EVALUASI KOMPARATIF EMPAT ALGORITMA *CLUSTERING* TERHADAP INDIKATOR PERFORMA DAN KESEJAHTERAAN PETANI PERKEBUNAN INDONESIA

Fathur Rachman[1*], Septian Dwi Arianto[2], Trifanny Salsabila[3]
[1,2]Program Studi Statistika, Universitas Terbuka, Tangerang Selatan, Indonesia
[3]Program Studi Sains Data, Universitas Terbuka, Tangerang Selatan, Indonesia

**Penulis korespondensi: 052846831@ecampus.ut.ac.id*

ABSTRAK

Kondisi sektor perkebunan di Indonesia masih menunjukkan perbedaan tingkat produktivitas dan kesejahteraan petani antar wilayah. Penelitian ini bertujuan membandingkan performa beberapa algoritma *clustering* dalam mengelompokkan karakteristik wilayah perkebunan di Indonesia. Analisis yang dilakukan menggunakan empat algoritma *clustering*, yaitu *K-Means*, *Ward's Hierarchical*, *Self-Organizing Maps*, dan *Gaussian Mixture Model* berdasarkan variabel produktivitas lahan, kontribusi produksi, indeks harga terima petani, dan nilai tukar petani. Hasil penelitian menunjukkan bahwa data perkebunan di Indonesia cenderung membentuk dua kelompok utama dengan karakteristik berbeda. Berdasarkan hasil evaluasi, algoritma *K-Means* memberikan performa *clustering* yang paling optimal dibandingkan metode lainnya. Kelompok pertama menunjukkan tingkat produktivitas dan kesejahteraan petani yang relatif lebih tinggi dibandingkan kelompok lainnya. Sementara itu, kelompok lainnya masih menunjukkan tingkat produktivitas dan kesejahteraan petani yang relatif lebih rendah. Dengan temuan ini, terlihat pentingnya pendekatan pembangunan berbasis wilayah dalam mendukung peningkatan kesejahteraan petani perkebunan di Indonesia.

Kata kunci: ekonomi pertanian, kesenjangan regional, *unsupervised machine learning*.

1 PENDAHULUAN

Sektor perkebunan menjadi salah satu sektor penting dalam mendukung perekonomian dan kesejahteraan masyarakat di Indonesia. Menurut Kementerian Pertanian (2023), sektor perkebunan memberikan kontribusi terbesar terhadap pendapatan domestik bruto pertanian, yaitu sebesar 3,88% pada tahun 2023 melalui komoditas unggulan seperti kelapa sawit, kopi, karet, dan kakao. Untuk mengukur kinerja subsektor ini secara menyeluruh, analisis tidak hanya bertumpu pada aspek produktivitas lahan, tetapi juga perlu mempertimbangkan tingkat kesejahteraan petani yang sering ditunjukkan melalui Nilai Tukar Petani (NTP) (Triwidia *et al*., 2024). Namun, kondisi sektor perkebunan di Indonesia masih menunjukkan perbedaan yang cukup besar antarwilayah. Beberapa provinsi memiliki tingkat produktivitas dan kesejahteraan petani yang relatif lebih tinggi dibandingkan wilayah lainnya. Di sisi lain, masih terdapat daerah yang menghadapi berbagai tantangan dalam pengembangan subsektor perkebunan. Perbedaan karakteristik tersebut menunjukkan bahwa pendekatan pembangunan yang seragam belum tentu sesuai untuk diterapkan pada seluruh daerah.

Perbedaan karakteristik antarwilayah menyebabkan pendekatan pembangunan yang seragam belum tentu dapat diterapkan secara efektif pada seluruh daerah. Salah satu pendekatan yang dapat digunakan untuk memahami keragaman tersebut adalah analisis *clustering*, yaitu teknik

pengelompokan wilayah berdasarkan kemiripan karakteristik tertentu (Ikotun *et al*., 2022). Berbagai metode *clustering* telah digunakan dalam penelitian untuk mengelompokkan data berdasarkan karakteristik yang dimiliki. Misalnya, algoritma *K-Means* memiliki keunggulan dalam efisiensi komputasi tetapi sensitif terhadap penentuan awal pusat data, sedangkan metode *Ward's Hierarchical* mampu meminimalkan variansi dalam *cluster* dan menyajikan struktur hierarki secara visual (Ikotun *et al*., 2022). Selain itu, algoritma *Self-Organizing Maps* (SOM) berbasis jaringan saraf tiruan sangat baik dalam memetakan data multidimensi yang kompleks (Ivansyah *et al*., 2023), sementara *Gaussian Mixture Model* (GMM) menggunakan pendekatan *soft clustering* berbasis probabilitas yang fleksibel dalam menangani distribusi data tumpang tindih (Paramarta *et al*., 2025). Di bidang pertanian dan perkebunan, metode ini telah dimanfaatkan untuk pemetaan wilayah pertanian, distribusi pangan, serta evaluasi capaian pembangunan daerah (Ivansyah *et al*., 2023; Yozza *et al*., 2024; Kartakoullis *et al*., 2025). Namun, sebagian besar penelitian terdahulu masih menggunakan satu atau dua algoritma *clustering* tanpa melakukan evaluasi komparatif secara menyeluruh, sehingga pemilihan metode terbaik sering kali belum dapat dibuktikan secara empiris (Ikotun *et al*., 2025). Selain itu, penelitian yang membandingkan beberapa algoritma *clustering* dengan karakteristik berbeda pada data performa dan kesejahteraan petani perkebunan di Indonesia masih sangat terbatas.

Berdasarkan kondisi tersebut, penelitian ini membandingkan empat algoritma *clustering* yang memiliki pendekatan berbeda dalam proses pengelompokan data, yaitu *K-Means*, *Ward's Hierarchical*, *Self-Organizing Maps* (SOM), dan *Gaussian Mixture Model* (GMM). Perbedaan pendekatan antar algoritma memungkinkan terbentuknya hasil *clustering* yang berbeda pada data perkebunan yang dianalisis. Oleh karena itu, penelitian ini bertujuan mengevaluasi performa keempat algoritma tersebut dalam mengelompokkan wilayah perkebunan di Indonesia berdasarkan indikator produktivitas dan kesejahteraan petani. Hasil penelitian diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai tipologi wilayah perkebunan di Indonesia serta menjadi bahan pertimbangan dalam penyusunan kebijakan pembangunan perkebunan yang lebih sesuai dengan kondisi masing-masing daerah.

2 METODE

2.1 Sumber Data dan Variabel

Penelitian ini menggunakan data sekunder yang bersumber dari Tabel Dinamis Badan Pusat Statistik. Data yang digunakan mencakup 36 provinsi di Indonesia pada tahun 2024, dengan mengecualikan Provinsi DKI Jakarta dan Papua Pegunungan karena keterbatasan ketersediaan data. Penelitian ini menggunakan empat variabel, yaitu produktivitas lahan (ton/ha, X1), kontribusi produksi provinsi terhadap total produksi nasional (%, X2), indeks harga terima petani (basis 2018 = 100, X3), dan nilai tukar petani (basis 2018 = 100, X4). Produktivitas lahan dihitung sebagai rasio antara produksi komoditas perkebunan dan luas lahan perkebunan. Kontribusi produksi provinsi dihitung sebagai persentase produksi komoditas perkebunan suatu provinsi terhadap total produksi perkebunan nasional. Sementara itu, indeks harga terima petani dan nilai tukar petani menggunakan tahun dasar 2018 = 100 sesuai standar yang ditetapkan oleh BPS.

