

PROBABILISTIC MODELING OF INDONESIA'S PAJALE PRODUCTION SYSTEM

PEMODELAN PROBABILISTIK SISTEM PRODUKSI PAJALE INDONESIA

Restu Puji Hidayat

Pusat Sosial Ekonomi dan Kebijakan Pertanian, Indonesia
corresponding author: restupuji@pertanian.go.id

Copyright © 2026 Restu Puji Hidayat. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Production structures and temporal dynamics of Indonesia's rice, maize, and soybean (Pajale) production system were analyzed using a probabilistic data-driven approach based on Gaussian Mixture Model (GMM). Harmonized provincial annual panel data on production and harvested area from 2005–2024 were used to identify latent production structures, temporal dynamics, uncertainty patterns, and regional efficiency differences. The analytical framework combined data harmonization, probabilistic clustering, temporal dynamics analysis, uncertainty assessment, and productivity-based regional typology classification. Commodity Production Signature (CPS) was employed to represent probabilistic relative structural contributions among commodities within the production system. The results identified four major production typologies with clear regional differentiation and high cluster stability. Temporal analysis revealed strong persistence and limited inter-cluster transitions, with transition patterns significantly different from random expectations, indicating structured temporal dynamics among production typologies. Integrated CPS and productivity analysis identified efficient core regions, structurally strong regions with improvement potential, emerging regions, and developing regions. A typology-based approach has the potential to support more adaptive agricultural development strategies aligned with regional characteristics.

Keywords: Agricultural Data Science, Commodity Production Signature, Gaussian Mixture Model, Probabilistic Clustering, Rice-Maize-Soybean Production System

Abstrak

Struktur produksi dan dinamika temporal sistem produksi padi, jagung, dan kedelai (Pajale) di Indonesia dianalisis menggunakan pendekatan probabilistik berbasis data melalui *Gaussian Mixture Model* (GMM). Data panel tahunan tingkat provinsi berupa produksi dan luas panen periode 2005–2024 digunakan untuk mengidentifikasi struktur produksi laten, dinamika temporal, pola ketidakpastian, dan perbedaan efisiensi wilayah. Kerangka analisis mengintegrasikan harmonisasi data, *clustering* probabilistik, analisis dinamika temporal, analisis ketidakpastian, serta klasifikasi tipologi wilayah berbasis produktivitas. *Commodity Production Signature* (CPS) digunakan untuk menggambarkan kontribusi struktural relatif antarkomoditas secara probabilistik dalam sistem produksi. Hasil analisis mengidentifikasi empat tipologi utama sistem produksi dengan perbedaan wilayah yang jelas dan stabilitas kluster yang tinggi. Analisis temporal menunjukkan persistensi yang kuat dan transisi antarkluster yang terbatas, dengan pola transisi yang berbeda secara signifikan.

dari ekspektasi acak, sehingga mengindikasikan adanya dinamika temporal yang terstruktur antartipologi produksi. Integrasi CPS dan produktivitas berhasil mengidentifikasi wilayah inti efisien (efficient core regions), wilayah dengan struktur kuat namun memerlukan peningkatan (structurally strong regions with improvement potential), wilayah berkembang (emerging regions), dan wilayah yang memerlukan penguatan lebih lanjut (developing regions). Pendekatan berbasis tipologi wilayah berpotensi mendukung pengembangan strategi pembangunan pertanian yang lebih adaptif dan sesuai dengan karakteristik regional.

Kata Kunci: Clustering Probabilistik, Commodity Production Signature, Data Sains Pertanian, Gaussian Mixture Model, Sistem Produksi Padi-Jagung-Kedelai

1. Pendahuluan

Sektor pertanian berperan penting dalam mendukung ketahanan pangan, stabilitas ekonomi, dan kesejahteraan masyarakat (Dermoredjo *et al.*, 2024). Komoditas padi, jagung, dan kedelai (Pajale) di Indonesia menjadi komponen utama sistem pangan nasional, dengan padi sebagai indikator utama ketahanan pangan (Pusat Data dan Sistem Informasi Pertanian, 2020; Rozaki *et al.*, 2023). Dalam konteks ketahanan pangan berkelanjutan, sistem produksi pangan menghadapi tantangan heterogenitas wilayah akibat variasi agroekologi, iklim, teknologi, dan faktor sosial-ekonomi (Food and Agriculture Organization, 2022; Reidsma *et al.*, 2023; Rulinawaty *et al.*, 2023; Hidayat, 2025).

Analisis sistem produksi pangan masih didominasi indikator berbasis produksi absolut, sementara integrasi struktur produksi, dinamika temporal, dan efisiensi dalam satu kerangka analisis masih terbatas (Jones *et al.*, 2017; Wu *et al.*, 2026). Meskipun demikian, pendekatan berbasis data multikomoditas berpotensi menangkap pola struktural dan dinamika produksi secara lebih komprehensif (Du, Yao and Hou, 2022; Becker-Reshef *et al.*, 2023; Yozza, Mukhlis and Maiyastri, 2024).

