CLUSTERING DESTINASI WISATA BERDASARKAN POPULARITAS DAN HARGA TIKET MENGGUNAKAN METODE K-MEANS

Faizal Dwi Saputra; Azizah Fatmawati, M.Sc., S.T.
Program Studi Teknik Informatika, Fakultas Komunikasi dan Informatika,
Universitas Muhammadiyah Surakarta

Abstrak

Yogyakarta merupakan kota dengan beragam daya tarik wisata yang menjadikanya salah satu destinasi wisata favorit. Seiring berkembangnya teknologi digital, semakin mudah untuk mendapatkan informasi tentang berbagai tempat wisata di internet. Permasalahan yang kerap muncul adalah kebingungan dalam memilih tujuan wisata yang akan dikunjungi karena dihadapkan pada banyaknya opsi destinasi wisata tersebut. Penelitian ini bertujuan untuk membantu mengelompokkan destinasi wisata di Yogyakarta sehingga dapat memudahkan wisatawan dalam memilih tujuan yang tepat. Untuk mengatasi hal ini, diperlukan salah satu teknis analisis yang lebih canggih seperti analisis klaster atau pengelompokan. Pada penelitian ini menerapkan metode clustering K-Means. Proses analisis dilakukan dengan memanfaatkan dataset dari Kaggle. Hasil analisis menunjukkan terdapat tiga klaster yaitu C0: wisata berbayar dengan popularitas tinggi sebanyak 188 destinasi wisata, C1: wisata berbayar dengan popularitas rendah sebanyak 131 destinasi wisata, C2: wisata gratis dengan popularitas sedang sebanyak 131 destinasi wisata. Hasil pengelompokan ini memberikan rekomendasi destinasi wisata yang terstruktur dan membantu wisatawan dalam menentukan tujuan wisata sesuai preferensi**.**

Kata Kunci: Clustering, data mining, destinasi wisata, k-means

Abstract

Yogyakarta is a city with diverse tourist attractions, making it a favorite destination. With the advancement of digital technology, it's increasingly easy to find information about various tourist attractions online. A common problem is confusion when choosing a destination due to the sheer number of options. This study aims to help categorize tourist destinations in Yogyakarta, making it easier for tourists to choose the right destination. To address this, a more sophisticated analytical technique, such as cluster analysis, is required. This study applies the K-Means clustering method. The analysis process utilizes datasets from Kaggle. The expected outcome of this study is the formation of tourist destination clusters based on specific characteristics, resulting in more structured recommendations and assisting tourists in making effective decisions. The analysis revealed three clusters: C0: paid tourism with high popularity of 188 destinations, C1: paid tourism with low popularity of 131 tourist destinations, and C2: free tourism with medium popularity of 131 tourist destinations. These clustering results provide structured destination recommendations and help travelers choose destinations based on their preferences.

Keywords: *Clustering, data mining, k-means, tourist destination*

1. PENDAHULUAN

Pariwisata adalah aktivitas bepergian yang dilakukan individu atau kelompok ke suatu destinasi di luar wilayah tempat tinggal sehari-hari dengan tujuan tertentu yang dapat mempengaruhi aspek ekonomi, lingkungan, sosial, dan budaya di tempat yang dikunjungi (Romero-castro & Miramontes-viña, 2022). Pariwisata menjadi salah satu faktor yang memiliki peran strategis dalam meningkatkan pertumbuhan ekonomi Indonesia. Potensi sektor ekonomi yang besar dapat mendorong peningkatan pendapatan masyarakat setempat serta membuka peluang kerja baru (Fadilla, 2024). Salah satu kota yang paling diminati adalah Yogyakarta. Yogyakarta dikenal sebagai daerah yang didukung oleh keberadaan berbagai objek wisata yang menarik dan beragam sehingga menjadi salah satu tujuan wisata yang memiliki daya tarik tinggi. Dari berbagai pilihan destinasi, mencakup wisata alam dan budaya hingga religius, kuliner, minat khusus, serta wisata belanja. Di antara berbagai jenis wisata tersebut, wisata budaya dikenal sebagai destinasi yang memiliki daya tarik tinggi bagi pengunjung di kota ini (Purnomo et al., 2021).

