107,192 dengan p-value $< 2,22 \times 10^{-16} < 0,05$, maka H_0 ditolak. Sehingga dapat disimpulkan bahwa variabel saling berkorelasi dan data layak untuk PCA.

PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component Analysis (PCA) memberikan *component score* untuk setiap kecamatan yang akan digunakan sebagai input dalam proses pengelompokan. Hasil PCA menunjukkan kontribusi tiap komponen terhadap total varians sebagai berikut:

Tabel 4. Proporsi Varians dan Proporsi Kumulatif PCA

	PC1	PC2	PC3	PC4	PC5	PC6	PC7
Standard Deviation	1,52	1,36	1,16	0,78	0,73	0,53	0,28
Proporsi Varians	0,33	0,26	0,19	0,09	0,08	0,04	0,01
Proporsi Kumulatif	0,33	0,59	0,79	0,87	0,95	0,99	1,00

Berdasarkan Tabel 4, menunjukkan bahwa menggunakan 3 komponen maka total varians yang dapat dijelaskan sebesar 78,6%, sedangkan 5 komponen menjelaskan 94,8% varians. *Component score* dari PC3 dan PC5 ini selanjutnya akan digunakan untuk melakukan pengelompokan provinsi menggunakan DBSCAN dan SNN Clustering.

DBSCAN PADA DATA LENGKAP

Pasangan parameter MinPts dan Epsilon untuk metode DBSCAN dalam pengelompokan pada data lengkap disajikan dalam Tabel 5.

Tabel 5. Pasangan Parameter dan SC dari DBSCAN untuk Data Lengkap

MinPts	Eps	SC
3	0,01	NA
3	0,02	NA
3	0,03	NA
⋮	⋮	⋮
3	0,66	0,6371
⋮	⋮	⋮
30	1,00	NA

Berdasarkan hasil pemilihan parameter optimal dapat dilihat bahwa kombinasi parameter terbaik pada data lengkap adalah MinPts = 3 dan Epsilon = 0,66, yang menghasilkan nilai *Silhouette Coefficient* (SC) sebesar 0,6371. Kombinasi parameter tersebut menghasilkan pembentukan *cluster* yang cukup baik dengan struktur pemisahan yang jelas antar *cluster*. Jumlah *cluster* yang terbentuk berdasarkan parameter optimal ini adalah sebanyak 3 *cluster*, dengan jumlah noise sebanyak 25 titik.

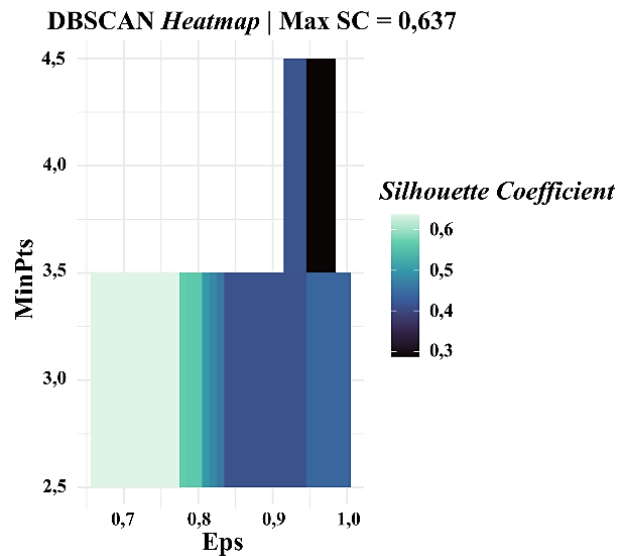
Rincian anggota *cluster* yang terbentuk oleh metode DBSCAN pada data lengkap ditampilkan di Tabel 6.

Tabel 6. Rincian Hasil *Cluster* dari DBSCAN untuk Data Lengkap

Cluster	Jumlah	Anggota (Provinsi)
0 (<i>noise</i>)	25	Aceh; Sumatera Utara; Sumatera Barat; Jambi; Bengkulu; Lampung; Kep. Riau; Jawa Tengah; DI Yogyakarta; Jawa Timur; Banten; Bali; Nusa Tenggara Barat; Nusa Tenggara Timur; Kalimantan Barat; Kalimantan Selatan; Sulawesi Selatan; Sulawesi Tenggara; Gorontalo; Sulawesi Barat; Maluku; Papua Barat; Papua; Papua Tengah; Papua Pegunungan
1	10	Riau; Sumatera Selatan; Kep. Bangka Belitung; Jawa Barat; Kalimantan Tengah; Kalimantan Timur; Kalimantan Utara; Sulawesi Utara; Papua Barat Daya; Papua Selatan
2	3	DKI Jakarta; Sulawesi Tengah; Maluku Utara

Pada data lengkap, Cluster 1 (10 provinsi) didominasi oleh wilayah dengan karakteristik produktivitas dan kondisi agroklimat yang relatif lebih homogen. Selanjutnya, Cluster 2 (3 provinsi) memiliki pola variabel yang berbeda secara signifikan dibanding cluster lainnya. Sementara itu, sebagian besar provinsi lainnya, yaitu 25 provinsi, berada pada kategori *noise*, menunjukkan bahwa persebaran data penuh memiliki tingkat densitas yang sangat bervariasi sehingga tidak mudah membentuk kelompok yang padat.