Penelitian ini menggunakan *Gaussian Mixture Model* (GMM) untuk mengidentifikasi struktur laten (tersembunyi) sistem produksi berdasarkan distribusi produksi dan luas panen sebagai pendekatan probabilistik yang mampu menangani ketidakpastian (Sadeghi, 2025), dengan standarisasi fitur untuk meningkatkan kualitas *clustering* (Wongoutong, 2024). Setelah struktur kluster probabilistik diperoleh, *Commodity Production Signature* (CPS) digunakan untuk menggambarkan kontribusi struktural relatif antarkomoditas terhadap pola laten produksi. Selanjutnya, dinamika temporal dianalisis menggunakan *persistence* dan *transition matrix* untuk menggambarkan stabilitas dan perubahan kluster antarperiode pengamatan. Secara konseptual, pendekatan ini digunakan untuk mengevaluasi dinamika perpindahan status dari waktu ke waktu (Stokes, 2024; Huzsvai *et al.*, 2025). Efisiensi produksi dianalisis menggunakan indikator produktivitas sebagai analisis lanjutan (Algarni *et al.*, 2025). Penelitian ini bertujuan untuk mengidentifikasi struktur produksi serta menganalisis dinamika temporal dan efisiensi sistem produksi Pajale

secara terintegrasi guna mendukung pengambilan keputusan pertanian berbasis data (Hidayat *et al.*, 2024).

Pendekatan yang digunakan yaitu mengintegrasikan identifikasi dominansi produksi, dinamika temporal, dan efisiensi dalam satu kerangka analisis multikomoditas. Penggunaan CPS memungkinkan penggambaran struktur kontribusi antarkomoditas secara lebih komprehensif dibandingkan pendekatan yang hanya menggunakan besaran produksi absolut. Integrasi GMM, CPS, dinamika temporal, dan tipologi produktivitas memberikan pendekatan yang lebih komprehensif dalam memahami heterogenitas regional dan dinamika temporal sistem produksi Pajale.

2. Metodologi

2.1 Data dan *Preprocessing*

Data produksi dan luas panen komoditas Pajale tingkat provinsi di Indonesia periode 2005–2024 digunakan dalam bentuk panel tahunan dengan unit analisis provinsi–tahun (Badan Pusat Statistik, 2025; Kementerian Pertanian Republik Indonesia, 2025).

Audit integritas data dilakukan melalui klasifikasi data menjadi *original*, *back-casted*, *imputed/interpolated*, dan *not applicable (pre-birth)*. Data *back-casted* direkonstruksi akibat pemekaran wilayah menggunakan proporsi kontribusi produksi dan luas panen pada tahun pertama ketika wilayah hasil pemekaran tersedia secara bersamaan dalam dataset, sedangkan *missing value* ditangani menggunakan interpolasi linier yang secara terbatas untuk menjaga konsistensi panel data, sehingga tidak dimaksudkan untuk menggambarkan secara penuh dinamika produksi pertanian yang bersifat nonlinear.

Dataset strategis kemudian dibentuk melalui *threshold-based filtering* sebelum dilakukan konstruksi fitur dan pemodelan GMM pada observasi panel provinsi–tahun (Han, Kamber and Pei, 2012). Kontribusi provinsi dihitung berdasarkan proporsi produksi dan luas panen terhadap total nasional, yang dirumuskan sebagai:

$$Contribution_p = \frac{Contribution_p^{prod} + Contribution_p^{area}}{\sum_{p=1}^P (Contribution_p^{prod} + Contribution_p^{area})} \quad (1)$$

Provinsi dikategorikan sebagai wilayah strategis apabila memenuhi:

$$Contribution_p \geq 0.05\% \quad (2)$$

Threshold 0,05% digunakan untuk mempertahankan wilayah dengan kontribusi yang relevan terhadap sistem produksi nasional Pajale dan diuji melalui analisis sensitivitas untuk memastikan stabilitas komposisi wilayah strategis.

2.2 Pemodelan Struktur Produksi dan *Commodity Production Signature (CPS)*

Struktur laten sistem produksi diidentifikasi menggunakan *Gaussian Mixture Model* (GMM) pada observasi panel provinsi–tahun (Bishop, 2006). Setiap observasi disajikan melalui enam fitur, yaitu produksi dan luas panen padi, jagung, dan kedelai yang telah distandarisasi menggunakan *z-score normalization* sebelum pemodelan GMM untuk menghindari bias perbedaan skala. GMM memodelkan distribusi data sebagai kombinasi beberapa distribusi Gaussian:

$$p(x) = \sum_{k=1}^K \pi_k N(x|\mu_k, \Sigma_k) \quad (3)$$

Parameter model diestimasi menggunakan algoritma *Expectation–Maximization (EM)* (Dempster, Laird and Rubin, 1977), sehingga menghasilkan *soft clustering* berupa probabilitas keanggotaan observasi terhadap setiap kluster. Jumlah kluster optimal ditentukan menggunakan *Bayesian Information Criterion (BIC)*, *Davies–Bouldin Index (DBI)*, dan *Silhouette Score* (Schwarz, 1978; Davies and Bouldin, 1979; Rousseeuw, 1987).