Berkat transformasi digital, sektor pariwisata Indonesia memiliki peluang besar dalam upaya memperkuat daya saing di kancah internasional dengan menawarkan pengalaman wisata yang inovatif dan menarik, serta memperkuat citra dan branding destinasi di seluruh dunia (Sibuea, 2025). Seiring berkembangnya teknologi digital, semakin mudah untuk mendapatkan informasi tentang berbagai tempat wisata di internet. Pada akhirnya, ini memengaruhi cara wisatawan memilih tempat wisata (Amellia et al., 2025). Banyaknya pilihan destinasi wisata di Yogyakarta baik wisatawan mancanegara maupun domestik, memiliki keterbatasan dari segi biaya dan waktu dalam melakukan perjalanan. Oleh karena itu, wisatawan yang ingin merencanakan liburan ke Yogyakarta perlu menentukan destinasi wisata yang mampu memenuhi kebutuhan serta minat pengunjung. Permasalahan yang kerap muncul adalah kebingungan dalam menentukan destinasi wisata yang akan dikunjungi karena dihadapkan pada banyak pilihan destinasi wisata tersebut (Utama et al., 2024). Mempertimbangkan faktor-faktor tersebut, dibutuhkan sebuah sistem yang dapat membantu mengolah juga mengelompokkan destinasi wisata berdasarkan popularitas dan harga tiket tertentu untuk menentukan destinasi mana yang dianggap populer dan berkualitas (Murzani & Arianto, 2023). Selain itu, guna untuk strategi promosi destinasi wisata (Rahmiyati & Siswantoro, 2025).

Pengelompokan destinasi wisata secara manual kurang efektif, terutama dengan jumlah data yang semakin besar. Sehingga, diperlukan salah satu teknik analisis yang lebih canggih seperti analisis klaster atau pengelompokan (Wang & Azizurrohman, 2025), khususnya metode *clustering*. *Clustering* ialah salah satu metode pengelompokan data yang memiliki karakteristik serupa ke dalam kelompok yang sama (Setiawan, 2015). *K-Means* merupakan salah satu metode yang banyak digunakan dalam

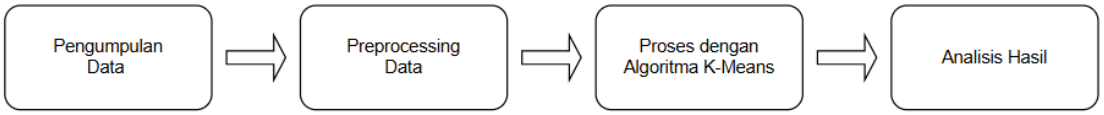
pengelompokan data (Murzani & Arianto, 2023). Keunggulan dari algoritma *K-Means* yaitu implementasi yang mudah dan proses eksekusi yang relatif cepat (Salsabilla, 2021). *K-Means* telah diimplementasikan pada beberapa studi sebelumnya untuk mengelompokkan destinasi wisata berdasarkan tingkat kunjungan wisatawan (Aldi et al., 2025). Penelitian tersebut belum menggabungkan faktor popularitas digital dengan harga tiket sebagai indikator dalam pengelompokan tempat wisata. Penelitian ini mengembangkan pengelompokan destinasi wisata yang lebih relevan bagi wisatawan dengan menggabungkan data popularitas dan harga tiket sebagai variabel utama yang diperoleh dari platform digital. Berdasarkan latar belakang tersebut, penelitian ini diharapkan mampu menghasilkan informasi yang relevan dan akurat serta membantu pengambilan keputusan dengan mengimplementasikan algoritma *K-Means* untuk melakukan pengelompokan destinasi wisata berdasarkan variabel-variabel tertentu. Penelitian ini juga diharapkan dapat membantu membangun sistem informasi pariwisata berbasis data.

2. METODE

Penelitian ini menggunakan metode data mining dan pendekatan kuantitatif untuk mengelompokkan destinasi wisata di Yogyakarta berdasarkan tingkat popularitas dan harga tiket masuk dengan menerapkan metode data mining dan teknik pengelompokan *K-means*. Data mining merupakan tahapan pengambilan informasi yang bernilai berasal dari sekumpulan data yang berukuran besar dan beragam melalui penerapan teknik statistik, algoritma, dan berbagai metode analisis. Tahapan ini bertujuan untuk mengidentifikasi karakteristik, korelasi, maupun kecenderungan tertentu dalam data sehingga menjadi informasi yang berguna, mudah dipahami, serta dapat dijadikan referensi dalam pengambilan keputusan (Nahjan et al., 2023). Penelitian ini menggunakan algoritma *K-Means* karena metode ini efektif dalam mengelompokkan objek berdasarkan tingkat kemiripan karakteristik antar data. Variabel yang digunakan dalam penelitian adalah popularitas destinasi wisata dan harga tiket yang merupakan data numerik sehingga sesuai dengan karakteristik algoritma *K-Means.*

