

CLUSTERING POTENSI PADI DI PROVINSI NUSA TENGGARA TIMUR MENGGUNAKAN *STANDARD K-MEANS* DAN *TRAJECTORY K-MEANS*

Jofri Ardo Tiganna Sembiring¹

¹Badan Pusat Statistik Kabupaten Sumba Barat Daya, Indonesia

*Korespondensi Penulis: jofrisembiring@bps.go.id

ARTIKEL INFO

Abstract

Article history:

Received 20 Oct, 2025

Revised 19 Nov, 2025

Accepted 10 Dec, 2025

Published 31 Dec, 2025

Introduction: The Agricultural sector is the mainstay of the economy of Nusa Tenggara Timur (NTT), contributing 28,87 percent to the regional GDP in 2024. From this amount, the food crop sub-sector contributed 22,94 percent, demonstrating the strategic role of food commodities in the regional economic structure. **Background Problem:** Rice, as the main staple food of the community, has high economic and social value because it directly affects food security and the welfare of the population. **Novelty:** This study aims to cluster districts/cities in NTT based on rice potential. **Research Method:** This analysis used Standard K-Means and Trajectory K-Means methods, utilizing indicators such as harvested area and production. **Result:** The analysis showed that the optimal number of clusters is four. The evaluation showed that Standard K-Means produces better grouping quality, while Trajectory K-Means remains relevant for identifying patterns of change over time. These findings confirm that region grouping based on rice potential can be the basis for more effective, efficient, and sustainable policy-making in strengthening the food sector in NTT.

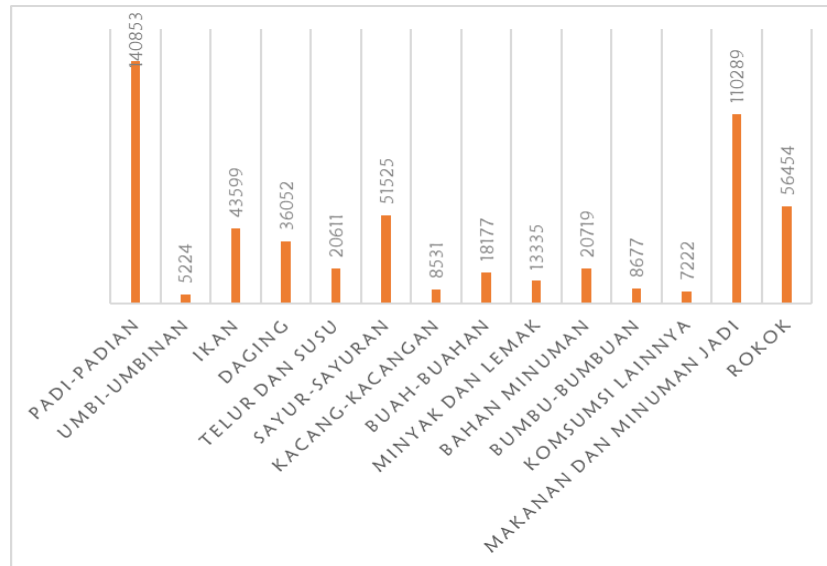
Keywords:

Clustering, K-Means, Trajectory, Rice, Production, Harvested Area, NTT

1. Pendahuluan

Sektor pertanian memiliki potensi yang signifikan terhadap perekonomian Nusa Tenggara Timur (NTT) sampai saat ini. Berdasarkan PDRB, sektor pertanian, kehutanan, dan perikanan memiliki kontribusi terbesar dalam perekonomian NTT sebesar 28,87 persen pada tahun 2024. Dilihat dari subsektor, subsektor tanaman pangan memiliki peranan yang cukup signifikan yakni sebesar 22,94 persen terhadap kategori pertanian, peternakan, pemburuan dan jasa pertanian. Pada tahun 2023 Badan Pusat Statistik melakukan Sensus Pertanian yang bertujuan untuk memenuhi kebutuhan data pertanian tingkat nasional dan internasional.

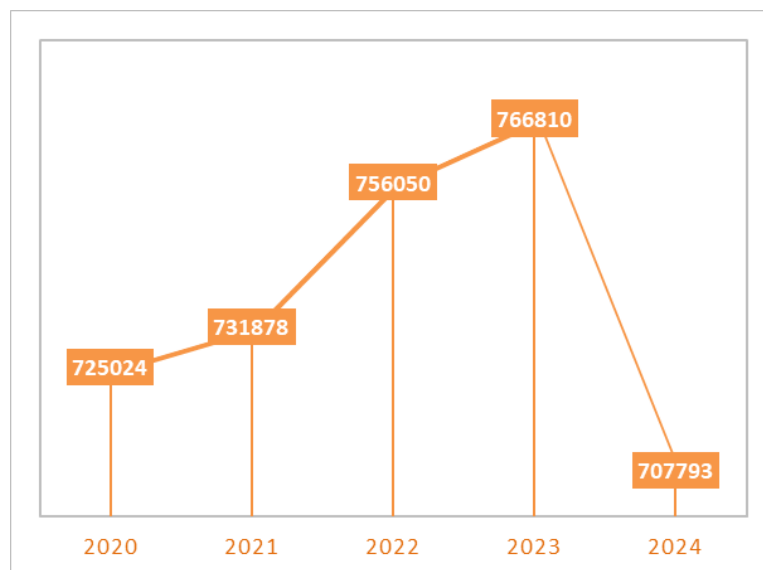
Padi-padian merupakan komoditas pertanian yang paling berperan di NTT. Berdasarkan Survei Sosial Ekonomi Nasional Maret 2024, pengeluaran rata-rata per kapita sebulan menurut kelompok makanan di Provinsi NTT didominasi oleh kelompok padi-padian yaitu sebesar 140.853 rupiah yang dapat dilihat pada Gambar 1.