Setelah setiap observasi ditetapkan ke dalam kluster berdasarkan probabilitas keanggotaan tertinggi hasil GMM (*hard assignment*), *Commodity Production Signature (CPS)* dihitung untuk menggambarkan kontribusi struktural relatif masing-masing komoditas terhadap komponen kluster terkait:

$$CPS_{j,p,t} = \frac{\phi(z_{j,p,t}; \mu_{k,j}, \sigma_{k,j})}{\sum_{j=1}^3 \phi(z_{j,p,t}; \mu_{k,j}, \sigma_{k,j})} \times 100 \quad (4)$$

dengan j sebagai komoditas Pajale, p menyatakan provinsi, t menyatakan tahun, dan k menyatakan komponen kluster GMM. Densitas Gaussian dihitung secara univariat pada masing-masing fitur produksi terstandarisasi untuk menggambarkan kontribusi struktural relatif setiap komoditas terhadap kluster GMM. Berbeda dengan pendekatan berbasis proporsi produksi, CPS mempertimbangkan karakteristik distribusi masing-masing komoditas dalam struktur kluster probabilistik hasil GMM. Pendekatan ini memungkinkan identifikasi kontribusi struktural relatif antarkomoditas secara lebih adaptif terhadap pola laten distribusi data, sehingga tidak hanya menggambarkan dominansi produksi secara absolut.

2.3 Analisis Dinamika Temporal dan Ketidakpastian

Dinamika temporal dianalisis berdasarkan hasil *clustering* untuk mengidentifikasi stabilitas dan perubahan struktur produksi antarwaktu menggunakan *persistence rate* dan *transition matrix* (Stokes, 2024; Huzsvai et al., 2025). *Persistence rate* dihitung sebagai proporsi observasi provinsi yang mempertahankan keanggotaan kluster dibandingkan tahun sebelumnya pada setiap periode pengamatan:

$$Persistence = \frac{(\sum_{p=1}^P I(c_{p,t} = c_{p,t+1}))}{P} \quad (5)$$

Sedangkan *transition matrix* menggambarkan probabilitas perpindahan antarkluster:

$$T_{ij} = \frac{N_{ij}}{\sum_{j=1}^K N_{ij}} \quad (6)$$

dengan N_{ij} sebagai jumlah observasi yang berpindah dari kluster i ke j dan K sebagai jumlah kluster.

Analisis ketidakpastian juga dilakukan menggunakan probabilitas keanggotaan hasil GMM untuk mengevaluasi tingkat keyakinan klasifikasi (Bishop, 2006). Observasi dikategorikan sebagai *low-confidence membership* apabila probabilitas tertinggi keanggotaannya pada suatu kluster kurang dari 0.80. Analisis sensitivitas terhadap *threshold low-confidence membership* dilakukan menggunakan beberapa nilai ambang (0.60, 0.70, 0.80, dan 0.90) untuk mengevaluasi stabilitas identifikasi observasi dengan ketidakpastian keanggotaan kluster. Tingkat ketidakpastian wilayah dihitung sebagai proporsi observasi *low-confidence membership* selama periode pengamatan untuk mengidentifikasi wilayah dengan kemungkinan zona transisi struktur produksi.

2.4 Analisis Produktivitas dan Klasifikasi Tipologi

Efisiensi produksi dianalisis menggunakan indikator produktivitas sebagai rasio antara produksi dan luas panen. Analisis ini dilakukan setelah identifikasi struktur produksi melalui GMM untuk menjaga pemisahan konseptual antara struktur produksi dan efisiensi. Hubungan antara CPS dan produktivitas dianalisis untuk mengidentifikasi keterkaitan struktur komoditas dan efisiensi wilayah, serta didukung melalui pemetaan produktivitas guna melihat variasi antarwilayah. CPS *strength* didefinisikan sebagai nilai CPS maksimum antar komoditas pada setiap observasi.

Struktur produksi berbasis CPS dan efisiensi berbasis produktivitas selanjutnya diintegrasikan untuk menghasilkan tipologi wilayah:

$$Typology_p = f(CPS\ Strength_p, Overall\ Productivity_p) \quad (7)$$

Klasifikasi dilakukan berdasarkan perbandingan nilai CPS *strength* dan produktivitas terhadap rata-rata nasional antarprovinsi. Rata-rata nasional digunakan sebagai titik acuan untuk membandingkan posisi relatif setiap provinsi terhadap kinerja agregat nasional. Pendekatan tersebut dipilih agar memudahkan identifikasi provinsi yang berada di atas maupun di bawah kinerja rata-rata sistem produksi pada tingkat nasional. Dengan demikian, tipologi yang dihasilkan menggambarkan posisi komparatif wilayah dalam sistem produksi Pajale. Wilayah dengan CPS dan produktivitas di atas rata-rata dikategorikan sebagai *efficient core regions*, sedangkan wilayah dengan CPS tinggi tetapi produktivitas rendah dikategorikan sebagai *structurally strong with improvement potential*. Wilayah dengan CPS rendah tetapi produktivitas tinggi dikategorikan sebagai *emerging regions*,

sementara wilayah dengan CPS dan produktivitas di bawah rata-rata dikategorikan sebagai *developing regions*.

