Journal of Informatics and Telecommunication Engineering
Sekretariat : Universitas Medan Area, Gedung Fakultas Teknik, Program Studi Informatika
Alamat : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1
E-mail : jite@uma.ac.id

Ref : 016/UMA/JITE/IV/2025 Medan, 21 April 2025
Subject : Letter of Acceptance

To :
Mr./Mrs. **Juni Irsan Farida**

Assalamu'alaikum Wr. Wb
We would like to express our sincere gratitude for your participation in submitting an article to the Journal of Informatics and Telecommunication Engineering (JITE). We hereby inform you that the article listed below:

Paper	**:**	**Pengelompokan Lokasi Pariwisata di Indonesia dengan Menggunakan Variasi Distance pada Algoritma K-Means**
Author	**:**	**Juni Irsan Farida, Andre Hasudungan Lubis**

Based on the review results, we are pleased to inform you that your submitted article has been ACCEPTED for publication in JITE Journal - **Special Issues 2025: Innovations in Predictive Analytics and Sentiment Analysis - Applications in Education, Healthcare, and Social Media**, ISSN: 2549-6247 (Print) ISSN: 2549-6255 (Online).

We would like to thank you for your attention and cooperation.

Wassalamu'alaikum, Wr.Wb.

Best Regards,

Muhathir, ST., M.Kom
Chief Editor

JITE, 4 (2) January 2021 ISSN 2549-6247 (Print) ISSN 2549-6255 (Online)

JITE (Journal of Informatics and Telecommunication Engineering)

Available online http://ojs.uma.ac.id/index.php/jite **DOI : 10.31289/jite.vxix.xxx**

Received: dd-mm-yyyy	Accepted: dd-mm-yyyy	Published: dd-mm-yyyy

Pengelompokan Lokasi Pariwisata di Indonesia dengan Menggunakan Variasi Distance pada Algoritma K-Means

Juni Irsan Farida[1], Andre Hasudungan Lubis[2]

[1,2]Fakultas Teknik, Universitas Medan Area, Indonesia

Juniirsan19@gmail.com, andrelubis2201@gmail.com

Abstrak

Indonesia memiliki kekayaan destinasi wisata yang beragam, namun pemetaan dan pengelompokannya masih menjadi tantangan dalam pengelolaan sektor pariwisata. Penelitian ini bertujuan untuk mengelompokkan lokasi wisata di Indonesia dengan menerapkan algoritma K-Means menggunakan tiga variasi metrik jarak, yaitu Euclidean Distance, Manhattan Distance, dan Canberra Distance. Data penelitian diperoleh dari sumber terbuka, kemudian melalui tahapan pra-pemrosesan, termasuk normalisasi data. Penentuan jumlah klaster optimal dilakukan menggunakan metode Elbow, sedangkan evaluasi hasil clustering dianalisis dengan Silhouette Score dan Davies-Bouldin Index. Hasil penelitian menunjukkan bahwa penggunaan Manhattan Distance menghasilkan nilai Silhouette Score tertinggi (0,321463), yang menunjukkan kualitas pengelompokan yang lebih baik dibandingkan dua metode lainnya. Hasil pengelompokan ini diharapkan dapat memberikan informasi yang lebih komprehensif bagi pemangku kepentingan dalam mendukung strategi promosi dan pengembangan infrastruktur pariwisata secara lebih efektif.

Kata Kunci: Pariwisata, Algoritma K-Means, Metrik Jarak, Clustering.

Abstract

Indonesia is home to a diverse range of tourist destinations, yet the classification and mapping of these locations remain a challenge in tourism management. This study aims to cluster tourist destinations in Indonesia by applying the K-Means algorithm with three distance metric variations: Euclidean Distance, Manhattan Distance, and Canberra Distance. The dataset was sourced from public data repositories and underwent preprocessing steps, including data normalization. The optimal number of clusters was determined using the Elbow Method, while the clustering results were evaluated using the Silhouette Score and Davies-Bouldin Index. The findings indicate that Manhattan Distance produced the highest Silhouette Score (0.321463), suggesting superior clustering performance compared to the other two metrics. The results of this study provide valuable insights for stakeholders in formulating strategic tourism promotion and infrastructure development efforts.

Keywords: Clustering, Tourism, K-Means Algorithm, Distance Metrics.

I. PENDAHULUAN

Indonesia, sebagai negara kepulauan terbesar di dunia, memiliki potensi pariwisata yang sangat beragam, mencakup keindahan alam, kekayaan budaya, serta warisan sejarah yang unik (Damayanti & Puspitasari, 2024). Sektor pariwisata berperan penting dalam perekonomian nasional dengan kontribusi signifikan terhadap Produk Domestik Bruto (PDB), penciptaan lapangan kerja, serta perolehan devisa negara (Aminuddin & Burhanuddin, 2023). Namun, pengelolaan destinasi wisata masih menghadapi berbagai tantangan, termasuk dalam hal pemetaan dan pengelompokan lokasi wisata yang efektif.