Sumber : Susenas 2024 Provinsi Nusa Tenggara Timur, BPS Nusa Tenggara Timur

Gambar 1. Pengeluaran Rata-rata per Kapita Sebulan menurut Kelompok Makanan di Provinsi NTT (Rupiah), 2024

Kelompok padi-padian terdiri atas beberapa komoditas, yaitu beras, beras ketan, jagung basah dengan kulit, jagung pipilan, tepung terigu, dan padi-padian lainnya (Direktorat Statistik Kesejahteraan Rakyat, 2023).



Sumber : BPS Nusa Tenggara Timur, diolah

Gambar 2. Produksi Padi di Provinsi NTT 2020–2024 (Ton-GKG)

Padi merupakan salah satu komoditas penting pada subsektor tanaman pangan serta berperan penting dalam kehidupan petani dan masyarakat NTT. Pada Gambar 2 dapat dilihat produksi padi di NTT cukup tinggi pada tahun 2023 sebesar 766.810 ton gabah kering giling. Namun, pada tahun 2024 terjadi penurunan yang sangat signifikan dengan total produksi sebesar 707.793 ton gabah kering giling dan menjadikannya titik terendah dalam lima tahun terakhir.

Swasembada pangan sangat penting untuk menjaga ketersediaan pangan di Indonesia, termasuk juga di NTT. Untuk mencapai hal tersebut perlu dipahami potensi komoditas padi di setiap kabupaten/kota di NTT. Penelitian mengenai potensi padi sudah cukup banyak dilakukan seperti *clustering* potensi padi di beberapa wilayah. Christiani (2024), *clustering* potensi padi menggunakan metode *K-Means* dengan variabel luas panen, produksi, dan produktivitas. *Clustering* dilakukan pada 22 kabupaten/kota dengan menggunakan data tahun 2023 di NTT dan menghasilkan tiga kluster berbeda. Pada kluster 1 terdiri dari lima kabupaten, kluster 2 dengan 10 kabupaten, dan kluster 3 dengan tujuh kabupaten. Penelitian lainnya, Arini (2023), *Clustering* ketahanan pangan metode *K-Means* dengan menggunakan satu periode data 2022 dengan variabel luas panen, produktivitas, dan produksi padi. Ketahanan pangan dapat dikelompokkan dengan 3 kluster di mana pada kluster 1 terdiri dari 23 kabupaten dan kluster 2 dan 3 masing-masing sebanyak sembilan kabupaten dan enam kabupaten.

Penelitian yang mempertimbangkan periode waktu pada potensi padi masih cukup sedikit dilakukan. Oleh karena itu, penelitian ini bertujuan mengelompokkan kabupaten berdasarkan dua indikator utama, yaitu luas panen dan produksi padi gabah kering giling. Analisis dilakukan dengan memperhitungkan dinamika waktu melalui penerapan *Standard K-Means* dan *Trajectory K-Means*. Penelitian ini juga membandingkan pengelompokan hasil *clustering* antar *Standard K-Means* dan *Trajectory K-Means*.

2. Metodologi

Penelitian ini menggunakan data sekunder hasil pengolahan data Ubinan Padi dan Kerangka Sampel Area tahun 2020–2024. Ubinan Padi merupakan kegiatan BPS yang bertujuan untuk mengukur produktivitas dari suatu komoditas. Sementara itu, Kerangka Sampel Area merupakan kegiatan yang bertujuan untuk mendapatkan luas wilayah pertanian baik panen, tanam, puso dan bukan lahan pertanian. Luas wilayah tersebut diperoleh berdasarkan amatan setiap bulan dari petugas Kerangka Sampel Area di masing-masing kabupaten/kota.

Analisis yang digunakan dalam penelitian ini meliputi analisis deskriptif dan analisis pengelompokan daerah kabupaten/kota yang berpotensi pada komoditas padi dengan menggunakan *Standard K-Means* (Agregat) dan *Trajectory K-Means* (DTW).