3. Hasil dan Pembahasan

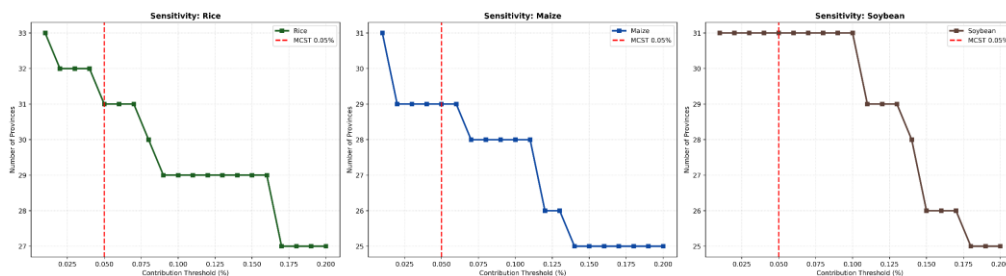
3.1 Kualitas Data dan Validasi Model

Hasil audit integritas data menunjukkan bahwa dataset strategis didominasi oleh data original sebesar 94.23%, sedangkan proporsi data *back-casted* dan *imputed/interpolated* relatif kecil (Tabel 1). Selain itu, kategori *Not Applicable (Pre-Birth)* tidak muncul pada dataset strategis, yang menunjukkan bahwa *preprocessing* berhasil menghasilkan dataset yang lebih bersih dan relevan.

Tabel 1: Audit integritas data

Kategori	Jumlah Data Global	Persentase Global (%)	Jumlah Data Strategis	Persentase Strategis (%)
<i>Back-casted</i>	204	8.95	86	4.73
<i>Imputed / Interpolated</i>	29	1.27	19	1.04
<i>Not Applicable (Pre-Birth)</i>	144	6.32	0	0.00
Original	1,903	83.46	1,715	94.23
Total	2,280	100.00	1,820	100.00

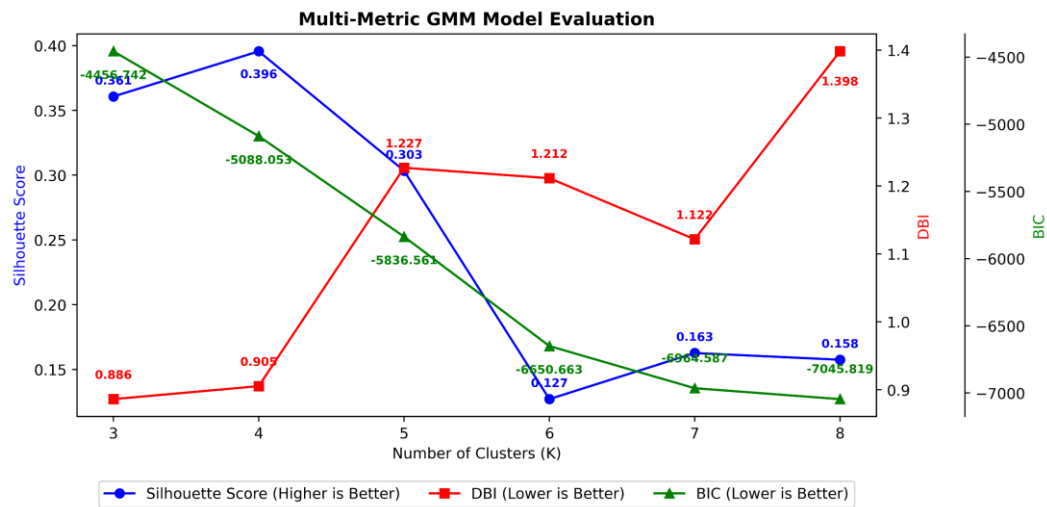
Analisis sensitivitas menghasilkan bahwa jumlah provinsi strategis relatif stabil pada threshold rendah hingga menengah sebelum menurun lebih tajam pada threshold yang lebih tinggi (Gambar 1). Threshold 0.05% dipilih karena mampu mempertahankan wilayah strategis utama sekaligus mengurangi noise dari wilayah dengan kontribusi sangat kecil. Analisis sensitivitas pada ketiga komoditas dilakukan untuk memastikan konsistensi threshold terhadap perbedaan distribusi produksi masing-masing komoditas.



Gambar 1: Analisis sensitivitas ambang batas kontribusi minimum

Evaluasi Model menggunakan BIC, DBI, dan *Silhouette Score* menunjukkan bahwa $K = 4$ merupakan jumlah kluster optimal (Gambar 2). *Silhouette Score* tertinggi diperoleh pada $K = 4$ sebesar 0.396, sedangkan DBI terendah diperoleh pada $K = 3$ sebesar 0.886 dan meningkat pada jumlah kluster yang lebih besar. Meskipun demikian, $K = 4$ dipilih karena memberikan keseimbangan terbaik antara kompleksitas model, kualitas

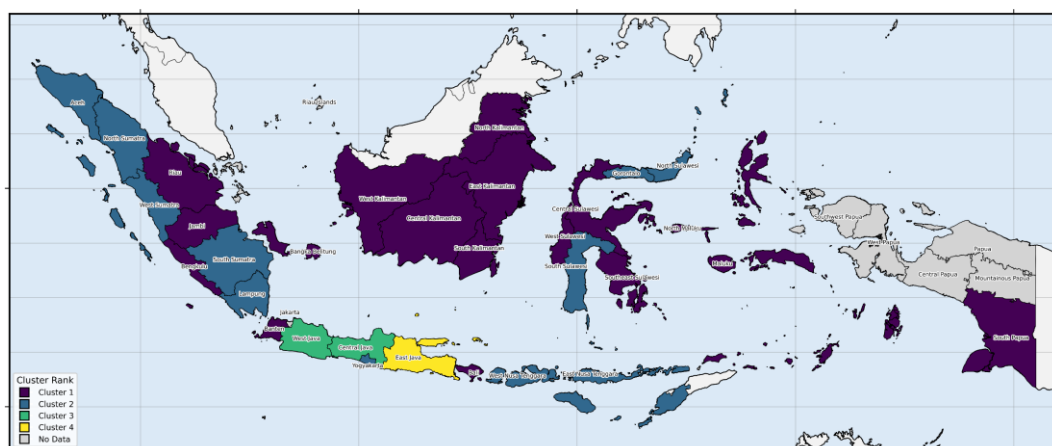
pemisahan kluster, dan homogenitas dalam kluster berdasarkan kombinasi ketiga metrik evaluasi tersebut.