Sistem klasifikasi yang sistematis dan berbasis data sangat diperlukan guna meningkatkan efisiensi perencanaan strategis, optimalisasi promosi, serta pemerataan pembangunan sektor pariwisata di seluruh

Indonesia (Ruja et al., 2024). Dalam era digital, pemanfaatan teknologi data mining dan machine learning menjadi solusi yang potensial untuk mengatasi permasalahan tersebut. Salah satu metode yang umum digunakan dalam pengelompokan data adalah algoritma K-Means, sebuah teknik clustering yang populer karena kemampuannya dalam mengelompokkan data ke dalam beberapa klaster berdasarkan kemiripan karakteristik (Sholekha et al., 2024).Kinerja algoritma ini sangat dipengaruhi oleh pemilihan distance metric yang digunakan, yang menentukan kedekatan antar data dalam ruang multidimensi. Beberapa distance metric yang umum digunakan dalam analisis clustering meliputi Euclidean Distance, Manhattan Distance, dan Cosine Similarity, di mana setiap metode memiliki keunggulan dan keterbatasan masing-masing dalam menangani jenis data tertentu (Ramdani et al., 2025).

Dalam konteks pemetaan destinasi wisata di Indonesia, pemilihan distance metric yang optimal dapat meningkatkan akurasi pengelompokan, sehingga menghasilkan klaster yang lebih representatif dan bermanfaat bagi pemangku kepentingan (Aminuddin & Burhanuddin, 2023). Misalnya, Euclidean Distance lebih sesuai untuk data dengan distribusi yang relatif seragam, sementara Manhattan Distance lebih efektif dalam menangani data yang mengandung nilai ekstrem atau outlier (Saniah & Edy Victor Haryanto, 2023). Studi yang dilakukan oleh menunjukkan bahwa variasi dalam pemilihan distance metric dapat memengaruhi kualitas hasil clustering secara signifikan, khususnya dalam konteks analisis data spasial (Sari & Indarti, 2023). Oleh karena itu, penelitian ini bertujuan untuk mengeksplorasi dan membandingkan performa berbagai distance metric dalam algoritma K-Means untuk mengelompokkan lokasi pariwisata di Indonesia. Dengan menerapkan metode clustering yang lebih akurat, diharapkan dapat dihasilkan peta pariwisata yang lebih terstruktur, sehingga dapat mendukung pengambilan keputusan strategis dalam pengembangan infrastruktur, penyusunan kebijakan pariwisata, serta optimalisasi promosi destinasi unggulan (Suraya et al., 2023).

Lebih jauh, penelitian ini juga diharapkan memberikan kontribusi dalam pengembangan metode clustering yang lebih adaptif terhadap karakteristik data spasial dan geografis. Temuan dari penelitian ini diharapkan tidak hanya bermanfaat bagi industri pariwisata tetapi juga memiliki implikasi luas dalam berbagai bidang lain, seperti perencanaan kota, manajemen sumber daya alam, dan pengembangan sistem informasi geografis. Dalam artikel ini, akan dibahas secara komprehensif mengenai implementasi algoritma K-Means dengan berbagai variasi distance metric serta evaluasi performa masing-masing metode dalam konteks pengelompokan lokasi pariwisata di Indonesia. Hasil dari penelitian ini diharapkan dapat menjadi referensi bagi akademisi, peneliti, serta praktisi dalam mengembangkan solusi berbasis data yang lebih efektif dan inovatif untuk pengelolaan destinasi wisata secara berkelanjutan.

II. METODE PENELITIAN

A. Algoritma K-means

K-means merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok (Fikri Sallaby et al., 2022). Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain (Sitepu et al., 2024). Adapun tujuan pengelompokkan data ini adalah untuk meminimalkan fungsi objektif yang diatur dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok (Lubis et al., 2023). Kemudian dihitung jarak antara setiap data dengan setiap pusat cluster Proses dasar algoritma k-means dapat dilihat di bawah ini :

1. Tentukan jumlah cluster yang ingin dibentuk dan tetapkan pusat cluster K.
2. Menggunakan jarak Euclidean kemudian hitung setiap data ke pusat cluster

$$d(i,k) = \sqrt{\sum_{i}^{m} (C_{\,i}j - C_{k}\,j)^2} \quad (1)$$

3. Kelompokkan data ke dalam cluster dengan jrak yang paling pendek dengan persamaan

$$min \sum_{k}^{i} -a_{ik} = \sqrt{\sum_{i}^{m} (C_{\ i}j - C_{k}j\ )^2} \quad (2)$$

