

ANALISIS KLASTER MULTIVARIAT KINERJA PASAR PARIWISATA KABUPATEN/KOTA DI NUSA TENGGARA TIMUR: PENDEKATAN INTEGRATIF *UNIFORM MANIFOLD APPROXIMATION AND PROJECTION (UMAP)* DAN *K-MEANS CLUSTERING*

Retno Fitriandari¹, Fadel Muhammad²

¹Badan Pusat Statistik Kota Kupang, Indonesia,

²Badan Pusat Statistik Provinsi Nusa Tenggara Timur, Indonesia

*Korespondensi Penulis: retnofit@bps.go.id, fadel.muhammad@bps.go.id

ARTIKEL INFO

Abstract

Article history:

Received 20 Oct, 2025

Revised 19 Nov, 2025

Accepted 1 Dec, 2025

Published 31 Dec, 2025

Introduction: Tourism plays a vital role in Indonesia's regional development, yet spatial disparities in tourism performance remain evident across Nusa Tenggara Timur (NTT). **Background Problem:** This study examines multidimensional tourism performance by integrating indicators of market demand, supply effectiveness, economic impact, and accessibility. The research addresses the problem of unequal regional tourism performance and asks: how can statistical clustering identify performance disparities among NTT's districts? **The novelty** of this study lies in applying unsupervised learning (K-Means clustering) at the district/city level. **Research Methods:** Analysis employed a combination of UMAP for dimensionality reduction and dual validation using the Silhouette Score and Adjusted Rand Index (ARI). Standardized secondary data (2021–2024) from Statistics Indonesia were used and analyzed using R 4.5.1. **Results** showed that the optimal number of clusters is three, with a Silhouette Score of 0.472 (weak structure/acceptable) and ARI of 0.813 (good/accepted well). Cluster 1 represents high-performing regions with superior accessibility and demand, Cluster 2 reflects transitional areas with strong capacity but weak utilization, and Cluster 3 includes underperforming regions. Centroid analysis revealed external access and market demand as key differentiators, providing an empirical basis for targeted tourism policy in NTT.

Keywords:

Adjusted Rand Index; Centroid Analysis; K-Means Clustering; Regional Disparity; Silhouette Score; Tourism Performance; UMAP

1. Pendahuluan

Sektor pariwisata memegang peranan signifikan dalam paradigma pembangunan nasional maupun regional. Di tingkat global dan nasional, sektor ini diakui sebagai katalisator utama penciptaan lapangan kerja dan kontributor fundamental terhadap Produk Domestik Bruto (PDB) (Anggarini, 2021). Sebelum terjadinya disrupsi pandemi, kontribusi pariwisata terhadap PDB Indonesia mencapai puncaknya hingga 4,97% pada tahun 2019, dan sektor ini secara konsisten menyerap tenaga kerja dalam jumlah besar, mencapai 22,89 juta jiwa pada tahun 2022 (Badan Pusat Statistik [BPS], 2021;

Kementerian Pariwisata dan Ekonomi Kreatif [Kemenparekraf], 2023). Lebih lanjut, sektor ini diproyeksikan berpotensi menjadi pendorong utama pertumbuhan ekonomi nasional, dengan target pertumbuhan hingga 8% pada tahun 2029 (Kementerian Koordinator Bidang Perekonomian [Kemenko Perekonomian], 2025).

Dalam konteks regional, Nusa Tenggara Timur (NTT) menjadi fokus perhatian nasional melalui penetapan Labuan Bajo (Kabupaten Manggarai Barat) sebagai salah satu Destinasi Pariwisata Super Prioritas (DPSP), sebagaimana diatur dalam Peraturan Presiden Republik Indonesia Nomor 18 Tahun 2020. Penetapan ini mencerminkan pengakuan pemerintah pusat atas kekayaan sumber daya maritim dan keunikan budaya sebagai modal utama pengembangan pariwisata. Optimalisasi sektor ini, yang berpusat di Labuan Bajo, selaras dengan visi Dasa Cita NTT Maju, Sehat, Cerdas, Sejahtera, dan Berkelanjutan, terutama pada pilar kesejahteraan dan pertumbuhan ekonomi berkelanjutan. Namun, di tengah pesatnya pembangunan pariwisata di wilayah super prioritas tersebut, pemerataan manfaat ekonomi ke seluruh wilayah NTT masih menjadi tantangan yang signifikan.

Data wisatawan nusantara (wisnus) di NTT menunjukkan pertumbuhan yang bervariasi antar kabupaten, dengan disparitas yang cukup signifikan. Kabupaten Manggarai Barat, sebagai DPSP, mencatat jumlah wisnus tertinggi, meningkat dari 472.708 pada tahun 2021 menjadi lebih dari satu juta pada tahun 2024, dengan jumlah akomodasi sebanyak 114 unit, menandakan kapasitas dan daya tarik yang kuat. Kabupaten Kupang juga menunjukkan peningkatan wisnus signifikan dari 151.365 menjadi 685.232, meskipun jumlah akomodasi tetap sangat terbatas, hanya sekitar 3 unit, mengindikasikan potensi pariwisata yang belum optimal dalam hal fasilitas pendukung. Di sisi lain, Kabupaten Sumba Tengah mencatat wisnus relatif rendah namun bertambah dari 51.700 menjadi 101.793, dengan ketersediaan akomodasi yang sangat minim yakni hanya 1 unit, menunjukkan kesenjangan besar antara permintaan dan kapasitas. Kabupaten Flores Timur memperlihatkan jumlah akomodasi yang lebih baik, yaitu sebanyak 15 unit, namun wisnus yang fluktuatif dan lebih rendah dibandingkan kabupaten lain, hanya meningkat dari 100.068 menjadi 366.375, menandakan tantangan dalam mempertahankan dan meningkatkan daya tarik wisatawan. Disparitas ini menegaskan perlunya strategi pengembangan pariwisata yang lebih merata agar manfaat ekonomi dapat dirasakan secara luas di seluruh wilayah NTT.

Kompleksitas dan ketidakseragaman data pariwisata di NTT menuntut pendekatan pengukuran yang bersifat komprehensif. Kinerja pariwisata merupakan fenomena multidimensi yang mencerminkan interaksi antara aspek penawaran, permintaan, dan dampak ekonomi. Pendekatan ini sejalan dengan model *destination competitiveness* yang mengintegrasikan sisi penawaran dan permintaan dalam analisis daya saing destinasi (Nuraeni & Hendriyani, 2022). Selain itu, konsumsi dan investasi pariwisata terbukti berkontribusi signifikan terhadap pertumbuhan ekonomi daerah (Hariyani, 2022). Dengan demikian, pengukuran kinerja pariwisata perlu dilakukan secara

terpadu melalui indikator kunjungan wisatawan, tingkat hunian akomodasi, infrastruktur dan aksesibilitas, serta kontribusi sektor akomodasi dan penyediaan makan minum terhadap PDRB.

Berbagai studi telah memanfaatkan metode statistik untuk menganalisis kinerja pariwisata di Indonesia. Suhardiman (2025) menggunakan *K-Means clustering* untuk memetakan tourism hotspot antarprovinsi berdasarkan data tamu hotel, namun analisisnya masih terbatas pada level provinsi. Amanda dan Setiawan (2022) juga menerapkan metode serupa untuk mengelompokkan daerah di Jawa Timur berdasarkan kunjungan wisatawan asing dan daya saing pariwisata, dengan temuan bahwa indikator permintaan menjadi faktor dominan. Masteriarsa dan Riyanto (2023) menambahkan dimensi *carrying capacity* dalam analisis kluster 34 provinsi, memperluas perspektif kinerja pariwisata meski tetap pada skala provinsi.