Heatmap nilai *Silhouette Coefficient* dari proses pengelompokan DBSCAN pada data lengkap Jeruk Siam di Indonesia tahun 2023 ditampilkan pada Gambar 1.



Gambar 1. *Heatmap* DBSCAN Data Lengkap

Berdasarkan Gambar 1, dapat dilihat bahwa nilai *Silhouette Coefficient* (SC) tertinggi sebesar 0,637 diperoleh pada kombinasi parameter MinPts = 3 dan nilai Eps berada pada 0,66 hingga 1. Pola warna yang semakin gelap pada area sekitar Eps > 0,8 menunjukkan penurunan nilai SC yang disebabkan oleh pengelompokan yang tidak sesuai.

DBSCAN PADA DATA PCA 80% (PC3)

Pasangan parameter MinPts dan Epsilon untuk metode DBSCAN dalam pengelompokan pada data lengkap disajikan dalam Tabel 7.

Tabel 7. Pasangan Parameter dan SC dari DBSCAN untuk Data Lengkap

<i>MinPts</i>	<i>Eps</i>	SC
3	0,01	NA
3	0,02	NA
3	0,03	NA
⋮	⋮	⋮
3	0,23	0,8547
⋮	⋮	⋮
30	1,00	NA

Berdasarkan hasil pada data PCA 3 komponen terlihat bahwa kombinasi parameter optimal adalah $\text{MinPts} = 3$ dan $\text{Epsilon} = 0,23$, dengan nilai $\text{SC} = 0,8547$. Nilai SC yang sangat tinggi ini menunjukkan bahwa struktur *cluster* yang terbentuk sangat baik, dengan pemisahan *cluster* yang kuat dan konsistensi internal yang tinggi. Berdasarkan parameter optimal tersebut, algoritma DBSCAN menghasilkan sebanyak 3 *cluster*, dengan *noise* sebanyak 29 titik.

Rincian anggota *cluster* yang terbentuk oleh metode DBSCAN pada data PCA 80% (PC3) ditampilkan di Tabel 8.

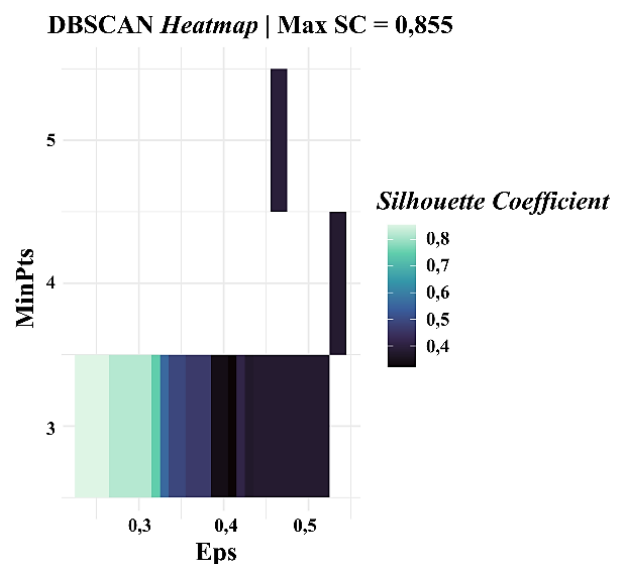
Tabel 8. Rincian Hasil Cluster dari DBSCAN untuk Data dengan PCA 80% (PC3)

Cluster	Jumlah	Anggota (Provinsi)
0 (<i>noise</i>)	29	Aceh; Sumatera Utara; Sumatera Barat; Riau; Jambi; Sumatera Selatan; Bengkulu; Lampung; Kep. Riau; Jawa Barat; Jawa Tengah; DI Yogyakarta; Jawa Timur; Banten; Bali; Nusa Tenggara Barat; Nusa Tenggara Timur; Kalimantan Barat; Kalimantan Selatan; Kalimantan Timur; Sulawesi Utara; Sulawesi Tenggara; Gorontalo; Sulawesi Barat; Maluku; Papua Barat; Papua; Papua Tengah; Papua Pegunungan
1	6	Kep. Bangka Belitung; Kalimantan Tengah; Kalimantan Utara; Sulawesi Selatan; Papua Barat Daya; Papua Selatan
2	3	DKI Jakarta; Sulawesi Tengah; Maluku Utara

Pada data dengan PCA 80% (PC3), Cluster 1 (6 provinsi), menunjukkan bahwa setelah reduksi dimensi, banyak provinsi memiliki pola kesamaan yang kuat dalam hal produktivitas dan faktor agroklimat. Cluster 2 (3 provinsi) kembali menunjukkan pola yang jauh berbeda dibanding

provinsi lainnya, konsisten dengan hasil pada data lengkap. Sedangkan, Cluster 0 menunjukkan setelah reduksi dimensi, banyak provinsi yang memiliki pola berbeda.

Nilai ϵ yang kecil (0,23) mencerminkan bahwa data PC3 lebih kompak sehingga cluster dapat dibentuk dengan radius yang kecil. Proses identifikasi *core point* dilakukan pada ruang PCA sehingga perbedaan antar provinsi lebih terstruktur dan terbukti menghasilkan cluster dengan kualitas sangat baik. *Heatmap* nilai *Silhouette Coefficient* dari proses pengelompokan DBSCAN pada data PCA 80% (PC3) Jeruk Siam di Indonesia tahun 2023 ditampilkan pada Gambar 2.