4. Hitung pusat cluster yang baru menggunakan persamaan

$$C_{kj} = \frac{\sum_{k}^{i} x_{ij}}{p} \quad (3)$$

5. Ulangi langkah dua sampai dengan empat sehingga sudah tidak ada lagi data yang berpindah ke cluster yang lain.

B. Metode Elbow

Metode Elbow digunakan untuk menentukan jumlah klaster optimal dengan melihat perubahan nilai inertia atau Within-Cluster Sum of Squares (WCSS) (Fitriyah et al., 2023). Metode Elbow muncul sebagai salah satu cara untuk mengevaluasi jumlah klaster yang optimal. Metode ini merupakan teknik klasik dalam evaluasi clustering dengan membentuk titik siku (elbow) pada suatu titik dan membandingkan nilai atau persentase dari sejumlah yang telah diuji (Winarta & Kurniawan, 2021). Grafik hubungan jumlah klaster dengan kesalahan yang semakin berkurang menunjukkan nilai pada kombinasi elbow menggunakan K-Means . Grafik ini akan menurun secara bertahap ketika nilai meningkat hingga akhirnya stabil. Metode ini menggunakan Sum Squared Error (SSE) untuk menilai jumlah klaster yang muncul selama pengujian dengan K-Means. Nilai SSE akan semakin menurun seiring bertambahnya jumlah yang diuji. SSE menggunakan jumlah anggota klaster terhadap pusatnya untuk memvalidasi klaster. Jumlah klaster yang terbaik adalah saat elbow terbentuk dengan sudut yang lebih besar (Suharti et al., 2022). Rumus untuk menghitung nilai SSE ditunjukkan dalam persamaan berikut:

$$SSE = \sum_{k=1}^{k} \sum_{X_i es_K} ||X_i - c_k||_2^2 \quad (4)$$

C. Silhouette Index

Silhouette Index adalah metode yang kuat untuk mengevaluasi hasil klaster selain metode Elbow Metode ini tidak bergantung pada data pelatihan, sehingga lebih cocok untuk tugas clustering (Satriatama et al., 2023).Efektivitas clustering, yang dikenal sebagai indeks silhouette, ditentukan oleh perbedaan antara rata-rata jarak dalam suatu klaster dan jarak minimum antar klister. Indeks ini dapat mengidentifikasi objek yang diposisikan dengan benar dalam klaster mereka dan objek yang berada di zona transisi antara klaster (Oop Sofiyah et al., 2023). Rumus untuk menghitung silhouette score:

$$SI(i) = \frac{b(i) - a\ (i)}{\max\{a(i) - b(i)\}} \quad (5)$$

D. Davies-Bouldin Index

Davies-Bouldin Index merupakan salah satu metode yang digunakan untuk mengukur validitas klaster dalam suatu metode pengelompokan (Nanda Shalsadilla et al., 2023). Kohesi dalam metode ini didefinisikan sebagai jumlah kedekatan data terhadap pusat klaster yang diikutinya, sedangkan separasi didasarkan pada jarak antara pusat klaster yang satu dengan yang lainnya (Muningsih et al., 2021). Pengukuran dengan Davies-Bouldin Index memaksimalkan jarak antar klaster dan serta berusaha meminimalkan jarak antara titik-titik dalam satu klaster. Jika jarak antar klaster maksimal, maka kesamaan karakteristik antara setiap klaster menjadi kecil sehingga perbedaan antara klaster lebih jelas. Sebaliknya, jika jarak intra-klaster minimum, maka setiap objek dalam klaster memiliki tingkat kesamaan karakteristik yang tinggi. Rumus untuk menghitung Davies-Bouldin score :

$$DBI_i = \frac{R_i}{\left|\frac{1}{C_1}\right| \sum_{j=1}^{K} d(C_i, C_j)} \quad (6)$$

E. Normalisasi Data

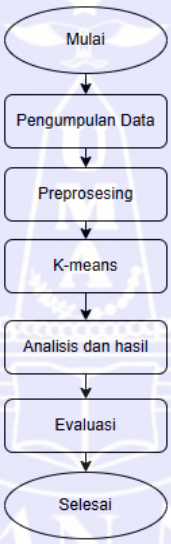
Normalisasi data adalah proses penyesuaian skala nilai atribut dalam dataset agar memiliki rentang nilai yang seragam, sehingga meminimalkan bias akibat perbedaan skala antar atribut (Ahmad Harmain et al., 2022). Normalisasi sangat penting dalam K-Means karena algoritma ini bergantung pada pengukuran jarak, di mana atribut dengan skala lebih besar dapat mendominasi hasil clustering (Mulyani et al., 2025). Salah satu metode normalisasi yang sering digunakan adalah Min-Max Scaling, yang merubah nilai fitur ke dalam rentang.