2.1. Bahan dan Sumber Data

Variabel yang digunakan dalam penelitian ini terdiri atas

X1 = Luas Panen (hektar)

X2 = Produksi Padi (ton-GKG)

Luas panen merupakan luas wilayah pertanian yang mengalami panen, panen dengan produksi paling sedikit 11 persen dari keadaan normal. Produksi padi merupakan jumlah padi dalam bentuk gabah kering giling yang dihasilkan di kabupaten/kota. Penelitian ini menggunakan periode waktu 2020–2024.

2.2. Metode Penelitian

Dasar *K-Means*

Metode yang digunakan dalam penelitian ini adalah *Standard K-Means Clustering* yang diterapkan pada data panel. Data panel merupakan kombinasi dimensi *cross-section* dan *time-series* (Baltagi, 2025). Pendekatan ini dipilih karena memungkinkan pengelompokan unit analisis dengan mempertimbangkan dinamika waktu.

Algoritma *K-Means* (MacQueen, 1967) merupakan salah satu metode *unsupervised clustering* yang bertujuan mempartisi n observasi ke dalam k kluster, dengan meminimalkan variasi intrakluster dan memaksimalkan variasi antarkluster. Secara matematis, fungsi objektif yang diminimalkan adalah:

$$J = \sum_{i=1}^n \sum_{j=1}^k r_{ij} |x_i - \mu_j|^2 \quad (1)$$

di mana:

- r_{ij} = indikator keanggotaan observasi ke- i pada kluster ke- j ;
- x_i = nilai variabel untuk unit ke- i ;
- μ_j = koordinat *centroid* kluster ke- j pada variabel

Penerapan *K-Means* pada data panel dilakukan dengan merepresentasikan setiap unit analisis sebagai vektor multidimensi yang terdiri atas nilai variabel pada beberapa periode waktu (Bonhomme & Manresa, 2015). Dengan demikian, metode tersebut dapat mengelompokkan unit analisis berdasarkan perkembangan dari waktu ke waktu, bukan hanya pada satu periode.

Pada penelitian ini dilakukan perbandingan antara dua pendekatan, yaitu *Standard K-Means* (Agregat) dan *Trajectory K-Means* (DTW). Pendekatan *Standard K-Means* (Agregat) menggunakan nilai rata-rata tahunan dari setiap variabel, sehingga dapat dibuat fungsi turunan:

$$J_{agregat} = \sum_{i=1}^n \sum_{j=1}^k r_{ij} \sum_{v=1}^p ||\bar{x}_{iv} - \mu_{jv}||^2 \quad (2)$$

di mana:

- r_{ij} = indikator keanggotaan observasi ke- i pada kluster ke- j ;
- \bar{x}_{iv} = nilai rata-rata variabel ke- v untuk unit ke- i ;
- μ_{jv} = koordinat centroid kluster ke- j pada variabel ke- v .

Sementara itu, pada pendekatan *Trajectory K-Means* (DTW), setiap unit analisis digambarkan sebagai vektor deret waktu multidimensi untuk seluruh variabel dan

periode. Melalui pendekatan ini, pola dinamis dari setiap unit analisis dapat dibandingkan menggunakan metode *Dynamic Time Warping* (DTW), dengan fungsi:

$$J_{trajectory} = \sum_{i=1}^n \sum_{j=1}^k r_{ij} DTW(x_i, \mu_j)^2 \quad (3)$$

Transformasi dan Standardisasi Data

Data awal yang digunakan berbentuk panel dengan dimensi kabupaten/kota dan tahun. Pada pendekatan *Standard K-Means* (Agregat), data ditransformasikan ke dalam format *wide* agar sesuai dengan algoritma *K-Means*, sehingga setiap kabupaten/kota direpresentasikan sebagai satu baris dengan kolom berisi nilai rata-rata variabel panen dan produksi padi.

Selanjutnya dilakukan standardisasi menggunakan *z-score* untuk menormalkan skala antar variabel. Pada pendekatan *Trajectory K-Means* (DTW) data tidak perlu ditransformasikan. Standardisasi diperlukan karena perbedaan skala dapat menyebabkan hasil *clustering* tidak optimal. Rumus standardisasi adalah:

$$z_{it} = \frac{x_{it} - \bar{x}_t}{s_t} \quad (4)$$

dengan x_{it} nilai variabel t untuk kabupaten i , \bar{x}_t rata-rata, dan s_t standar deviasi.

Penentuan Jumlah Kluster Optimal

Menentukan jumlah kluster yang optimal merupakan langkah penting dalam penerapan *K-Means*. Dalam penelitian ini digunakan beberapa indeks validitas internal:

- **Calinski–Harabasz Index (CH):** semakin besar nilainya, semakin baik struktur kluster (Caliński & Harabasz, 1974).

$$CH = \frac{B/(k-1)}{W/(n-k)} \quad (5)$$

- **Davies–Bouldin Index (DB):** semakin kecil nilainya, semakin baik pemisahan antarkluster (Davies & Bouldin, 1979).

