

PERBANDINGAN ALGORITMA K-MEANS DAN HIERARCHICAL CLUSTERING DALAM SEGMENTASI KABUPATEN/KOTA DI JAWA TIMUR BERDASARKAN DATA PERJALANAN DAN PERGERAKAN WISATAWAN

Sella Brina Winona Rizky
Sistem Informasi, Universitas Merdeka Malang
Jalan Terusan Dieng, Malang, Indonesia
rizky.sella@student.unmer.ac.id

ABSTRAK

Pariwisata merupakan sektor penting bagi perekonomian Jawa Timur, namun distribusi kunjungan wisatawan antar kabupaten/kota masih timpang. Promosi wisata yang tidak tersegmentasi menyebabkan beberapa daerah tertinggal dalam pengembangan destinasi. Penelitian ini bertujuan melakukan segmentasi wilayah di Jawa Timur berdasarkan jumlah perjalanan dan pergerakan wisatawan nusantara tahun 2023-2024, sebagai dasar perumusan strategi promosi yang lebih efektif. Data diperoleh dari BPS dan portal opendata.jatimprov.go.id, diproses melalui pembersihan, agregasi, transformasi, dan normalisasi menggunakan *Min-Max Scaler*. Dua algoritma *clustering*, yaitu *K-Means* dan *Hierarchical Clustering* dengan pendekatan *Agglomerative Ward Linkage*, digunakan untuk segmentasi wilayah. Jumlah kluster optimal ditentukan sebanyak tiga menggunakan *Elbow Method* dan *dendrogram*. Evaluasi hasil segmentasi menunjukkan bahwa *Hierarchical Clustering* lebih unggul berdasarkan nilai *Silhouette Score* (0.5689) dan *Dunn Index* (0.2396), sedangkan *K-Means* lebih tinggi pada *Calinski-Harabasz Index* (39.23). Segmentasi ini mengidentifikasi Kota Surabaya sebagai *outlier*, serta membagi wilayah lain ke dalam kluster sedang dan rendah, yang masing-masing direkomendasikan strategi promosi berbeda.

Kata kunci : segmentasi, clustering, K-Means, Hierarchical Clustering, Jawa Timur, Pariwisata

1. PENDAHULUAN

Pertumbuhan sektor pariwisata Jawa Timur mengalami dinamika pesat dalam beberapa tahun terakhir, namun distribusi kunjungan wisatawan masih timpang; beberapa daerah sangat ramai, sementara daerah lain sepi. Kebijakan promosi yang cenderung umum, tanpa memperhatikan karakteristik lokal, menyebabkan potensi wisata di tingkat kabupaten/kota belum dimaksimalkan. Oleh karena itu, diperlukan pendekatan *data-driven segmentation* untuk mengelompokkan wilayah berdasarkan pola kunjungan mereka. Jawa Timur sendiri memiliki banyak potensi wisata yang perlu dipelajari lebih dalam agar dapat menjadi daya tarik ekonomi yang signifikan [1].

Permasalahan yang muncul adalah belum adanya segmentasi wilayah berbasis data kunjungan wisatawan yang digunakan sebagai dasar strategi promosi pariwisata daerah. Segmentasi semacam ini sebenarnya kerap diterapkan dalam bidang pemasaran dan pariwisata (*market segmentation*), namun di Jawa Timur, studi yang memanfaatkan model *unsupervised learning* untuk segmentasi wilayah wisata masih terbatas.

Penelitian ini bertujuan untuk mengelompokkan kabupaten/kota di Jawa Timur berdasarkan jumlah perjalanan dan pergerakan wisatawan nusantara menggunakan metode *clustering*. Dengan pendekatan ini, diharapkan pola kunjungan antar wilayah dapat dianalisis secara objektif dan menjadi dasar penyusunan strategi promosi yang lebih tepat sasaran.

Sebagai rencana penyelesaian masalah, penelitian ini menggunakan dua metode *clustering*,

yaitu *K-Means* dan *Hierarchical Clustering*. *K-Means* dikenal memiliki keunggulan dalam hal efisiensi [2]. Sementara *Hierarchical Clustering* unggul dalam fleksibilitas visualisasi struktur data melalui *dendrogram*. Data tahunan wisatawan domestik tahun 2023-2024 diperoleh dari BPS dan portal opendata.jatimprov.go.id, yang awalnya berbentuk data bulanan. Data tersebut diolah melalui tahapan *data cleaning*, agregasi, transformasi, dan normalisasi *Min-Max Scaler*. Penentuan jumlah kluster dilakukan menggunakan *Elbow Method* dan *dendrogram*, dan hasil evaluasi segmentasi diuji dengan *Silhouette Score*, *Calinski-Harabasz Index*, dan *Dunn Index*.

Hasil akhir diharapkan mampu mengelompokkan wilayah wisata secara efektif, sehingga dapat membantu pemerintah dalam menganalisis pola kunjungan dan menyusun strategi promosi yang lebih terarah.

2. TINJAUAN PUSTAKA

Beberapa penelitian terdahulu telah membahas penggunaan algoritma *clustering* dalam segmentasi wisata. Maulida et al, [3] menerapkan algoritma *K-Means* untuk segmentasi wisatawan di Desa Lon Malang, Madura dan menemukan bahwa hasil kluster dapat membantu menyusun strategi promosi berbasis karakteristik pengunjung. Penelitian oleh Zuhal, [4] membandingkan performa *K-Means* dan *Hierarchical Clustering* pada lima dataset berbeda, dan menemukan bahwa *Hierarchical Clustering* menghasilkan nilai *DBI (Davies-Bouldin Index)* lebih rendah, menunjukkan hasil kluster lebih optimal. Kemudian penelitian oleh Azizah et al, [5] mengkaji komparasi

algoritma *K-Means* dan *Hierarchical Clustering* pada data kehadiran karyawan, dan menemukan bahwa *K-Means* menghasilkan metrik evaluasi yang lebih baik dalam konteks tersebut.

