

Analisis Segmentasi Negara Asal Wisatawan Mancanegara ke Indonesia Menggunakan Metode K-Means Clustering

Dista Dian Saputri¹, Angel Tesalonika², Desta Aditya³, Satya Yoma⁴, Aprilisa Arum⁵

¹⁻⁵Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Duta Bangsa, Surakarta, Indonesia

Email: ¹230103131@mhs.udb.ac.id, ²230103259@mhs.udb.ac.id, ³230103129@mhs.udb.ac.id,
⁴230103839@mhs.udb.ac.id, ⁵aprilisa_arumsari@mhs.udb.ac.id

Abstrak—Jumlah kunjungan wisatawan mancanegara ke Indonesia menunjukkan perbedaan yang cukup beragam jika dilihat berdasarkan negara asalnya. Perbedaan tersebut mencerminkan adanya karakteristik kunjungan yang tidak sama sehingga diperlukan proses segmentasi untuk memperoleh gambaran mengenai kelompok negara dengan pola kunjungan yang serupa. Penelitian ini bertujuan melakukan segmentasi negara asal wisatawan mancanegara menggunakan algoritma K-Means Clustering dengan memanfaatkan data kunjungan wisatawan periode 2020–2025 yang diperoleh dari Badan Pusat Statistik (BPS). Tahapan penelitian meliputi preprocessing data, integrasi dataset, pembentukan empat fitur utama, yaitu total kunjungan, rata-rata kunjungan, standar deviasi, dan growth rate, kemudian dilanjutkan dengan transformasi logaritmik, normalisasi menggunakan MinMax Scaler, serta penentuan jumlah cluster menggunakan Elbow Method dan Silhouette Score. Hasil evaluasi menunjukkan bahwa jumlah cluster yang paling sesuai adalah tiga cluster dengan nilai Silhouette Score sebesar 0,5413. Proses clustering menghasilkan tiga kelompok yang terdiri atas 72 negara, 83 negara, dan 38 negara dengan karakteristik kunjungan yang berbeda. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan negara asal wisatawan secara cukup baik sehingga dapat dimanfaatkan sebagai informasi pendukung dalam penyusunan strategi promosi dan pengembangan pasar pariwisata Indonesia.

Kata Kunci: K-Means Clustering; Data Mining; Pariwisata; Wisatawan Mancanegara; Segmentasi.

Abstract—The number of international tourist arrivals to Indonesia varies considerably depending on the tourists' countries of origin. These differences indicate distinct visitation characteristics, making market segmentation necessary to better understand groups of countries with similar travel patterns. This study aims to segment the countries of origin of international tourists using the K-Means Clustering algorithm based on tourist arrival data from 2020 to 2025 obtained from Statistics Indonesia (BPS). The research process included data preprocessing, dataset integration, feature construction consisting of total arrivals, average arrivals, standard deviation, and growth rate, followed by logarithmic transformation, MinMax normalization, and the determination of the optimal number of clusters using the Elbow Method and Silhouette Score. The evaluation results indicated that three clusters provided the optimal grouping, with a Silhouette Score of 0.5413. The clustering process produced three groups consisting of 72, 83, and 38 countries, respectively, each representing different tourist arrival characteristics. The findings demonstrate that the K-Means algorithm is capable of identifying meaningful patterns among countries of origin and can serve as a valuable reference for supporting tourism market segmentation and the development of more targeted tourism promotion strategies in Indonesia.

Keywords: K-Means Clustering; Data Mining; Tourism; International Tourists; Segmentation.

1. PENDAHULUAN

Kontribusi sektor pariwisata terhadap perekonomian tidak hanya terlihat dari meningkatnya jumlah wisatawan, tetapi juga dari dampaknya terhadap berbagai aktivitas ekonomi yang saling berkaitan. Pertumbuhan sektor ini mampu mendorong perkembangan usaha di bidang transportasi, akomodasi, kuliner, perdagangan, dan industri kreatif. Bagi Indonesia, kondisi tersebut menjadikan pariwisata sebagai salah satu sektor yang terus diprioritaskan untuk mendukung peningkatan devisa serta memperkuat daya saing destinasi wisata di tingkat internasional (Liyanto et al., 2024; Simorangkir et al., 2024).