Menurut penelitian terdahulu, algoritma K-Means banyak digunakan dalam analisis *clustering* karena mudah diimplementasikan dan efisiensinya dalam pengolahan data, terutama dalam mengolah data dalam jumlah besar. Menurut penelitian (Ramadhan et al., 2025) *K-Means* mampu menghasilkan pemisahan klaster yang optimal berdasarkan nilai evaluasi seperti *Silhouette Score.* Selain itu penelitian (Apriyanto, 2024) menyatakan bahwa *K-Means* lebih unggul dari *Fuzzy C-Means* dalam hal kecepatan komputasi, meskipun *Fuzzy C-Means* lebih mudah menangani data yang bercampur. Oleh karena itu, *K-Means* digunakan karena dinilai mampu memberikan hasil *clustering* yang optimal dengan proses komputasi yang efisien dan sesuai dengan karakteristik data yang digunakan.

Guna memperoleh jumlah klaster yang baik dapat menerapkan metode *Elbow,* melalui perhitungan nilai *Within-Cluster Sum of Squares* (WCSS) pada berbagai variasi nilai k atau jumlah klaster (Nahjan et al., 2023) dan dapat dilihat melalui *persentase* setiap klaster yang menghasilkan titik belok pada titik tertentu. Metode Elbow biasa digambarkan dalam bentuk grafik guna memperjelas pola siku yang terbentuk (Maori, 2023). Hasil *clustering* diharapkan mampu memberikan gambaran segmentasi destinasi wisata Yogyakarta yang lebih terstruktur dan informatif. Untuk memenuhi kebutuhan lainnya dalam penelitian, dataset yang digunakan dipilih melalui proses seleksi setelah diunduh dari platform Kaggle. Gambar 1 menampilkan proses pembentukan klaster dengan metode *K-Means* (Wasesha & Syafrianto, 2025).



Gambar 1. Proses Clusteing

2.1. Pengumpulan Data

Tahap awal penelitian yaitu mencari data. Data penelitian ini didapatkan dari platform Kaggle. Dataset yang didapat berisi informasi destinasi wisata Yogyakarta dengan jumlah 450 data yang diunduh sekitar satu tahun lalu. Data yang didapat berupa nama_wisata, kategori, rating, jumlah_voting, htm_weekday, htm_weekend, latitude, longitude, image, dan description. Atribut yang digunakan meliputi rating dan jumlah_voting sebagai indikator popularitas destinasi wisata, sedangkan htm_weekday dan htm_weekend sebagai indikator harga tiket berdasarkan hari kunjungan. Setelah proses pengumpulan data, langkah berikutnya pemilihan dan *preprocessing* data guna memastikan bahwa atribut yang digunakan mengacu pada tujuan penelitian serta siap dianalisis. Tabel dataset terdapat pada Tabel 1.

Tabel 1. Dataset Destinasi Wisata Yogyakarta

No	Nama_wisata	Rating	Jumlah Voting	Htm_weekday	Htm_Weekend
1	Candi Prambanan	4,7	108803	50000	50000
2	Candi Borobudur	4,7	103839	150000	150000
3	Tebing Breksi	4,5	66707	10000	10000
4	Nol Kilometer Jl.Malioboro	4,8	58129	0	0
…	…	…	…	…	…
450	Wisata Air Wanatirta	5	1	10000	15000

2.2 Preprocessing Data

Preprocessing data ialah tahapan memperbarui data mentah ke bentuk data yang lebih teratur,

terstruktur, dan konsisten, sehingga siap digunakan untuk proses analisis (Tjortjis, 2025). Proses ini dilakukan dengan menghapus data yang tidak konsisten maupun yang tidak relevan, seperti duplikasi, nilai kosong, normalisasi data, dan outlier (Efendi et al., 2025). Normalisasi data merupakan metode praproses data dengan mengubah nilai suatu atribut ke dalam rentang 0 hingga 1 berdasarkan nilai terkecil dan nilai tertinggi, sehingga menghasilkan skala data yang lebih seimbang untuk proses analisis (Dwididanti et al., 2019). Nilai kosong merupakan kondisi ketika nilai atribut tidak ditemukan atau hilang pada beberapa entri data. Jika tidak dikelola dengan baik, akan berpotensi menyebabkan penurunan kualitas dan hasil analisis (Palanivinayagam & Damaševiˇ, 2023).