Gambar 2. Heatmap DBSCAN Data PCA 80% (PC3)

Berdasarkan Gambar 2, dapat dilihat bahwa nilai *Silhouette Coefficient* nilai SC paling besar adalah 0,854, yang muncul pada $\text{MinPts} = 3$ dan rentang Eps sekitar 0,23. Ketika nilai Eps bergerak mendekati 0,4 atau lebih, nilai SC mulai menurun, yang berarti *cluster* menjadi kurang terpisah karena peningkatan radius Eps yang menyebabkan titik-titik dari lokasi berbeda bergabung, menurunkan kualitas pengelompokan. Hasil ini menunjukkan bahwa reduksi dimensi dengan PCA 80% meningkatkan kualitas cluster, karena pola data menjadi lebih padat dan mudah dipisahkan.

DBSCAN PADA DATA PCA 90% (PC5)

Pasangan parameter MinPts dan Epsilon untuk metode DBSCAN dalam pengelompokan pada data PCA 90% (PC5) disajikan dalam Tabel 5.

Tabel 9. Pasangan Parameter dan SC dari DBSCAN untuk Data PCA 90% (PC5)

<i>MinPts</i>	<i>Eps</i>	SC
3	0,01	NA
3	0,02	NA
3	0,03	NA
⋮	⋮	⋮
3	0,39	0,801
⋮	⋮	⋮
30	1,00	NA

Berdasarkan hasil pemilihan parameter optimal dapat dilihat bahwa kombinasi parameter terbaik adalah $MinPts = 3$ dan $Epsilon = 0,39$, yang menghasilkan nilai *Silhouette Coefficient* (SC) sebesar 0,801. Nilai SC ini termasuk tinggi, sehingga *cluster* yang terbentuk dapat dikatakan sangat baik dengan pemisahan densitas yang cukup jelas. Algoritma DBSCAN menghasilkan sebanyak 3 *cluster*, dengan *noise* sebanyak 32 titik.

Rincian anggota *cluster* yang terbentuk oleh metode DBSCAN pada data PCA 90% (PC5) ditampilkan di Tabel 10.

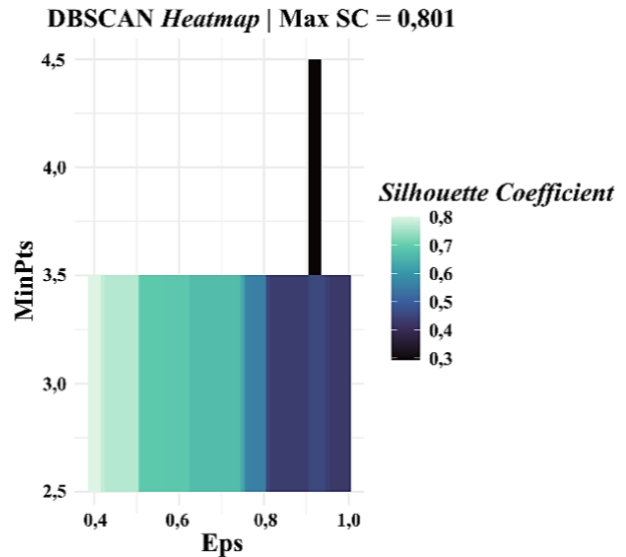
Tabel 10. Rincian Hasil Cluster dari DBSCAN untuk Data PCA 90% (PC5)

Cluster	Jumlah	Anggota (Provinsi)
0 (<i>noise</i>)	32	Aceh; Sumatera Utara; Sumatera Barat; Riau; Jambi; Sumatera Selatan; Bengkulu; Lampung; Kep. Riau; Jawa Barat; Jawa Tengah; DI Yogyakarta; Jawa Timur; Banten; Bali; Nusa Tenggara Barat; Nusa Tenggara Timur; Kalimantan Barat; Kalimantan Selatan; Kalimantan Timur; Kalimantan Utara; Sulawesi Utara; Sulawesi Selatan; Sulawesi Tenggara; Gorontalo; Sulawesi Barat; Maluku; Papua Barat; Papua; Papua Selatan; Papua Tengah; Papua Pegunungan
1	3	Kep. Bangka Belitung; Kalimantan Tengah; Papua Barat Daya
2	3	DKI Jakarta; Sulawesi Tengah; Maluku Utara

Pada data lengkap, Cluster 1 (3 provinsi) memiliki pola berbeda dari cluster mayoritas. Selanjutnya, Cluster 2 (3 provinsi) menunjukkan konsistensi hasil bahwa ketiga provinsi ini memiliki karakteristik yang unik dibanding provinsi lain dalam variabel penelitian. Sementara itu, 32 provinsi, berada pada kategori *noise*, menunjukkan bahwa persebaran data

penyuh memiliki tingkat densitas yang sangat bervariasi sehingga tidak mudah membentuk kelompok yang padat.

Heatmap nilai *Silhouette Coefficient* dari proses pengelompokan DBSCAN pada data PCA 90% (PC5) Jeruk Siam di Indonesia tahun 2023 ditampilkan pada Gambar 3.