Indonesia memiliki banyak destinasi wisata yang bisa dianggap unggulan dan tersebar di berbagai wilayah, mulai dari wisata alam, wisata budaya, wisata sejarah, hingga wisata bahari. Keberagaman wisata menjadikan daya tarik bagi wisatawan mancanegara dari berbagai negara. Namun demikian, karakteristik wisatawan yang berasal dari berbagai negara tidak selalu sama. Setiap negara memiliki pola kunjungan, tingkat pertumbuhan, dan kontribusi yang berbeda terhadap jumlah wisatawan yang datang ke Indonesia (Fitriandari & Muhammad, 2025).

Dalam beberapa tahun terakhir, khususnya setelah pandemi COVID-19, pola kunjungan wisatawan mancanegara mengalami perubahan yang cukup signifikan. Pembatasan mobilitas internasional menyebabkan penurunan jumlah wisatawan pada tahun 2020 dan 2021. Seiring dengan pemulihan sektor pariwisata global, jumlah kunjungan wisatawan mulai menunjukkan peningkatan pada tahun-tahun berikutnya. Kondisi tersebut menghasilkan variasi pola kunjungan yang menarik untuk dianalisis lebih lanjut (Aldi & Fatah, 2025).

Kemajuan teknologi informasi menyebabkan jumlah data yang tersedia terus meningkat dan dapat dimanfaatkan sebagai dasar dalam proses pengambilan keputusan. Agar data tersebut menghasilkan informasi yang bernilai, diperlukan proses analisis menggunakan teknik data mining. Melalui data mining, pola maupun hubungan yang tersembunyi di dalam kumpulan data dapat ditemukan sehingga menghasilkan informasi yang bermanfaat. Salah satu metode yang sering digunakan adalah clustering, yaitu teknik yang mengelompokkan objek berdasarkan kesamaan karakteristik sehingga objek dalam kelompok yang sama memiliki tingkat kemiripan yang lebih tinggi dibandingkan dengan kelompok lainnya. (Aldi & Fatah, 2025).

Di antara berbagai metode clustering, K-Means menjadi salah satu algoritma yang banyak digunakan karena proses perhitungannya relatif sederhana serta mudah diterapkan pada berbagai jenis data. Algoritma ini membentuk sejumlah kelompok dengan menentukan titik pusat (centroid) sebagai acuan pengelompokan. Setiap data akan ditempatkan pada cluster yang memiliki jarak paling dekat dengan centroid, sehingga objek dengan karakteristik yang mirip cenderung berada pada kelompok yang sama, sedangkan objek yang memiliki karakteristik berbeda akan terpisah ke cluster lainnya (Liyanto et al., 2024; Yolandari et al., 2024).

Penelitian mengenai segmentasi wisatawan menggunakan teknik clustering telah banyak dilakukan. Namun sebagian besar penelitian masih menggunakan data dalam rentang waktu yang terbatas atau hanya mempertimbangkan jumlah kunjungan pada satu periode tertentu. Pada penelitian ini, data kunjungan wisatawan mancanegara selama enam tahun digunakan untuk memperoleh gambaran yang lebih komprehensif mengenai karakteristik negara asal wisatawan (Hesti & Karyati, 2024; Fitriandari & Muhammad, 2025).

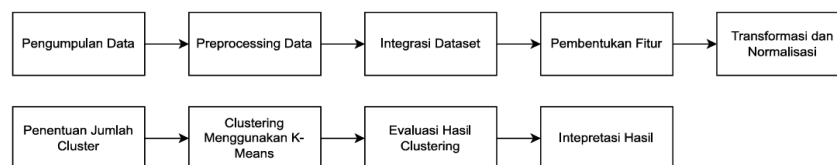
2. METODE PENELITIAN

2.1 Sumber Data

Data yang digunakan dalam penelitian ini merupakan data jumlah kunjungan wisatawan mancanegara ke Indonesia berdasarkan negara asal yang diperoleh dari Badan Pusat Statistik (BPS). Data tersedia dalam format CSV untuk setiap tahun pengamatan, yaitu tahun 2020 hingga 2025 (Badan Pusat Statistik, 2025).