Temuan-temuan tersebut menguatkan bahwa kedua algoritma memiliki keunggulan masing-masing, tergantung pada karakteristik data dan tujuan segmentasi. Oleh karena itu, penelitian ini membandingkan langsung dua algoritma tersebut dalam konteks segmentasi promosi wisata berbasis data.

2.1. Pariwisata dan Wisatawan Nusantara

Pariwisata merupakan aktivitas rekreasi di luar tempat tinggal yang memanfaatkan berbagai layanan. Sektor ini tidak hanya berkontribusi pada ekonomi nasional, tetapi juga memenuhi kebutuhan fisik, spiritual, dan intelektual wisatawan. Indonesia memiliki daya tarik wisata yang tinggi berkat kekayaan budaya dan alamnya [6].

Wisatawan nusantara (wisnus) didefinisikan sebagai individu yang melakukan perjalanan domestik selama kurang dari enam bulan, tanpa tujuan pendidikan atau pekerjaan. Perjalanan tersebut mencakup kunjungan ke objek wisata, akomodasi komersial, serta menempuh jarak minimal 100 km pulang-pergi.

Faktor yang memengaruhi pergerakan wisnus antara lain meningkatnya tren perjalanan (sekitar 2% per tahun), berwisata sebagai gaya hidup, kemajuan teknologi informasi yang membuat destinasi mudah dikenal, serta peningkatan aksesibilitas transportasi.

2.2. Strategi Promosi Wisata

Strategi merupakan seni merancang langkah-langkah untuk mencapai tujuan tertentu. Dalam konteks pemasaran, promosi berfungsi sebagai alat komunikasi persuasif kepada pelanggan potensial untuk mendorong terjadinya transaksi antara perusahaan dan konsumen [7].

Strategi promosi pemasaran merupakan serangkaian tindakan terencana yang memanfaatkan elemen-elemen promosi seperti periklanan, promosi penjualan, hubungan masyarakat, dan penjualan langsung. Tujuannya adalah menarik minat konsumen untuk melakukan pembelian atau pertukaran produk dan jasa.

Dalam industri pariwisata, promosi memiliki peran krusial. Keberhasilan suatu destinasi sangat bergantung pada efektivitas promosi yang dilakukan, sehingga diperlukan strategi promosi yang tepat dan terarah [8]. Saat ini, aktivitas manusia yang tak terpisahkan dari internet menjadikan media sosial sebagai sarana utama berbagi momen perjalanan, yang secara tidak langsung menjadi sumber informasi bagi wisatawan lainnya [9].

2.3. Segmentasi Wilayah Pada Pariwisata

Segmentasi wilayah dalam pariwisata merupakan pendekatan strategis untuk mengelompokkan daerah

berdasarkan karakteristik aktivitas wisata, seperti jumlah kunjungan dan potensi daya tarik. Pendekatan ini penting dalam menyusun strategi promosi yang lebih tepat sasaran dan berbasis data.

Segmentasi dapat dimanfaatkan untuk mengidentifikasi daerah dengan tingkat kunjungan yang berbeda, mengenali wilayah potensial yang belum tergarap optimal, serta merancang strategi pengembangan wisata sesuai karakteristik masing-masing wilayah.

Penelitian ini menerapkan segmentasi wilayah berdasarkan data perjalanan dan pergerakan wisatawan nusantara di kabupaten/kota Provinsi Jawa Timur selama tahun 2023-2024. Kedua variabel tersebut digunakan untuk membentuk kelompok wilayah dengan tingkat intensitas kunjungan yang serupa. Hasil segmentasi diharapkan mampu memberikan gambaran distribusi wisatawan yang lebih akurat sebagai dasar pengambilan keputusan dalam promosi dan pengembangan pariwisata daerah.

2.4. Data Mining

Data mining merupakan proses klasifikasi dan pencarian pola dari kumpulan data berukuran besar [10]. Proses ini menjadi bagian penting dalam tahapan *Knowledge Discovery in Databases (KDD)*, yang mencakup seleksi, transformasi, analisis, dan evaluasi data. Penelitian ini menerapkan teknik data mining untuk mengelompokkan kabupaten/kota berdasarkan jumlah perjalanan dan pergerakan wisatawan.

2.5. Algoritma K-Means Clustering

K-Means merupakan algoritma *clustering* yang populer karena kesederhanaannya dan efisiensi komputasi yang tinggi. Algoritma ini membagi data ke dalam k klaster berdasarkan kedekatan terhadap pusat klaster (*centroid*), dengan proses iteratif hingga posisi *centroid* stabil.

Kelebihan *K-Means* terletak pada kemampuannya menangani data dalam skala besar dengan cepat dan efisien. Algoritma ini efektif digunakan dalam segmentasi wilayah berbasis data numerik maupun spasial, seperti distribusi kunjungan wisatawan.