2.2 Analisis Deskriptif dan Korelasi

Analisis deskriptif digunakan untuk menggambarkan karakteristik data dari 36 provinsi yang diteliti. Statistik deskriptif yang dihitung meliputi rata-rata, standar deviasi, nilai minimum, nilai maksimum, serta kuartil pertama (Q1), median (Q2), dan kuartil ketiga (Q3) untuk setiap variabel. Sebaran data juga divisualisasikan menggunakan boxplot untuk setiap variabel. Visualisasi

tersebut digunakan untuk mengidentifikasi kemungkinan adanya nilai pencilan pada data. Selanjutnya, analisis korelasi Pearson digunakan untuk mengevaluasi hubungan linear antarvariabel penelitian. Hasil korelasi divisualisasikan dalam bentuk *heatmap* untuk mempermudah identifikasi pola hubungan antarvariabel.

2.3 Analisis *Clustering*

2.3.1 Pra-pemrosesan Data

Sebelum analisis *clustering* dilakukan, data terlebih dahulu melalui tahap pra-pemrosesan. Pada penelitian ini, pra-pemrosesan difokuskan pada standarisasi seluruh variabel menggunakan metode Z-score normalization. Tahap ini diperlukan karena keempat variabel yang digunakan memiliki satuan pengukuran dan rentang nilai yang berbeda. Tanpa standarisasi, variabel dengan skala yang lebih besar berpotensi mendominasi perhitungan jarak dalam proses *clustering*. Metode Z-score normalization mentransformasikan setiap nilai pengamatan menjadi nilai standar yang memilki rata-rata 0 dan simpangan baku 1.

2.3.2 Uji Kecenderungan *Cluster* dan Penentuan Jumlah *Cluster* Optimal

Sebelum analisis *clustering* dilakukan, data terlebih dahulu duhu untuk memastikan adanya kecenderungan pembentukan kelompok secara statistik. Uji statistik Hopkins digunakan untuk membandingkan distribusi jarak antar observasi dalam data dengan distribusi acak seragam. Nilai Hopkins yang (*H*) yang melebihi 0,5 atau mendekati 1 menunjukkan adanya kecenderungan pengelompokan yang kuat (Wright, 2022). Selain uji statistik Hopkins, kecenderungan pengelompokan juga dievaluasi secara visual menggunakan metode *Visual Assessment of Cluster Tendency* (VAT). Metode ini menyusun ulang matriks ketidaksamaan (*dissimilarity matrix*) antar provinsi ke dalam bentuk visualisasi *heatmap* (Bezdek, 2002).

Penentuan jumlah *cluster* optimal (*k*) dilakukan sebelum proses *clustering* untuk memastikan jumlah kelompok yang dibentuk mampu merepresentasikan sttruktur data secara baik. Pada penelitian ini, jumlah *cluster* ditentukan menggunakan tiga pendekatan, yaitu metode *Silhouette Method* (Kaufman & Rousseeuw, 2009), *Elbow Method* (Madhulatha, 2012), dan *Gap Statistics* (Tibshirani *et al*., 2001).

2.3.3 Algoritma *clustering*

Penelitian ini menggunakan empat algoritma *clustering* yang berbeda-beda, yaitu *K-Means clustering*, *Ward's Hierarchical Clustering*, *Self-Organizing Map* (SOM), dan *Gaussian Mixture Model* (GMM). Empat algoritma *clustering* tersebut dipilih karena memiliki pendekatan yang berbeda-beda. *K-Means clustering* bekerja dengan meminimalkan *Within-Cluster Sum of Squares* (WCSS), yaitu total jarak kuadrat antara setiap observasi dan *centroid cluster* tempat observasi tersebut berada (Ikotun *et al*., 2022). Di sisi lain, *Ward's Hierarchical Clustering* didasarkan pada meminimalkan peningkatan *Sum of Squares Error* (SSE) pada setiap tahap penggabungan *cluster*. Selanjutnya, *Self-Organizing Map* (SOM) merupakan metode *clustering* berbasis jaringan saraf tiruan tak terawasi yang mampu memetakan data multidimensi ke dalam representasi berdimensi lebih rendah dengan tetap mempertahankan hubungan kedekatan antar observasi (Kohonen, 2013). Terakhir, *Gaussian Mixture Model* (GMM) merupakan pendekatan *clustering* berbasis probabilitas yang mengasumsikan bahwa data berasal dari kombinasi beberapa distribusi Gaussian, sehingga setiap observasi memiliki peluang menjadi anggota lebih dari satu *cluster* (McLachlan & Peel, 2000).

2.3.4 Komparasi Metode *Clustering* dan Analisis Profil *Cluster*

Setelah seluruh algoritma *clustering* diterapkan, evaluasi dilakukan untuk menentukan metode yang menghasilkan kualitas *cluster* terbaik. Dalam penelitian ini, empat indeks validitas internal digunakan untuk membandingkan keempat algoritma *clustering*, yaitu *Silhouette Index*, *Dunn Index*, *Davies-Bouldin Index* (DBI), dan *Connectivity Index*. Penggunaan beberapa indeks secara bersamaan bertujuan untuk memperoleh penilaian yang lebih komprehensif terhadap kualitas *cluster* yang terbentuk.

Selanjutnya, hasil *clustering* dari metode terbaik digunakan untuk menyusun profil *cluster* berdasarkan produktivitas lahan, kontribusi produksi, indeks harga terima petani, dan nilai tukar petani guna mengidentifikasi karakteristik wilayah perkebunan di Indonesia. Penyusunan profil *cluster* dilakukan menggunakan statistik deskriptif untuk masing-masing variabel. Selain itu, uji uji-*t* dua sampel independen digunakan untuk melihat perbedaan nilai rata-rata keempat variabel antara dua *cluster* yang terbentuk. Sebelum uji-*t* dilaksanakan, uji homogenitas ragam dilakukan terlebih dahulu.

3 HASIL DAN PEMBAHASAN

3.1 Analisis Deskriptif dan Korelasi

Analisis deskriptif dilakukan untuk memberikan gambaran umum mengenai karakteristik distribusi dan sebaran data dari empat variabel performa dan kesejahteraan petani perkebunan di Indonesia. Analisis deskriptif disajikan dalam Tabel 1, dan visualisasi diagram kotak-garis (*boxplot*) pada Gambar 1.

Produktivitas lahan (X1) menunjukkan variasi yang cukup besar antarprovinsi (Tabel 1). Sebaran data yang relatif lebar serta keberadaan beberapa pencilan pada bagian atas distribusi mengindikasikan bahwa kapasitas produksi perkebunan di Indonesia masih bersifat heterogen (Gambar 1). Perbedaan tersebut dapat dipengaruhi oleh kondisi agroekologi, tingkat adopsi teknologi budidaya, kualitas bibit, intensitas pemeliharaan tanaman, serta dukungan infrastruktur pertanian (Tamba, 2025). Temuan ini menunjukkan bahwa keunggulan produktivitas perkebunan masih terkonsentrasi pada wilayah-wilayah tertentu yang memiliki sumber daya dan dukungan teknologi yang lebih baik.

Selain itu, keragaman yang lebih ekstrem terlihat pada variabel kontribusi produksi (X2). Rataan kontribusi produksi provinsi terhadap nasional tercatat sebesar 2,78%, sedangkan median (Q2) hanya sebesar 0,85%. Perbedaan yang cukup besar antara nilai rata-rata dan median mengindikasikan distribusi yang tidak simetris dan cenderung menceng ke kanan (*positive skewness*) (Gambar 1). Pola ini menunjukkan bahwa sebagian provinsi memiliki kontribusi produksi yang relatif rendah, sementara hanya sebagian kecil provinsi yang berperan sebagai sentra utama produksi perkebunan nasional. Kondisi tersebut sejalan dengan studi Simangunsong *et al*. (2022) yang menunjukkan bahwa mayoritas provinsi berada pada kelompok produksi rendah, sedangkan hanya sebagian kecil yang termasuk kelompok produksi tinggi. Kondisi yang sama juga terlihat dalam data Kementerian Pertanian RI (2023), yang menunjukkan bahwa sentra produksi perkebunan nasional masih didominasi oleh beberapa provinsi di Sumatera dan Kalimantan. Hasil ini mengindikasikan adanya konsentrasi spasial produksi perkebunan yang cukup kuat di Indonesia.