$$s_j = \frac{1}{n_j} \sum_{i \in C_j} \|x_i - \mu_j\| \quad (6)$$

$$R_{jl} = \frac{s_j + s_l}{\|\mu_j - \mu_l\|} \quad (7)$$

$$DB = \frac{1}{k} \sum_j \max_{l \neq j} R_{jl} \quad (8)$$

- **Ray–Turi Index (RT):** digunakan untuk mengukur perbandingan variasi intraklaster terhadap jarak antarsentroid (Ray & Turi, 1999).

$$RT = \frac{W}{\sum_{j < l} |\mu_j - \mu_l|^2} \quad (9)$$

- **Silhouette Coefficient (s(i)):** mengukur tingkat kemiripan antaranggota dalam satu klaster dibandingkan dengan klaster lainnya. Nilai *Silhouette* berkisar antara -1 hingga 1. Nilai semakin mendekati 1 menunjukkan klaster yang terpisah dengan baik, sedangkan nilai yang mendekati 0 menunjukkan adanya tumpang tindih antarklaster (Rousseeuw, 1987).

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (10)$$

3. Hasil dan Pembahasan

3.1. Analisis Deskriptif

Rata-rata luas panen padi di NTT periode 2020–2024 memperlihatkan pola distribusi yang tidak merata antarkabupaten/kota. Pada Gambar 3 terlihat bahwa kabupaten dengan kontribusi terbesar terhadap luas panen padi berada di wilayah Manggarai Barat, Manggarai, dan Manggarai Timur. Kondisi ini sejalan dengan karakteristik geografis ketiga kabupaten yang relatif memiliki areal pertanian yang luas.



Sumber : BPS Nusa Tenggara Timur, diolah

Gambar 3. Sebaran Rata-rata Luas Panen Padi di Provinsi NTT 2020–2024

Di sisi lain, terdapat kelompok kabupaten yang memiliki luas panen di atas rata-rata provinsi, meskipun tidak sebesar kelompok teratas. Kabupaten-kabupaten ini tetap memberikan kontribusi signifikan terhadap total luas panen di NTT, namun variasinya lebih beragam dan cenderung dipengaruhi oleh kondisi geografis maupun pola pemanfaatan lahan.

Sementara itu, beberapa wilayah lain memiliki luas panen rendah dibandingkan kabupaten lainnya seperti Alor, Lembata, Sabu Raijua, beserta sejumlah wilayah dengan kapasitas lahan yang terbatas menunjukkan kontribusi relatif rendah terhadap total luas panen provinsi NTT. Hal ini dapat dihubungkan dengan keterbatasan lahan subur, topografi yang lebih menantang, serta adanya tekanan dari sektor non-pertanian yang bersaing dalam penggunaan lahan.



Sumber : BPS Nusa Tenggara Timur, diolah

Gambar 4. Sebaran Rata-rata Produksi Padi di Provinsi NTT 2020–2024

Jika dilihat dari sisi produksi padi, pola distribusi antarkabupaten/kota di NTT juga menunjukkan kecenderungan yang serupa dengan rata-rata luas panen. Pada Gambar 4, Kabupaten Manggarai Barat, Manggarai, dan Manggarai Timur menempati posisi dengan kontribusi produksi terbesar, sejalan dengan dominasi luas lahan yang dimiliki. Namun, tidak semua kabupaten dengan lahan panen tinggi mampu menghasilkan produksi tinggi. Terdapat wilayah yang memiliki luas lahan di atas rata-rata, tetapi tingkat produksinya tidak sebanding, yang mengindikasikan adanya perbedaan dalam tingkat produktivitas dan efisiensi penggunaan lahan. Sebaliknya, beberapa kabupaten dengan luasan lahan relatif kecil, seperti Belu atau Sikka, tetap mampu menjaga kontribusi produksi pada level moderat, yang menunjukkan pemanfaatan lahan lebih intensif. Secara keseluruhan, kondisi ini memperlihatkan bahwa ketersediaan lahan merupakan faktor penting, tetapi bukan

satu-satunya penentu produksi, karena teknologi budidaya, pola tanam, dan kualitas pengelolaan juga memainkan peran krusial.

3.2. Evaluasi Jumlah *Klaster*

Hasil evaluasi model *clustering* dengan tiga indeks, yaitu CH, DBI, dan RTI, menunjukkan variasi kualitas hasil pada jumlah klaster yang berbeda. Berdasarkan Tabel 1, pada skenario dua klaster, nilai CH mencapai 58,28 dengan DBI sebesar 0,497 dan RTI sebesar 1,638. Meskipun nilai CH cukup tinggi, nilai RTI yang besar mengindikasikan bahwa pemisahan antarklaster masih belum optimal. Pada skenario tiga klaster, nilai CH menurun menjadi 56,50, DBI stabil pada 0,494, namun RTI turun signifikan ke 0,815. Hal ini menandakan adanya perbaikan dalam kedekatan internal antaranggota klaster meskipun pemisahan antarklaster tidak sebaik model dua klaster.