Gambar 3. *Heatmap* DBSCAN Data PCA 90% (PC5)

Berdasarkan Gambar 3, dapat dilihat bahwa nilai *Silhouette Coefficient* (SC) maksimum adalah 0,801, yang dicapai pada $MinPts = 3$ dan rentang Eps sekitar 0,39. Area terang pada heatmap mengindikasikan bahwa pemilihan Eps kecil hingga sedang (0,3-0,4) memberikan kualitas *cluster* yang optimal, sedangkan nilai Eps yang lebih besar ($\geq 0,5$) menandakan menurunnya SC yang disebabkan oleh terbentuknya *cluster* yang terlalu besar atau bercampur, sehingga memperburuk pemisahan antar *cluster*. Hal ini menunjukkan bahwa pada data PCA 90%, struktur data tetap memerlukan radius kecil agar relasi titik yang benar-benar berdekatan dapat dipertahankan.

SNN PADA DATA LENGKAP

Pasangan parameter k , $MinPts$, dan $Epsilon$ untuk metode SNN *clustering* pada data lengkap ditampilkan dalam Tabel 11.

Tabel 11. Pasangan Parameter dan SC dari DBSCAN untuk Data Lengkap

k	<i>MinPts</i>	<i>Eps</i>	SC
3	3	3	NA
3	4	3	NA

3	5	3	NA
⋮	⋮	⋮	⋮
10	3	10	0,7403
	⋮	⋮	⋮
30	30	30	NA

Berdasarkan hasil pengujian parameter SNN pada data lengkap, kombinasi optimal diperoleh pada $k = 10$, $MinPts = 3$, dan $Epsilon = 10$ dengan $SC = 0,7403$. Parameter tersebut menghasilkan struktur cluster terbaik setelah melalui proses pembentukan k -nearest neighbors dan SNN graph, kemudian penentuan titik inti menggunakan MinPts dan Epsilon.

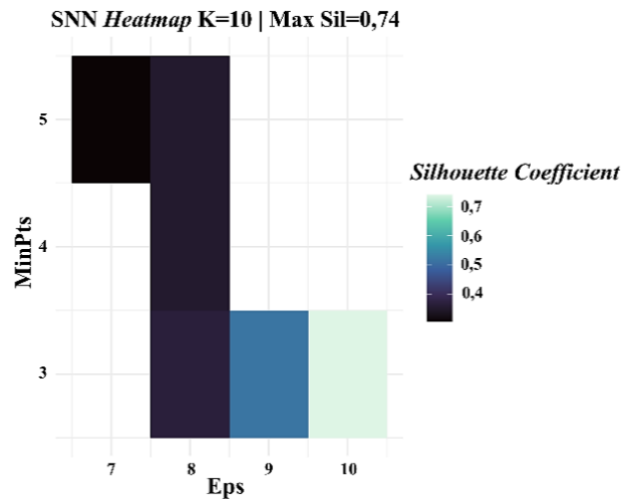
Algoritma SNN menghasilkan sebanyak 3 cluster, dengan jumlah noise sebanyak 31 titik. Rincian anggota cluster yang terbentuk oleh metode SNN pada data lengkap ditampilkan pada Tabel 12.

Tabel 12. Rincian Hasil Cluster dari SNN untuk Data Lengkap

Cluster	Jumlah	Anggota (Provinsi)
0 (noise)	31	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Riau, Jawa Barat, Jawa Tengah, DKI Jakarta, Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Papua Barat, Papua Barat Daya, Papua, Papua Tengah, Papua Pegunungan
1	4	Kep. Bangka Belitung, Kalimantan Tengah, Kalimantan Utara, Papua Selatan
2	3	DKI Jakarta; Sulawesi Tengah; Maluku Utara

Proses pengelompokan menggunakan algoritma SNN pada data lengkap menghasilkan tiga kelompok utama. Cluster 0 merupakan kelompok terbesar yang berisi sebagian besar provinsi di Indonesia. Cluster 1 terdiri dari empat provinsi yang membentuk kelompok tersendiri. Sementara itu, Cluster 2 memiliki kedekatan struktur data yang berbeda dari dua cluster lainnya.

Heatmap nilai Silhouette Coefficient dari proses pengelompokan SNN clustering untuk $k = 10$ pada data lengkap Jeruk Siam di Indonesia tahun 2023 ditampilkan pada Gambar 4.



Gambar 4. Heatmap SNN Clustering pada Data Lengkap

Berdasarkan Gambar 4, pola umum yang terlihat menunjukkan bahwa nilai Eps yang rendah menghasilkan kualitas cluster yang kurang baik, karena hubungan antar titik dalam graph menjadi lemah. Selain itu, MinPts bernilai kecil (1-2) pada berbagai nilai Eps menghasilkan nilai SC yang lebih rendah dibandingkan MinPts yang lebih tinggi. Hal ini menunjukkan bahwa peningkatan MinPts membuat struktur cluster menjadi lebih padat dan stabil, sehingga nilai SC meningkat.

Ketika nilai Eps mendekati nilai 10, SC menunjukkan kecenderungan meningkat. Nilai Eps yang besar memperluas radius hubungan antar titik, sehingga kedekatan antar titik lebih kuat dan memberikan kualitas clustering yang lebih baik. Konfigurasi optimal pada data lengkap dihasilkan ketika hubungan graph cukup padat dan kuat.