Rumus Min-Max Scaling:

$$x^1 = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (7)$$

F. Alur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang sistematis untuk memastikan bahwa proses pengelompokan lokasi pariwisata di Indonesia menggunakan algoritma K-Means dengan variasi distance metric dapat dilakukan secara efektif dan menghasilkan output yang akurat. Tahapan penelitian ini dirancang untuk memudahkan pemahaman alur kerja dan memastikan bahwa setiap langkah dilakukan dengan cermat.



Gambar 1. Alur Penelitian

Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari sumber publik yang tersedia di https://www.kaggle.com/datasets/farazbeniqnomf/indonesiaecotourism. Dataset ini berisi informasi mengenai lokasi-lokasi wisata di Indonesia, yang mencakup 182 entri dengan 13 atribut. Dataset ini digunakan sebagai dasar untuk melakukan pengelompokan lokasi wisata berdasarkan algoritma K-Means dengan berbagai metrik jarak. Data yang tersedia akan diproses lebih lanjut, termasuk pembersihan dan transformasi data, sebelum digunakan dalam proses clustering.

Preprocessing

Pada tahap ini, dilakukan serangkaian proses pembersihan dan transformasi data guna memastikan kualitas data yang optimal sebelum digunakan dalam analisis lebih lanjut. Langkah-langkah yang

diterapkan meliputi penanganan data yang hilang dengan berbagai metode seperti imputasi nilai rata-rata, median, atau penggunaan model prediktif. Selain itu, dilakukan normalisasi nilai numerik untuk memastikan skala data yang seragam, sehingga dapat meningkatkan kinerja algoritma yang sensitif terhadap perbedaan skala.

G. K-means Clustering

Pada penelitian ini pengelompokan data menggunakan algoritma K-Means dengan pendekatan tiga jenis metode perhitungan jarak. Dalam algoritma K-Means Clustering, jarak antara titik data dan pusat klaster sangat berpengaruh dalam menentukan hasil pengelompokan (Febrianty et al., 2023). Beberapa metode yang umum digunakan untuk mengukur jarak antara titik data adalah Euclidean Distance, Manhattan Distance, dan Canberra Distance.

Euclidean Distance

Euclidean Distance adalah metode pengukuran jarak paling umum yang digunakan dalam algoritma K-Means. Jarak ini dihitung berdasarkan teorema Pythagoras, yaitu panjang garis lurus antara dua titik dalam ruang multidimensi.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y)^2} \quad (8)$$

Manhattan Distance

Manhattan Distance, juga dikenal sebagai L1 norm, mengukur jarak antara dua titik dengan hanya memperhitungkan perbedaan absolut pada setiap dimensi tanpa mempertimbangkan arah diagonal. Jarak ini disebut juga Taxicab Distance, karena menyerupai cara taksi bergerak dalam jaringan jalan kota yang berbentuk grid.

$$d(x,y) = \sum_{i=1}^{n}|X_i - y_i| \quad (9)$$

Canberra Distance

Canberra Distance adalah metode pengukuran jarak yang lebih sensitif terhadap perbedaan kecil dalam nilai fitur dibandingkan Euclidean dan Manhattan. Jarak ini dihitung dengan membandingkan nilai absolut relatif dari setiap fitur antara dua titik.

$$d(x,y) = \sum_{i=1}^{n}\frac{|x_i - y_i|}{x_i - y_i} \quad (10)$$

H. Evaluasi

Evaluasi hasil clustering merupakan tahap penting untuk menilai kualitas pengelompokan yang dihasilkan oleh algoritma K-Means. Dalam penelitian ini, tiga metrik evaluasi digunakan untuk memastikan bahwa klaster yang terbentuk memiliki struktur yang optimal, yaitu Elbow Method, Silhouette Index, dan Davies-Bouldin Index. Elbow Method membantu menentukan jumlah klaster optimal dengan menganalisis perubahan nilai *Within-Cluster Sum of Squares* (WCSS), di mana titik tekukan (*elbow point*) menunjukkan jumlah klaster yang paling efisien. Silhouette Index digunakan untuk mengevaluasi kepadatan dan pemisahan antar klaster dengan membandingkan rata-rata jarak titik data dalam klaster dengan jarak ke klaster terdekat. Nilai Silhouette yang lebih tinggi menunjukkan bahwa objek berada dalam klaster yang tepat, sementara nilai yang lebih rendah menunjukkan kemungkinan adanya kesalahan pengelompokan.