2.2 Tahapan Penelitian

Penelitian dilakukan melalui beberapa tahapan yang ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

2.3 Preprocessing Data

Tahap preprocessing bertujuan untuk mempersiapkan data agar memiliki kualitas yang baik sebelum digunakan pada proses clustering. Langkah pertama adalah menghapus data agregasi wilayah seperti ASEAN, Eropa, Amerika, Afrika, dan kategori lainnya karena penelitian berfokus pada negara asal wisatawan secara individual.

Selain itu dilakukan penyesuaian nama kolom agar lebih mudah diproses pada tahap analisis berikutnya. Setelah proses pembersihan selesai, diperoleh sekitar 230 negara yang digunakan sebagai objek clustering.

2.4 Integrasi Dataset

Dataset tahunan yang telah dibersihkan kemudian digabungkan menggunakan atribut negara sebagai kunci utama. Hasil integrasi menghasilkan satu dataset yang memuat informasi jumlah kunjungan wisatawan untuk setiap negara selama enam tahun pengamatan.

Proses integrasi memungkinkan analisis dilakukan secara longitudinal sehingga pola kunjungan wisatawan dapat diamati dalam rentang waktu yang lebih panjang.

2.5 Pembentukan Fitur

Empat fitur utama digunakan dalam penelitian ini, yaitu:

a. Total Kunjungan

Merupakan jumlah seluruh kunjungan wisatawan dari suatu negara selama periode 2020–2025.

b. Rata-rata Kunjungan

Merupakan rata-rata jumlah kunjungan wisatawan setiap tahun.

c. Standar Deviasi

Digunakan untuk mengukur tingkat fluktuasi jumlah kunjungan wisatawan selama periode pengamatan.

d. Growth Rate

Digunakan untuk mengukur tingkat pertumbuhan jumlah wisatawan.

$$Growth Rate = \frac{Total_{2025} - Total_{2020}}{Total_{2020}} \times 100\%$$

2.6 Transformasi Data

Distribusi jumlah kunjungan wisatawan memiliki rentang yang sangat besar sehingga sebagian besar data bersifat right skewed. Untuk mengurangi pengaruh nilai ekstrem dilakukan transformasi logaritmik menggunakan fungsi $\log_{10}()$.

Transformasi ini diterapkan pada variabel Total Kunjungan, Rata-rata Kunjungan, dan Standar Deviasi sehingga distribusi data menjadi lebih seimbang sebelum dilakukan proses clustering.

2.7 Normalisasi Data

Setelah transformasi logaritmik, dilakukan normalisasi menggunakan metode MinMax Scaler agar seluruh variabel memiliki rentang nilai antara 0 hingga 1.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Normalisasi diperlukan karena algoritma K-Means menggunakan perhitungan jarak Euclidean sehingga variabel dengan nilai yang lebih besar dapat mendominasi proses clustering apabila tidak dinormalisasi (Amellia & Kurnia, 2025).

2.8 Penentuan Jumlah Cluster dan Clustering

Penentuan jumlah cluster dilakukan menggunakan dua pendekatan, yaitu Elbow Method dan Silhouette Score.

Pada metode Elbow dihitung nilai Sum of Squared Error (SSE) untuk jumlah cluster antara 2 hingga 10. Titik siku pada grafik menunjukkan jumlah cluster yang paling optimal.

Selanjutnya dilakukan evaluasi menggunakan Silhouette Score untuk mengukur kualitas pemisahan antar cluster. Nilai silhouette yang semakin mendekati 1 menunjukkan kualitas clustering yang semakin baik.

Setelah diperoleh jumlah cluster optimal, proses pengelompokan dilakukan menggunakan algoritma K-Means dengan parameter $random_state = 42$ dan $n_init = 10$ agar hasil pengelompokan lebih stabil.

Setiap negara kemudian dikelompokkan berdasarkan kemiripan karakteristik kunjungan wisatawan yang dimilikinya.