Sebagian besar analisis statistik pariwisata di Indonesia masih berfokus pada level provinsi, sementara studi di tingkat kabupaten/kota terutama di wilayah dengan kondisi geografis kompleks seperti NTT masih terbatas. Penelitian terdahulu cenderung menitikberatkan pada indikator permintaan seperti kunjungan wisatawan, tingkat hunian hotel, dan kapasitas tampung, tanpa mengintegrasikan aspek penawaran maupun dampak ekonomi terhadap Produk Domestik Regional Bruto (PDRB). Kesenjangan ini menimbulkan kebutuhan mendesak untuk pendekatan yang lebih komprehensif dan inovatif dalam memahami dinamika pariwisata di tingkat lokal.

Untuk itu, penelitian ini menggunakan pendekatan *unsupervised learning*, khususnya analisis kluster, guna mengidentifikasi pola kemiripan kinerja daya tarik pasar pariwisata antar kabupaten/kota di NTT berdasarkan indikator multidimensi. Dengan menggabungkan berbagai variabel yang mencerminkan aspek permintaan, penawaran, dan dampak ekonomi, penelitian ini bertujuan memberikan klasifikasi yang lebih holistik dan akurat. Hasilnya diharapkan dapat menjadi dasar empiris yang kuat bagi perumusan kebijakan pengembangan pariwisata yang lebih terarah, efektif, dan berbasis data, sekaligus memperkaya penerapan metode statistik modern dalam analisis sektoral pariwisata daerah yang selama ini belum banyak digarap. Penelitian ini sangat penting untuk menjembatani kesenjangan informasi dan mendukung pemerataan manfaat ekonomi dari pariwisata di wilayah NTT yang beragam dan kompleks.

2. Metodologi

2.1. Bahan dan Sumber Data

Penelitian ini menggunakan data sekunder yang diperoleh dari BPS Provinsi NTT. Data dikumpulkan untuk 22 Kabupaten/Kota di NTT selama periode waktu 2021-2024. Data yang dikumpulkan mewakili enam dimensi utama kinerja pariwisata, yang mencakup aspek permintaan, penawaran, dampak ekonomi, dan aksesibilitas. Daftar variabel disajikan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Nama Variabel	Deskripsi Data	Dimensi
(1)	(2)	(3)	(4)
x1	Jumlah Kunjungan Wisnus	Angka total kunjungan Wisatawan Nusantara per tahun	Permintaan Pasar
x2	TPK Total	Tingkat Penghunian Kamar (TPK) Total (hotel bintang dan non-bintang) per tahun	Efektivitas Penawaran
x3	Jumlah Akomodasi Total	Total unit usaha akomodasi (hotel/losmen/penginapan) per tahun	Kapasitas Penawaran
x4	PDRB Akomodasi & Makanan Minuman	Nilai PDRB Nominal Sektor I (Akomodasi dan Makanan Minuman) per tahun	Dampak Ekonomi
x5	Indeks Aksesibilitas Eksternal	Indeks komposit Z-Score dari frekuensi angkutan Udara + Kapal	Akses Eksternal
x6	Kepadatan Jaringan Jalan	Rasio Panjang Jalan (km) per Luas Wilayah (km ²) per tahun	Akses Internal

Dimensi permintaan diwakili oleh jumlah kunjungan wisnus per kabupaten/kota yang menggambarkan tingkat daya tarik aktual suatu daerah terhadap pasar domestik. Dari sisi penawaran, digunakan Tingkat Penghunian Kamar (TPK) total untuk menunjukkan efektivitas pemanfaatan kapasitas akomodasi yang tersedia, serta jumlah unit usaha akomodasi (hotel, penginapan, dan sejenisnya) sebagai proksi ketersediaan fasilitas fisik pariwisata di tiap wilayah. Kedua indikator ini memberikan gambaran tentang keseimbangan antara kapasitas penawaran dan daya serap pasar. Dampak ekonomi sektor pariwisata diukur melalui PDRB sektor penyediaan akomodasi dan makanan minuman pada harga berlaku, yang mencerminkan kontribusi langsung aktivitas pariwisata terhadap perekonomian daerah.

Sementara itu, aspek aksesibilitas dibagi menjadi dua dimensi. Akses eksternal direpresentasikan oleh frekuensi angkutan udara dan laut (jumlah penerbangan dan pelayaran per tahun) yang menunjukkan intensitas konektivitas antardaerah dan kemudahan kedatangan wisatawan dari luar wilayah. Akses internal diukur melalui kepadatan jaringan jalan, yaitu rasio panjang jalan terhadap luas wilayah yang memberikan gambaran tentang kemudahan mobilitas wisatawan di dalam daerah. Secara keseluruhan, keenam variabel ini membentuk kerangka yang komprehensif untuk menangkap hubungan antara permintaan wisata, kapasitas penawaran, dampak ekonomi, serta infrastruktur pendukung yang menentukan kinerja pariwisata di NTT.

Dalam penyajian hasil analisis, digunakan kode administrasi wilayah pada beberapa grafik dan tabel. Penggunaan kode tersebut dipilih untuk menjaga

keringkasan tampilan sekaligus memastikan konsistensi penandaan antarunit analisis. Daftar kode wilayah beserta nama kabupaten/kota disajikan pada Tabel 2.

Tabel 2. Daftar Kode Administrasi Wilayah dan Nama Kabupaten/Kota di Provinsi NTT

Kode Administrasi Wilayah (1)	Nama Kabupaten/Kota (2)
5301	Kabupaten Sumba Barat
5302	Kabupaten Sumba Timur
5303	Kabupaten Kupang
5304	Kabupaten Timor Tengah Selatan
5305	Kabupaten Timor Tengah Utara
5306	Kabupaten Belu
5307	Kabupaten Alor
5308	Kabupaten Lembata
5309	Kabupaten Flores Timur
5310	Kabupaten Sikka
5311	Kabupaten Ende
5312	Kabupaten Ngada
5313	Kabupaten Manggarai
5314	Kabupaten Rote Ndao
5315	Kabupaten Manggarai Barat
5316	Kabupaten Sumba Tengah
5317	Kabupaten Sumba Barat Daya
5318	Kabupaten Nagekeo
5319	Kabupaten Manggarai Timur
5320	Kabupaten Sabu Raijua
5321	Kabupaten Malaka
5371	Kota Kupang

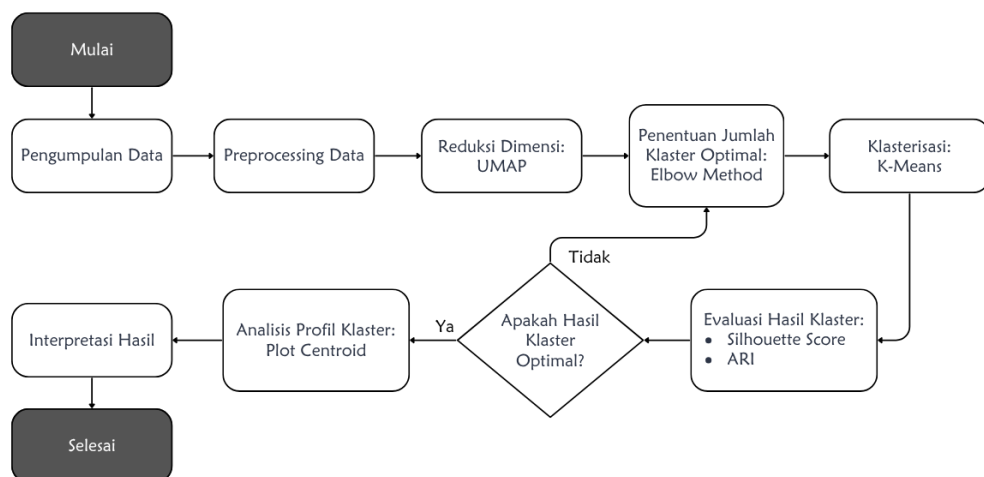
2.2. Metode Penelitian

Penelitian ini menggunakan pendekatan eksploratori kuantitatif dengan teknik *unsupervised learning*, yang bertujuan untuk mengidentifikasi pola kemiripan kinerja pariwisata antarwilayah. Tahapan analisis dalam penelitian ini dilakukan secara sistematis melalui beberapa prosedur yang ditunjukkan pada Gambar 1. Proses analisis dimulai dari pengumpulan dan praproses data, di mana seluruh data diolah menjadi nilai tunggal per kabupaten/kota. Nilai tersebut diperoleh dari rata-rata data periode 2021–2024, sehingga mencerminkan kondisi kinerja pariwisata yang lebih stabil antarwaktu. Selanjutnya, data yang telah dirata-ratakan, distandardisasi menggunakan metode *Z-Score* untuk menghilangkan bias akibat perbedaan satuan pengukuran dan memastikan seluruh variabel memiliki skala yang setara dalam analisis kluster.