Gambar 2: Evaluasi model

3.2 Struktur Produksi dan *Commodity Production Signature (CPS)*

Clustering GMM mengidentifikasi empat tipologi utama sistem produksi Pajale dengan perbedaan antarwilayah (Gambar 3 dan Tabel 2). Kluster 1 menggambarkan wilayah dengan struktur produksi campuran namun masih menunjukkan dominasi relatif komoditas tertentu, sedangkan Kluster 2 menunjukkan komposisi multikomoditas yang lebih seimbang. Kluster 3 dan Kluster 4 merupakan pusat produksi utama dengan dominasi padi yang kuat. Seluruh kluster memiliki *mean membership probability* sangat tinggi (0.99–1.00), yang menunjukkan bahwa struktur kluster bersifat stabil dengan pemisahan kluster yang signifikan.

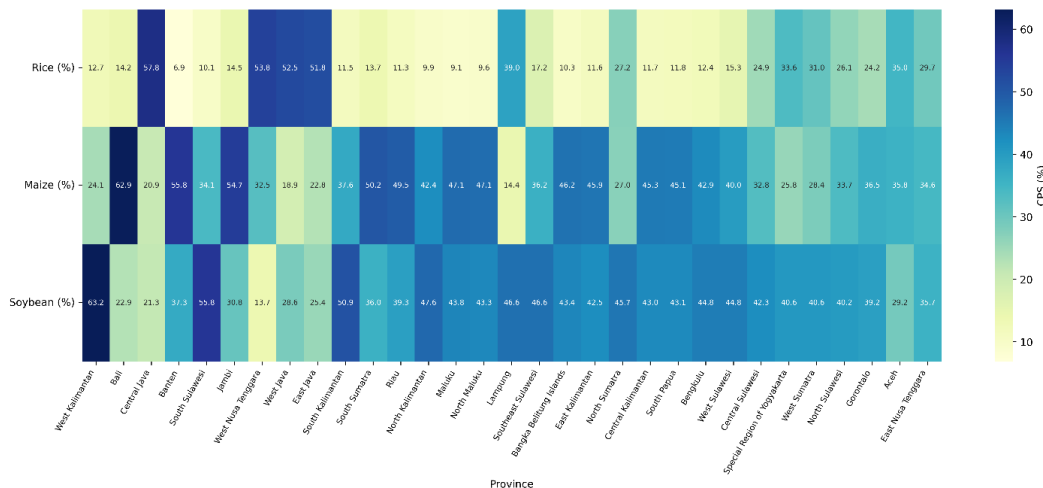


Gambar 3: Peta kluster GMM

Tabel 2: Karakteristik utama kluster GMM

Variabel	Kluster 1	Kluster 2	Kluster 3	Kluster 4
Provinsi	17	11	2	1
<i>Mean membership probability</i>	0.99	1.00	1.00	1.00
Produksi padi (ton)	650,334	1,861,333	10,044,245	11,242,874
Produksi jagung (ton)	50,281	732,132	2,128,897	5,385,217
Produksi kedelai (ton)	2,619	18,157	86,350	309,357
Luas panen padi (ha)	157,149	383,525	1,772,504	1,950,237
Luas panen jagung (ha)	12,973	154,467	376,901	1,188,299
Luas panen kedelai (ha)	2,007	14,091	54,111	214,866
CPS padi (%)	10.94	31.84	56.83	44.96
CPS jagung (%)	46.14	29.64	17.83	30.80
CPS kedelai (%)	42.93	38.51	25.34	24.24
Dominan absolut	Padi	Padi	Padi	Padi
Dominan struktural (CPS)	Jagung	Kedelai	Padi	Padi

Analisis CPS menunjukkan bahwa dominansi produksi absolut tidak secara langsung mencerminkan struktur produksi sebenarnya. Meskipun padi mendominasi secara nasional, kontribusi struktural probabilitas jagung dan kedelai tetap signifikan di banyak wilayah, terutama di luar Jawa. *Heatmap* CPS pada Gambar 4 memperlihatkan bahwa sistem produksi Pajale memiliki karakteristik multikomoditas dengan tingkat heterogenitas yang cukup tinggi antarwilayah.