Selain itu, Davies-Bouldin Index digunakan untuk mengukur validitas clustering dengan membandingkan rasio rata-rata jarak dalam klaster terhadap jarak antar klaster. Semakin rendah nilai

Davies-Bouldin Index, semakin baik kualitas klaster yang dihasilkan, karena menunjukkan bahwa klaster yang terbentuk lebih terpisah dengan jelas dan memiliki tingkat kekompakan yang tinggi. Evaluasi menggunakan ketiga metrik ini memastikan bahwa hasil clustering tidak hanya sesuai secara matematis, tetapi juga memiliki interpretasi yang bermakna dalam konteks pengelompokan lokasi pariwisata. Dengan demikian, analisis ini dapat memberikan wawasan yang lebih mendalam terkait pola wisata di Indonesia, serta membantu pengambilan keputusan dalam pengembangan sektor pariwisata berdasarkan karakteristik klaster yang terbentuk.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini berisi informasi mengenai lokasi pariwisata di Indonesia dengan berbagai atribut penting seperti identitas tempat wisata, lokasi, kategori, harga tiket, rating pengunjung, serta gambar dan peta lokasi. Informasi ini dikumpulkan dari sumber yang telah diverifikasi untuk memastikan validitas dan kelengkapan data. Atribut utama dalam dataset meliputi place_id, place_name, place_description, category, city, price, rating, description_location, place_img, gallery_photo_img, dan place_map. Untuk analisis clustering dengan metode K-Means, atribut yang paling relevan adalah category, city, price, dan rating, karena dapat membantu dalam mengelompokkan tempat wisata berdasarkan karakteristik yang mirip. Sebelum dilakukan analisis, dataset mengalami tahap preprocessing yang mencakup pembersihan data, normalisasi nilai numerik, serta pemilihan fitur yang relevan guna meningkatkan akurasi dan efektivitas proses clustering.

B. Preprosesing Data

Tahap pemrosesan data bertujuan untuk menyiapkan dataset sebelum dilakukan analisis clustering menggunakan algoritma K-Means. Pada tebel 1 Data siap di menunjukkan contoh data yang digunakan dalam penelitian ini, yang terdiri dari beberapa atribut utama, yaitu place_id sebagai identitas unik lokasi wisata, place_name sebagai nama tempat wisata, category yang menunjukkan kategori tempat wisata, city yang merepresentasikan lokasi tempat wisata berdasarkan kode kota, price yang menunjukkan harga tiket masuk dalam satuan mata uang, dan rating yang mencerminkan ulasan pengunjung terhadap tempat wisata tersebut. Pada tahap preprocessing, dilakukan pembersihan data untuk menghapus duplikasi dan menangani nilai yang hilang,

Tabel 1. Data Siap

Place_id	Place_name	category	city	Price	rating
1	Taman Nasional Gunung Leuser	10	0	25000.0	45
2	Desa Wisata Munduk	15	1	10000.0	4.5
3	Desa Wisata Penglipuran	8	1	25000.0	4.8
4	Taman Nasional Bali Barat	19	1	15000.0	4.5
5	Bukit Jamur	11	2	12000.0	4.2

C. Normalisasi Data

Normalisasi data dilakukan untuk memastikan bahwa semua atribut numerik memiliki skala yang seragam sehingga tidak ada atribut yang mendominasi dalam proses clustering. Pada tabel 2 Hasil Normalisasi data menunjukkan hasil dari proses normalisasi menggunakan metode StandardScaler, yang mengubah nilai dari setiap fitur menjadi distribusi dengan rata-rata nol dan standar deviasi satu. Atribut yang dinormalisasi dalam penelitian ini meliputi rating, price, city, dan category.

Proses normalisasi ini penting karena atribut seperti harga tiket memiliki skala yang jauh lebih besar dibandingkan rating atau kategori. Jika tidak dinormalisasi, algoritma K-Means dapat memberikan bobot yang lebih besar pada atribut dengan skala tinggi, sehingga dapat mempengaruhi hasil pengelompokan. Dengan normalisasi, semua atribut memiliki kontribusi yang seimbang dalam

perhitungan jarak antar data, sehingga hasil clustering menjadi lebih akurat dan representatif terhadap pola dalam dataset.