Setelah semua data distandarisasi, selanjutnya dilakukan reduksi dimensi menggunakan *Uniform Manifold Approximation and Projection* (UMAP). Selanjutnya, dilakukan penentuan jumlah kluster optimal dengan menggunakan metode *Elbow*, yang kemudian menjadi dasar untuk penerapan algoritma *K-Means* dalam proses pengelompokan data. Evaluasi terhadap hasil kluster mengacu pada *Silhouette Score* (Skor Siluet) dan *Adjusted Rand Index* (ARI). Apabila hasil evaluasi menunjukkan tingkat optimalitas yang memadai, maka dilanjutkan dengan analisis profil kluster melalui plot profil *centroid* untuk menginterpretasikan karakteristik masing-masing kluster.

Seluruh analisis dilakukan menggunakan perangkat lunak R (versi 4.5.1) dan melibatkan tiga tahapan utama. Reduksi dimensi dilakukan dengan algoritma UMAP, diimplementasikan menggunakan *package umap*. Selanjutnya, klusterisasi *K-Means* dijalankan menggunakan fungsi *kmeans()* dari *package stats* (bawaan R), setelah menentukan jumlah kluster optimal (k) melalui Metode Elbow.

Validasi internal kluster diukur dengan Koefisien *Silhouette*, dihitung menggunakan fungsi *silhouette()* dari *package cluster* dan divisualisasikan dengan *fviz_silhouette()* dari *package factoextra*. Sementara itu, stabilitas kluster diuji melalui prosedur *bootstrapping* dan diukur menggunakan ARI yang disediakan oleh *package aricode*. Seluruh visualisasi data dan hasil dibuat menggunakan *package ggplot2* dan *ggrepel*.



Gambar 1. Diagram Alur Tahapan Analisis

***Uniform Manifold Approximation and Projection* (UMAP)**

UMAP adalah teknik reduksi dimensi nonlinear yang memetakan data berdimensi tinggi ke ruang berdimensi rendah dengan mempertahankan struktur lokal dan global data. UMAP membangun graf k -tetangga terdekat menggunakan bobot keterhubungan (McInnes dkk., 2018). UMAP unggul dalam menjaga keseimbangan antara pelestarian struktur lokal dan global serta skalabilitas pada dataset besar, dengan parameter $n_neighbors$ dan min_dist yang dapat disesuaikan (Suhandi, 2025; Mertayasa, 2022).

Elbow Method

Elbow method merupakan metode yang digunakan untuk menentukan jumlah kluster terbaik, yaitu dengan cara melihat persentase setiap kluster yang akan membentuk siku pada titik tertentu. Tujuan dari metode *Elbow* adalah untuk memilih nilai k yang kecil dan memiliki nilai *withinss* (*SSE*) yang rendah (Maori & Evanita, 2023).

K-Means Clustering

K-Means adalah algoritma klusterisasi yang bertujuan membagi data ke dalam kluster berdasarkan kemiripan (Ahmed, 2020). Setiap data akan dikelompokkan ke kluster dengan *centroid* (titik pusat) terdekat, yang dihitung menggunakan jarak *Euclidean* (Mukhtar dkk., 2024). Berikut adalah rumus untuk jarak *Euclidean* (Siagian dkk., 2025):

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2}$$

Keterangan:

$d(x, c)$ = jarak antardata dan *centroid* kluster

x_i = nilai atribut dari data ke- i

c_i = nilai *centroid* dari atribut ke- i

n = jumlah atribut dalam dataset

Skor Siluet (Silhouette Score)

Skor siluet menggambarkan seberapa baik suatu objek sesuai dengan klasternya (*cohesion*) dibandingkan dengan kluster lain (*separation*) (Tan dkk., 2014). Indeks ini dinyatakan melalui *silhouette coefficient* dengan rentang -1 hingga 1 , di mana nilai mendekati 1 menunjukkan pengelompokan yang baik, sedangkan nilai mendekati -1 menandakan objek lebih cocok berada pada kluster lain (Saputra dkk., 2020). Interpretasi nilai *silhouette coefficient* merujuk pada klasifikasi umum sebagaimana dijelaskan oleh Rousseeuw (1987) dan dikembangkan lebih lanjut dalam Kaufman & Rousseeuw (2005), disajikan pada Tabel 3.

Tabel 3. Interpretasi Skor Siluet

Rentang Skor Siluet	Interpretasi	Arti Praktis	Kelayakan Kluster
(1)	(2)	(3)	(4)
0.71 – 1.00	Struktur sangat kuat	Klaster sangat jelas terpisah	Diterima Sangat Baik
0.52 – 0.70	Struktur wajar / baik	Klaster terstruktur dengan baik, kompak, dan terpisah	Diterima Baik
0.26 – 0.50	Struktur lemah	Struktur klaster masih ditemukan, tetapi batas-batasnya kurang jelas	Dapat Diterima
< 0.25	Tidak ditemukan struktur berarti	Klastering cenderung acak atau tidak bermakna	Ditolak

Adjusted Random Index (ARI)

ARI merupakan metrik validasi eksternal yang menilai kesesuaian hasil clustering dengan *gold standard* atau label acuan. Sebagai pengembangan dari *Rand Index*, ARI mengoreksi pengaruh kebetulan sehingga hasilnya lebih akurat (Costa, 2013). Berbagai penelitian dan dokumentasi perangkat lunak (misalnya *CrossClustering package R*, serta studi oleh Romano dkk., 2016) menggunakan interpretasi yang disajikan pada Tabel 4.

Tabel 4. Interpretasi Nilai *Adjusted Random Index*

Rentang Nilai ARI	Interpretasi	Arti Praktis	Kelayakan Kluster
(1)	(2)	(3)	(4)
≥ 0.90	Sangat baik	Klaster hampir identik dengan referensi (<i>ground truth</i>)	Diterima Sangat Baik
0.80 – 0.89	Baik	Klaster cukup sesuai dengan target	Diterima Baik
0.65 – 0.79	Sedang	Beberapa kesalahan pengelompokan terjadi	Dapat Diterima
< 0.65	Kurang baik	Struktur klaster lemah/tidak stabil	Kurang Layak
≈ 0	Setara dengan acak	Tidak ada struktur berarti	Ditolak
< 0	Lebih buruk dari acak	Pembagian klaster keliru	Ditolak