2.3 Proses Clustering Menggunakan K-Means

K-Means adalah algoritma clustering non-hierarki yang digunakan untuk mengelompokkan data ke dalam sejumlah klaster berdasarkan tingkat kemiripan karakteristik antar data (Hartono et al., 2023). Algoritma K-Means memerlukan penetapan jumlah kelompok (k). Bekerja berulang kali untuk mengubah posisi centroid, pusat klaster, hingga menghasilkan hasil pengelompokan yang stabil dan ideal (Ahmed et al., 2020). Proses ini menempatkan setiap data ke kelompok yang memiliki centroid terdekat berdasarkan perhitungan jarak, lalu diperbarui secara iteratif hingga konvergen (Park & Choi, 2022). Menurut (Syamfithriani et al., 2023), K-Means memiliki kelebihan berupa proses komputasi yang relatif cepat, sederhana untuk diimplementasikan, serta mampu menghasilkan kualitas pengelompokan yang baik pada data berukuran besar. K-Means juga memiliki beberapa keterbatasan seperti hasil pengelompokan sangat dipengaruhi oleh penentuan centroid awal dan jumlah klaster (k) yang harus ditentukan sebelum proses pengelompokan dilakukan. Selain itu, K-Means juga sensitif terhadap outlier sehingga memengaruhi posisi centroid dan kualitas hasil pengelompokan. Tahapan dalam penyelesaian masalah melalui penerapan teknik *K-Means Clustering* pada proses data mining meliputi (Yetri, 2024):

a. Menentukan jumlah klaster
b. Memilih centroid awal secara manual dari data
c. Menghitung jarak data dari pusat klaster menggunakan rumus *Euclidean Distance.* Persamaan *Euclidean Distance* terdapat pada Persamaan 1.

$$d = \sqrt{\{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2\}} \quad (1)$$

Keterangan:

d = jarak antara dua data

x = nilai variabel pada dimensi pertama

y = nilai variabel pada dimensi kedua

z = nilai variabel pada dimensi ketiga

d. Pusat klaster dihitung menggunakan keanggotaan klaster baru

e. Proses klaster selesai jika tidak berubah, apabila berubah, proses *clustering* dilanjutkan hingga pusat klaster tidak berubah (Permana et al., 2023).

2.4 Analisis Hasil

Hasil pengolahan data menggunakan algoritma *K-Means* digunakan untuk mengidentifikasi karakteristik masing-masing klaster berdasarkan rating, jumlah voting, dan harga tiket. Analisis ini bertujuan untuk mengelompokkan data berdasarkan tingkat popularitas dan harga tiket, serta mempermudah pengunjung memilih destinasi wisata yang sesuai preferensi dan anggaran. Untuk memastikan kualitas hasil pengelompokan, dilakukan evaluasi terhadap model *K-Means.* Evaluasi ini dilakukan untuk menilai tingkat kesamaan data dalam setiap klaster serta perbedaan antar klaster. Metode yang umum digunakan antara lain *Silhouette Score* untuk menilai kualitas klaster dan penentuan klaster yang baik menggunakan metode *Elbow* (Sudrajat et al., 2025). Pada *Elbow Method,* penentuan jumlah klaster dilakukan dengan melihat nilai *Within-Cluster Sum of Squares* (WCSS). Persamaan WCSS terdapat pada Persamaan 2 (Widyawati et al., 2025).

$$\text{WCSS} = \sum_{i-1}^{k} \sum_{j=1}^{n} (x_j - c_i)^2 \quad (2)$$

Keterangan:

x_j= data ke-j

c_i = centroid klaster ke-i

k = jumlah klaster

Selain itu, kualitas klaster juga dapat dievaluasi menggunakan *Silhouette Score. Silhouete Score* merupakan metode evaluasi internal *clustering* yang mengukur seberapa baik objek berada pada klasternya dibandingkan dengan klaster lain. Nilai *Silhouette Score* berada pada rentang -1 hingga 1, di mana nilai yang mendekati 1 menunjukkan bahwa objek berada pada klaster yang semakin baik. Sebaliknya, nilai yang mendekati 0 menunjukkan adanya tumpang tindih antar klaster sehingga objek berada di antara klaster, sedangkan nilai negatif menunjukkan kemungkinan objek ditempatkan pada klaster yang kurang sesuai (Shutaywi, 2021). (Mutawalli & Fadli, 2023) mengatakan bahwa *Silhouette Score* digunakan mengukur sejauh mana objek berada pada suatu klaster. Persamaan *Silhouette Score* terdapat pada Persamaan 3.