Indeks Harga Terima Petani (X3) digunakan untuk menggambarkan perkembangan harga komoditas yang diterima petani dibandingkan tahun dasar. Secara umum, seluruh provinsi menunjukkan nilai indeks di atas tahun dasar (2018 = 100), yang mengindikasikan adanya peningkatan harga komoditas perkebunan dalam periode pengamatan (Tabel 1). Meskipun

demikian, sebaran data yang relatif lebar pada Gambar 1 menunjukkan bahwa peningkatan harga yang diterima petani tidak berlangsung secara merata antarprovinsi.

Tabel 1. Statistik deskriptif empat variabel performa dan kesejahteraan petani perkebunan di Indonesia

Variabel	Min.	Q1	Q2	Q3	Maks.	Rataan	Standar deviasi
Produktivitas lahan (ton per ha)	0,46	0,92	1,45	2,31	4,36	1,66	0,88
Kontribusi produksi (%)	0,08	0,25	0,85	2,68	17,90	2,78	4,30
Indeks harga terima petani	109,0	127,0	150,0	189,0	255,0	161,0	41,10
Nilai tukar petani	96,5	106,0	126,0	156,0	204,0	134,0	33,20

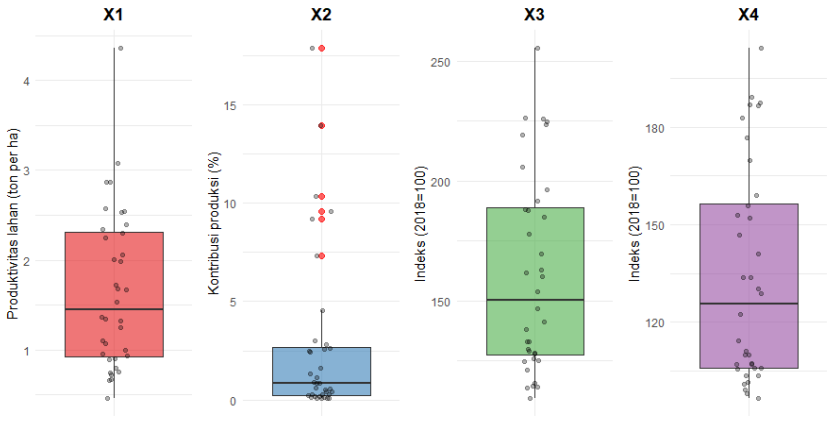
Perbedaan tersebut dapat dipengaruhi oleh variasi komoditas unggulan regional, akses pasar, keterhubungan dengan rantai pasok, serta efisiensi sistem pemasaran komoditas perkebunan. Temuan ini sejalan dengan Prayitno *et al*. (2024) yang menunjukkan bahwa perkembangan harga yang diterima petani sangat dipengaruhi oleh dukungan infrastruktur, teknologi, dan kondisi pasar lokal. Selain itu, Wahyuni *et al*. (2025) menunjukkan bahwa wilayah dengan integrasi pasar yang lebih baik cenderung memperoleh manfaat harga yang lebih besar dibandingkan wilayah yang masih menghadapi keterbatasan akses pemasaran. Oleh karena itu, tingginya nilai Indeks Harga Terima Petani pada beberapa provinsi dapat menunjukkan tingkat integrasi wilayah tersebut dalam jaringan perdagangan komoditas perkebunan yang lebih kompetitif.

Meskipun demikian, tingginya harga yang diterima petani belum tentu sejalan dengan tingginya tingkat kesejahteraan. Hal ini terlihat pada variabel Nilai Tukar Petani (X4), yang merupakan indikator komprehensif karena membandingkan harga yang diterima petani dengan harga yang harus dibayar untuk konsumsi rumah tangga dan input produksi. Secara umum, sebagian besar provinsi memiliki nilai NTP di atas 100, yang menunjukkan bahwa petani perkebunan berada dalam kondisi surplus ekonomi dibandingkan tahun dasar. Namun demikian, masih terdapat beberapa wilayah yang berada pada kondisi mendekati bahkan di bawah batas kesejahteraan (NTP = 100), yang mengindikasikan bahwa peningkatan pendapatan belum sepenuhnya mampu mengimbangi kenaikan biaya yang harus ditanggung petani.

Temuan ini menunjukkan bahwa peningkatan harga komoditas yang diterima petani tidak selalu berujung pada peningkatan kesejahteraan yang setara. Hasil tersebut konsisten dengan Salahuddin *et al*. (2023) yang menemukan bahwa petani perkebunan di Konawe Selatan mengalami NTP di bawah 100 akibat kenaikan biaya produksi yang lebih cepat dibandingkan peningkatan harga jual. Temuan serupa juga dilaporkan oleh Ningsih *et al*. (2025), yang menunjukkan bahwa kenaikan harga komoditas belum tentu meningkatkan kesejahteraan apabila diikuti oleh peningkatan biaya produksi dan konsumsi rumah tangga petani. Oleh karena itu, hubungan antara Indeks Harga Terima Petani dan Nilai Tukar Petani tidak selalu bersifat linier karena kesejahteraan petani pada akhirnya ditentukan oleh keseimbangan antara pendapatan dan pengeluaran.

Secara keseluruhan, analisis deskriptif menunjukkan bahwa keempat variabel memiliki tingkat heterogenitas yang tinggi, distribusi yang tidak sepenuhnya normal, serta keberadaan pencilan terutama pada variabel kontribusi produksi (X2). Karakteristik tersebut mengindikasikan adanya ketimpangan spasial yang cukup kuat dalam sistem perkebunan Indonesia, baik dari aspek produktivitas, konsentrasi produksi, harga yang diterima petani, maupun tingkat kesejahteraan petani. Temuan ini memperkuat relevansi penggunaan pendekatan *clustering* dalam penelitian ini. Menurut Fitriana *et al*. (2024), analisis *cluster* sangat efektif untuk mengidentifikasi pola

kemiripan antar wilayah yang memiliki karakteristik pertanian berbeda-beda. Di sisi lain, Mujahidin & Hasanah (2025) menjelaskan bahwa heterogenitas data dan keberadaan pencilan merupakan kondisi yang sesuai untuk mengevaluasi kinerja berbagai algoritma *clustering*.



Gambar 1. Sebaran data empat variabel performa dan kesejahteraan petani perkebunan di Indonesia.

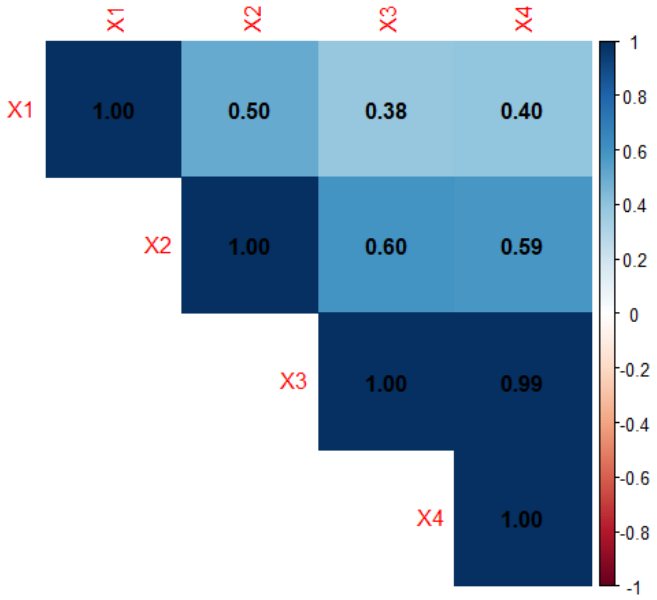
Setelah karakteristik distribusi masing-masing variabel dieksplorasi melalui statistik deskriptif dan *boxplot*, analisis korelasi dilakukan untuk memahami keterkaitan antara indikator performa produksi dan kesejahteraan petani perkebunan. Gambar 2 menyajikan matriks korelasi Pearson yang menggambarkan arah dan kekuatan hubungan linear antara produktivitas lahan (X1), kontribusi produksi (X2), Indeks Harga Terima Petani (X3), dan Nilai Tukar Petani (X4). Secara umum, seluruh variabel menunjukkan korelasi positif, yang mengindikasikan Namun demikian, kekuatan hubungan tersebut bervariasi sehingga memberikan informasi penting mengenai faktor-faktor yang paling berpengaruh terhadap kesejahteraan petani perkebunan.