Tebel 2. Hasil Normalisasi Data

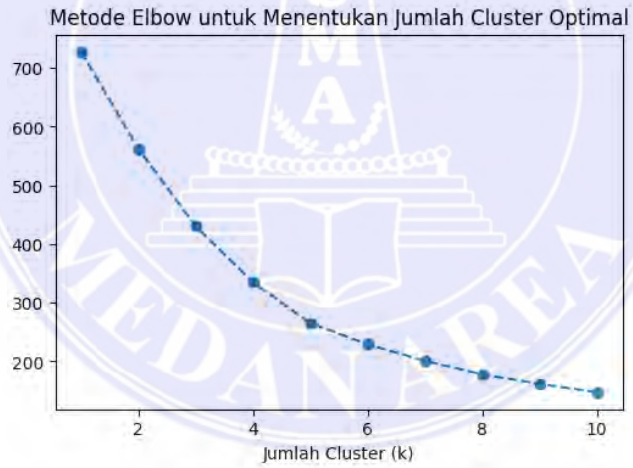
Rating	Price	City	Category
0.335925	-0.064039	-1.416130	0.254833
0.335925	-0.241964	-1.360439	1.164236
1.674719	-0.064039	-1.360439	-0.108929
0.335925	-0.182656	-1.360439	1.891758
-1.002870	-0.218241	-1.304749	0.436713

D. Menentukan Cluster

Menentukan jumlah cluster yang optimal merupakan langkah penting dalam analisis clustering menggunakan algoritma K-Means. Dalam penelitian ini, metode Elbow digunakan untuk menentukan jumlah cluster yang tepat. Metode ini bekerja dengan menghitung Sum of Squared Errors (SSE) untuk berbagai jumlah cluster (k).

E. Metode Elbow

Berdasarkan grafik metode Elbow pada gambar 2 Hasil Metode Elbow dapat kita lihat yang mendekati berdasarkan metode siku adalah angka 4 dan 5, dalam hal ini penulis melihat Jumlah cluster optimal yang dipilih adalah 4, karena pada titik ini terjadi perubahan signifikan dalam nilai Within-Cluster Sum of Squares (WCSS) sebelum grafik melandai. Pemilihan ini memastikan bahwa hasil clustering cukup representatif tanpa menyebabkan overfitting atau segmentasi berlebihan. Selain itu, keputusan ini didukung oleh evaluasi menggunakan Silhouette Score dan Davies-Bouldin Index, yang menunjukkan bahwa jumlah cluster tersebut memberikan hasil pengelompokan yang optimal dalam mengelompokkan lokasi pariwisata berdasarkan karakteristiknya.



Gambar 2. Hasil Metode Elbow

F. Menghitung Cluster dan Centroid

Pada tabel 3 Cluster dan Centroid menunjukkan hasil perhitungan centroid dari 4 cluster yang terbentuk dalam proses K-Means clustering, di mana setiap baris merepresentasikan titik tengah dari masing-masing cluster, dan kolom C1, C2, C3, dan C4 menunjukkan nilai rata-rata atribut yang telah dinormalisasi. Centroid ini berfungsi sebagai representasi utama dari karakteristik kelompok data, di mana nilai yang lebih tinggi atau lebih rendah pada suatu atribut mencerminkan kecenderungan dalam kelompok tersebut. Misalnya, cluster dengan nilai tinggi pada suatu atribut menunjukkan bahwa lokasi wisata dalam cluster tersebut memiliki karakteristik yang lebih menonjol pada atribut tersebut dibandingkan cluster lainnya. Dengan informasi ini, pola dalam pengelompokan dapat dianalisis lebih lanjut untuk memahami perbedaan antar kelompok, seperti menentukan mana yang memiliki harga tiket lebih tinggi, kategori wisata yang lebih populer, atau lokasi dengan rating terbaik berdasarkan ulasan pengunjung.

Tabel 3. Cluster dan Centroid

Cluster	C1	C2	C3	C4
0	0.057818	-0.100571	-1.103779	0.370815
1	-0.019574	-0.031168	0.653858	-1.138558
2	-0.101758	-0.136672	0.715103	0.842447
3	1.228454	7.942606	0.198895	-1.109272

G. Hasil Akhir

Pada tabel 4 Hasil Akhir menunjukkan hasil perhitungan centroid dari 4 cluster yang terbentuk dalam proses K-Means clustering, di mana setiap baris merepresentasikan titik tengah dari masing-masing cluster. Centroid ini berfungsi sebagai representasi utama dari karakteristik kelompok data, di mana nilai yang lebih tinggi atau lebih rendah pada suatu atribut mencerminkan kecenderungan dalam kelompok tersebut. Misalnya, cluster dengan nilai tinggi pada suatu atribut menunjukkan bahwa lokasi wisata dalam cluster tersebut memiliki karakteristik yang lebih menonjol pada atribut tersebut dibandingkan cluster lainnya. Dengan informasi ini, pola dalam pengelompokan dapat dianalisis lebih lanjut untuk memahami perbedaan antar kelompok, seperti menentukan mana yang memiliki harga tiket lebih tinggi, kategori wisata yang lebih populer, atau lokasi dengan rating terbaik berdasarkan ulasan pengunjung.