2.9 Evaluasi Hasil Clustering

Kualitas hasil pengelompokan pada penelitian ini dievaluasi menggunakan Silhouette Score. Metode evaluasi ini digunakan untuk mengetahui seberapa baik setiap objek ditempatkan pada cluster yang sesuai. Penilaian dilakukan dengan membandingkan tingkat kedekatan suatu objek terhadap anggota dalam cluster yang sama (cohesion) serta jaraknya terhadap cluster terdekat lainnya (separation).

Nilai Silhouette Score berada pada rentang -1 sampai 1. Nilai yang semakin mendekati 1 menunjukkan bahwa objek memiliki tingkat kemiripan yang tinggi dengan anggota pada cluster tempatnya berada dan memiliki perbedaan yang jelas dengan cluster lainnya. Sebaliknya, nilai yang berada di sekitar 0 mengindikasikan bahwa batas antarcluster masih saling berdekatan sehingga pemisahan kelompok belum optimal. Apabila nilai yang diperoleh bernilai negatif, kondisi tersebut menunjukkan bahwa suatu objek kemungkinan lebih sesuai ditempatkan pada cluster lain daripada cluster yang ditempatinya saat ini (Komparasi Hasil Segmentasi Metode K-Means dan Agglomerative Hierarchical, 2025).

Pada penelitian ini, proses evaluasi dilakukan dengan menghitung nilai Silhouette Score untuk beberapa alternatif jumlah cluster, yaitu $K = 2$ hingga $K = 10$. Nilai yang paling tinggi dipilih sebagai dasar dalam menentukan jumlah cluster yang digunakan pada proses segmentasi negara asal wisatawan mancanegara.

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

Dengan:

$a(i)$ = rata-rata jarak objek ke seluruh anggota pada cluster yang sama.

$b(i)$ = rata-rata jarak objek ke cluster lain yang memiliki jarak paling dekat.

$s(i)$ = nilai Silhouette Score untuk objek ke- i

Selanjutnya, nilai Silhouette Score keseluruhan diperoleh dari rata-rata seluruh nilai $s(i)$ pada setiap objek dalam dataset. Nilai rata-rata tersebut digunakan sebagai indikator kualitas hasil clustering.

3. ANALISA DAN PEMBAHASAN

3.1 Hasil Integrasi Dataset

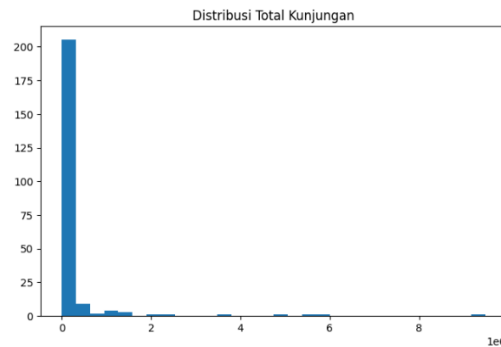
Tahap pertama setelah proses preprocessing adalah menggabungkan seluruh data kunjungan wisatawan mancanegara dari tahun 2020 hingga 2025 menjadi satu dataset utama. Proses penggabungan dilakukan berdasarkan atribut Negara sehingga setiap negara hanya memiliki satu baris data yang berisi jumlah kunjungan wisatawan pada setiap tahun pengamatan.

Selanjutnya dilakukan pembentukan empat fitur yang akan digunakan pada proses clustering, yaitu Total Kunjungan, Rata-rata Kunjungan, Standar Deviasi, dan Growth Rate. Keempat fitur tersebut dipilih karena mampu menggambarkan karakteristik masing-masing negara, baik dari sisi jumlah kunjungan, rata-rata kunjungan per tahun, tingkat fluktuasi, maupun pertumbuhan kunjungan wisatawan selama periode penelitian.

Sebelum dilakukan proses clustering, data ditransformasikan menggunakan fungsi $\log_{1p}()$ dan dinormalisasi menggunakan MinMax Scaler agar seluruh fitur memiliki rentang nilai yang sama sehingga tidak terjadi dominasi salah satu variabel pada proses perhitungan jarak Euclidean.