Temuan menarik terlihat pada hubungan antara Indeks Harga Terima Petani (X3) dan Nilai Tukar Petani (X4) yang memiliki koefisien korelasi sebesar 0,99. Nilai ini menunjukkan hubungan positif yang sangat kuat dan mengindikasikan bahwa perubahan harga komoditas yang diterima petani merupakan faktor dominan yang memengaruhi perubahan tingkat kesejahteraan petani perkebunan. Secara teoritis, temuan ini dapat dijelaskan karena NTP adalah rasio antara indeks harga yang diterima petani dan indeks harga yang dibayar petani. Oleh karena itu, ketika harga komoditas perkebunan meningkat secara signifikan, NTP cenderung meningkat selama kenaikan tersebut tidak diimbangi oleh peningkatan biaya produksi dan konsumsi rumah tangga yang lebih besar (Ningsih *et al*., 2025).

Hubungan yang cukup menarik juga terlihat pada variabel kontribusi produksi (X2). Variabel ini memiliki korelasi sebesar 0,60 dengan Indeks Harga Terima Petani (X3) dan 0,59 dengan Nilai Tukar Petani (X4). Kekuatan hubungan tersebut termasuk dalam kategori sedang hingga kuat, yang menunjukkan bahwa wilayah dengan kontribusi produksi lebih besar cenderung memiliki tingkat harga terima dan tingkat kesejahteraan petani yang lebih tinggi. Hasil ini mengindikasikan adanya keuntungan ekonomi yang diperoleh oleh daerah sentra produksi. Provinsi pusat produksi perkebunan umumnya memiliki infrastruktur logistik yang lebih baik, akses pasar yang lebih luas, volume perdagangan yang lebih tinggi, serta posisi tawar yang lebih kuat dalam rantai nilai komoditas (Wahyuni *et al*., 2025; Prayitno *et al*., 2024). Kondisi tersebut memungkinkan petani memperoleh harga jual yang relatif lebih kompetitif dibandingkan wilayah dengan skala produksi

yang kecil. Meskipun demikian, kekuatan korelasi X2 dengan X3 dan X4 yang hanya berada pada tingkat moderat menunjukkan bahwa besarnya kontribusi produksi tidak selalu menjamin tingginya kesejahteraan petani. Oleh karena itu, peningkatan produksi perlu diiringi dengan perbaikan tata niaga dan penguatan posisi tawar petani dalam rantai pasok komoditas perkebunan.

Berbeda dengan variabel lainnya, produktivitas lahan (X1) menunjukkan korelasi yang relatif lebih rendah terhadap seluruh indikator kesejahteraan. Korelasi antara produktivitas lahan dan kontribusi produksi tercatat sebesar 0,50, sedangkan korelasinya dengan Indeks Harga Terima Petani (X3) dan Nilai Tukar Petani (X4) masing-masing sebesar 0,38 dan 0,40. Hal ini menunjukkan bahwa peningkatan produktivitas fisik lahan tidak secara otomatis menghasilkan peningkatan kesejahteraan petani dalam proporsi yang sama. Kesejahteraan petani perkebunan tidak hanya ditentukan oleh jumlah produksi yang dihasilkan per satuan luas lahan, tetapi juga oleh faktor-faktor ekonomi yang terjadi setelah proses produksi berlangsung. Produktivitas yang tinggi dapat kehilangan dampak ekonominya apabila harga jual rendah, biaya input meningkat, atau sistem pemasaran tidak berjalan secara efisien (Tamba, 2025). Dengan kata lain, produktivitas merupakan syarat penting bagi peningkatan pendapatan, tetapi bukan satu-satunya faktor penentu kesejahteraan. Strategi pembangunan perkebunan yang hanya berorientasi pada peningkatan produktivitas berisiko menghasilkan pertumbuhan produksi tanpa diikuti peningkatan kesejahteraan petani secara proporsional.



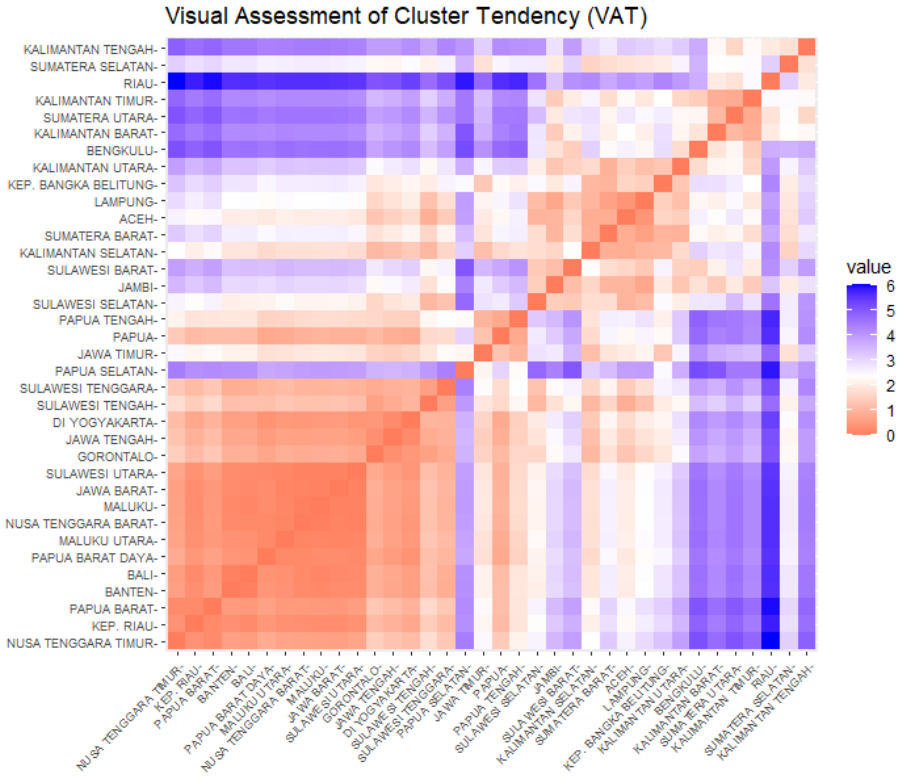
Gambar 2. Korelasi antar empat variabel performa dan kesejahteraan petani perkebunan di Indonesia.

Gambar 2 menunjukkan adanya korelasi yang sangat kuat antara Indeks Harga Terima Petani (X3) dan Nilai Tukar Petani (X4) (r = 0,99). Kondisi ini mengindikasikan adanya redudansi informasi yang berpotensi memengaruhi representasi jarak pada algoritma *clustering* berbasis Euclidean. Meskipun demikian, korelasi yang tinggi antarvariabel bukan merupakan pelanggaran asumsi dalam analisis *clustering*, melainkan faktor yang dapat memengaruhi pembentukan struktur kelompok dan kontribusi relatif masing-masing variabel dalam proses pengelompokkan. Gniazdowski & Kaliszewski (2019) menunjukkan bahwa variabel yang berkorelasi tetap dapat digunakan dalam analisis *clustering* selama masih memberikan informasi yang relevan terhadap struktur data. Namun demikian, penelitian selanjutnya dapat mempertimbangkan penggunaan

Principal Component Analysis (PCA), seleksi variabel, atau ukuran jarak yang mempertimbangkan struktur kovarians data untuk mengurangi potensi redundansi informasi antar indikator (Brown *et al*., 2022).