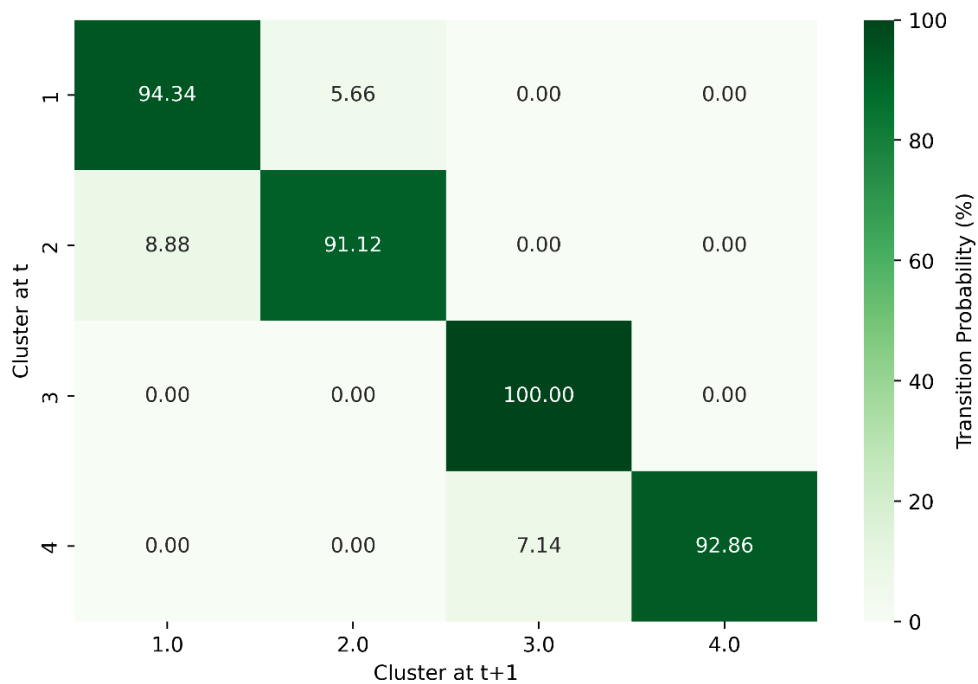
Gambar 4: *Heatmap Commodity Production Signature (CPS)*

3.3 Dinamika Temporal dan Ketidakpastian

Analisis dinamika temporal menunjukkan bahwa sistem produksi Pajale relatif stabil sepanjang periode 2005–2024. Hal ini ditunjukkan oleh *transition matrix* pada Gambar 5, di mana probabilitas bertahan pada masing-masing kluster sangat tinggi, yaitu 94.34% pada Kluster 1, 91.12%

pada Kluster 2, 100% pada Kluster 3, dan 92.86% pada Kluster 4. Transisi antarkluster terjadi secara terbatas dan umumnya berlangsung antarkluster dengan karakteristik produksi yang relatif serupa.

Untuk mengevaluasi pola transisi kluster terjadi secara acak atau tidak, dilakukan uji Chi-Square terhadap matriks transisi. Hasil pengujian menunjukkan distribusi transisi berbeda secara signifikan dari pola independen ($X^2 = 1,554.35$; $p < 0.001$), yang mengindikasikan adanya pola persistensi dan transisi yang terstruktur. Temuan ini konsisten dengan tingginya tingkat persistensi kluster yang diamati pada sebagian besar wilayah.



Gambar 5. *Transition matrix* kluster produksi Pajale.

Analisis ketidakpastian menunjukkan bahwa sebagian besar provinsi memiliki tingkat keanggotaan kluster yang stabil dengan proporsi *low-confidence membership* yang rendah. Ketidakpastian relatif lebih tinggi ditemukan pada beberapa provinsi transisional di Sumatera, Kalimantan, Sulawesi, Bali, dan Yogyakarta, yang mengindikasikan adanya dinamika struktur produksi antarwaktu.

Analisis sensitivitas pada Tabel 3 menunjukkan bahwa proporsi *low-confidence* tetap rendah pada seluruh nilai *threshold* yang diuji. Jumlah provinsi yang teridentifikasi juga relatif stabil sehingga interpretasi ketidakpastian keanggotaan kluster tidak sensitif terhadap pemilihan *threshold*.

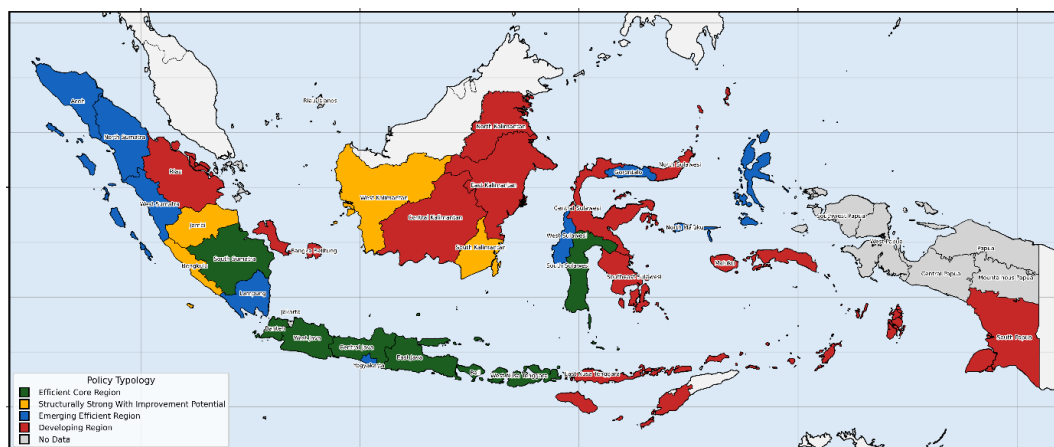
Tabel 3: Analisis sensitivitas threshold low-confidence membership