Tebel 4. Hasil Akhir

Cluster	Total Data	Percentage	Kategori
C1	69	37.91%	Bagus
C2	59	32.42%	Cukup
C3	52	28.57%	Kurang
C4	2	1.10%	Sangat Kurang

H. Evaluasi

Bagian evaluasi bertujuan untuk mengukur kualitas hasil clustering dengan menggunakan dua metrik utama, yaitu Silhouette Score dan Davies-Bouldin Index. Pada tabel 5 Hasil Evaluasi menunjukkan hasil evaluasi clustering menggunakan tiga jenis metrik jarak yang berbeda: Euclidean, Manhattan, dan Canberra. Silhouette Score mengukur sejauh mana objek dalam satu cluster memiliki kemiripan yang lebih tinggi dengan cluster mereka sendiri dibandingkan dengan cluster lain. Nilai tertinggi diperoleh pada jarak Manhattan (0,321463), menunjukkan bahwa metode ini menghasilkan clustering yang lebih baik dibandingkan dengan Euclidean dan Canberra. Davies-Bouldin Index (DBI), yang mengukur kualitas pemisahan antar-cluster, memberikan nilai yang sama untuk ketiga metode (1,014618), menunjukkan bahwa dalam hal pemisahan cluster, tidak ada perbedaan yang signifikan antara ketiga pendekatan jarak ini. Berdasarkan hasil ini, metode Manhattan Distance lebih direkomendasikan karena memiliki Silhouette Score yang lebih tinggi, yang berarti pengelompokan yang lebih jelas dan terpisah dibandingkan dengan dua metode lainnya.

Tabel 5 Hasil Evaluasi

Variasi	Silhouette Score	Davies-Bouldin Index
Euclidean	0.303952	1.014618
Manhattan	0.321463	1.014618
Canberra	0.204389	1.014618

IV. SIMPULAN

Penelitian ini mengkaji penerapan algoritma K-Means dengan berbagai metrik jarak dalam pengelompokan destinasi wisata di Indonesia. Hasil analisis menunjukkan bahwa pemilihan metrik jarak berpengaruh terhadap kualitas clustering, dengan Manhattan Distance menunjukkan performa terbaik

dalam menjaga keterpisahan antar klaster dan keseragaman dalam tiap kelompok. Pengelompokan yang dihasilkan dapat dimanfaatkan dalam berbagai aspek perencanaan pariwisata, termasuk optimalisasi strategi pemasaran dan pengelolaan destinasi berbasis data. Ke depan, penelitian serupa dapat dikembangkan dengan mempertimbangkan variabel tambahan seperti preferensi wisatawan, faktor musiman, serta pendekatan algoritma lain yang lebih adaptif terhadap karakteristik data spasial.

DAFTAR PUSTAKA

Ahmad Harmain, Paiman, P., Kurniawan, H., Kusrini, K., & Dina Maulina. (2022). Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas. *TEKNIMEDIA: Teknologi Informasi Dan Multimedia*, *2*(2), 83–89. https://doi.org/10.46764/teknimedia.v2i2.49

Aminuddin, M. A., & Burhanuddin, A. (2023). Potensi Kekayaan Dan Keberagaman Maritim Di Wilayah Papua Dalam Upaya Mendorong Kesejahteraan Rakyat. *Mandub : Jurnal Politik, Sosial, Hukum Dan Humaniora*, *1*(4), 157–176. https://doi.org/10.59059/mandub.v1i4.607

Damayanti, R. A., & Puspitasari, A. Y. (2024). Kajian Potensi Daya Tarik Wisata Heritage di Indonesia. *Jurnal Kajian Ruang*, *4*(1), 13. https://doi.org/10.30659/jkr.v4i1.36639

Febrianty, E., Awalina, L., & Rahayu, W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail. *Jurnal Teknologi Dan Informasi (JATI)*, *13*(2), 122–137. https://doi.org/10.34010/jati.v13i2

Fikri Sallaby, A., Tri Alinse, R., Novita Sari, V., & Ramadani, T. (2022). Pengelompokan Barang Menggunakan Metode K-Means Clustering Berdasarkan Hasil PenjualanDi Toko Widya Bengkulu PENGELOMPOKAN BARANG MENGGUNAKAN METODE K-MEANS CLUSTERING BERDASARKAN HASIL PENJUALAN DI TOKO WIDYA BENGKULU. *Jurnal Media Infotama*, *18*(1), 2022.