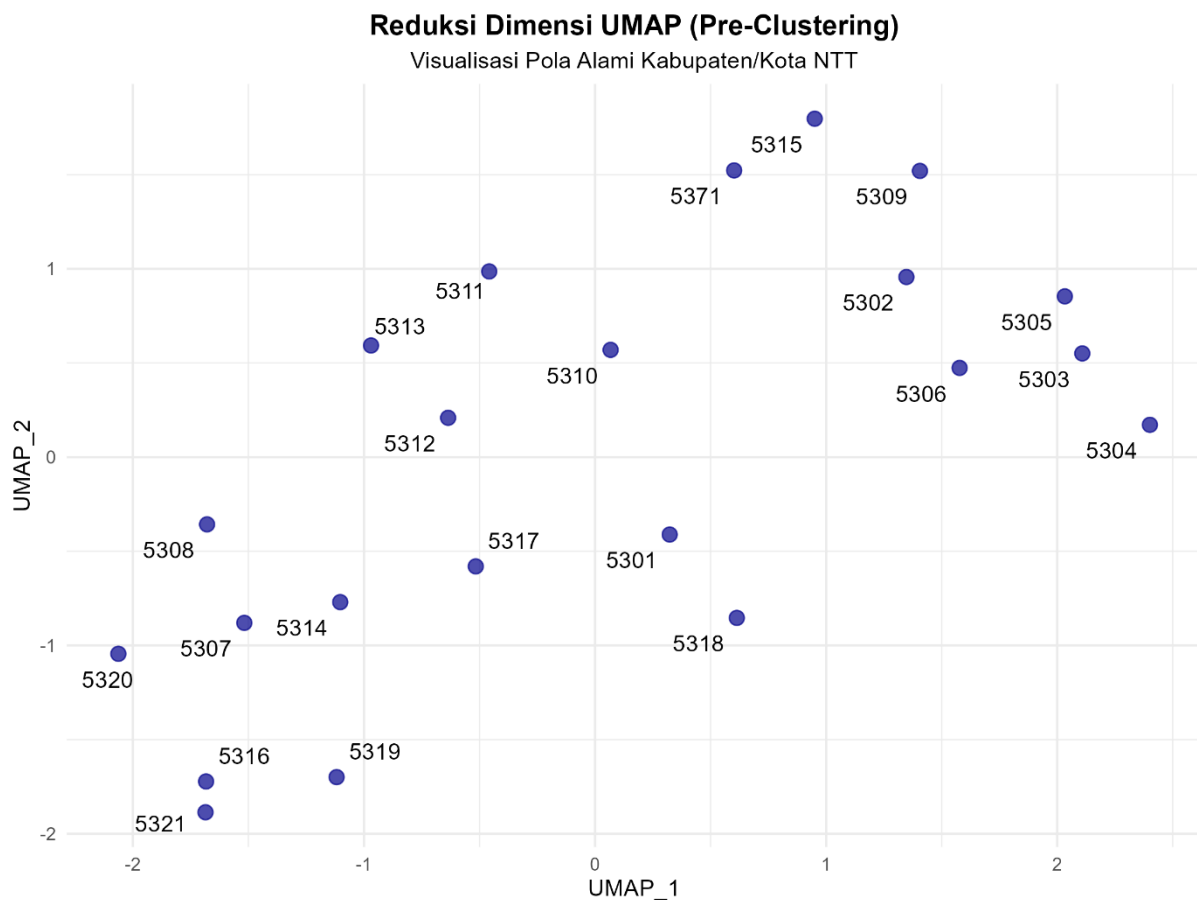
Plot Profil Centroid

Dalam analisis kluster, *centroid* merupakan representasi dari titik pusat suatu kluster yang dihitung berdasarkan nilai rata-rata seluruh objek di dalam kluster tersebut. Pada metode *K-Means clustering*, *centroid* mencerminkan karakteristik umum atau profil tipikal dari setiap kelompok data yang terbentuk (Jain, 2010).

3. Hasil dan Pembahasan

Proses pengolahan data diawali dengan standarisasi enam variabel menggunakan metode *Z-Score* guna menghilangkan perbedaan skala antarindikator sehingga seluruh variabel memiliki kontribusi yang setara dalam analisis. Data yang telah distandarisi selanjutnya digunakan sebagai dasar untuk tahap reduksi dimensi menggunakan UMAP.

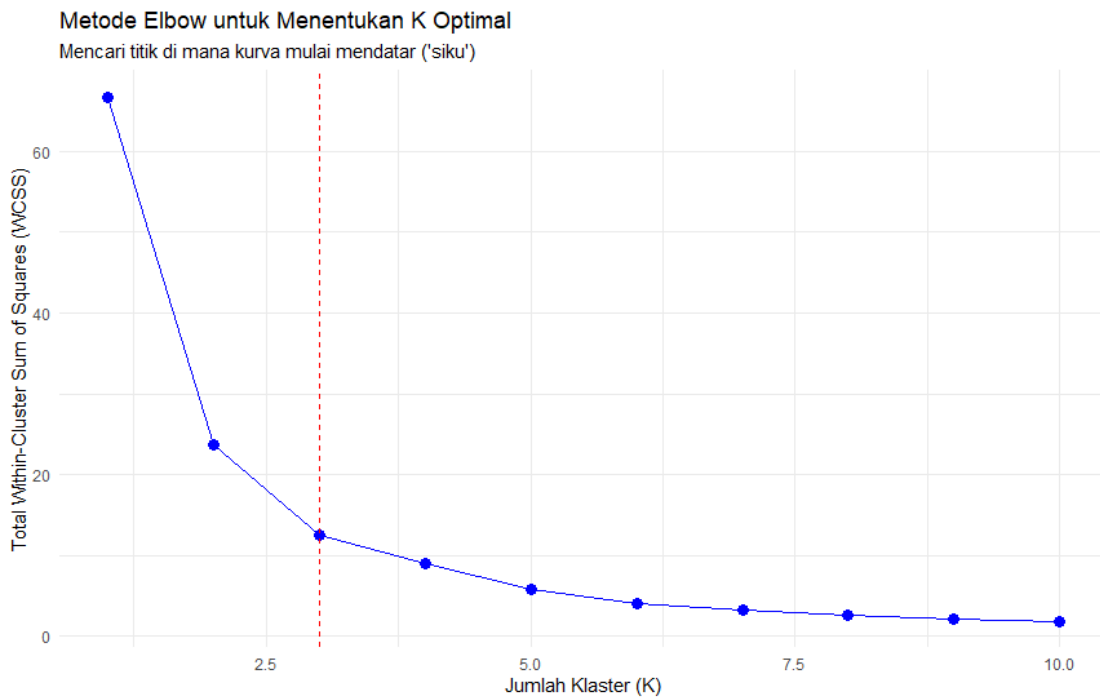
Pada tahap ini, data kinerja pariwisata direduksi ke dalam dua komponen utama (UMAP_1 dan UMAP_2) dengan parameter $n_neighbors = 7$ dan $min_dist = 0,1$ yang bertujuan untuk memvisualisasikan kemiripan dan kedekatan struktural antar kabupaten/kota berdasarkan kinerja pariwisata mereka sebelum klusterisasi dilakukan.



Gambar 2. Visualisasi Hasil Reduksi Dimensi UMAP (Pre-Clustering)

Reduksi dimensi menggunakan UMAP menghasilkan representasi dua dimensi dari enam variabel pariwisata yang telah distandarisi. Titik-titik mewakili masing-masing kabupaten/kota di NTT, di mana kedekatan posisi menunjukkan kemiripan karakteristik kinerja pariwisata. Secara visual tampak kecenderungan pembentukan beberapa kelompok alami, yang mengindikasikan adanya variasi pola kinerja antarwilayah dan menjadi dasar bagi proses klusterisasi selanjutnya.

Sebelum melakukan proses klusterisasi dengan *K-Means*, perlu ditentukan terlebih dahulu jumlah klaster (k) yang akan dibentuk. Penentuan jumlah klaster optimal dilakukan menggunakan Metode Siku (*Elbow Method*).



Gambar 3. Metode *Elbow* untuk Menentukan k Optimal

Gambar 3 menunjukkan bahwa nilai WCSS menurun secara drastis dari $k = 1$ ke $k = 2$. Penurunan signifikan kedua terjadi pada titik $k = 3$, di mana kurva mulai menunjukkan kemiringan yang melambat dan membentuk "siku". Berdasarkan kaidah metode ini, tiga (3) kluster ditetapkan sebagai jumlah pengelompokan yang paling optimal, memberikan keseimbangan terbaik antara minimalisasi variasi internal dan efisiensi model kluster. Langkah selanjutnya adalah menerapkan algoritma *K-Means clustering* dengan $k = 3$ untuk memperoleh segmentasi wilayah yang lebih jelas. Tabel 5 menyajikan nilai koordinat UMAP dan hasil klasterisasi *K-Means* $k = 3$ untuk 22 kabupaten/kota di Provinsi NTT.