2.6. Algoritma Hierarchical Clustering

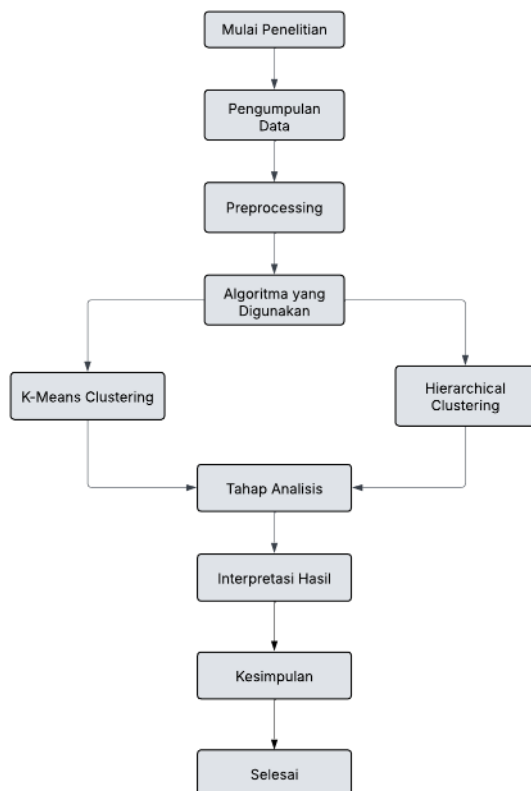
Hierarchical clustering merupakan metode pengelompokan data secara bertingkat yang membentuk struktur klaster dalam bentuk hierarki. Terdapat dua pendekatan utama, yaitu *Agglomerative* (bottom-up) yang menggabungkan objek dari individu menjadi satu klaster besar, dan *Divisive* (top-down) yang memecah satu kelompok besar menjadi klaster-klaster kecil.

Penelitian ini menggunakan metode *Agglomerative Hierarchical Clustering* dengan pendekatan jarak *Euclidean Distance*. Pada tahap awal, setiap data dianggap sebagai satu klaster tersendiri. Selanjutnya, dua klaster dengan jarak terdekat digabung secara bertahap hingga seluruh data

membentuk satu kluster tunggal. Proses ini divisualisasikan dalam bentuk *dendrogram* untuk menunjukkan struktur hierarki penggabungan.

Untuk menentukan jarak antar kluster setelah penggabungan, digunakan metode *Ward*, yang bertujuan meminimalkan total *sum of squared errors* (SSE) dalam setiap langkah penggabungan. Metode ini mempertimbangkan peningkatan terkecil dalam SSE, sehingga menghasilkan kluster yang homogen di dalam dan heterogen antar kluster.

3. METODE PENELITIAN



Gambar 1. Diagram alur penelitian

Penelitian ini merupakan studi kuantitatif dengan pendekatan komparatif yang bertujuan membandingkan performa algoritma *K-Means* dan *Hierarchical Clustering* dalam segmentasi kabupaten/kota berdasarkan jumlah perjalanan dan pergerakan wisatawan nusantara. Segmentasi ini digunakan untuk mengidentifikasi pola kunjungan yang relevan dalam perumusan strategi promosi pariwisata. Penelitian bersifat eksploratif dengan memanfaatkan data sekunder dari Badan Pusat Statistik (BPS). Seperti yang terlihat pada Gambar 1, diagram alur penelitian menggambarkan tahapan proses yang dilakukan mulai dari pengumpulan data, pra-pemrosesan, penerapan algoritma klustering, hingga evaluasi hasil segmentasi.

3.1. Data Preprocessing

Pra-pemrosesan data merupakan tahap awal yang penting dalam *data mining* untuk memastikan bahwa

data yang digunakan bersih, relevan, dan siap dianalisis. Proses ini dilakukan melalui empat langkah utama: *cleaning*, *agregasi*, *transformasi*, dan *normalisasi*.

Data cleaning bertujuan meningkatkan kualitas data dengan menghapus duplikat, memperbaiki format tidak seragam, serta menangani nilai kosong pada kolom penting seperti *Kabupaten/Kota*, *Jumlah Perjalanan*, dan *Jumlah Pergerakan*. Karena data berasal dari sumber sekunder seperti BPS, proses ini penting untuk memastikan keakuratan analisis.

Data bulanan dari tahun 2023 dan 2024 kemudian diagregasi menjadi nilai total tahunan per kabupaten/kota. Agregasi dilakukan dengan menjumlahkan seluruh data perjalanan dan pergerakan per bulan, sehingga menghasilkan satu nilai kumulatif per wilayah. Nilai agregat ini menjadi fitur utama dalam proses segmentasi.

Transformasi data dilakukan dengan menyatukan dua variabel utama jumlah perjalanan dan jumlah pergerakan ke dalam satu tabel utama. Data yang semula dalam format *time series* diubah menjadi format per wilayah, dengan struktur dua kolom numerik dan satu kolom kategori. Format yang seragam ini penting agar dapat dibaca dan diproses oleh algoritma *clustering* di Python.

Normalisasi dilakukan menggunakan metode *Min-Max Scaling* untuk mengubah skala nilai ke dalam rentang 0-1. Langkah ini diperlukan agar tidak ada variabel yang mendominasi perhitungan jarak pada algoritma *K-Means* dan *Hierarchical Clustering*. Proses ini dilakukan dengan Python melalui pustaka *Scikit-learn*.

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

X = nilai asli

X_{min} = nilai minimum

X_{max} = nilai maksimum

X_{norm} = nilai setelah normalisasi

Dengan data yang telah dibersihkan, diagregasi, ditransformasikan, dan dinormalisasi, seluruh dataset siap digunakan dalam proses analisis *clustering*.

3.2. Penerapan Algoritma K-Means Clustering

Clustering adalah proses mengelompokkan data berdasarkan kemiripan. Dalam penelitian ini, algoritma *K-Means* digunakan untuk mengelompokkan kabupaten/kota di Jawa Timur berdasarkan jumlah perjalanan wisatawan nusantara [11]. Tujuannya adalah mengidentifikasi pola kunjungan wisatawan sebagai dasar strategi promosi.

Pemilihan jumlah kluster dilakukan dengan metode *Elbow* dan *Silhouette Score*. Setelah jumlah kluster ditentukan, *K-Means* diterapkan menggunakan jarak *Euclidean* secara iteratif hingga *centroid* stabil. Hasil segmentasi dianalisis untuk melihat wilayah dengan kunjungan tinggi, sedang, dan rendah, serta divisualisasikan dalam bentuk grafik.