Tabel 1. *Output* Evaluasi Jumlah Klaster dengan Empat Indeks

klaster	Calinski Harabasz	Davies Bouldin	Ray Turi	Silhouette
2	58,28	0,497	11,638	0.650
3	56,50	0,494	0,815	0.605
4	91,71	0,503	0,353	0.554

Skenario empat klaster menghasilkan nilai CH tertinggi yaitu 91,71, DBI meningkat ke 0,503. Namun, RTI justru turun drastis ke 0,353, yang menunjukkan kualitas pemisahan klaster yang lebih baik dan kedekatan internal yang semakin kuat. Berdasarkan teori, nilai CH yang tinggi mengindikasikan klaster yang lebih terpisah dengan baik (Caliński & Harabasz, 1974), sementara DBI yang rendah merefleksikan bahwa klaster lebih kompak dan berbeda satu sama lain (Davies & Bouldin, 1979). Di sisi lain, RTI yang lebih kecil menandakan struktur klaster semakin optimal dalam hal kepadatan internal dibandingkan jarak antarklaster (Ray & Turi, 1999). Sementara itu, indeks *Silhouette*, nilai pada setiap klaster masih dikategorikan sebagai *good clustering structure* karena berada pada rentang nilai 0.5–0.7 (Kaufman & Rousseeuw, 1990). Selain mempertimbangkan indeks kuantitatif, pemilihan klaster juga mempertimbangkan aspek parsimoni dan interpretabilitas model. Tibshirani dkk, (2001) menjelaskan dalam pemilihan jumlah klaster harus mempertimbangkan keseimbangan dan kesederhanaan model dan kemampuan klaster untuk merepresentasikan struktur data sebenarnya. Penggunaan klaster yang sedikit dan besar dapat menyembunyikan pola penting data dan pembagian data yang tidak bermakna.

Dengan demikian, hasil ini menunjukkan bahwa pengelompokan dengan empat klaster memberikan kombinasi terbaik antara pemisahan antar klaster dan kekompakan internal, sehingga dapat dianggap sebagai model yang paling optimal dalam menggambarkan potensi padi di Provinsi NTT.

3.3. *Standard K-Means* (Agregat)

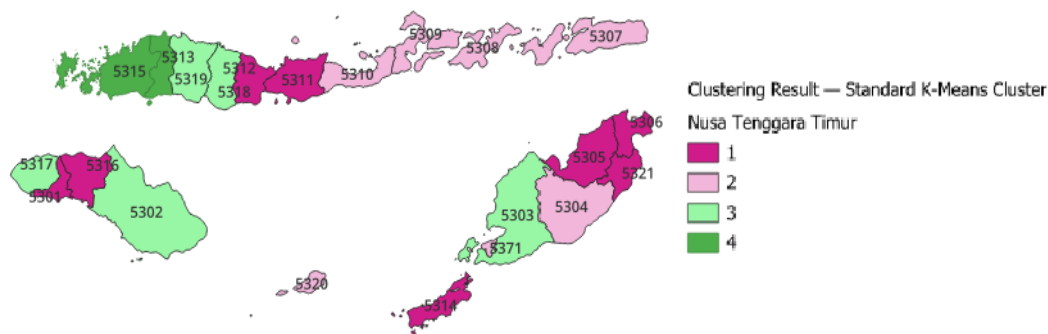
Hasil analisis klaster dengan metode *Standard K-Means* (Agregat) pada Tabel 2 menunjukkan bahwa pembagian data ke dalam empat kelompok mampu menjelaskan sebesar 93,9 persen dari total variasi. Nilai ini menandakan bahwa struktur data sudah dapat dipetakan dengan baik melalui proses pengelompokan

tersebut. Hal ini menunjukkan bahwa hampir seluruh variasi yang terdapat di dalam data berhasil diakomodasi oleh pemisahan ke dalam empat klaster. Dengan demikian, hasil *clustering* ini dapat dianggap cukup kuat serta representatif dalam menggambarkan potensi padi dan dapat menjadi dasar yang valid untuk interpretasi lebih lanjut maupun pengambilan keputusan berbasis klaster.

Tabel 2. Hasil *Standard K-Means Clustering* (Agregat) Berdasarkan *Within-Cluster Sum of Squares*

Klaster	Rata-rata Luas Panen	Rata-rata Produksi		
1	-0,2448241	-0,2723692		
2	-0,9710373	-0,9057255		
3	0,9457069	0,7991196		
4	2,0136597	2,2617168		
Clustering vector: [1] 2 1 1 2 2 3 2 1 4 4 3 1 3 1 2 2 1 3 1 3 2 1				
Within cluster sum of squares by cluster:				
[1]	0,4828680	0,5015757	1,0643647	0,5302901
(between SS / total SS = 93,9 %)				

Hasil analisis *Standard K-Means Clustering* (Agregat) dengan empat klaster menghasilkan distribusi ukuran klaster yang cukup seimbang, yakni klaster 1 beranggotakan delapan kabupaten, klaster 2 beranggotakan tujuh kabupaten, klaster 3 terdiri dari lima kabupaten, dan klaster 4 mencakup dua kabupaten.