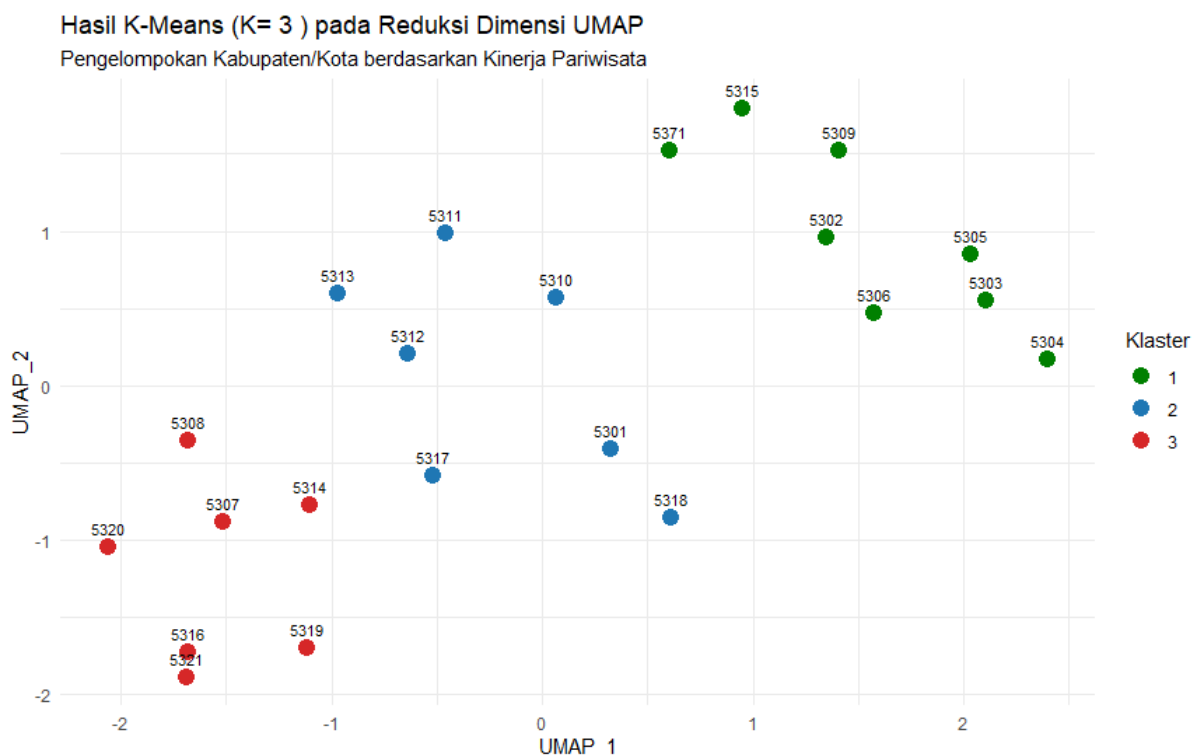
Tabel 5. Nilai Koordinat UMAP dan Hasil Klasterisasi

Koordinat UMAP dan Hasil Klasterisasi $k = 3$							
UMAP_1	UMAP_2	Kabupaten/Kota	Klaster	UMAP_1	UMAP_2	Kabupaten/Kota	Klaster
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
0,32	-0,41	5301	2	1,58	0,47	5306	1
1,35	0,96	5302	1	-1,52	-0,88	5307	3
2,11	0,55	5303	1	-1,68	-0,36	5308	3
2,40	0,17	5304	1	1,41	1,52	5309	1
2,03	0,85	5305	1	0,07	0,57	5310	2

Tabel 5. Nilai Koordinat UMAP dan Hasil Klasterisasi (lanjutan)

Koordinat UMAP dan Hasil Klasterisasi $k = 3$							
UMAP_1	UMAP_2	Kabupaten/Kota	Klaster	UMAP_1	UMAP_2	Kabupaten/Kota	Klaster
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
-0,46	0,99	5311	2	-0,52	-0,58	5317	2
-0,64	0,21	5312	2	0,61	-0,85	5318	2
-0,97	0,59	5313	2	-1,12	-1,70	5319	3
-1,10	-0,77	5314	3	-2,06	-1,04	5320	3
0,95	1,80	5315	1	-1,69	-1,89	5321	3
-1,68	-1,72	5316	3	0,60	1,52	5371	1

Hasil klasterisasi *K-Means* kemudian diproyeksikan kembali ke dalam ruang dimensi UMAP untuk visualisasi hasil pengelompokan, seperti yang disajikan pada Gambar 4. Visualisasi tersebut menunjukkan bagaimana observasi-observasi yang sebelumnya dekat dalam proyeksi UMAP kini secara definitif telah dikelompokkan ke dalam Klaster 1, Klaster 2, dan Klaster 3. Klaster-klaster yang dihasilkan terdistribusi dengan pemisahan spasial yang nyata dalam ruang dua dimensi, mengindikasikan bahwa algoritma *K-Means* berhasil mengorganisir data ke dalam kelompok yang berbeda secara struktural.



Gambar 4. Hasil *K-Means Clustering*

Klaster 1 (hijau) terkonsentrasi di kuadran kanan atas, mencerminkan posisi superior yang menonjol. Anggota klaster ini adalah 5371 (Kota Kupang), 5315 (Manggarai Barat), 5309 (Flores Timur), 5302 (Sumba Timur), 5306 (Belu), 5305 (Timor Tengah Utara), 5303 (Kabupaten Kupang), dan 5304 (Timor Tengah Selatan). Secara visual, posisi klaster yang menonjol di bagian atas dan kanan grafik mengindikasikan bahwa kelompok ini cenderung memiliki nilai-nilai variabel yang lebih tinggi dibandingkan dua

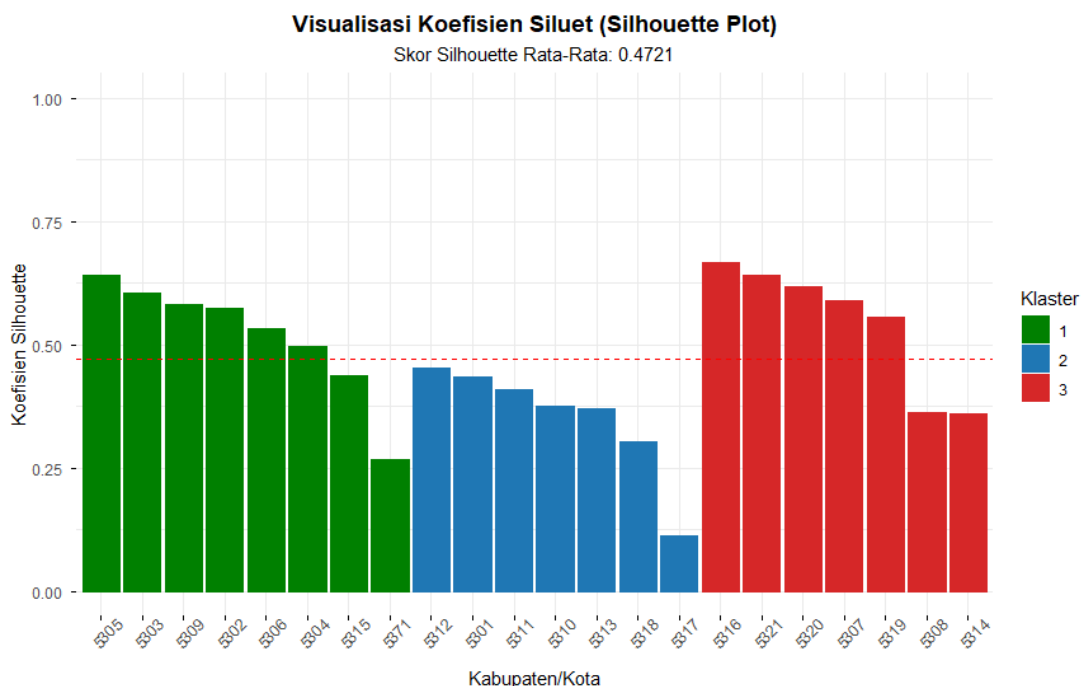
klaster lainnya, sehingga dapat diasosiasikan sebagai wilayah dengan kinerja pariwisata yang relatif lebih unggul.