3.2 Analisis Distribusi Data

Sebelum dilakukan proses clustering, distribusi data dianalisis untuk mengetahui karakteristik dataset yang digunakan. Hasil visualisasi distribusi total kunjungan wisatawan ditunjukkan pada Gambar 2.

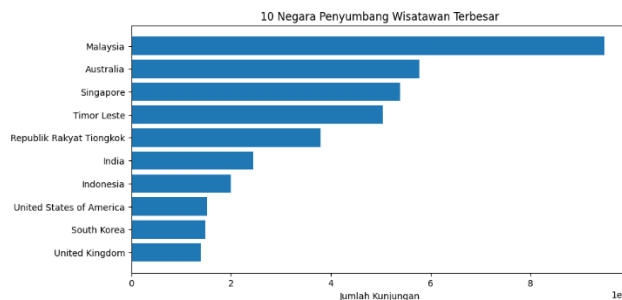


Gambar 2. Distribusi Total Kunjungan

Berdasarkan Gambar 2 terlihat bahwa distribusi data memiliki bentuk right skewed atau condong ke kanan. Sebagian besar negara memiliki jumlah kunjungan yang relatif rendah, sedangkan hanya beberapa negara yang memiliki jumlah kunjungan sangat tinggi. Kondisi tersebut menunjukkan bahwa terdapat perbedaan skala data yang cukup besar antarnegara. Apabila data langsung digunakan dalam proses clustering tanpa transformasi, negara dengan jumlah kunjungan yang sangat besar akan mendominasi proses pembentukan cluster. Oleh karena itu dilakukan transformasi logaritmik menggunakan fungsi $\log_{10}()$ sebelum proses normalisasi dan clustering.

3.3 Analisis Negara Penyumbang Wisatawan Terbesar

Untuk memperoleh gambaran awal mengenai karakteristik data, dilakukan analisis terhadap sepuluh negara dengan jumlah kunjungan wisatawan terbesar selama periode 2020–2025.



Gambar 3. Negara Penyumbang Wisatawan Terbesar

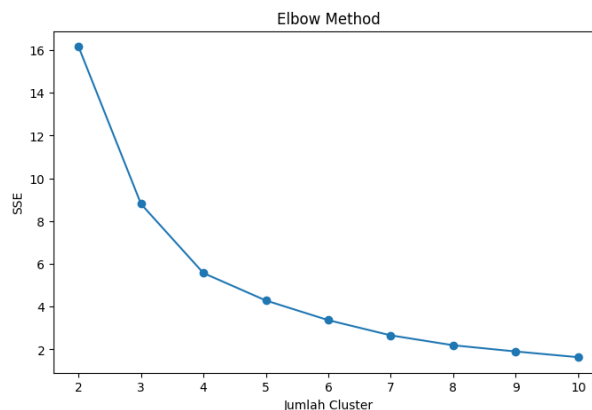
Berdasarkan Gambar 3, Malaysia menjadi negara dengan jumlah kunjungan wisatawan terbesar ke Indonesia selama periode penelitian dengan total 9.484.962 kunjungan. Posisi berikutnya ditempati oleh Australia sebanyak 5.771.991 kunjungan, Singapura sebanyak 5.382.176 kunjungan, serta Timor Leste sebanyak 5.041.771 kunjungan. Selain itu terdapat beberapa negara lain yang juga memberikan kontribusi besar terhadap jumlah wisatawan mancanegara, seperti Republik Rakyat Tiongkok, India, Amerika Serikat, Korea Selatan, Inggris, dan Jepang.

Dominasi negara-negara tersebut menunjukkan bahwa kawasan Asia Pasifik masih menjadi pasar utama pariwisata Indonesia. Faktor kedekatan geografis, hubungan ekonomi, kemudahan akses transportasi, serta kerja sama antarnegara menjadi beberapa faktor yang diduga memengaruhi tingginya jumlah wisatawan dari kawasan tersebut.