Gambar 5. Peta *Clustering* Potensi Padi di Provinsi Nusa Tenggara Timur dengan *Standard K-Means* (Agregat)

Klaster 1 adalah klaster dengan nilai rata-rata luas panen dan produksi sedikit di bawah rata-rata keseluruhan yang menunjukkan bahwa wilayah pada klaster 1 memiliki potensi produksi dan luas panen yang relatif sedang. Kabupaten yang berada di klaster 1 adalah Sumba Barat (5301), Timor Tengah Utara (5305), Belu (5306), Ende (5311), Rote Ndao (5314), Sumba Tengah (5316), Nagekeo (5318) dan Malaka (5321). Klaster 2 adalah klaster dengan nilai rata-rata luas panen dan produksi paling rendah. Hal ini menunjukkan bahwa wilayah pada klaster 2 memiliki potensi produksi dan luas panen yang rendah. Kabupaten yang berada di klaster 2 adalah Timor Tengah Selatan (5304), Alor (5307), Lembata (5308), Flores Timur (5309), Sikka (5310), Sabu Raijua (5320), dan Kota Kupang (5371).

Sebaliknya, klaster 3 menunjukkan nilai positif dengan rata-rata luas panen dan produksi yang lebih tinggi dibandingkan dengan klaster 1 ataupun 2. Kabupaten yang berada di klaster 3 adalah Sumba Timur (5302), Kupang (5303), Ngada (5312), Sumba Barat Daya (5317), Manggarai Timur (5319). Klaster yang memiliki potensi luas panen dan produksi paling besar adalah klaster 4 karena memiliki nilai paling besar yang mengidentifikasi daerah tersebut memiliki potensi pertanian padi paling unggul di NTT. Kabupaten yang berada pada klaster 4 adalah Manggarai (5313) dan Manggarai Barat (5315).

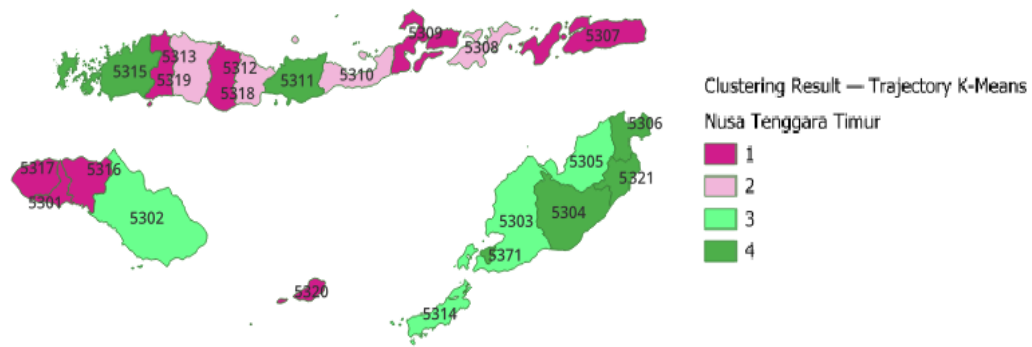
3.4. *Trajectory K-Means (DTW)*

Trajectory K-Means dengan empat klaster menggunakan jarak *Dynamic Time Warping (DTW)* dan *centroid* berbasis *Dynamic Barycenter Averaging (DBA)* menunjukkan struktur pengelompokan yang cukup jelas berdasarkan kesamaan pola perubahan waktu. Berdasarkan hasil *clustering*, ukuran masing-masing klaster bervariasi, yaitu klaster 1 memiliki delapan anggota dengan nilai *Av_Dist* sebesar 7,07. Meskipun merupakan klaster terbesar, nilai jaraknya relatif tinggi, yang menunjukkan bahwa anggota di klaster ini memiliki variasi pola yang cukup besar atau heterogen. Klaster 2 berisi empat anggota dengan *Av_Dist* sebesar 5,71, nilai terkecil di antara semua klaster. Hal ini menunjukkan bahwa klaster 2 memiliki tingkat kesamaan pola yang paling tinggi dan memiliki *trajektori* yang paling mirip satu sama lain. Klaster 3 juga memiliki empat anggota dengan *Av_Dist* sebesar 6,83, menunjukkan tingkat kesamaan sedang. Pola dalam klaster ini cukup mirip, tetapi tidak sekompak klaster 2. Klaster 4 terdiri dari enam anggota dengan *Av_Dist* sebesar 7,14, nilai tertinggi di antara semua klaster, sehingga dapat dikatakan bahwa anggota klaster ini memiliki variasi pola *trajektori* paling beragam. Secara keseluruhan, hasil ini mengindikasikan bahwa klaster 2 adalah yang paling kompak dan seragam, sedangkan klaster 4 paling beragam.