Klaster 2 (biru) tersebar di area tengah, menunjukkan anggota klaster ini memiliki profil kinerja yang bervariasi tetapi berada dalam kisaran menengah antara kelompok unggul dan kelompok tertinggal (Klaster 3). Anggota klaster ini adalah 5311 (Ende), 5313 (Manggarai), 5312 (Ngada), 5310 (Sikka), 5301 (Sumba Barat), dan 5318 (Nagekeo).

Klaster 3 (merah) terkonsentrasi di kuadran kiri bawah, mencerminkan kemiripan profil kinerja yang rendah dan tertinggal. Anggota klaster ini adalah 5308 (Lembata), 5320 (Sabu Raijua), 5307 (Alor), 5314 (Rote Ndao), 5316 (Sumba Tengah), 5321 (Malaka), dan 5319 (Manggarai Timur). Pemisahan spasial ini menunjukkan bahwa perbedaan tingkat kinerja pariwisata antardaerah cukup signifikan, dengan kelompok unggul, menengah, dan tertinggal yang terbentuk secara jelas.

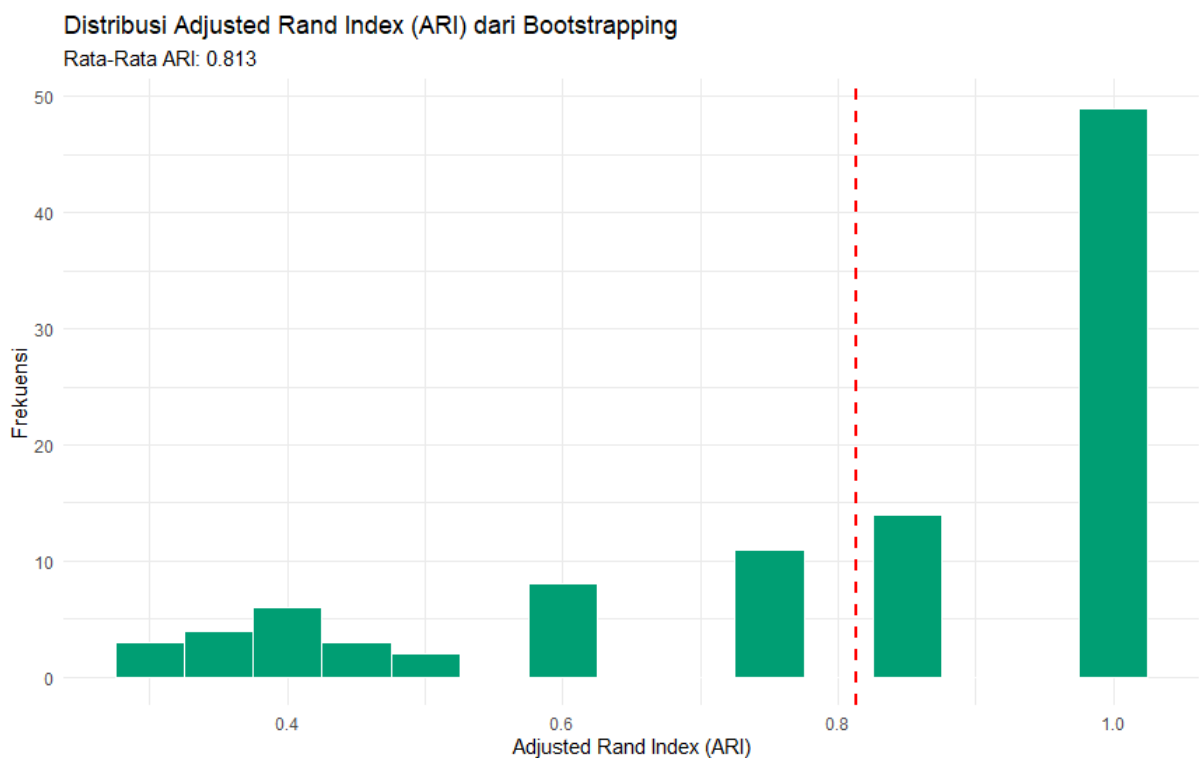
Setelah pengelompokan diperoleh, uji stabilitas dan konsistensi diterapkan untuk memverifikasi kualitas dan reliabilitas hasil *K-Means clustering*. Uji ini penting untuk memastikan bahwa pembagian kelompok yang dihasilkan memiliki signifikansi struktural dan bukan merupakan hasil fluktuasi data atau kebetulan inisialisasi algoritma.

Evaluasi kualitas klaster dilakukan menggunakan koefisien siluet (*silhouette coefficient*) dan *Adjusted Rand Index* (ARI). Kedua metrik ini secara kolektif berfungsi sebagai tolok ukur kuantitatif terhadap kualitas pengelompokan (kohesivitas dan separasi) dan konsistensi hasil klasterisasi saat data diuji melalui *bootstrapping*. Kualitas internal klaster dievaluasi menggunakan koefisien siluet untuk mengukur kepaduan (*cohesion*) dan keterpisahan (*separation*) antar klaster.



Gambar 5. Visualisasi Koefisien Siluet (*Silhouette Plot*)

Nilai rata-rata koefisien siluet sebesar 0,472 menunjukkan bahwa pemisahan antar-klastrer relatif lemah, namun masih berada pada tingkat yang dapat diterima untuk analisis eksploratori. Gambar 5 menyajikan distribusi nilai skor siluet untuk setiap kabupaten/kota (ditampilkan dalam bentuk batang) pada masing-masing klastrer. Setiap objek diurutkan berdasarkan nilai skor siluetnya dalam setiap klastrer. Berdasarkan Gambar 5, dapat dilihat bahwa sebagian besar kabupaten/kota pada Klastrer 1 dan Klastrer 3 memiliki nilai skor siluet yang lebih tinggi dari rata-rata skor keseluruhan. Hal ini mengindikasikan bahwa anggota pada kedua klastrer tersebut memiliki tingkat kesesuaian yang baik terhadap klasternya masing-masing. Sementara kabupaten/kota pada Klastrer 2 berada di bawah rata-rata skor siluet yang berarti kualitas klastrer ini paling rendah diantara ketiganya.



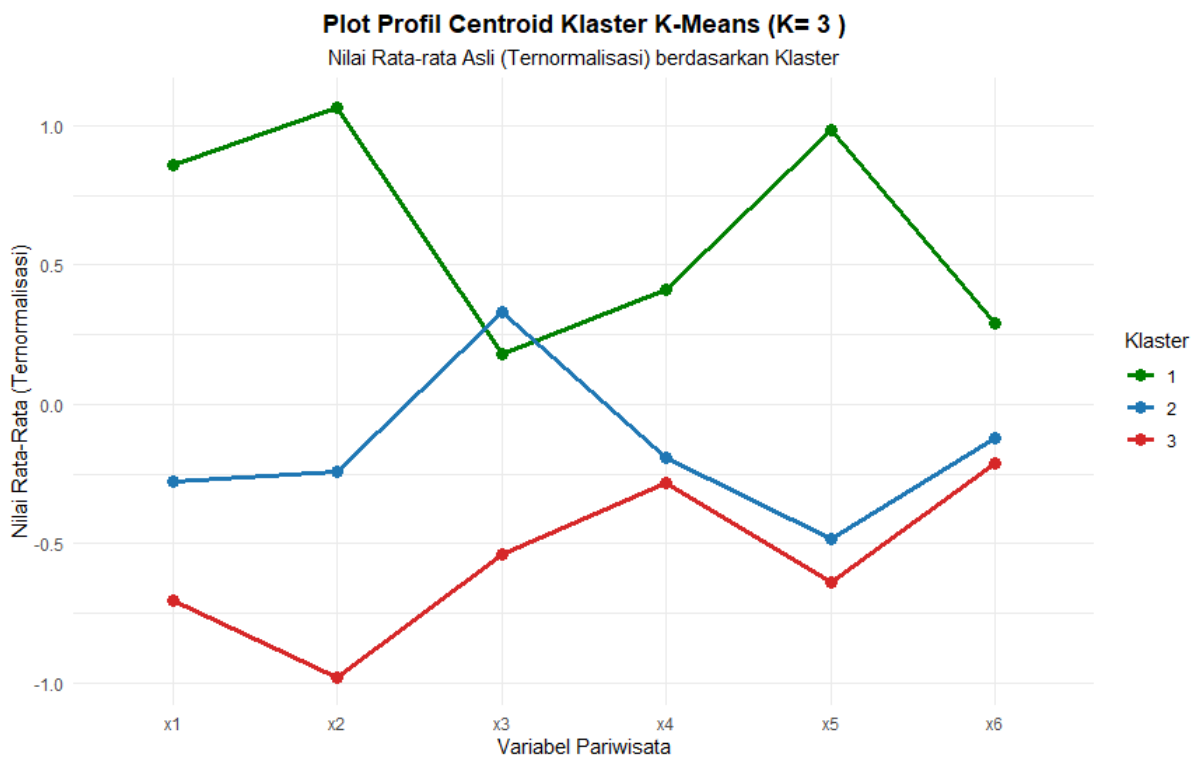
Gambar 6. Distribusi *Adjusted Rand Index* (ARI)