3.2 Analisis *Clustering*

Sebelum menerapkan algoritma *clustering*, pengujian awal diperlukan untuk memastikan bahwa data memiliki kecenderungan membentuk kelompok yang bermakna. Dalam penelitian ini, evaluasi kecenderungan *cluster* dilakukan menggunakan kombinasi statistik Hopkins dan *Visual Assessment of Cluster Tendency* (VAT) yang disajikan pada Gambar 3.

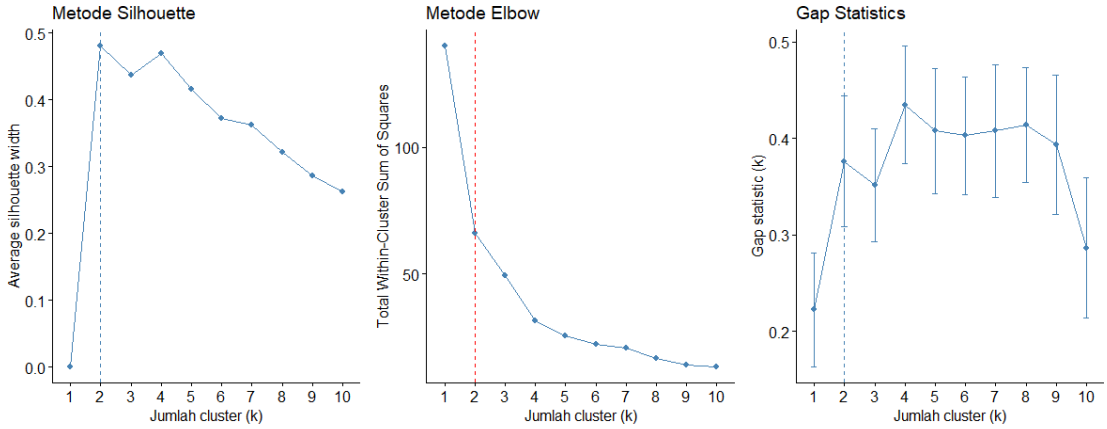


Gambar 3. *Visual assessment of cluster tendency* berdasarkan empat variabel performa dan kesejahteraan petani perkebunan di Indonesia (Hopkins = 0,8456, *p-value* = 0,002)

Hasil pengujian menunjukkan bahwa data memiliki nilai Hopkins sebesar 0,8456 dengan *p-value* sebesar 0,002. Nilai Hopkins yang mendekati 1 mengindikasikan adanya kecenderungan pengelompokan yang kuat (Wright, 2022). Dengan demikian, struktur data performa dan kesejahteraan petani perkebunan di Indonesia tidak terbentuk secara acak, melainkan memiliki pola pengelompokan alami yang jelas. Hal ini juga dikonfirmasi melalui *heatmap* VAT pada Gambar 3. Secara visual, blok-blok yang terbentuk pada diagonal utama menunjukkan tingkat kemiripan internal (*within-cluster similarity*) yang tinggi dan tingkat perbedaan antarblok (*between-cluster dissimilarity*) yang relatif besar. Pola ini menunjukkan bahwa provinsi-provinsi di Indonesia memiliki karakteristik performa perkebunan dan kesejahteraan petani yang heterogen, sehingga layak untuk dianalisis lebih lanjut menggunakan pendekatan *clustering*.

Gambar 4 menyajikan hasil penentuan jumlah *cluster* optimal (*k*) melalui tiga pendekatan, yaitu metode *Silhouette Method*, *Elbow Method*, dan *Gap Statistics*. Penentuan jumlah *cluster* merupakan tahapan paling penting dalam analisis *clustering* karena memengaruhi kualitas

pengelompokkan dan interpretasi hasil yang diperoleh. Pemilihan jumlah *cluster* yang terlalu sedikit berpotensi menyederhanakan struktur data yang sebenarnya kompleks (*under-partitioning*), sedangkan jumlah *cluster* yang terlalu banyak dapat menghasilkan kelompok yang sulit diinterpretasikan dan kurang relevan secara substantif (Utami & Ispriyanti, 2023). Berdasarkan hasil analisis, jumlah *cluster* optimal yang digunakan adalah dua kelompok. Jumlah kelompok tersebut selanjutnya digunakan pada keempat algoritma *clustering* berikutnya. Visualisasi hasil *clustering* pada Gambar 5 menunjukkan bahwa seluruh algoritma berhasil mengidentifikasi dua kelompok utama sebagaimana yang telah diprediksi pada tahap analisis kecenderungan *cluster* dan penentuan jumlah *cluster* optimal. Namun demikian, pola pemisahan yang dihasilkan masing-masing metode menunjukkan karakteristik yang berbeda.



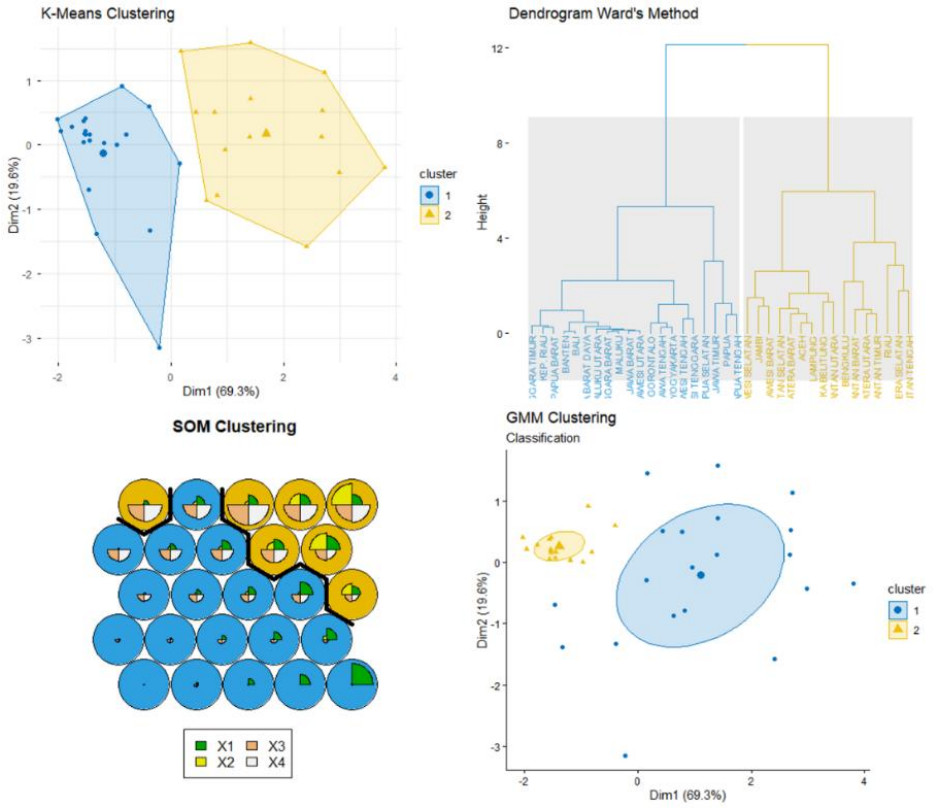
Gambar 4. Penentuan jumlah *cluster* berdasarkan metode Silhouette, Elbow, dan Gap Statistics

Pada *K-Means*, kedua kelompok terlihat terpisah secara jelas pada ruang komponen utama dengan tingkat tumpang tindih yang rendah (Gambar 5). Hasil ini menunjukkan adanya struktur *cluster* yang cukup kompak sehingga sebagian besar provinsi berada di sekitar pusat kelompok masing-masing. Sementara itu, *Ward's Hierarchical Clustering* menghasilkan struktur pengelompokan yang serupa, tetapi divisualisasikan dalam bentuk *dendrogram* yang memperlihatkan tingkat kemiripan antar provinsi secara bertahap (Gambar 5). Dendrogram menunjukkan bahwa beberapa provinsi memiliki kedekatan karakteristik yang tinggi, sedangkan provinsi lainnya bergabung pada tingkat ketidakmiripan yang lebih besar.