Tabel 3. Hasil *Trajectory K-Means Clustering* Berdasarkan *Average Intra-Cluster Distance*

Klaster	Size	Av_Dist
1	8	7,072563
2	4	5,706628
3	4	6,832918
4	6	7,143947

Gambar 6 menunjukkan hasil *clustering* dengan metode *Trajectory K-Means*. Kabupaten Sumba Barat (5301), Alor (5307), Flores Timur (5309), Ngada (5312), Manggarai (5313), Sumba Tengah (5316), Sumba Barat Daya (5317), dan Sabu Raijua (5320) berada pada klaster 1 yang memiliki variasi pola yang cukup besar atau heterogen. Kabupaten yang berada di klaster 2 ada Lembata (5308), Sikka (5310), Nagekeo (5318), dan Manggarai Timur (5319). Klaster 2 terdiri dari kabupaten yang memiliki pola paling tinggi dan paling mirip antar kabupaten dibandingkan klaster lainnya. Sedangkan klaster yang memiliki tingkat kesamaan sedang adalah klaster 3 yang terdiri dari Kabupaten Sumba Timur (5302), Kupang (5303), Timor Tengah Utara (5305), Rote Ndao (5314). Sebaliknya, klaster dengan tingkat kesamaan paling rendah adalah klaster 4 yaitu Kabupaten Timor Tengah Selatan (5404), Belu (5306), Ende (5311), Manggarai Barat (5315), Malaka (5321), dan Kota Kupang (5371).



Gambar 6. Peta *Clustering* Potensi Padi di Provinsi Nusa Tenggara Timur dengan *Trajectory K-Means* (DTW)

3.5. Perbandingan *Standard K-Means* (Agregat) dan *Trajectory K-Means* (DTW)

Berdasarkan Tabel 4, hasil evaluasi *Standard K-Means* (Agregat) dan *Trajectory K-Means* (DTW), terlihat adanya perbedaan yang cukup signifikan antara metode *Standard K-Means* (Agregat) dan *Trajectory K-Means* (DTW) dalam hal kualitas pembentukan kluster. Nilai CH pada metode *Standard K-Means* (Agregat) mencapai 91,709, jauh lebih tinggi dibandingkan *Trajectory K-Means* (DTW) yang hanya 3,308, menunjukkan bahwa metode *Standard K-Means* (Agregat) menghasilkan kluster yang lebih terpisah dan kompak. Selanjutnya, nilai DB pada *Standard K-Means Clustering* (Agregat) sebesar 0,503, yang menunjukkan tingkat kemiripan antarkluster rendah. Sebaliknya, *Trajectory K-Means* (DTW) menunjukkan nilai DB sebesar 2,063, menandakan adanya tumpang tindih yang lebih besar antarkluster.

Tabel 4. Hasil Evaluasi *Standard K-Means* (Agregat) dengan *Trajectory K-Means* (DTW)

Metode	Calinski Harabasz	Davies Bouldin	Ray Turi	Silhouette
<i>Standard K-Means</i> (Agregat)	91,709	0,503	0,353	0,554
<i>Trajectory K-Means</i> (DTW)	3,308	2,063	2,649	0,036

Pada indikator RTI, *Standard K-Means* (Agregat) kembali menunjukkan performa lebih baik karena memiliki nilai RTI lebih kecil yang menandakan struktur kluster semakin optimal dalam hal kepadatan internal dibandingkan jarak antarkluster (Ray & Turi, 1999). Nilai RTI *Standard K-Means* (Agregat) sebesar 0,353. Sebaliknya, nilai RTI sebesar 2,649 pada *Trajectory K-Means* (DTW) menunjukkan model tidak optimal. Sementara itu, pada metode *Silhouette*, metode *Standard K-Means* (Agregat) kembali menunjukkan performa lebih baik dengan nilai 0,554, menandakan bahwa data relatif lebih sesuai dengan kluster masing-masing.

Sebaliknya, nilai *Silhouette* sebesar 0.036 pada *Trajectory K-Means* (DTW) menunjukkan adanya inkonsistensi dalam pembentukan kluster. Metode ini kurang optimal dalam memisahkan pola data pada kasus ini. Secara keseluruhan, hasil ini mengindikasikan bahwa *Standard K-Means* (Agregat) lebih unggul dan lebih sesuai untuk digunakan dalam *clustering* potensi padi di NTT dibandingkan *Trajectory K-Means* (DTW).

4. Simpulan dan Saran

Metode *Standard K-Means* (Agregat) mengelompokkan kluster berdasarkan rata-rata produksi dan luas panen tanpa mempertimbangkan dinamika waktu. Metode ini membagi menjadi 4 kluster yaitu rendah, sedang, tinggi, sangat tinggi. Kluster dengan potensi padi sangat tinggi adalah kluster 4, ditandai oleh rata-rata luas panen dan produksi tertinggi dibandingkan dengan kluster lainnya. Kabupaten Manggarai (5313) dan Manggarai Barat (5315) menjadi kabupaten dengan potensi padi paling tinggi di Nusa Tenggara Timur berdasarkan hasil *Standard K-Means* (Agregat).