Evaluasi jumlah klastrer menggunakan ARI menghasilkan nilai rata-rata ARI sebesar 0,813. Nilai ini tergolong tinggi (mendekati 1) sehingga menunjukkan tingkat kesesuaian yang baik antara hasil klasterisasi dengan data referensi (*ground truth*). Temuan ini mengindikasikan bahwa pembentukan tiga klastrer yang diperoleh memiliki konsistensi yang kuat serta struktur pengelompokan yang tidak terjadi secara acak. Dengan demikian, hasil klasterisasi dapat dikatakan stabil dan dapat direplikasi secara andal dalam analisis yang serupa.

Konsistensi klastrer (ARI) yang kuat membuktikan bahwa pengelompokan ini kokoh meskipun terdapat sedikit ambiguitas pada batas klastrer. Kembali merujuk pada Gambar 5, skor siluet yang rendah pada Klastrer 2, yang merupakan klastrer sekunder/menengah, merefleksikan realitas bahwa kabupaten/kota di Klastrer 2 adalah daerah transisi yang

memiliki profil *hybrid*, artinya tidak cukup unggul (Klaster 1) dan tidak cukup tertinggal (Klaster 3). Secara statistik, hal ini menyebabkan posisi anggota Klaster 2 berada lebih dekat dengan batas antar-klaster, sehingga menghasilkan nilai siluet yang lebih rendah. Oleh karena itu, skor siluet yang lemah namun dapat diterima (rata-rata koefisien siluet=0,472) ini justru memvalidasi interpretasi teoretis tentang adanya kelompok menengah dalam struktur spasial kinerja pariwisata di NTT.

Meskipun skor siluet menunjukkan bahwa pemisahan antar klaster relatif lemah, nilai ARI yang tinggi mengindikasikan bahwa struktur klaster yang dihasilkan tetap konsisten dengan pola referensi. Oleh karena itu, analisis selanjutnya dilakukan untuk menelaah karakteristik masing-masing klaster melalui Plot Profil *Centroid K-Means*. Gambar 7 menunjukkan bahwa disparitas kinerja pariwisata di NTT didorong oleh perbedaan ekstrem pada Akses Eksternal (x5) dan Permintaan Pasar (x1).



Gambar 7. Plot Profil Centroid Klaster

Klaster 1 (hijau) dengan jelas menduduki posisi sebagai kelompok unggul atau berkinerja tinggi. Klaster ini, yang diwakili oleh kabupaten/kota seperti Manggarai Barat, Flores Timur, dan Kota Kupang, dicirikan oleh nilai rata-rata ternormalisasi yang tinggi pada hampir semua dimensi, terutama pada Efektivitas Penawaran (x2) dan Akses Eksternal (x5), mengindikasikan bahwa daerah-daerah ini berhasil menarik Permintaan Pasar (x1) yang tinggi dan mengoptimalkan tingkat hunian akomodasi, didukung oleh konektivitas yang sangat baik.

Kontras dengan Klaster 1, Klaster 3 (merah) mewakili kelompok tertinggal atau berkinerja rendah. Anggota klaster ini seperti Sabu Raijua, Malaka, dan Sumba Tengah, menunjukkan nilai rata-rata untuk klaster pada titik terendah (negatif) di seluruh

variabel, terutama pada Efektivitas Penawaran (x2). Profil ini menunjukkan bahwa wilayah-wilayah di Klaster 3 menghadapi tantangan mendasar di seluruh rantai nilai pariwisata, mulai dari kurangnya Permintaan Pasar (x1), rendahnya Tingkat Penghunian Kamar (x2), hingga terbatasnya Akses Eksternal (x5).

Sementara itu, Klaster 2 (biru) yang diwakili oleh kabupaten seperti Ende, Manggarai, dan Sumba Barat berperan sebagai kelompok menengah atau daerah transisi yang menunjukkan profil *hybrid*. Secara umum, kinerja mereka berada di antara klaster unggul dan tertinggal. Namun, klaster ini menunjukkan anomali penting, yaitu memiliki nilai rata-rata yang relatif tinggi pada Kapasitas Penawaran (x3), bahkan melebihi Klaster 1, tetapi Efektivitas Penawaran (x2) mereka tetap rendah. Realitas ini merefleksikan bahwa kabupaten/kota di Klaster 2 telah melakukan investasi yang signifikan pada sisi penawaran (akomodasi) yang besar, namun belum mampu diimbangi oleh permintaan dan optimalisasi tingkat hunian. Profil *hybrid* ini sejalan dengan ambiguitas statistik yang tercermin dalam skor Siluet Klaster 2 (Gambar 5), di mana penugasan mereka kurang definitif karena profilnya tidak murni unggul dan juga tidak sepenuhnya tertinggal. Secara keseluruhan, plot profil *centroid* ini memberikan dasar yang kuat untuk perumusan kebijakan yang ditargetkan pada kebutuhan spesifik masing-masing klaster.

4. Simpulan dan Saran

Penelitian ini bertujuan mengklasifikasikan 22 kabupaten/kota di NTT berdasarkan indikator multidimensi kinerja pariwisata untuk memberikan dasar empiris bagi kebijakan pembangunan yang lebih terarah. Menggunakan analisis *K-Means clustering* pada enam variabel yang mencakup aspek Permintaan Pasar, Efektivitas dan Kapasitas Penawaran, Dampak Ekonomi, serta Aksesibilitas, ditemukan bahwa kabupaten/kota di NTT dapat dikelompokkan menjadi tiga klaster dengan profil karakteristik yang sangat berbeda. Temuan utama dari analisis klaster ini adalah adanya ketimpangan spasial yang signifikan dan terstruktur dalam kinerja pariwisata daerah.

1. Klaster 1 (unggul): kelompok ini merepresentasikan *top performer* pariwisata. Kabupaten/kota yang masuk dalam klaster ini adalah Kota Kupang, Manggarai Barat, Flores Timur, Sumba Timur, Belu, Timor Tengah Utara, Kabupaten Kupang, dan Timor Tengah Selatan. Anggotanya dicirikan oleh kinerja tertinggi di hampir seluruh variabel, khususnya Efektivitas Penawaran (TPK) dan Akses Eksternal. Klaster ini menunjukkan ekosistem pariwisata yang matang, di mana permintaan pasar tinggi bertemu dengan aksesibilitas dan operasional akomodasi yang optimal, menghasilkan dampak ekonomi yang superior.
2. Klaster 3 (tertinggal): kelompok ini adalah *bottom performer* dengan kinerja terendah di semua dimensi. Kabupaten yang masuk dalam klaster ini adalah Lembata, Sabu Raijua, Alor, Rote Ndao, Sumba Tengah, Malaka, dan Manggarai Timur. Skor rata-rata yang sangat rendah, terutama pada Efektivitas Penawaran dan Permintaan Pasar, mengindikasikan adanya keterbatasan mendasar pada

seluruh rantai nilai pariwisata. Klaster ini secara fundamental tertinggal dan memerlukan intervensi pembangunan pariwisata dari tahap awal.