Hasil SOM juga menunjukkan terbentuknya dua kelompok utama yang relatif konsisten dengan *K-Means* dan *Ward's*. Namun, beberapa provinsi masih berada pada area transisi sehingga batas antar *cluster* terlihat kurang tegas (Gambar 5). Kondisi ini mengindikasikan bahwa SOM mampu menangkap variasi lokal dan hubungan nonlinier antar variabel yang mungkin tidak sepenuhnya terdeteksi oleh metode berbasis *centroid*. Sementara itu, GMM menghasilkan pola serupa dengan batas *cluster* yang lebih fleksibel. Visualisasi berbentuk elips menunjukkan bahwa setiap kelompok direpresentasikan sebagai distribusi probabilitas sehingga suatu provinsi dapat memiliki peluang keanggotaan pada lebih dari satu *cluster* (Gambar 5). Temuan ini mengindikasikan bahwa meskipun keempat algoritma menghasilkan struktur dua *cluster* yang relatif konsisten, terdapat perbedaan dalam cara masing-masing metode merepresentasikan wilayah yang memiliki karakteristik transisional.

Untuk menentukan algoritma yang menghasilkan kualitas pengelompokan terbaik, dilakukan evaluasi menggunakan empat indeks validasi internal, yaitu *Silhouette Index*, *Dunn Index*, *Davies-*

Bouldin Index (DBI), dan *Connectivity Index* (Tabel 2). Keempat indeks tersebut dipilih karena merepresentasikan dimensi kualitas *cluster* yang berbeda. *Silhouette Index* mengevaluasi keseimbangan antara kohesi internal dan separasi antar kelompok, *Dunn Index* mengukur rasio antara jarak minimum antar *cluster* dan dispersi maksimum dalam *cluster, Davies-Bouldin Index* menilai tingkat kemiripan antar kelompok, sedangkan *Connectivity Index* mengukur kemampuan algoritma dalam mempertahankan hubungan antar objek secara lokal.



Gambar 5. Visualisasi *clustering* provinsi di Indonesia berdasarkan performa dan kesejahteraan petani perkebunan di Indonesia menggunakan empat algoritma berbeda

Berdasarkan Tabel 2, algoritma *K-Means* menunjukkan performa terbaik secara keseluruhan dengan *Silhouette* tertinggi (0,4808) dan DBI terendah (0,9434). Hasil tersebut menunjukkan bahwa *K-Means* menghasilkan *cluster* yang relatif paling homogen secara internal sekaligus memiliki pemisahan yang paling jelas dibandingkan algoritma lainnya. Nilai *Silhouette* yang mendekati 0,5 mengindikasikan bahwa sebagian besar provinsi telah ditempatkan pada *cluster* yang sesuai dengan karakteristiknya. Di sisi lain, nilai DBI yang paling kecil menunjukkan bahwa variasi dalam kelompok relatif rendah dibandingkan jarak antar kelompok yang terbentuk. Temuan ini mengindikasikan bahwa struktur data performa dan kesejahteraan petani perkebunan Indonesia cenderung membentuk kelompok yang relatif kompak dan terpisah dengan cukup baik.

Di sisi lain, *Ward's Hierarchical Clustering* menunjukkan performa yang relatif sebanding dengan *K-Means*. Nilai *Silhouette* yang diperoleh hanya sedikit lebih rendah (0,4781), sementara metode ini menghasilkan nilai *Connectivity* terbaik (6,0528). Hasil tersebut menunjukkan bahwa *Ward's* mampu mempertahankan struktur kedekatan antar objek dengan baik selama proses pembentukan *cluster*. Selain menghasilkan kualitas pengelompokan yang baik, *Ward's* juga

memberikan informasi tambahan mengenai struktur hubungan antarprovinsi melalui representasi hierarkisnya

Sementara itu, algoritma SOM menghasilkan nilai *Dunn Index* tertinggi (0,2430), yang menunjukkan bahwa metode ini mampu membentuk *cluster* dengan tingkat separasi yang relatif baik. Namun demikian, nilai *Silhouette* dan DBI yang masih berada di bawah *K-Means* menunjukkan bahwa keunggulan tersebut belum diikuti oleh kualitas pemisahan kelompok yang lebih baik secara keseluruhan. Meskipun algoritma SOM mampu menghasilkan struktur *cluster* yang cukup baik, performanya masih berada di bawah *K-Means* dan *Ward's* berdasarkan evaluasi gabungan seluruh indeks validasi.

Berbeda dengan ketiga metode lainnya, GMM menunjukkan performa yang relatif lebih rendah pada sebagian besar indikator evaluasi. Hasil ini menunjukkan bahwa pendekatan probabilistik yang digunakan GMM tidak memberikan keuntungan pada struktur data yang dianalisis. Hal ini diduga karena jumlah observasi yang relatif terbatas, sehingga estimasi parameter distribusi probabilistik menjadi kurang stabil. Selain itu, karakteristik data yang menunjukkan dua kelompok yang relatif tegas menyebabkan pendekatan *soft clustering* yang digunakan GMM tidak memberikan manfaat tambahan yang signifikan. Temuan ini sejalan dengan Maulidya *et al*. (2024) dan Putri *et al*. (2025) yang menunjukkan bahwa metode berbasis probabilitas cenderung menghasilkan kualitas pengelompokan yang lebih rendah dibandingkan metode dengan pembagian kelompok yang jelas.

Tabel 2. Perbandingan empat metode *clustering* provinsi di Indonesia berdasarkan tipologi performa dan kesejahteraan petani perkebunan

Metode *clustering*	Silhouette	Dunn Index	DBI	Connectivity
K-Means	**0,4808**	0,1860	**0,9434**	9,6794
Ward's	0,4781	0,1899	0,9455	**6,0528**
SOM	0,4494	**0,2430**	0,9675	11,6151
GMM	0,3966	0,1329	0,9736	8,6702

Meskipun setiap algoritma memiliki keunggulan pada indikator tertentu, tidak terdapat satu metode yang unggul pada seluruh indikator validasi. Kondisi ini menunjukkan bahwa kualitas *clustering* bersifat multidimensional dan sangat dipengaruhi oleh karakteristik data yang dianalisis. Namun demikian, metode *K-Means* unggul pada dua indikator, yaitu *Silhouette Index* dan *Davies–Bouldin Index*. Selain itu, visualisasi hasil *clustering* memperlihatkan bahwa *K-Means* menghasilkan batas kelompok yang lebih jelas dan mudah diinterpretasikan dibandingkan metode lainnya. Dengan demikian, hasil pengelompokan yang diperoleh dari algoritma *K-Means* selanjutnya digunakan sebagai dasar dalam analisis profil *cluster*.

Hasil pengelompokan pada Tabel 3 menunjukkan bahwa 36 provinsi di Indonesia terbagi ke dalam dua kelompok utama yang memiliki karakteristik performa dan kesejahteraan petani perkebunan yang berbeda secara signifikan. Berdasarkan uji statistik *t* dua sampel independen, seluruh variabel menunjukkan nilai *p-value* < 0,05. Hal ini mengindikasikan adanya perbedaan rata-rata produktivitas lahan, kontribusi produksi, Indeks Harga Terima Petani, dan Nilai Tukar Petani (NTP) berbeda secara signifikan antara kedua *cluster*. Hasil ini menunjukkan bahwa pembagian kelompok yang dihasilkan oleh algoritma *K-Means* memiliki makna yang kuat dalam menggambarkan profil wilayah perkebunan Indonesia.