3.4 Penentuan Jumlah Cluster

Penentuan jumlah cluster dilakukan menggunakan Elbow Method dan Silhouette Score. Elbow Method digunakan untuk melihat perubahan nilai Sum of Squared Error (SSE) terhadap jumlah cluster, sedangkan Silhouette Score digunakan untuk mengevaluasi kualitas hasil pengelompokan.

3.4.1 Elbow Method



Gambar 4. Elbow Method

Berdasarkan Gambar 4 terlihat bahwa nilai SSE mengalami penurunan yang cukup tajam dari $K = 2$ hingga $K = 4$. Setelah itu penurunan mulai melandai sehingga penambahan jumlah cluster tidak memberikan perubahan yang terlalu signifikan.

Kondisi tersebut menunjukkan bahwa jumlah cluster yang terlalu banyak tidak selalu menghasilkan pengelompokan yang lebih baik. Oleh karena itu diperlukan evaluasi tambahan menggunakan Silhouette Score untuk menentukan jumlah cluster yang paling optimal.

3.4.2 Silhoutte Score

Pengujian dilakukan terhadap nilai $K = 2$ hingga $K = 10$. Hasil pengujian ditunjukkan pada Tabel 1.

Tabel 1. Silhoutte Score

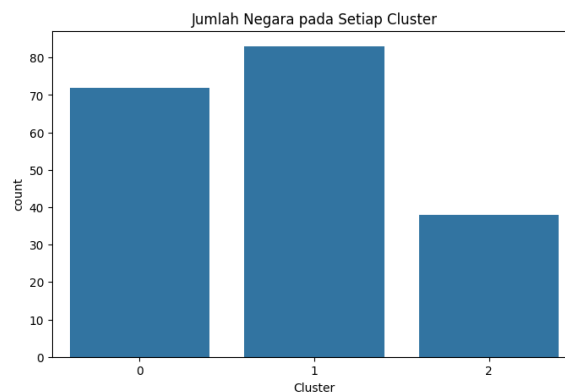
Jumlah Cluster	Silhouette Score
2	0,5386
3	0,5413
4	0,5290
5	0,4902
6	0,5055
7	0,4765
8	0,4759
9	0,4559
10	0,4459

Berdasarkan Tabel 3 terlihat bahwa nilai Silhouette Score tertinggi diperoleh pada $K = 3$ dengan nilai sebesar 0,5413. Nilai tersebut menunjukkan bahwa hasil pengelompokan memiliki tingkat pemisahan yang cukup baik dibandingkan jumlah cluster lainnya.

Walaupun selisih nilai antara $K = 2$ dan $K = 3$ tidak terlalu besar, penggunaan tiga cluster dinilai lebih mampu menggambarkan variasi karakteristik negara asal wisatawan sehingga dipilih sebagai jumlah cluster yang digunakan dalam penelitian ini.

3.5 Hasil Clustering Menggunakan K-Means

Berdasarkan hasil penentuan jumlah cluster, proses clustering dilakukan menggunakan algoritma K-Means dengan jumlah cluster sebanyak tiga. Jumlah anggota pada masing-masing cluster dapat dilihat pada Gambar 5.



Gambar 5. Jumlah Negara tiap Cluster

Berdasarkan Gambar 5 diperoleh bahwa cluster 0 terdiri dari 72 negara., cluster 1 terdiri dari 83 negara, dan cluster 2 terdiri dari 38 negara. Hasil tersebut menunjukkan bahwa sebagian besar negara berada pada kelompok dengan jumlah kunjungan rendah hingga menengah, sedangkan hanya sebagian kecil negara yang memiliki jumlah kunjungan sangat tinggi.

3.6 Karakteristik Setiap Cluster

Karakteristik masing-masing cluster dianalisis berdasarkan nilai rata-rata dari setiap fitur. Berdasarkan Tabel 3 terlihat bahwa Cluster 2 memiliki rata-rata jumlah kunjungan paling tinggi dibandingkan cluster lainnya. Hal ini menunjukkan bahwa negara-negara pada cluster tersebut merupakan penyumbang utama wisatawan mancanegara ke Indonesia. Sebaliknya, Cluster 1 memiliki rata-rata jumlah kunjungan paling rendah. Meskipun demikian, nilai growth rate pada cluster ini masih menunjukkan adanya peningkatan jumlah kunjungan sehingga kelompok ini tetap memiliki peluang untuk dikembangkan pada masa mendatang. Sementara itu Cluster 0 berada pada kategori menengah. Negara-negara dalam cluster ini memiliki jumlah kunjungan yang cukup besar namun belum sebesar cluster prioritas. Dengan strategi promosi yang tepat, kelompok ini berpotensi berkembang menjadi pasar utama pariwisata Indonesia.