3.3. Penerapan Algoritma Hierarchical Clustering
Hierarchical Clustering adalah metode *unsupervised learning* yang membentuk struktur klaster secara bertingkat dalam bentuk dendrogram. Tidak seperti *K-Means*, metode ini tidak memerlukan jumlah klaster di awal dan membentuk klaster secara bertahap melalui pendekatan *agglomerative*.

Dalam penelitian ini, *Hierarchical Clustering* digunakan sebagai pembandingan terhadap *K-Means* untuk segmentasi kabupaten/kota di Jawa Timur berdasarkan jumlah perjalanan wisatawan nusantara. Data yang telah dinormalisasi diproses menggunakan jarak *Euclidean* dan metode penggabungan *Ward linkage*. *Dendrogram* kemudian digunakan untuk menentukan jumlah klaster optimal.

Hasil segmentasi dibandingkan dengan *K-Means* untuk menilai konsistensi dan kualitas pengelompokan wilayah.

3.4. Teknik Analisis Data

Peneliti menggunakan tiga metrik evaluasi untuk membandingkan performa algoritma *K-Means* dan *Hierarchical Clustering*, yaitu *Silhouette Score*, *Calinski-Harabasz Index*, dan *Dunn Index*. Tujuannya adalah untuk menentukan metode yang menghasilkan segmentasi wilayah paling optimal dalam konteks promosi pariwisata. *Silhouette Score* mengukur konsistensi data dalam klaster, dengan nilai antara -1 hingga 1; semakin tinggi, semakin baik. *Calinski-Harabasz Index* menilai kualitas klaster berdasarkan rasio variansi antar dan dalam klaster, sedangkan *Dunn Index* mengevaluasi jarak minimum antar klaster dan kekompakan dalam klaster. Nilai tinggi pada ketiga metrik menunjukkan segmentasi yang lebih baik dan terpisah jelas.

3.5. Interpretasi Segmentasi

Hasil klusterisasi dengan *K-Means* dan *Hierarchical Clustering* menghasilkan tiga kelompok wilayah di Jawa Timur berdasarkan jumlah perjalanan dan pergerakan wisatawan nusantara tahun 2023-2024. Setiap klaster mencerminkan pola kunjungan yang berbeda. Klaster pertama berisi daerah dengan kunjungan tinggi, cocok untuk strategi retensi dan promosi berkelanjutan. Klaster kedua mencakup wilayah dengan kunjungan sedang, yang membutuhkan penguatan promosi lokal. Sementara itu, klaster ketiga terdiri dari daerah dengan kunjungan rendah dan perlu intervensi berupa promosi masif dan pengembangan infrastruktur. Segmentasi ini diharapkan menjadi dasar strategi promosi yang lebih terarah dan efisien.

4. HASIL DAN PEMBAHASAN

4.1. Data Preprocessing

Pada proses data cleaning, data dibersihkan menggunakan Microsoft Excel untuk memastikan data bebas dari duplikat, nilai kosong, dan inkonsistensi format pada kolom penting seperti nama wilayah, jumlah perjalanan, dan jumlah pergerakan. Data hasil

pembersihan ini kemudian digunakan dalam tahap analisis.

Gambar 2. Data mentah jumlah perjalanan wisatawan nusantara tahun 2023

Gambar 2 menunjukkan data mentah kunjungan wisatawan tahun 2023 sebelum dilakukan proses pembersihan data.

Gambar 3. Data mentah jumlah perjalanan wisatawan nusantara tahun 2024

Gambar 3 menampilkan kondisi awal data kunjungan wisatawan tahun 2024 yang belum dibersihkan.

Gambar 4. Data mentah jumlah pergerakan wisatawan nusantara tahun 2023 dan 2024

Gambar 4 menggambarkan data pergerakan wisatawan selama dua tahun yang masih dalam bentuk mentah.

Gambar 5. Hasil proses cleaning data jumlah perjalanan wisatawan tahun 2023

Gambar 5 menampilkan data jumlah perjalanan tahun 2023 yang telah melalui proses *data cleaning* untuk menghilangkan nilai ganda, kosong, dan inkonsistensi.

Gambar 6. Hasil proses cleaning data jumlah perjalanan wisatawan tahun 2024

Gambar 6 menunjukkan hasil pembersihan data perjalanan wisatawan tahun 2024 yang telah siap untuk dianalisis.

	A	B	C	D	E
1	No	Kabupaten/Kota	Jumlah Pergerakan	Bulan	Tahun
2	1	BANGKALAN	159267	JANUARI	2023
3	2	BANYUWANGI	235570	JANUARI	2023
4	3	BLITAR	258054	JANUARI	2023
5	4	BOJONEGORO	48377	JANUARI	2023
6	5	BONDOWOSO	48298	JANUARI	2023
7	6	GRESIK	374913	JANUARI	2023
8	7	JEMBER	94120	JANUARI	2023
9	8	JOMBANG	148911	JANUARI	2023
10	9	KEDIRI	247796	JANUARI	2023
11	10	LAMONGAN	391333	JANUARI	2023
12	11	LUMAJANG	70733	JANUARI	2023
13	12	MADIUN	60025	JANUARI	2023

Gambar 7. Hasil proses cleaning data jumlah pergerakan wisatawan tahun 2023 dan 2024

Gambar 7 memperlihatkan data pergerakan wisatawan tahun 2023 dan 2024 setelah melalui tahapan pembersihan data. Data bulanan tahun 2023 dan 2024 kemudian diagregasi menjadi total tahunan per kabupaten/kota. Proses agregasi dilakukan menggunakan fitur pengelompokan dan penjumlahan di Excel, sehingga diperoleh satu nilai total perjalanan dan pergerakan wisatawan per wilayah untuk dua tahun.