Cluster 1 merupakan kelompok terbesar yang terdiri atas 21 provinsi, meliputi sebagian besar wilayah Jawa, Bali, Nusa Tenggara, Maluku, dan Papua. Sebaliknya, *Cluster* 2 terdiri atas

15 provinsi yang didominasi oleh wilayah Sumatra, Kalimantan, serta sebagian Sulawesi. Perbedaan komposisi geografis ini mengindikasikan bahwa struktur sektor perkebunan Indonesia masih menunjukkan konsentrasi spasial yang kuat pada kawasan-kawasan tertentu yang telah lama berkembang sebagai sentra komoditas perkebunan nasional. Perbedaan tersebut terlihat jelas pada indikator produktivitas lahan. *Cluster* 2 memiliki produktivitas rata-rata yang jauh lebih tinggi dibandingkan *Cluster* 1. Hal ini menunjukkan bahwa provinsi-provinsi pada *Cluster* 2 memiliki kapasitas produksi yang lebih baik, baik karena faktor agroekologi, ketersediaan lahan, tingkat adopsi teknologi, maupun keberadaan komoditas perkebunan yang lebih kompetitif. Temuan ini sejalan dengan penelitian Tamba (2025), yang menunjukkan bahwa produktivitas pertanian antar provinsi di Indonesia sangat dipengaruhi oleh tingkat adopsi teknologi, karakteristik subsektor pertanian, dan kapasitas pengembangan wilayah.

Tabel 3. Perbandingan dua *cluster* berdasarkan metode *K-Means* pada empat variabel tipologi performa dan kesejahteraan petani perkebunan

Variabel	*Cluster* 1	*Cluster* 2	*p-value*
Jumlah provinsi	21	15	
Provinsi	Kepulauan Riau, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Selatan, Sulawesi Utara, Sulawesi Tengah, Sulawesi Tenggara, Gorontalo, Maluku, Maluku Utara, Papua Barat, Papua Barat Daya, Papua, Papua Selatan, Papua Tengah	Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kalimantan Barat, Kalimantan Tengah, Kalimantan Timur, Kalimantan Utara, Sulawesi Selatan, Sulawesi Barat	
Produktivitas lahan	1,32 ± 0,89	2,13 ± 0,63	<0,05
Kontribusi produksi	0,59 ± 0,77	5,84 ± 5,30	<0,05
Indeks harga terima petani	131,0 ± 15,3	202,0 ± 26,1	<0,05
Nilai tukar petani	110,0 ± 11,0	168,0 ± 21,5	<0,05

Selain produktivitas, kesenjangan antar *cluster* juga terlihat pada kontribusi produksi terhadap total produksi nasional. *Cluster* 2 memiliki kontribusi produksi rata-rata lebih besar daripada provinsi pada *Cluster* 1. Hal ini memperlihatkan pola konsentrasi spasial sektor perkebunan yang telah lama menjadi karakteristik pembangunan agribisnis Indonesia, terutama pada wilayah Sumatra dan Kalimantan yang menjadi pusat pengembangan komoditas strategis seperti kelapa sawit, karet, dan berbagai komoditas ekspor lainnya. Temuan ini sejalan dengan Mujahidin & Hasanah (2025) yang menemukan bahwa provinsi-provinsi dengan produksi pertanian tinggi cenderung membentuk kelompok tersendiri karena memiliki kapasitas produksi dan kontribusi nasional yang jauh lebih besar dibandingkan wilayah lainnya.

Kesenjangan performa produksi tersebut ternyata diikuti oleh perbedaan yang sangat jelas pada indikator kesejahteraan petani. *Cluster* 2 memiliki rata-rata Indeks Harga Terima Petani yang jauh lebih tinggi dibandingkan *Cluster* 1. Perbedaan serupa juga terlihat pada variabel Nilai Tukar Petani. Hal ini menunjukkan bahwa keunggulan produksi yang dimiliki *Cluster* 2 tidak hanya ditunjukkan dalam aspek produksi, tetapi juga dalam hal keuntungan ekonomi yang lebih besar

bagi petani. Sebaliknya, petani pada *Cluster* 1 berada pada kondisi yang relatif lebih rentan. Kondisi ini mengindikasikan bahwa sebagian besar provinsi dalam *Cluster* 1 masih menghadapi berbagai keterbatasan struktural, baik dalam aspek produktivitas, akses pasar, efisiensi rantai pasok, maupun kemampuan menciptakan nilai tambah dari komoditas perkebunan yang dihasilkan. Hasil ini memperkuat temuan Fitriana *et al*. (2024) yang menunjukkan bahwa sektor pertanian Indonesia memiliki tingkat heterogenitas yang tinggi dan secara alami membentuk kelompok wilayah dengan karakteristik yang berbeda.

Hasil *clustering* ini menunjukkan adanya pola pembagian secara spasial dalam sektor perkebunan Indonesia. *Cluster* 2 dapat dikategorikan sebagai kelompok wilayah dengan performa produksi dan kesejahteraan petani yang relatif tinggi, sedangkan *Cluster* 1 merepresentasikan kelompok wilayah dengan performa dan kesejahteraan yang relatif lebih rendah. Hal ini menunjukkan bahwa ketimpangan sektor perkebunan tidak hanya terjadi pada aspek produksi, tetapi juga terlihat pada tingkat kesejahteraan petani yang dihasilkan. Yusman & Berliana (2025) menunjukkan bahwa keberadaan kelompok wilayah yang memiliki karakteristik berbeda memerlukan pendekatan pembangunan yang berbeda pula. Dalam penelitian ini, provinsi-provinsi pada *Cluster* 2 memerlukan kebijakan yang berorientasi pada peningkatan daya saing, hilirisasi industri, sertifikasi keberlanjutan, dan penguatan akses pasar global. Sebaliknya, provinsi-provinsi pada *Cluster* 1 lebih membutuhkan kebijakan yang berfokus pada peningkatan produktivitas, perbaikan infrastruktur logistik, penguatan kelembagaan petani, akses pembiayaan, serta efisiensi tata niaga komoditas.

4 KESIMPULAN

Hasil penelitian menunjukkan bahwa data performa dan kesejahteraan petani perkebunan di Indonesia memiliki kecenderungan pengelompokan yang kuat dan secara optimal terbagi ke dalam dua *cluster*, di mana algoritma *K-Means* merupakan algoritma *clustering* terbaik. *Cluster* yang terbentuk mengidentifikasi adanya dualisme struktural sektor perkebunan Indonesia, yaitu kelompok provinsi dengan performa produksi dan kesejahteraan petani yang relatif tinggi, serta kelompok provinsi dengan performa dan kesejahteraan yang relatif lebih rendah. Temuan ini menegaskan pentingnya penerapan kebijakan pembangunan perkebunan yang berbasis tipologi wilayah agar intervensi yang dilakukan lebih tepat sasaran sesuai dengan karakteristik dan kebutuhan masing-masing kelompok provinsi.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Fakultas Sains dan Teknologi Universitas Terbuka atas dukungan akademik yang diberikan selama penyusunan makalah ini. Apresiasi tinggi juga disampaikan kepada Badan Pusat Statistik atas penyediaan data sekunder dinamis yang menjadi dasar utama dalam analisis penelitian ini. Selain itu, ucapan terima kasih ditujukan kepada seluruh rekan sejawat serta pihak lain yang telah meluangkan waktu untuk memberikan saran dan diskusi konstruktif hingga artikel ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

Bezdek, J. C., & Hathaway, R. J. (2002). VAT: a tool for visual assessment of (*cluster*) tendency [Conference paper]. *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No.02CH37290)*, *3*, 2225-2230vol.3. https://doi.org/10.1109/IJCNN.2002.1007487