3. Klaster 2 (menengah/transisi): kelompok ini memiliki profil *hybrid* yang berfungsi sebagai zona transisi dan menunjukkan ambiguitas statistik yang valid. Kabupaten yang masuk dalam klaster ini adalah Ende, Manggarai, Ngada, Sikka, Sumba Barat, dan Nagekeo. Klaster ini teridentifikasi memiliki Kapasitas Penawaran (Jumlah Akomodasi) yang relatif tinggi, namun gagal mengkonversi kapasitas tersebut menjadi Efektivitas Penawaran yang optimal (TPK rendah). Realitas ini menegaskan bahwa daerah ini telah berinvestasi pada *supply side* tetapi menghadapi tantangan besar pada *demand side* dan manajemen operasional.

Berdasarkan temuan tersebut, adapun saran-saran yang direkomendasikan adalah sebagai berikut:

1. Penting bagi pemerintah daerah untuk menghindari penerapan kebijakan pariwisata yang seragam. Sebaliknya, alokasi sumber daya dan prioritas program perlu ditentukan secara efektif dengan mempertimbangkan perbedaan mendasar pada profil setiap klaster, yaitu:
 - Untuk kabupaten pada klaster 2, kebijakan harus difokuskan pada peningkatan kualitas layanan dan penciptaan permintaan untuk memanfaatkan kapasitas penawaran (x3) yang sudah tersedia.
 - Untuk kabupaten pada Klaster 3, anggaran harus diprioritaskan untuk pemenuhan prasyarat dasar pariwisata, yaitu peningkatan aksesibilitas eksternal dan konektivitas internal.
 - Untuk kabupaten/kota pada Klaster 1, kebijakan harus diarahkan pada aspek diversifikasi produk, pengembangan SDM, dan menjaga daya dukung pariwisata agar pertumbuhan yang sudah dicapai tetap berkelanjutan.
2. Penelitian berikutnya disarankan untuk mengembangkan model klastering yang lebih mendalam dengan memisahkan data akomodasi berdasarkan peringkat bintang, memasukkan data kunjungan wisatawan mancanegara (Wisman), data Objek Daya Tarik Wisata (ODTW), dan *Tourism Satellite Account* (TSA) guna memahami peran devisa dan efisiensi segmen pasar premium secara lebih rinci.

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] Ahmed, M. (2020). *The k-means Algorithm: A Comprehensive Survey and Performance Evaluation*. *Electronics*, 9(8), 1295. <https://doi.org/10.3390/electronics9081295>
- [2] Amanda, L. F. P., & Setiawan, A. B. (2022). *Grouping the Number of Foreign Tourist Visits and Analysis of Tourism Competitiveness in East Java*. *Efficient: Indonesian Journal of Development Economics*, 5(3), 103–114.
- [3] Anggarini, F. (2021). *Peran sektor pariwisata terhadap pertumbuhan ekonomi di ASEAN 4 (Thailand, Singapore, Malaysia, dan Indonesia)*. *Jurnal Industri Pariwisata*, 6(2), 1-13. <https://doi.org/10.36441/pariwisata.v7i1.2059>
- [4] Badan Pusat Statistik. (2021). *Indonesia Tourism Satellite Accounts (TSA) 2016-2019*. Jakarta: Badan Pusat Statistik.
- [5] Costa, A. C. da R. (2013). *Toward a Formal Reconstruction of Kelsen's Theory of Legal Systems*. *2013 2nd Workshop-School on Theoretical Computer Science*, 165–171. <https://doi.org/10.1109/WEIT.2013.23>
- [6] Hariyani, H. F. (2022). *Tourism Sector Performance on Indonesia's Economic Growth*. *Jurnal Ekonomi Pembangunan*, 20(2), 145–158. <https://doi.org/10.22219/jep.x16i1.8184>
- [7] Jain, A. K. (2010). *Data clustering: 50 years beyond K-Means*. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- [8] Kaufman, L., & Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- [9] Kementerian Koordinator Bidang Perekonomian Republik Indonesia. (2025, Maret 13). *Mendongkrak kinerja sektor pariwisata, pemerintah siapkan berbagai program lintas sektor*. Kementerian Koordinator Bidang Perekonomian Republik Indonesia. <https://www.ekon.go.id/publikasi/detail/6253/mendongkrak-kinerja-sektor-pariwisata-pemerintah-siapkan-berbagai-program-lintas-sektor>
- [10] Kementerian Pariwisata dan Ekonomi Kreatif (Kemenparekraf). (2023). *Outlook Pariwisata dan Ekonomi Kreatif Indonesia 2023/2024*. Jakarta: Kemenparekraf.
- [11] Maori, N. A. & Evanita, E. (2023). *Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering*. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 14(2), 277-288. <https://doi.org/10.24176/simet.x14i2.9630>
- [12] Masteriarsa, M. F., & Riyanto, R. (2023). *Tourism destination mapping based on tourism characteristics and carrying capacity of province in Indonesia*. *Jurnal Perencanaan Pembangunan: The Indonesian Journal of Development Planning*, 7(3), 344-361. <https://doi.org/10.36574/jpp.v7i3.460>

- [13] McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform manifold approximation and projection for dimension reduction*. arXiv preprint arXiv:1802.03426. <https://arxiv.org/abs/1802.03426>
- [14] Mertayasa, I. K. T., & Atmaja, I. D. M. B. (2022). *Pemodelan topik pada ulasan hotel menggunakan metode BERTopic dengan prosedur c-TF-IDF*. *Jurnal Nasional Teknologi Informasi dan Aplikasinya (JNATIA)*, 1(1), 307–316. <https://doi.org/10.24843/JNATIA.2022.v01.i01.p37>
- [15] Mukhtar, H., Syafriadi, S., & Rahmadani, R. (2024). *Algoritma K-Means untuk pengelompokan perilaku customer*. *SEIS: Journal of Software Engineering and Information Systems*, 4(2), 45–54. <https://doi.org/10.37859/seis.x4i2.7615>
- [16] Nuraeni, S., & Hendriyani, I. (2022). *A Competitive Advantage Model for Indonesia's Sustainable Tourism Destinations from Supply and Demand Side Perspectives*. *Sustainability*, 14(24), 16398. MDPI. <https://doi.org/10.3390/su142416398>
- [17] Romano, S., Vinh, N. X., Bailey, J., & Verspoor, K. (2016). *Adjusting for chance clustering comparison measures*. *Journal of Machine Learning Research*, 17(1), 1–32. <https://doi.org/10.48550/arXiv.1512.01286>
- [18] Rousseeuw, P. J. (1987). *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- [19] Saputra, D. M., Saputra, D., & Oswari, L. D. (2020). *Effect of distance metrics in determining K-value in K-means clustering using elbow and silhouette method*. In *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)* (pp. 341–346). Atlantis Press. <https://doi.org/10.2991/aisr.k.200424.051>
- [20] Siagian, A. R., & Wulandari, D. (2025). *Penerapan jarak Euclidean pada pengelompokan data menggunakan algoritma K-Means* [Skripsi, Universitas Islam Negeri Sultan Syarif Kasim Riau]. UIN SUSKA Repository. <https://repository.uin-suska.ac.id/85603/>
- [21] Suhandi, N., Gustriansyah, R., & Destria, A. (2025). *Klasifikasi penyakit TBC menggunakan metode UMAP dan K-NN*. *Bit-Tech (Binary Digital - Technology)*, 7(3). <https://doi.org/10.32877/bt.v7i3.2227>
- [22] Suhardiman, M. Y. (2025). *Clustering Tourism Hotspots in Indonesia Using Hotel Guest Data and the K-Means Algorithm*. *IJRESM*, 8(6), 106-111.
- [23] Tan, P.-N., Steinbach, M., & Kumar, V. (2014). *Introduction to data mining*. Pearson Education Limited.