SNN PADA DATA PCA 80% (PC3)

Pasangan parameter k , $MinPts$, dan $Epsilon$ untuk metode SNN clustering pada data lengkap ditampilkan dalam Tabel 13.

Tabel 13. Pasangan Parameter dan SC dari DBSCAN untuk Data PCA 80% (PC3)

k	$MinPts$	Eps	SC
3	3	3	NA
3	4	3	NA
3	5	3	NA
⋮	⋮	⋮	⋮
15	3	15	0,8719
	⋮	⋮	⋮
30	30	30	NA

Berdasarkan hasil pengujian parameter pada data PCA tiga komponen, parameter optimal ditunjukkan oleh $k = 15$, $MinPts = 3$, dan $Epsilon = 15$ dengan $SC = 0,8719$. Kombinasi ini memberikan kualitas pengelompokan tertinggi setelah pembentukan daftar ketetanggaan dan SNN *graph*, serta penentuan *core point* berdasarkan nilai $MinPts$ dan $Epsilon$.

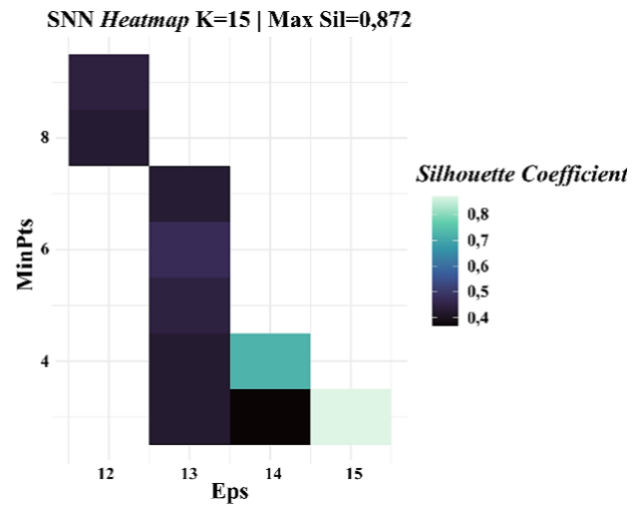
Algoritma SNN menghasilkan sebanyak 3 cluster, dengan jumlah noise sebanyak 0 titik. Rincian anggota cluster yang terbentuk oleh metode SNN pada data PCA 80% (PC3) ditampilkan pada Tabel 14.

Tabel 14. Rincian Hasil Cluster dari SNN untuk Data PCA 80% (PC3)

Cluster	Jumlah	Anggota (Provinsi)
0 (noise)	32	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Sulawesi Utara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Papua Barat, Papua Barat Daya, Papua, Papua Tengah, Papua Pegunungan
1	3	DKI Jakarta, Sulawesi Tengah, Maluku Utara
2	3	Kalimantan Utara, Sulawesi Selatan, Papua Selatan

Proses pengelompokan menggunakan algoritma SNN pada data lengkap menghasilkan tiga kelompok utama. **Cluster 0** tetap menjadi kelompok terbesar dengan komposisi provinsi yang mirip dengan data lengkap. Cluster 1 terdiri dari 3 provinsi yang pada data lengkap merupakan Cluster 2. Sementara itu, Cluster 2 diisi oleh sisanya. Struktur ini menunjukkan bahwa transformasi PCA 80% mempertahankan pola pengelompokan utama namun memberikan sedikit pergeseran pada beberapa provinsi.

Heatmap nilai *Silhouette Coefficient* dari proses pengelompokan SNN *clustering* untuk $k = 15$ pada data PCA 80% Jeruk Siam di Indonesia tahun 2023 ditampilkan pada Gambar 5.



Gambar 5. *Heatmap* SNN *Clustering* pada Data PCA 80% (PC3)

Berdasarkan Gambar 5, pola *heatmap* pada PC3 menunjukkan bahwa nilai $MinPts$ yang tinggi menghasilkan *cluster* yang lebih baik, sementara $MinPts$ kecil sering menghasilkan SC yang rendah, menandakan struktur *cluster* yang kurang padat. Performa terbaik muncul ketika nilai Eps mendekati nilai 15, yang menunjukkan bahwa hubungan antar titik yang lebih luas dan kuat dibutuhkan untuk menghasilkan struktur *cluster* optimal pada ruang PC3. Hal ini juga menggambarkan bahwa reduksi dimensi menjadi 3 komponen justru meningkatkan separabilitas *cluster*, terbukti dari SC tertinggi (0,872) yang lebih baik dibandingkan data lengkap.

SNN PADA DATA PCA 90% (PC5)

Pasangan parameter k , $MinPts$, dan $Epsilon$ untuk metode SNN *clustering* pada data lengkap ditampilkan dalam Tabel 15.