Threshold	Observasi low-confidence (%)	Jumlah provinsi terdampak
0.6	0.65	4
0.7	0.81	4
0.8	1.61	7
0.9	2.10	8

3.4 Produktivitas, Efisiensi, dan Tipologi Wilayah

Analisis produktivitas menunjukkan bahwa penguatan struktur produksi tidak selalu diikuti oleh peningkatan efisiensi secara linear antarwilayah. Produktivitas padi relatif stabil, sedangkan jagung menunjukkan variasi efisiensi yang lebih tinggi dan kedelai cenderung memiliki produktivitas lebih rendah.

Integrasi CPS dan produktivitas menghasilkan empat tipologi wilayah sebagaimana ditunjukkan pada Gambar 6, yaitu *efficient core regions*, *structurally strong with improvement potential*, *emerging regions*, dan *developing regions*. Wilayah inti efisien terkonsentrasi di kawasan Jawa–Bali–NTB serta beberapa provinsi di Sumatera dan Sulawesi, sedangkan wilayah berkembang cenderung tersebar di berbagai kawasan Indonesia.



Gambar 6. Tipologi wilayah berbasis CPS dan produktivitas.

Hasil ini menunjukkan bahwa sistem produksi Pajale memiliki karakteristik yang beragam antarwilayah, sehingga pendekatan berbasis tipologi berpotensi mendukung pengembangan pertanian yang lebih adaptif melalui penguatan wilayah inti, optimalisasi wilayah potensial, serta penyesuaian strategi antarwilayah berdasarkan struktur produksi dan efisiensi. Klasifikasi tipologi menggunakan rata-rata nasional sebagai titik acuan sehingga hasilnya bersifat relatif terhadap kondisi agregat nasional dan dapat dipengaruhi oleh provinsi dengan skala produksi yang tinggi.

4. Simpulan dan Saran

Sistem produksi Pajale di Indonesia menunjukkan heterogenitas regional yang kuat. Pendekatan *Gaussian Mixture Model* (GMM) dan *Commodity Production Signature* (CPS) mampu mengidentifikasi struktur laten sistem produksi secara komprehensif melalui integrasi produksi dan luas panen. Analisis temporal menunjukkan stabilitas kluster yang tinggi dengan transisi antarkluster yang terbatas dan terstruktur, sedangkan analisis produktivitas mengungkapkan perbedaan efisiensi antarwilayah. Integrasi struktur produksi dan produktivitas menghasilkan tipologi wilayah yang menggambarkan karakteristik utama sistem produksi Pajale di Indonesia.

Pendekatan berbasis tipologi wilayah berpotensi mendukung pengembangan strategi pembangunan pertanian yang lebih adaptif dan sesuai dengan karakteristik regional. *Efficient core regions* memerlukan penguatan keberlanjutan dan stabilitas produksi, *structurally strong with improvement potential regions* memerlukan peningkatan teknologi dan efisiensi pemanfaatan sumber daya, *emerging regions* memiliki peluang penguatan penguatan struktur produksi dan efisiensi, sedangkan *developing regions* memerlukan dukungan kapasitas produksi, infrastruktur, dan akses teknologi.

Penelitian ini menggunakan pendekatan *feature-based probabilistic clustering* pada tingkat provinsi dan belum memodelkan ketergantungan spasial antarwilayah secara eksplisit, sehingga hasil lebih menggambarkan heterogenitas regional daripada hubungan spasial formal. Pengembangan selanjutnya dapat mengintegrasikan data multidimensi, resolusi yang lebih rinci, dan pendekatan statistik spasial. Meskipun uji Chi-Square menunjukkan pola transisi yang berbeda signifikan dari kondisi acak, interpretasi hasil tetap perlu mempertimbangkan keterbatasan jumlah transisi antarkluster, frekuensi harapan pada beberapa sel, serta pilihan parameter yang digunakan dalam analisis.

Daftar Pustaka

- Algarni, S.M. *et al.* (2025) 'Analyzing the impact of climate change on rice production and strategies for enhancing efficiency, sustainability, and global food security', *International Journal of Innovative Research and Scientific Studies*, 8(2), pp. 2946–2957. Available at: <https://doi.org/10.53894/ijirss.v8i2.5888>.
- Badan Pusat Statistik (2025) *Statistik Tanaman Pangan (data produksi dan luas panen padi, jagung, dan kedelai tingkat provinsi)*. Jakarta: Badan Pusat Statistik (BPS).
- Becker-Reshef, I. *et al.* (2023) 'Crop Type Maps for Operational Global Agricultural Monitoring', *Scientific Data*, 10(1), pp. 1–12. Available at: <https://doi.org/10.1038/s41597-023-02047-9>.
- Bishop, C.M. (2006) *Pattern Recognition and Machine Learning*. New York: Springer.