3.7 Analisis Anggota Setiap Cluster

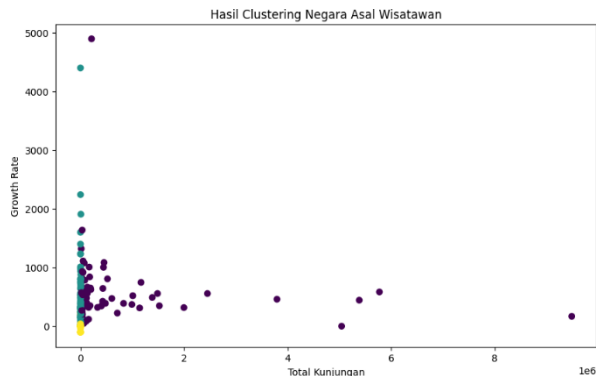
Berdasarkan hasil pengelompokan, negara-negara dalam Cluster 2 didominasi oleh Malaysia, Australia, Singapura, Timor Leste, Republik Rakyat Tiongkok, India, Amerika Serikat, Korea Selatan, Inggris, Prancis, Jepang, Jerman, Belanda, Filipina, Rusia, Taiwan, Selandia Baru, Arab Saudi, dan Spanyol. Negara-negara tersebut memiliki jumlah kunjungan yang jauh lebih tinggi dibandingkan negara lainnya sehingga dapat dikategorikan sebagai pasar prioritas bagi sektor pariwisata Indonesia.

Cluster 0 berisi negara-negara dengan jumlah kunjungan menengah, seperti Ukraina, Rumania, Pakistan, Republik Ceko, Brunei Darussalam, Mesir, Hongaria, Bangladesh, Kazakhstan, Meksiko, Finlandia, Maroko, dan beberapa negara lainnya. Kelompok ini memiliki kontribusi yang cukup baik, namun masih memerlukan peningkatan promosi agar jumlah kunjungan wisatawan terus meningkat.

Sementara itu, Cluster 1 didominasi oleh negara-negara dengan jumlah kunjungan yang relatif kecil, seperti Ghana, Bhutan, Moldova, Kosta Rika, Makedonia, Mozambik, Uganda, Ethiopia, Guatemala, Tajikistan, Turkmenistan, Kuba, Afghanistan, Kamerun, Senegal, Suriname, Montenegro, Madagaskar, Zambia, dan Panama. Rendahnya jumlah kunjungan pada kelompok

ini menunjukkan bahwa kontribusi negara-negara tersebut terhadap sektor pariwisata Indonesia masih terbatas.

3.8 Visualisasi Hasil Clustering



Gambar 6. Hasil Clustering

Berdasarkan Gambar 6 terlihat bahwa ketiga cluster memiliki persebaran yang cukup berbeda. Cluster 2 berada pada area dengan nilai Total Kunjungan yang tinggi, sedangkan Cluster 1 terkonsentrasi pada jumlah kunjungan yang rendah. Cluster 0 berada di antara kedua kelompok tersebut sehingga dapat dikatakan sebagai kelompok negara dengan karakteristik kunjungan menengah.