	A	B	C	D
1	No	Kabupaten/Kota	Jumlah Perjalanan	Tahun
2	1	Kabupaten Pacitan	5729284	2023
3	2	Kabupaten Ponorogo	5328060	2023
4	3	Kabupaten Trenggalek	6826622	2023
5	4	Kabupaten Tulungagung	4366574	2023
6	5	Kabupaten Blitar	5500933	2023
7	6	Kabupaten Kediri	6269490	2023
8	7	Kabupaten Malang	11138805	2023
9	8	Kabupaten Lumajang	6002527	2023
10	9	Kabupaten Jember	9264929	2023
11	10	Kabupaten Banyuwangi	8092664	2023
12	11	Kabupaten Bondowoso	3986544	2023

Gambar 8. Hasil agregasi jumlah perjalanan wisatawan nusantara tahun 2023 per kabupaten/kota

Gambar 8 menampilkan total akumulasi perjalanan wisatawan nusantara sepanjang tahun 2023 yang telah dijumlahkan berdasarkan masing-masing kabupaten/kota.

	A	B	C	D
1	No	Kabupaten/Kota	Jumlah Perjalanan	Tahun
2	1	Kabupaten Pacitan	2743332	2024
3	2	Kabupaten Ponorogo	3612230	2024
4	3	Kabupaten Trenggalek	2989379	2024
5	4	Kabupaten Tulungagung	4523017	2024
6	5	Kabupaten Blitar	5086653	2024
7	6	Kabupaten Kediri	7832201	2024
8	7	Kabupaten Malang	14629462	2024
9	8	Kabupaten Lumajang	3472167	2024
10	9	Kabupaten Jember	6886000	2024
11	10	Kabupaten Banyuwangi	5430929	2024
12	11	Kabupaten Bondowoso	2272587	2024
13	12	Kabupaten Situbondo	2938730	2024

Gambar 9. Hasil agregasi jumlah perjalanan wisatawan nusantara tahun 2024 per kabupaten/kota

Gambar 9 menunjukkan hasil agregasi data perjalanan wisatawan tahun 2024 yang dihitung secara total untuk setiap wilayah administratif.

	A	B	C
1	No	Kabupaten/Kota	Jumlah Pergerakan Tahun 2023-2024
2	1	JEMBER	1513011
3	2	BOJONEGORO	971743
4	3	BANYUWANGI	4936672
5	4	GRESIK	7644466
6	5	JOMBANG	3329090
7	6	BANGKALAN	9300481
8	7	BLITAR	5768284
9	8	BONDOWOSO	1029243
10	9	KEDIRI	5148419
11	10	KOTA BATU	14544072
12	11	KOTA BLITAR	4275123
13	12	KOTA KEDIRI	1178740

Gambar 10. Hasil agregasi jumlah pergerakan wisatawan nusantara tahun 2023-2024

Gambar 10 memperlihatkan hasil penjumlahan total pergerakan wisatawan selama dua tahun terakhir untuk masing-masing kabupaten/kota di Jawa Timur.

Data hasil agregasi kemudian ditransformasikan menjadi satu tabel utama. Dua variabel, yaitu jumlah perjalanan dan jumlah pergerakan, digabung berdasarkan nama kabupaten/kota sebagai kunci utama. Tabel akhir terdiri dari tiga kolom: *Kabupaten/Kota*, *Jumlah Perjalanan*, dan *Jumlah Pergerakan*, dengan format yang disederhanakan dan kolom tidak relevan dihapus.

	A	B	C	D
1	No	Kabupaten/Kota	Jumlah Perjalanan Tahun 2023-2024	Jumlah Pergerakan Tahun 2023-2024
2	1	Kabupaten Bangkalan	9645244	9300481
3	2	Kabupaten Banyuwangi	13523593	4936672
4	3	Kabupaten Blitar	10587586	5768284
5	4	Kabupaten Bojonegoro	10584435	971743
6	5	Kabupaten Bondowoso	6259131	1029243
7	6	Kabupaten Gresik	14408732	7644466
8	7	Kabupaten Jember	16150929	1513011
9	8	Kabupaten Jombang	11208615	3329090
10	9	Kabupaten Kediri	14101691	5148419
11	10	Kabupaten Lamongan	10537064	8249401
12	11	Kabupaten Lumajang	9474694	1770594
13	12	Kabupaten Madiun	7144725	826590

Gambar 11. Hasil transformasi data jumlah perjalanan dan pergerakan wisatawan tahun 2023-2024 per kabupaten/kota

Gambar 11 menunjukkan hasil transformasi data setelah variabel jumlah perjalanan dan pergerakan digabungkan ke dalam satu tabel utama berdasarkan kabupaten/kota, sehingga data siap digunakan dalam proses analisis klustering.

Data yang telah ditransformasi kemudian dinormalisasi menggunakan metode *Min-Max Scaling* dengan *Python* di *Google Colab*. Proses ini menyetarakan skala antar variabel dalam rentang 0 hingga 1, agar tidak ada variabel yang mendominasi dalam proses *clustering*.

```
# normalisasi data
from sklearn.preprocessing import MinMaxScaler

# Ambil kolom numerik
X = df[['Jumlah Perjalanan Tahun 2023-2024', 'Jumlah Pergerakan Tahun 2023-2024']]

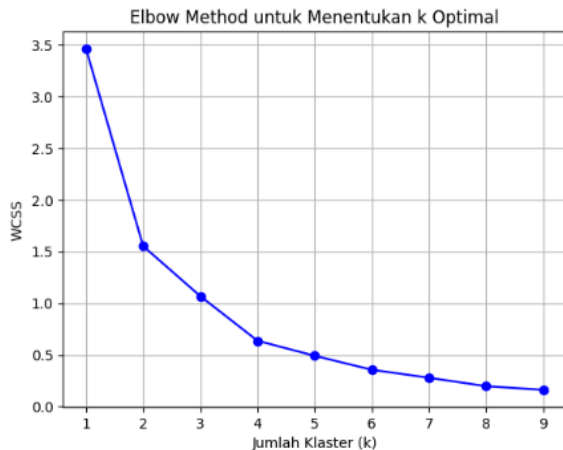
# Normalisasi
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)
```

Gambar 12. Proses normalisasi data dengan min-max scaler

Gambar 12 menampilkan hasil proses normalisasi menggunakan teknik *Min-Max Scaling*, di mana semua nilai data dikonversi ke dalam skala 0 hingga 1 untuk menjaga kesetaraan bobot antar variabel dalam perhitungan jarak klustering.