Tabel 15. Pasangan Parameter dan SC dari DBSCAN untuk Data PCA 90% (PC5)

k	$MinPts$	Eps	SC
3	3	3	NA
3	4	3	NA
3	5	3	NA
⋮	⋮	⋮	⋮
12	3	12	0,801
⋮	⋮	⋮	⋮
30	30	30	NA

Berdasarkan hasil pengelompokan menggunakan data PCA 5 komponen menghasilkan parameter optimal berupa $k = 12$, $MinPts = 3$, dan $Epsilon = 12$, dengan *Silhouette Coefficient* (SC) sebesar 0,801. Nilai SC ini menunjukkan bahwa *cluster* yang terbentuk

berada pada kategori baik dengan pemisahan antar *cluster* yang jelas.

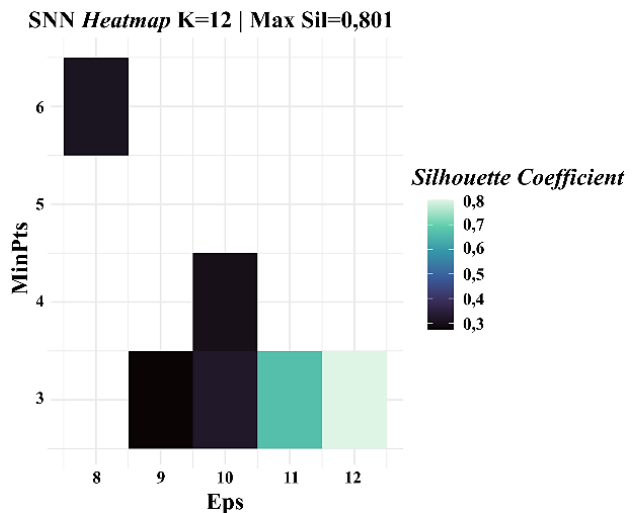
Algoritma SNN menghasilkan sebanyak 3 *cluster*, dengan jumlah *noise* sebanyak 31 titik. Rincian anggota *cluster* yang terbentuk oleh metode SNN pada data lengkap ditampilkan pada Tabel 16.

Tabel 16. Rincian Hasil Cluster dari SNN untuk Data PCA 90% (PC5)

Cluster	Jumlah	Anggota (Provinsi)
0 (<i>noise</i>)	32	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Papua Barat, Papua, Papua Selatan, Papua Tengah, Papua Pegunungan
1	3	Kep. Bangka Belitung, Kalimantan Tengah, Papua Barat Daya
2	3	DKI Jakarta; Sulawesi Tengah; Maluku Utara

Pada data PCA sebesar 90% (PC5), algoritma SNN kembali menghasilkan tiga kelompok. Cluster 0 mengelompokkan sebagian besar provinsi, menunjukkan konsistensi pola dengan hasil sebelumnya. Cluster 1 terdiri dari 3 provinsi yang berbeda dengan hasil PC3. Cluster 2 terdiri dari 3 provinsi yang menandakan kestabilan pengelompokan untuk ketiga provinsi tersebut meskipun data direduksi hingga 5 komponen utama.

Heatmap nilai *Silhouette Coefficient* dari proses pengelompokan SNN *clustering* untuk $k = 12$ pada data PCA 90% Jeruk Siam di Indonesia tahun 2023 ditampilkan pada Gambar 6.



Gambar 4. Heatmap SNN Clustering pada Data PCA 90% (PC5)

Berdasarkan Gambar 4, pola yang muncul pada PC5 mirip dengan data sebelumnya, di mana nilai Eps yang terlalu kecil menghasilkan kualitas *cluster* yang buruk, sementara peningkatan Eps memberikan hubungan *graph* yang lebih kuat. Selain itu, MinPts kecil menghasilkan *cluster* yang tidak stabil, dan nilai SC yang dihasilkan lebih rendah.

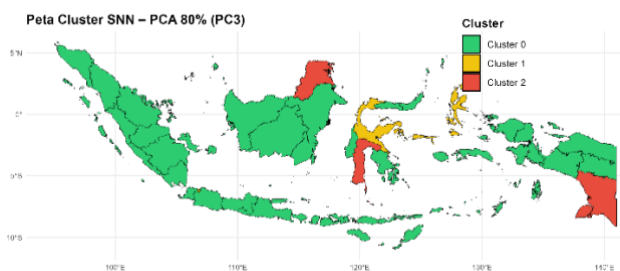
Kombinasi terbaik terjadi ketika nilai Eps mendekati nilai 12, menandakan bahwa PC5 membutuhkan tingkat kepadatan *graph* yang cukup tinggi untuk membentuk *cluster* yang baik. PC5 memuat lebih banyak komponen PCA, nilai SC terbaiknya lebih rendah dibanding PC3, menunjukkan bahwa penambahan komponen justru menambah *noise* yang mengurangi kejernihan pola *cluster*.

PEMBAHASAN

Berdasarkan proses pengelompokan menggunakan DBSCAN dan SNN pada data lengkap serta data hasil transformasi PCA, diperoleh hasil yang cukup berbeda. Algoritma DBSCAN menghasilkan *silhouette coefficient* yang bervariasi, yaitu sebesar 0,637 pada data lengkap, 0,855 pada data PCA 80% (PC3), dan 0,801 pada data PCA 90% (PC5). Algoritma SNN menghasilkan hubungan *cluster* yang lebih stabil dan kuat. *Silhouette Coefficient* SNN mencapai 0,740 pada data lengkap, 0,872 pada data PCA PC3, dan 0,801 pada data PCA PC5. Berdasarkan SC disimpulkan bahwa algoritma SNN lebih unggul dalam menangani data dibandingkan DBSCAN.