Visualisasi tersebut menunjukkan bahwa algoritma K-Means mampu memisahkan negara berdasarkan karakteristik kunjungan wisatawan secara cukup baik. Hal ini juga didukung oleh nilai Silhouette Score sebesar 0,5413, yang mengindikasikan bahwa hasil pengelompokan memiliki kualitas yang cukup baik untuk digunakan sebagai dasar segmentasi pasar wisatawan mancanegara.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma K-Means Clustering dapat diterapkan untuk mengelompokkan negara asal wisatawan mancanegara berdasarkan data kunjungan ke Indonesia selama periode 2020–2025. Proses penelitian diawali dengan pembersihan dan penggabungan data, dilanjutkan dengan pembentukan fitur, transformasi logaritmik, serta normalisasi menggunakan MinMax Scaler. Penentuan jumlah cluster dilakukan melalui Elbow Method dan Silhouette Score, kemudian diperoleh tiga cluster sebagai konfigurasi terbaik dengan nilai Silhouette Score sebesar 0,5413, yang menunjukkan bahwa hasil pengelompokan memiliki kualitas yang cukup baik.

Hasil segmentasi menghasilkan tiga kelompok negara dengan karakteristik kunjungan yang berbeda. Cluster 2 beranggotakan 38 negara dengan tingkat kunjungan tertinggi sehingga dapat dianggap sebagai kelompok negara yang memberikan kontribusi terbesar terhadap kunjungan wisatawan mancanegara ke Indonesia. Cluster 0 terdiri atas 72 negara dengan tingkat kunjungan menengah yang masih memiliki peluang untuk ditingkatkan melalui strategi promosi yang lebih terarah. Sementara itu, Cluster 1 mencakup 83 negara dengan jumlah kunjungan relatif rendah, namun beberapa negara pada kelompok ini masih menunjukkan potensi pertumbuhan sehingga dapat dipertimbangkan sebagai pasar yang prospektif pada masa mendatang.

Berdasarkan hasil yang diperoleh, penerapan algoritma K-Means mampu memberikan gambaran mengenai karakteristik negara asal wisatawan sehingga dapat mendukung proses pengambilan keputusan dalam penyusunan strategi pengembangan pariwisata. Segmentasi yang dihasilkan diharapkan dapat menjadi informasi pendukung dalam menentukan prioritas promosi sesuai karakteristik masing-masing kelompok negara. Pada penelitian berikutnya, analisis dapat dikembangkan dengan menambahkan variabel lain yang berkaitan dengan aktivitas wisatawan serta membandingkan hasil K-Means dengan metode clustering yang berbeda agar diperoleh hasil segmentasi yang lebih mendalam.

REFERENCES

- Aldi, M. A., & Fatah, Z. (2025). Implementasi K-Means Clustering dalam Pengelompokan Data Kunjungan Wisatawan Asing di Indonesia.
- Amellia, S., & Kurnia, Y. (2025). Perbandingan Algoritma K-Means dan Hierarchical Clustering dalam Sistem Rekomendasi Destinasi Wisata di Indonesia.
- Badan Pusat Statistik. (2025). Statistik Kunjungan Wisatawan Mancanegara ke Indonesia.
- Dharmawan, K. S., Sasmita, G. M. A., & Piarsa, I. N. (2025). Clustering Tourism Destinations in Denpasar City Based on Visitor Preferences Using K-Means and DBSCAN.
- Fitriandari, R., & Muhammad, F. (2025). Analisis Klaster Multivariat Kinerja Pasar Pariwisata Kabupaten/Kota di NTT.
- Hesti, L., & Karyati, C. M. K. (2024). Komparasi Algoritma K-Means dan K-Medoids untuk Klasterisasi Data Mining Perjalanan Wisatawan Nusantara.
- Komparasi Hasil Segmentasi Metode K-Means dan Agglomerative Hierarchical Terhadap Provinsi di Indonesia Berdasarkan Profil Perjalanan Wisata Tahun 2024. (2025).
- Liyanto, E. W., Homaidi, A., & Lutfi, A. (2024). Implementasi K-Means Clustering Menggunakan RapidMiner dalam Pengelompokan Data Kunjungan Wisatawan Asing di Provinsi Jawa Timur.
- Simorangkir, L., Sany, E., & Feraldi, M. (2024). Penerapan Metode K-Means untuk Pengelompokan Data Kunjungan Wisata.
- Yolandari, N. A., et al. (2024). Analisis Perbandingan K-Means dan DBSCAN Menggunakan Evaluasi Silhouette Index.