Fitriyah, H., Safitri, E. M., Muna, N., Khasanah, M., Aprilia, D. A., & Nurdiansyah, D. (2023). Implementasi Algoritma Clustering Dengan Modifikasi Metode Elbow Untuk Mendukung Strategi Pemerataan Bantuan Sosial Di Kabupaten Bojonegoro. *Jurnal Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika*, *4*(3), 1598–1607. https://doi.org/10.46306/lb.v4i3.453

Lubis, A. H., Utami, W. R., & Lubis, J. H. (2023). Implementation of k-means clustering for the job provision in urban village. *Jurnal Matematika Dan Ilmu Pengetahuan Alam LLDikti Wilayah 1 (JUMPA)*, *3*(1), 21–31. https://doi.org/10.54076/jumpa.v3i1.312

Mulyani, A., Khoerunisa, S., & Kurniadi, D. (2025). *Perbandingan Kinerja Algoritma KNN dan SVM Menggunakan SMOTE untuk Klasifikasi Penyakit Diabetes*. *14*, 25–34. https://doi.org/10.22146/jnteti.v14i1.15198

Muningsih, E., Maryani, I., & Handayani, V. R. (2021). Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa. *Jurnal Sains Dan Manajemen*, *9*(1), 96.

Nanda Shalsadilla, Shantika Martha, & Hendra Perdana. (2023). Penentuan Jumlah Cluster Optimum Menggunakan Davies Bouldin Index dalam Pengelompokan Wilayah Kemiskinan di Indonesia. *STATISTIKA Journal of Theoretical Statistics and Its Applications*, *23*(1), 63–72. https://doi.org/10.29313/statistika.v23i1.1743

Oop Sofiyah, S., R., N., & Danar Dana, R. (2023). Analisis Efektivitas Pelayanan Publik Menggunakan K-Means Clustering Di Kecamatan Sukagumiwang. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(2), 1291–1296. https://doi.org/10.36040/jati.v7i2.6536

Ramdani, R., Suarna, N., Ali, I., & Efendi, D. I. (2025). PENERAPAN ALGORITMA K-MEANS DALAM ANALISIS DATA KEPENDUDUKAN UNTUK OPTIMALISASI PENGELOMPOKAN DI DESA PASAWAHAN. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(1).

Ruja, I. N., Wahyuningtyas, N., Adi, K. R., Putri, Z. M., Noer, V., & Hidayat, A. (2024). Sinergi Stakeholder dalam Mengembangkan Desa Wisata Berkelanjutan: Pemetaan dan Peran di Desa Samar Tulungagung. *Jurnal Humanitas: Katalisator Perubahan Dan Inovator Pendidikan*, *10*(4), 634–652.

Saniah, S., & Edy Victor Haryanto. (2023). Rancang Bangun Aplikasi Pencarian Lokasi Vaksin Pada Puskesmas Di Kota Medan Menggunakan Metode Euclidean Distance Berbasis Android. *Jurnal Ilmiah Sistem Informasi Dan Ilmu Komputer*, *3*(1), 01–11. https://doi.org/10.55606/juisik.v3i1.365

Sari, I., & Indarti, D. (2023). *TEXT MINING: Praktek Klasifikasi dan Pemodelan Topik dengan Phyton*. uwais inspirasi indonesia.

Satriatama, A. E., Wibowo, A. P., Arnold, I. G. N., Pratama, R. B., Masyhuda, T. A., Agusti, Y. A., Purwanti, E., & Werdiningsih, I. (2023). Analisis Klaster Data Pasien Diabetes untuk Identifikasi Pola dan

Karakteristik Pasien. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *5*(3), 172–182. https://doi.org/10.47233/jteksis.v5i3.828

Sholekha, E., Irawan, B., & Bahtiar, A. (2024). Analisis Penjualan Produk Snack Dan Minuman Menggunakan Metode K-Means Pada Dataset Transaksi Penjualan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(3), 2533–2539. https://doi.org/10.36040/jati.v8i3.9310

Sitepu, E., Suryowati, K., & Pratiwi, N. (2024). Pengelompokan Kabupaten/ Kota Di Provinsi Sumatera Utara Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2022 Dengan Metode Fuzzy C-Means Dan K- Medoids. *Jurnal Statistika Industri Dan Komputasi*, *9*(2), 1–10. https://doi.org/10.34151/statistika.v9i2.4848

Suharti, P. H., Suryandari, A. S., & Amalia, R. N. (2022). Analisis Kinerja Modul Pengendali Tekanan Udara Pct-14 Berbasis Plc Dengan Berbagai Metoda Tuning. *Sebatik*, *26*(2), 420–427. https://doi.org/10.46984/sebatik.v26i2.2134

Suraya, Sholeh, M., & Andayati, D. (2023). Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa. *Skanika*, *6*(1), 51–60. https://doi.org/10.36080/skanika.v6i1.2982

Winarta, A., & Kurniawan, W. J. (2021). Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python. *JTIK (Jurnal Teknik Informatika Kaputama)*, *5*(1), 113–119. https://doi.org/10.59697/jtik.v5i1.593