4.2. Penerapan Algoritma K-Means Clustering

Algoritma *K-Means* digunakan untuk mengelompokkan kabupaten/kota di Jawa Timur berdasarkan data jumlah perjalanan dan pergerakan wisatawan nusantara tahun 2023-2024.



Gambar 13. Grafik elbow method

Gambar 13 menunjukkan grafik Elbow Method yang digunakan untuk menentukan jumlah kluster optimal. Terlihat bahwa titik siku berada pada $k = 3$, yang menunjukkan bahwa tiga kluster memberikan hasil yang paling efisien berdasarkan nilai WCSS. Nilai Within-Cluster Sum of Squares (WCSS) dari $k = 1$ hingga 9 ditampilkan dalam Tabel 1 berikut:

Tabel 1. Nilai WCSS untuk setiap jumlah kluster (k)

Jumlah Kluster (k)	WCSS
1	3.46
2	1.55
3	1.07
4	0.64
5	0.49
6	0.36
7	0.28
8	0.20
9	0.16

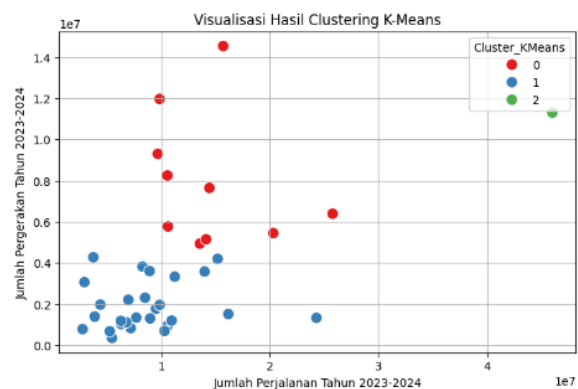
Dari Tabel 1 dapat dilihat bahwa nilai WCSS terus menurun seiring bertambahnya jumlah kluster. Namun, penurunan yang paling signifikan terjadi pada $k = 3$, yang menguatkan hasil dari Elbow Method bahwa pemilihan tiga kluster adalah yang paling optimal untuk segmentasi wilayah berdasarkan data wisatawan.

Setelah kluster ditentukan, K-Means diterapkan dan menghasilkan 3 kluster. Distribusi wilayah pada masing-masing kluster ditampilkan pada Tabel 2.

Tabel 2. Jumlah kabupaten/kota pada setiap kluster

Kluster	Jumlah Wilayah
0	10
1	27
2	1

Dari Tabel 2 dapat dilihat bahwa mayoritas wilayah berada pada Kluster 1, sedangkan Kluster 2 hanya berisi satu wilayah yang bersifat ekstrem atau outlier. Visualisasi hasil pengelompokan ditampilkan dalam Gambar 14. Setiap titik mewakili satu kabupaten/kota dengan warna berbeda untuk tiap kluster.



Gambar 14. Visualisasi hasil clustering k-means

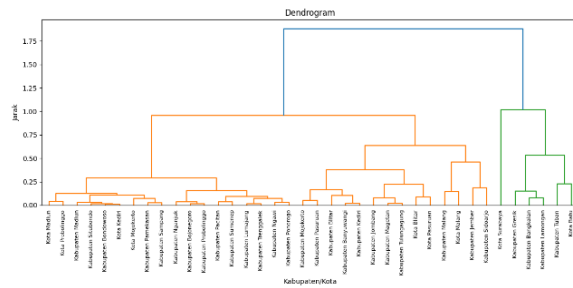
Gambar 14 menunjukkan distribusi wilayah dalam tiga kluster berdasarkan hasil algoritma K-Means. Visualisasi ini menggambarkan posisi relatif masing-masing kabupaten/kota terhadap pusat kluster (centroid) berdasarkan jumlah perjalanan dan pergerakan wisatawan. Hasil menunjukkan bahwa:

- Kluster 0: Wilayah dengan aktivitas wisata sedang, cocok untuk penguatan promosi dan daya tarik lokal.
- Kluster 1: Wilayah dengan aktivitas rendah, menjadi target peningkatan promosi dan pengembangan destinasi.
- Kluster 2: Hanya berisi Kota Surabaya, sebagai *outlier* positif dengan jumlah perjalanan dan pergerakan tertinggi, membutuhkan pendekatan promosi khusus.

Segmentasi ini membantu menyusun strategi promosi wisata yang lebih tepat sasaran berdasarkan karakteristik masing-masing wilayah.

4.3. Penerapan Algoritma Hierarchical Clustering

Hierarchical Clustering digunakan untuk mengelompokkan kabupaten/kota di Jawa Timur berdasarkan data jumlah perjalanan dan pergerakan wisatawan tahun 2023-2024. Metode yang digunakan adalah Agglomerative dengan pendekatan Ward linkage, yang membentuk struktur pohon kluster melalui dendrogram yang terdapat pada Gambar 15.