Sementara itu, metode *Trajectory K-Means* (DTW) mengelompokkan data dengan mempertimbangkan pola perubahan dari waktu ke waktu. Metode ini membagi kabupaten/kota di Provinsi NTT menjadi empat kluster. Hasil dari metode *Trajectory K-Means* (DTW) menunjukkan kluster 2 menjadi kluster yang memiliki pola yang paling mirip. Pada kluster 2 terdiri dari empat kabupaten yaitu kabupaten Lembata (5308), Sikka (5310), Nagekeo (5318), dan Manggarai Timur (5319). Kluster ini mencerminkan kabupaten-kabupaten memiliki pola pertumbuhan yang stabil dan relatif sama dalam tren peningkatan atau penurunan.

Berdasarkan hasil evaluasi kedua metode, *Standard K-Means* (Agregat) lebih berfokus pada tingkat potensi padi di NTT, sedangkan *Trajectory K-Means* (DTW) memberikan wawasan mendalam terkait kesamaan pola potensi padi dari waktu ke waktu. Penelitian ini menunjukkan bahwa *Standard K-Means* (Agregat) lebih unggul dalam sisi pemisahan karena dapat mengidentifikasi kelompok wilayah dengan perbedaan produksi dan luas panen yang signifikan. Namun, *Trajectory K-Means* (DTW) tetap relevan digunakan dalam pengambilan keputusan dikarenakan dengan metode ini dapat dilihat kesamaan pola temporal dari kabupaten-kabupaten yang mengalami dinamika produksi seperti faktor iklim, pasar, dan kebijakan lintas waktu.

Berdasarkan hasil penelitian ini, rekomendasi yang dapat diberikan kepada pemerintah daerah NTT sebagai pembuat kebijakan adalah mempertimbangkan dua hal yaitu perbedaan potensi padi antar wilayah dan kemiripan pola potensi padi dari waktu ke waktu. Dua hal ini sangat penting agar potensi padi tidak hanya meningkat pada tahun tersebut saja melainkan dapat terus berkembang dan berkelanjutan. Rekomendasi lain yang dapat diberikan dalam bentuk dukungan seperti modernisasi alat pertanian, pengembangan infrastruktur, dan akses *input* untuk meningkatkan potensi padi di Provinsi NTT. Selain itu, pemerintah perlu berfokus pada kebijakan stabilisasi seperti penguatan irigasi dan pemberdayaan petani dalam menghadapi dampak perubahan iklim, sehingga petani tetap siap menanggapi berbagai dinamika

yang terjadi dari waktu ke waktu. Dengan mempertimbangkan hasil kedua metode ini, pemerintah dapat merumuskan kebijakan yang tidak hanya berfokus pada peningkatan produksi tetapi juga memastikan keberlanjutan dan stabilitas pertanian dalam jangka panjang.

Penelitian selanjutnya disarankan untuk mengembangkan analisis dengan mempertimbangkan faktor spasial. Dengan mempertimbangkan kedekatan antar kabupaten hasil *clustering*, potensi padi dapat dianalisis lebih optimal dan akurat. Selain itu, masih terdapat berbagai metode *clustering* serta variabel eksternal lainnya seperti curah hujan, harga pupuk, dan sebagainya. Penggunaan periode data yang lebih panjang juga direkomendasikan agar *clustering* tidak hanya bersifat deskriptif tetapi mampu memberikan dasar prediktif dan kebijakan berbasis bukti.

Ethics Approval

Seluruh prosedur dalam penelitian ini telah sesuai dengan prinsip-prinsip serta etika publisitas.

Conflict of Interest

Penulis menyatakan tidak memiliki konflik kepentingan terkait penelitian ini.

Daftar Pustaka

- [1] Arini, E. R. (2023). Penerapan K-Means Cluster di Provinsi Jawa Timur. *Journal of Science Nusantara*, 3(1), 32 - 36.
- [2] Baltagi, B. H. (2025). *Econometric Analysis of Panel Data*, Third edition. John Wiley & Sons Ltd.
- [3] Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *The Econometric Society*, 82(3), 1147–1184.
- [4] Caliński, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1 - 27. <https://doi.org/https://doi.org/10.1080/03610927408827101>.
- [5] Christiani, N. V. (2024). Analisis Daerah Potensi Produksi Padi Di Provinsi Nusa Tenggara Timur Menggunakan K-Means Cluster. *Jurnal Statistik Terapan*, 4(1), 1-11.
- [6] Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224 - 227. <https://doi.org/https://doi.org/10.1109/TPAMI.1979.4766909>.
- [7] Direktorat Statistik Kesejahteraan Rakyat. (2023). *Konsep dan Definisi SUSENAS Maret 2024*. Badan Pusat Statistik.

- [8] Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.
- [9] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Dalam J. MacQueen, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (hal. 1967). University of California Press.
- [10] Ray, S., & Turi, R. (1999). Determination of number of clusters in k-means clustering and application in colour image segmentation. *Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques*.
- [11] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20(1), 53 - 65. [https://doi.org/https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/https://doi.org/10.1016/0377-0427(87)90125-7).
- [12] Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating The Number of Clusters in a Data Set Via The Gap Statistic. *Journal Of The Royal Statistical Society: Series B*, 63(2), 411 - 423. <https://doi.org/10.1111/1467-9868.00293>.