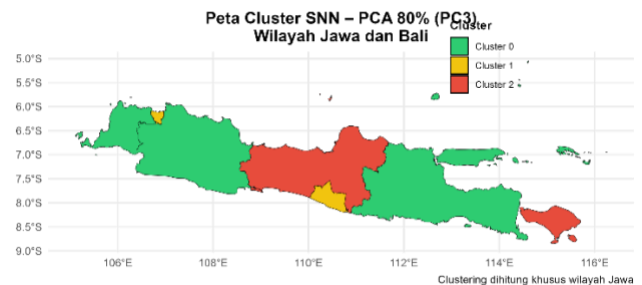
Struktur kelompok yang dihasilkan oleh SNN pada data lengkap menunjukkan kemiripan dengan hasil DBSCAN pada data PCA 80% dan 90%, menegaskan bahwa PCA berperan penting dalam mereduksi dimensi sekaligus meningkatkan kinerja kedua algoritma. PCA dengan 3 komponen utama dan 5 komponen utama (PC5) meningkatkan kualitas cluster pada SNN, tetapi DBSCAN tetap menunjukkan sensitivitas terhadap struktur densitas lokal sehingga cluster optimal tidak selalu terbentuk. Ketidakmampuan DBSCAN membentuk cluster optimal dapat dikaitkan dengan kurangnya variasi yang cukup untuk validasi cluster tunggal menggunakan *Silhouette Coefficient*.

Evaluasi titik *noise* dan cluster menunjukkan bahwa algoritma SNN menghasilkan pengelompokan yang lebih konsisten dibanding DBSCAN. Cluster terbagi menjadi tiga *cluster*. Cluster 0 berisi sebagian besar provinsi penghasil jeruk dengan skala produksi besar dan kondisi agroklimat yang beragam, sehingga membentuk kelompok dominan dalam data. Cluster 1, yang terdiri dari DKI Jakarta, Sulawesi Tengah, dan Maluku Utara, memperlihatkan pola berbeda karena memiliki produksi kecil namun pertumbuhan yang sangat tinggi, sehingga terpisah dari sentra produksi utama. Cluster 2 mencakup Kalimantan Utara, Sulawesi Selatan, dan Papua Selatan, yang dikelompokkan bersama karena memiliki pola agroklimat serupa dan skala produksi yang relatif kecil hingga mendekati nol. Provinsi dengan produksi sangat rendah atau tidak ada produksi tetap teridentifikasi sebagai *noise*, mengindikasikan posisinya sebagai titik outlier dalam keseluruhan struktur data. Hasil pengelompokan ini divisualisasikan pada peta provinsi Indonesia berdasarkan produktivitas jeruk siam ditunjukkan pada Gambar 7.



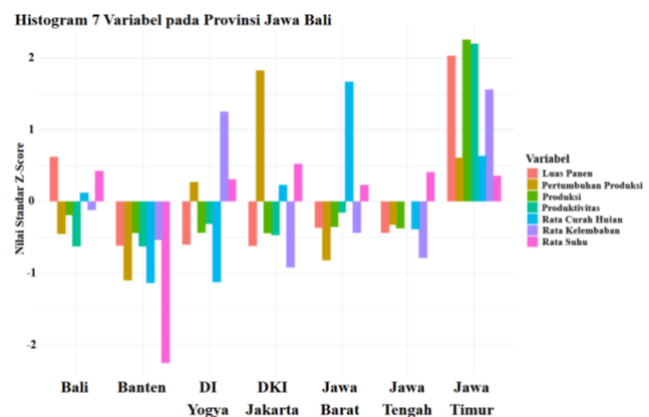
Gambar 7. Peta Wilayah Indonesia Berdasarkan Produktivitas Jeruk Siam

Gambar 7 menunjukkan bahwa beberapa provinsi dalam Cluster 1 terlihat berdekatan secara geografis, yang menunjukkan kemungkinan adanya efek spasial terhadap pola produksi, di mana provinsi yang berdekatan cenderung memiliki kesamaan dalam produktivitas jeruk siam. Jika dilihat pada Provinsi Jawa Bali, diperoleh SNN *clustering* dengan PCA 80% (PC3) pada Gambar 8.



Gambar 8. Peta Wilayah Jawa Bali

Jika dilihat melalui wilayah Jawa Bali, terbentuk *cluster* yang berbeda dengan pengelompokan pada seluruh provinsi. Hal ini dapat dilihat melalui histogram berikut:



Gambar 9. Histogram Provinsi Jawa Bali terhadap Variabel Penelitian

Berdasarkan histogram, dapat ditunjukkan bahwa Cluster 0 (Banten, Jawa Barat, dan Jawa Timur) membentuk kelompok dengan kondisi produksi dan karakteristik agroklimat yang relatif serupa dan stabil. Cluster 1 (DKI Jakarta dan DI Yogyakarta), terpisah karena bukan sentra produksi, sehingga pola variabel pertaniannya berbeda dari provinsi lain. Sementara, Cluster 2 (Jawa Tengah dan Bali) menunjukkan kemiripan karakteristik terkait produksi, produktivitas, dan kondisi iklim, sehingga

keduanya teridentifikasi sebagai kelompok wilayah dengan performa pertanian yang relatif lebih baik.