Brown, P. O., Chiang, M. C., Guo, S., Jin, Y., Leung, C. K., Murray, E. L., Pazdor, A. G. M., & Cuzzocrea, A. (2022). Mahalanobis distance based *K-Means clustering*. In *Data Warehousing and Knowledge Discovery* (pp. 314–327). Springer. https://doi.org/10.1007/978-3-031-12670-3_23

Fitriana, I. N. L., Leviany, F., Faulina, R., Nuramaliyah, N., & Safitri, E. (2024). Mapping indonesia's agricultural diversity: *clustering* provinces with self-organizing maps. *Jurnal Lebesgue Jurnal Ilmiah Pendidikan Matematika Matematika Dan Statistika*, *5*(3), 2178–2196. https://doi.org/10.46306/lb.v5i3.844

Gniazdowski, Z., & Kaliszewski, D. (2019). On the *clustering* of correlated random variables. *Zeszyty Naukowe Warszawskiej Wyższej Szkoły Informatyki*, *18*, 45–58. https://doi.org/10.26348/ZNWWSI.18.45

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2022). *K-Means clustering* algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178–210. https://doi.org/10.1016/j.ins.2022.11.139

Ikotun, A. M., Habyarimana, F., & Ezugwu, A. E. (2025). *Cluster* validity indices for automatic *clustering*: A comprehensive review. *Heliyon, 11*(2). doi:10.1016/j.heliyon.2025.e41953

Ivansyah, A.R., Fitri, F., Kurniawati, Y., & Mukhti, T.O. (2023). Implementation self organizing maps method in *cluster* analysis based on achievement suistainable development goal/SDG's West Sumatera Province. *UNP Journal of Statistics and Data Science, 1*(5), 480–487. https://doi.org/10.24036/ujsds/vol1-iss5/118

Kartakoullis, A., Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2025). Gaussian mixture model *clustering* allows accurate semantic image segmentation of wheat kernels from near-infrared hyperspectral images. *Chemometrics and Intelligent Laboratory Systems*, *259*, 105341. https://doi.org/10.1016/j.chemolab.2025.105341

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, John Wiley & Sons.

Kementerian Pertanian. (2023). Analisis PDB Sektor Pertanian 2023. In *https://satudata.pertanian.go.id/details/publikasi/489*. Pusat Data dan Sistem Informasi Pertanian. Retrieved June 6, 2026, from https://satudata.pertanian.go.id/details/publikasi/489

Kohonen, T. (2012). Essentials of the self-organizing map. *Neural Networks*, *37*, 52–65. https://doi.org/10.1016/j.neunet.2012.09.018

Madhulatha, T. S. (2012). An overview on *clustering* methods. *IOSR Journal of Engineering*, *2*(4), 719-725. https://doi.org/10.48550/arXiv.1205.1117

Maulidya, A., Sitorus, Z., Siahaan, A. P. U., & Iqbal, M. (2024). Analysis of increasing student service satisfaction using k-means clustering algorithm and gaussian mixture models (GMM). *International Journal Of Computer Sciences and Mathematics Engineering*, *3*(1), 29-35.

McLachlan, G. J., & Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons.

Mujahidin, I., & Hasanah, S. H. (2025). Comparative analysis of *K-Means*, k-medoids, and fuzzy c-means for *clustering* provinces in indonesia based on rice production in 2024. *Jurnal Gaussian, 14*(2), 356-365. https://doi.org/10.14710/j.gauss.14.2.356-365

Ningsih, A. C., Mardiyanti, S., Amruddin, A., & Natsir, M. (2025). Dinamika luas panen dan harga komoditas terhadap nilai tukar petani subsektor tanaman pangan di Provinsi Sulawesi Tengah: analisis time series. *SENTRI: Jurnal Riset Ilmiah*, *4*(12), 4035–4042. https://doi.org/10.55681/sentri.v4i12.5054

Paramarta, V., Rahman, A., Priska, L., Gurning, R.R., & Purwati, W.A. (2025). Comparative analysis of *K-Means* and gaussian mixture model in *clustering* global CO2 emissions. *Bit-Tech*, *8*(1), 998–1008. https://doi.org/10.32877/bt.v8i1.2805

Prayitno, G., Anwar, R., Indrayanto, T., Permana, M., & Purnomowati, R. (2024). Comparison of farmer welfare in Kediri Regency and Probolinggo Regency based on farmer exchange rates (FER). *Journal of Regional and Rural Studies*, *2*(1), 15–26. https://doi.org/10.21776/rrs.v2i1.28

Putri, T. S., Diyasa, I. G. S. M., & Junaidi, A. (2025). Performance comparison of gaussian mixture model, hierarchical clustering, and k-medoids in passenger data clustering. *bit-Tech*, *8*(2), 1615-1624.

Salahuddin, S., Limi, M. A., Zani, M., & Abdullah, S. (2023). Penyusunan nilai tukar petani (NTP) pada tanaman pangan dan perkebunan di Kabupaten Konawe Selatan. *Jurnal Ilmiah Penyuluhan dan Pengembangan Masyarakat*, *3*, 49-58. https://doi.org/10.56189/jippm.v3i0.46306

Simangunsong, A. A., Gunawan, I., & Nasution, Z. M. (2022). Pengelompokkan hasil produksi tanaman perkebunan berdasarkan provinsi menggunakan metode k-means. *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, I(4), 2828–9099. https://doi.org/10.55123/jomlai.v1i4.1661

Tamba, I. M. (2025). Determinant of productivity in the Indonesian agricultural sector. *International Journal of Agricultural Economics*, *10*(6), 387-401. https://doi.org/10.11648/j.ijae.20251006.15

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of *cluster*s in a data set via the gap statistic. *Journal of the royal statistical society: series b (statistical methodology)*, *63*(2), 411-423.

Triwidia, E., Nuraini, I., Boedirochminarni, A., & Firmansyah, A. (2024). Analisis pengaruh produktivitas padi, indeks harga yang dibayar petani dan produksi padi terhadap kesejahteraan petani di Indonesia. *JSHP: Jurnal Sosial Humaniora dan Pendidikan*, *8*(1). https://jurnal.poltekba.ac.id/index.php/jsh/article/view/2086

Utami, I. T., & Ispriyanti, D. (2023). *K-Means cluster* count optimization with silhouette index validation and Davies Bouldin index (Case study: Coverage of pregnant women, childbirth, and postpartum health services in Indonesia in 2020). *Barekeng: Jurnal Ilmu Matematika dan Terapan*, *17*(2), 707–716. https://doi.org/10.30598/barekengvol17iss2pp0707-0716

Wahyuni, S., Ilham, N., Sahyuti, Azis, M., Suharyono, S. (2025). The strategy of enhancing plantation production in the food estate area of Central Kalimantan. In: Nawawi, Alami, A.N., Bautista, J., Aung-Thwin, M., Simandjuntak, D., Prasetyo, Y.E. (eds) *Managing Disruption and Developing Resilience for a Better Southeast Asia*. SEASIA 2022. Springer Proceedings in Humanities and Social Sciences. Springer, Singapore. https://doi.org/10.1007/978-981-96-2116-3_24

Wright, K. (2022). Will the real hopkins statistic please stand up? *The R Journal*, *14*(3), 282–292. https://doi.org/10.32614/RJ-2022-055

Yozza, H., Mukhlis, A., & Maiyastri. (2024). *Clustering* provinces in Indonesia based on the main food crop production using the Spatial Fuzzy C-Means. *EKSAKTA: Berkala Ilmiah Bidang MIPA*, *25*(04), 498–507. https://doi.org/10.24036/eksakta/vol25-iss04/385

Yusman, A., & Berliana, S. M. (2025). *Unveiling regional disparities in Indonesia: Clustering provinces by development indicators*. Proceedings of the International Conference on Data Science and Official Statistics. https://doi.org/10.34123/icdsos.v2025i1.645