Gambar 15. Visualisasi dendrogram hierarchical clustering

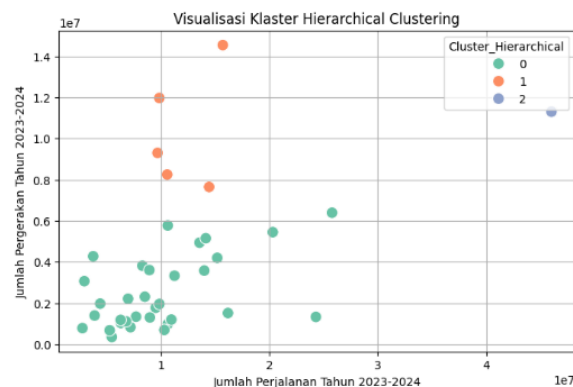
Gambar 15 menunjukkan struktur pohon kluster yang terbentuk dari proses hierarchical clustering. Dari dendrogram tersebut ditentukan bahwa jumlah kluster optimal adalah tiga. Distribusi hasil kluster terdapat pada Tabel 3 berikut:

Tabel 3. Jumlah Kabupaten/Kota pada Setiap Kluster (Hierarchical Clustering)

Kluster	Jumlah Wilayah
0	32
1	5
2	1

Tabel 3 memperlihatkan hasil pembagian wilayah ke dalam tiga kluster berdasarkan pendekatan hierarkis, dengan Kluster 0 mendominasi jumlah wilayah yang terklasternkan.

Visualisasi hasil kluster berdasarkan dua variabel juga ditampilkan dalam bentuk scatter plot terlihat pada Gambar 16.



Gambar 16. Visualisasi hasil clustering hierarchical

Gambar 16 memperlihatkan sebaran kabupaten/kota dalam tiga kluster hasil Hierarchical Clustering, berdasarkan jumlah perjalanan dan pergerakan wisatawan.

Hasil menunjukkan:

- Kluster 0 mencakup mayoritas wilayah dengan aktivitas wisata sedang.
- Kluster 1 berisi wilayah dengan kunjungan wisata rendah.
- Kluster 2 hanya terdiri dari Kota Surabaya, yang muncul sebagai *outlier positif* karena memiliki

jumlah perjalanan dan pergerakan wisatawan tertinggi.

Metode *Hierarchical Clustering* berhasil mengelompokkan mayoritas wilayah ke dalam satu kluster besar, sekaligus mengidentifikasi wilayah ekstrem secara jelas, sehingga relevan untuk dasar segmentasi promosi wisata.

4.4. Teknik Analisis Data

Penelitian ini mengevaluasi hasil segmentasi kabupaten/kota di Jawa Timur menggunakan algoritma *K-Means* dan *Hierarchical Clustering*, berdasarkan dua variabel: jumlah perjalanan dan pergerakan wisatawan nusantara tahun 2023-2024. Evaluasi dilakukan menggunakan tiga metrik utama: *Silhouette Score*, *Calinski-Harabasz (CH) Index*, dan *Dunn Index*. Hasil dari evaluasi terdapat pada Tabel 4 berikut:

Tabel 4. Hasil Evaluasi Metrik Clustering

Metrik Evaluasi	K-Means	Hierarchical
Silhouette Score	0.5230	0.5689
CH Index	39.23	34.11
Dunn Index	0.0950	0.2396

Tabel 4 menunjukkan bahwa *Hierarchical Clustering* unggul dalam metrik *Silhouette Score* dan *Dunn Index*, yang menandakan bahwa kluster yang terbentuk lebih kompak dan terpisah dengan jelas. Sementara itu, algoritma *K-Means* unggul dalam *Calinski-Harabasz Index*, yang menunjukkan struktur antar kluster yang lebih menyebar.

Berdasarkan hasil evaluasi tersebut, *Hierarchical Clustering* lebih direkomendasikan sebagai metode segmentasi wilayah dalam konteks promosi wisata. Metode ini mampu menghasilkan kluster yang lebih stabil, representatif terhadap pola perjalanan wisatawan, dan efektif dalam mengidentifikasi wilayah dengan karakteristik ekstrem seperti Kota Surabaya.

4.5. Interpretasi Segmentasi

Hasil *clustering* membagi kabupaten/kota di Jawa Timur ke dalam tiga kluster berdasarkan jumlah perjalanan dan pergerakan wisatawan nusantara (2023-2024), menggunakan algoritma *K-Means* dan *Hierarchical Clustering*.

Pada kedua metode, Kota Surabaya konsisten muncul sebagai *outlier positif*, menandakan peran strategisnya sebagai pusat pergerakan wisatawan.

K-Means:

- Kluster 0: Wilayah dengan aktivitas wisata sedang (contoh: Banyuwangi, Kota Batu), cocok untuk promosi digital dan event lokal.
- Kluster 1: Wilayah rendah, perlu dukungan infrastruktur dan promosi intensif.
- Kluster 2: Hanya Surabaya dengan kunjungan tertinggi.

Hierarchical Clustering:

- Klaster 0: Mayoritas wilayah dengan aktivitas sedang. ideal untuk strategi regional *tourism*.
- Klaster 1: Wilayah mirip aktivitasnya (contoh: Bangkalan. Gresik). cocok untuk promosi tematik terfokus.
- Klaster 2: Tetap hanya Surabaya sebagai *outlier*.

Perbedaan hasil antar algoritma (misalnya. Kota Batu dan Gresik berpindah klaster) menunjukkan bahwa pilihan metode memengaruhi arah strategi promosi. Berdasarkan nilai evaluasi (*Silhouette Score* dan *Dunn Index*). *Hierarchical Clustering* dinilai lebih stabil dan representatif. serta lebih sensitif terhadap *outlier*.