PENUTUP

SIMPULAN

Berdasarkan hasil analisis, algoritma SNN pada data PCA dengan 3 komponen utama (PC3) yang mempertahankan 78,5% variasi menghasilkan pengelompokan terbaik. Cluster 0 mencakup wilayah penghasil jeruk nasional dengan produksi besar dan karakteristik agroklimat yang beragam, sehingga membentuk kelompok dominan. Cluster 1 menunjukkan produksi kecil namun pertumbuhan yang tinggi, sehingga terpisah dari sentra produksi utama. Sementara itu, Cluster 2 terbentuk akibat kesamaan pola agroklimat dan skala produksi yang relatif kecil hingga nol, sehingga membentuk kelompok tersendiri di luar provinsi dengan produksi besar. SNN menunjukkan stabilitas dan kualitas validasi yang lebih baik dibandingkan DBSCAN, sementara penerapan PCA terbukti meningkatkan kinerja kedua metode.

SARAN

Penggunaan *Silhouette Coefficient* sebagai validasi *cluster* memiliki keterbatasan, sehingga eksplorasi metode lain, seperti *Density-Based Clustering Validation* (DBCV) dapat dipertimbangkan untuk meningkatkan akurasi pengelompokan. Hasil pengelompokan ini dapat menjadi acuan bagi pemerintah atau pihak terkait dalam merancang strategi pengembangan jeruk siam, dengan mendukung provinsi berproduktivitas tinggi dan meningkatkan produktivitas provinsi berproduktivitas menengah hingga rendah agar distribusi produksi lebih merata di seluruh wilayah.

DAFTAR PUSTAKA

- Badan Pusat Statistik. (2025). *Produksi jeruk siam di Indonesia 2023*. Jakarta: BPS.
- Badan Pusat Statistik & Kementerian Pertanian. (2024). *Statistik Pertanian 2024*. SatuData Pertanian.
- Campello, R. J. G. B., Moulavi, D., & Sander, J. (2021). *Density-based clustering: A review*. ACM Transactions on Knowledge Discovery from Data.
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226–231.
- FAO. (2021). *The state of horticulture production and productivity in Southeast Asia*. Rome: Food and Agriculture Organization.
- Fauziyah, L., & Rahmawati, D. (2021). Analisis faktor-faktor yang memengaruhi produktivitas hortikultura di Indonesia. *Jurnal Ilmu Pertanian Indonesia*, 10(2), 45–58.
- Houle, M., Kriegel, H.-P., Kröger, P., Schubert, E., & Zimek, A. (2010). Can shared-neighbor distances defeat the curse of dimensionality? *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 965–973.
- Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques* (4th ed.). Burlington, MA: Morgan Kaufmann.
- Khosasih, E. (2021). Perbandingan algoritma DBSCAN dan Shared Nearest Neighbor (SNN) pada penyebaran COVID-19 di Indonesia. *Jurnal Teknologi dan Sistem Komputer*, 9(3), 215–222.
- Misra, P., Singh, A., & Singh, R. (2020). A review on density-based clustering algorithms for big data analysis. *Journal of King Saud University – Computer and Information Sciences*, 32(1), 21–32.
- Moore, D. S., McCabe, G. P., & Craig, B. A. (2017). *Introduction to the practice of statistics* (9th ed.). New York, NY: W. H. Freeman and Company.
- Mustakim, M., Sari, D. P., & Rahmat, R. F. (2021). Comparison of DBSCAN and PCA-DBSCAN for earthquake clustering. *Jurnal Teknologi dan Sistem Informasi*, 7(2), 89–96.
- Putra, A., & Wulandari, R. (2021). Penerapan analisis spasial untuk pengelolaan produktivitas komoditas pertanian di Indonesia. *Jurnal Geografi dan Lingkungan*, 19(1), 12–25.
- Purwanto, A., Sari, D., & Nugraha, A. (2021). Analisis spasial produktivitas komoditas pertanian berbasis GIS. *Jurnal Agribisnis dan Teknologi*, 8(2), 77–88.
- Schubert, E., & Zimek, A. (2019). Shared-neighbor distances for correlation clustering and outlier detection. *Machine Learning*, 108, 719–761.
- Shah, S., & Kusiak, A. (2021). *Data clustering for spatial and spatiotemporal analysis: Methods and applications*. Information Sciences, 563, 1–20.

- Wahyuni, S., & Yuliana, N. (2022). Variasi produktivitas komoditas pertanian di Indonesia: Analisis spasial. *Jurnal Agroekonomi*, 40(3), 150–162.
- Xu, R. (2020). *Clustering quality evaluation: A review and new perspectives*. *Pattern Recognition Letters*, 138, 228–235.
- Zhao, Y., & Qin, X. (2021). *Density-based clustering algorithms and applications: A survey*. *IEEE Access*, 9, 120702–120720.
- Astuti, A. B., Guci, A. N., Alim, V. I. A., Azizah, L. N., Putri, M. K., & Ngabu, W. (2023). Non Hierarchical K-Means Analysis To Clustering Priority Distribution of Fuel Subsidies in Indonesia. *Barekeng: Jurnal Ilmu Matematika Dan Terapan*, 17(3), 1663-1672.