- Davies, D.L. and Bouldin, D.W. (1979) 'A Cluster Separation Measure', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), pp. 224–227. Available at: <https://doi.org/10.1109/TPAMI.1979.4766909>.
- Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) 'Maximum Likelihood from Incomplete Data Via the EM Algorithm', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 39(1), pp. 1–22. Available at: <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.
- Dermoredjo, S.K. *et al.* (2024) 'National food development policies in Indonesia: An analysis of food sustainability and security', *BIO Web of Conferences*. Edited by A. Suryana *et al.*, 119, p. 05006. Available at: <https://doi.org/10.1051/bioconf/202411905006>.
- Du, G., Yao, L. and Hou, D. (2022) 'The spatio-temporal changes of cropping patterns in the black soil area of China: Lessons from Wangkui County', *Frontiers in Environmental Science*, 10(October), pp. 1–15. Available at: <https://doi.org/10.3389/fenvs.2022.981721>.
- Food and Agriculture Organization (2022) *Repurposing Food and Agricultural Policies to Make Healthy Diets More Affordable, The State of Food Security and Nutrition in the World 2022*. Rome. Available at: <https://doi.org/https://doi.org/10.4060/cc0639en>.
- Han, J., Kamber, M. and Pei, J. (2012) *Data Mining: Concepts and Techniques*. Third Edit, *Data Mining: Concepts and Techniques*. Third Edit. Morgan Kaufmann. Available at: <https://doi.org/10.1016/C2009-0-61819-5>.
- Hidayat, R.P. *et al.* (2024) 'Support Vector Regression Algorithm for Predicting Rice Production In West Java Province', in *2024 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*. IEEE, pp. 978–983. Available at: <https://doi.org/10.1109/ICIMCIS63449.2024.10957446>.
- Hidayat, R.P. (2025) 'Regional Stability and Dynamics of Rice Production in West Java through Spatiotemporal Clustering', *Jurnal Masyarakat Informatika*, 16(2), pp. 229–246. Available at: <https://doi.org/10.14710/jmasif.16.2.76056>.
- Huzsvai, L. *et al.* (2025) 'Discrete-Time Markov Chain Method for Predicting Probability of Crop Yield Variability', *Earth*, 6(4), p. 142. Available at: <https://doi.org/10.3390/earth6040142>.
- Jones, J.W. *et al.* (2017) 'Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science', *Agricultural Systems*, 155, pp. 269–288. Available at: <https://doi.org/10.1016/j.agsy.2016.09.021>.

Kementerian Pertanian Republik Indonesia (2025) 'Basis Data Statistik Pertanian (BDSP): Data produksi dan luas panen komoditas pangan'. Kementerian Pertanian Republik Indonesia.

Pusat Data dan Sistem Informasi Pertanian (2020) *Outlook padi komoditas pertanian subsektor tanaman pangan tahun 2020*. Edited by A.A. Susanti and A. Supriyatna. Kementerian Pertanian.

Reidsma, P. *et al.* (2023) 'Alternative systems and strategies to improve future sustainability and resilience of farming systems across Europe: from adaptation to transformation', *Land Use Policy*, 134, p. 106881. Available at: <https://doi.org/https://doi.org/10.1016/j.landusepol.2023.106881>.

Rousseeuw, P.J. (1987) 'Silhouettes: A graphical aid to the interpretation and validation of cluster analysis', *Journal of Computational and Applied Mathematics*, 20, pp. 53–65. Available at: [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7).

Rozaki, Z. *et al.* (2023) 'Chapter Two - Food security, diversification, and inequality: Indonesia in the era of economic recovery and high price trends', in M.J.B.T.-A. in F.S. and S. Cohen (ed.) *Advances in Food Security and Sustainability*. Elsevier, pp. 43–94. Available at: <https://doi.org/https://doi.org/10.1016/bs.af2s.2023.07.007>.

Rulinawaty *et al.* (2023) 'Optimalisasi Kebijakan Pangan Berbasis Big Data: Model Manajemen Produksi dan Distribusi Beras di Indonesia', *Ranah Research : Journal of Multidisciplinary Research and Development*, 5(4), pp. 378–392. Available at: <https://doi.org/10.38035/rrj.v7i3.1593>.

Sadeghi, B. (2025) 'Clustering in geo-data science: Navigating uncertainty to select the most reliable method', *Ore Geology Reviews*, 181, p. 106591. Available at: <https://doi.org/10.1016/j.oregeorev.2025.106591>.

Schwarz, G. (1978) 'Estimating the Dimension of a Model', *The Annals of Statistics*, 6(2). Available at: <https://doi.org/10.1214/aos/1176344136>.

Stokes, J.R. (2024) 'A Markov chain model of crop conditions and intrayear crop yield forecasting', *Journal of Forecasting*, 43(3), pp. 583–592. Available at: <https://doi.org/10.1002/for.3052>.

Wongoutong, C. (2024) 'The impact of neglecting feature scaling in k-means clustering', *PLOS ONE*. Edited by N. Aunsri, 19(12), p. e0310839. Available at: <https://doi.org/10.1371/journal.pone.0310839>.

Wu, Q. *et al.* (2026) 'Exploring the influencing factors and applying practices for the coordinated integration of digital economy and agricultural systems: evidence from China', *Asia-Pacific Journal of Regional Science*, 10(1), p. 19. Available at: <https://doi.org/10.1007/s41685-026-00422-z>.

Yozza, H., Mukhlis, A. and 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), pp. 498–507. Available at: <https://doi.org/10.24036/eksakta/vol25-iss04/385>.