Rekomendasi Strategi Promosi:

- Klaster tinggi: *Branding* dan aktivasi event unggulan.
- Klaster sedang: Integrasi antarwilayah dalam paket wisata.
- Klaster rendah: Promosi digital. pelatihan SDM. dan pengembangan destinasi.

Strategi ini sejalan dengan kebijakan RPJMD Jatim 2019-2024 dan arah Kemenparekraf dalam promosi berbasis data dan pemberdayaan ekonomi kreatif. Efektivitas algoritma bergantung pada konteks data; dalam kasus ini. ***Hierarchical Clustering lebih cocok*** untuk segmentasi wilayah berbasis pariwisata.

5. KESIMPULAN DAN SARAN

Penelitian melakukan segmentasi kabupaten/kota di Jawa Timur berdasarkan jumlah perjalanan dan pergerakan wisatawan nusantara tahun 2023-2024 dengan menggunakan algoritma *K-Means* dan *Hierarchical Clustering*. Jumlah klaster optimal ditentukan sebanyak tiga berdasarkan *Elbow Method* dan *dendrogram*. Setelah melalui tahapan *data cleaning*. agregasi. transformasi. dan normalisasi. *K-Means* menghasilkan tiga klaster: rendah (kunjungan kurang dari 10 juta). sedang (10-30 juta). dan tinggi (Kota Surabaya sebagai *outlier* dengan lebih dari 55 juta kunjungan). *Hierarchical Clustering* juga menghasilkan tiga klaster dengan pemisahan struktur yang lebih stabil. Berdasarkan evaluasi. *Hierarchical* unggul dalam *Silhouette Score* (0.5689) dan *Dunn Index* (0.2396). sedangkan *K-Means* sedikit lebih tinggi pada *CH Index* (39.23). Segmentasi ini dapat menjadi dasar strategi promosi wisata yang lebih terarah: klaster tinggi difokuskan pada branding dan event unggulan. klaster sedang pada integrasi antar wilayah. dan klaster rendah pada promosi digital serta pengembangan destinasi. Penelitian ini mendukung strategi promosi wisata berbasis data seperti yang tercantum dalam RPJMD Jawa Timur 2019-2024 dan kebijakan Kemenparekraf. Adapun saran untuk penelitian selanjutnya adalah menggunakan pendekatan serupa untuk mendukung kebijakan Dinas Pariwisata. menambahkan variabel baru seperti lama tinggal wisatawan atau data sosial ekonomi untuk

meningkatkan akurasi. serta mengembangkan sistem berbasis *data mining* seperti *dashboard* atau visualisasi interaktif *clustering*.

DAFTAR PUSTAKA

- [1] E. W. Liyanto, A. Homaidi, and A. Lutfi, "Data Mining IMPLEMENTASI K-MEANS CLUSTERING DALAM PENGELOMPOKAN DATA KUNJUNGAN WISATAWAN ASING DI PROVINSI JAWA TIMUR," *E-Link J. Tek. Elektro dan Inform.*, vol. 19, no. 2, p. 205, 2024, doi: 10.30587/e-link.v19i2.8224
- [2] A. Utama, W. Imama Sabilla, and R. Wakhidah, "Sistem Rekomendasi Tempat Wisata di Malang Raya Menggunakan Metode K-Means Clustering," *J. Inform. dan Multimed.*, vol. 16, no. 1, pp. 1–15, 2024, doi: 10.33795/jtim.v16i1.5178
- [3] P. Rahma Maulida, F. A. Semnasti, and B. K. K. Semnasti, "Penentuan Segmentation, Targeting, dan Positioning pada wisatawan desa wisata Lon Malang Madura menggunakan K-means Clustering," *Waluyo Jatmiko Proceeding*, vol. 16, no. 1, pp. 111–120, 2023, doi: 10.33005/wj.v16i1.49
- [4] N. K. Zuhail, "Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering," *Pros. Semin. Nas. Teknol. Dan Sains*, vol. 1, pp. 200–205, 2022
- [5] F. Putri Azizah, S. Shofiah Hilabi, and A. Hananto, "Perbandingan Algoritma K-Means dan Hierarchical Untuk Klasterisasi Data Kehadiran Karyawan," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, pp. 351–361, 2025
- [6] M. T. Sarira *et al.*, "Pola Perjalanan Wisatawan Di Kepulauan Spermonde," *Geogr. J. Kajian, Penelit. dan Pengemb. Pendidik.*, vol. 11, no. 1, p. 119, 2023, doi: 10.31764/geography.v11i1.14083
- [7] R. N. Nugraha, P. D. Rosa, and M. Ivanka, "Metaverse Tourism Sebagai Strategi Promosi Wisata Museum Bank Indonesia," *J. Ilm. Wahana Pendidik.*, vol. 9, no. 8, pp. 731–744, 2023, [Online]. Available: <https://doi.org/10.5281/zenodo.7901659>
- [8] C. S. Akasse and R. Ramansyah, "Strategi Promosi Pariwisata melalui Media Sosial dalam Meningkatkan Pengunjung di Desa Wisata," *J. Socius J. Sociol. Res. Educ.*, vol. 10, no. 1, pp. 52–60, 2023, doi: 10.24036/scs.v10i1.457
- [9] R. K. W. Astuti, M. Andarwati, "Urgensi Informasi Pemasaran Dalam Pengembangan Desa Wisata di Kota Batu, Jawa Timur," *Semin. Nas. Kepariwisata*, vol. 3, p. 121, 2022
- [10] F. Abdul Kadir, F. Natsir, and S. Suaedah, "Penerapan Data Mining Untuk Identifikasi Kinerja Technical Support Di Pt Maxindo Mitra Solusi Dengan Metode Naïve Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 6, pp. 11545–11553, 2024, doi: 10.36040/jati.v8i6.11509
- [11] R. Hidayati, "Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering," *Techno.COM*, 2021