$$\text{S} = \frac{b-a}{\max(a,b)} \quad (3)$$

Keterangan:

S = nilai *Silhouette Score* untuk suatu data

α = rata-rata jarak suatu data terhadap klaster yang sama

b = rata-rata jarak data terhadap klaster terdekat lainnya

Memilih jumlah klaster yang tepat merupakan hal penting karena dapat memengaruhi kualitas hasil pengelompokan. Pemilihan yang salah dapat menyebabkan *over-clustering* atau *under-clustering*

(Hartono & Lusiana, 2026). Oleh karena itu, kombinasi beberapa model evaluasi diperlukan agar hasil *clustering* lebih optimal dan representatif.

Berdasarkan hasil evaluasi yang dilakukan, diperoleh jumlah klaster optimal sebanyak tiga klaster (k=3). Selanjutnya, masing-masing klaster diinterpretasikan berdasarkan karakteristik atribut rating, jumlah voting, dan harga tiket yang diperoleh dari nilai centroid hasil *K-Means*. Penamaan setiap klaster dilakukan berdasarkan kombinasi karakteristik ketiga variabel tersebut sehingga dapat menggambarkan perbedaan karakteristik antar kelompok destinasi wisata. Penelitian ini juga terdapat sistem visualisasi berbasis web yang digunakan untuk menampilkan hasil *clustering*. Sistem ini memungkinkan pengguna untuk melihat informasi hasil pengelompokan destinasi wisata dengan lebih jelas dan mudah dipahami.

3. HASIL DAN PEMBAHASAN

Pengelompokan destinasi wisata di Yogyakarta dilakukan dengan menerapkan teknik *K-Means Clustering* berdasarkan karakteristik yang dimiliki oleh setiap destinasi. Pemilihan Variabel dilakukan berdasarkan tujuan penelitian, yaitu mengelompokkan destinasi wisata berdasarkan popularitasnya. Variabel rating digunakan untuk menggambarkan tingkat kepuasan pengunjung, jumlah voting menunjukkan banyaknya partisipasi pengguna dalam memberikan penilaian sehingga mencerminkan tingkat eksposur destinasi, sedangkan harga tiket digunakan sebagai karakteristik ekonomi destinasi yang memungkinkan analisis hubungan antara biaya kunjungan dan tingkat popularitas. Pemilihan variabel yang relevan, proses *clustering* mampu menghasilkan insight yang sesuai dengan tujuan penelitian, yaitu mengidentifikasi kelompok destinasi wisata berdasarkan karakteristik yang dimiliki. Hasil pengelompokan memberikan gambaran mengenai kelompok-kelompok destinasi wisata yang dapat dimanfaatkan untuk memudahkan dalam menentukan tujuan wisata yang mampu memenuhi kebutuhan dan minat pengunjung. Setiap klaster yang terbentuk kemudian diinterpretasikan untuk memberikan informasi yang lebih mudah dipahami oleh wisatawan.

3.1 Pemrosesan data

	nama_wisata	kategori	rating	jumlah_voting	htm_weekday	htm_weekend	latitude	longitude	deskripsi
0	Candi Prambanan	Budaya_Dan_Sejarah	4.7	108803	50000	50000	-7.751.834.561	1.104.915.318	<p>candi prambanan adalah kompleks candi hindu...
1	Candi Borobudur	Budaya_Dan_Sejarah	4.7	103839	150000	150000	-7.607.086.854	1.102.036.226	<p>candi yang pernah masuk sebagai salah satu ...
2	Tebing Breksi	Alam	4.5	66707	10000	10000	-7.781.476.547	1.105.045.757	<p>tebing breksi merupakan tempat wisata yang ...
3	Nol Kilometer Jl.Malioboro	Budaya_Dan_Sejarah	4.8	58129	0	0	-78.013.803	1.103.647.652	<p>titik pusat kota yogyakarta yang menjadi te...
4	Gembira Loka Zoo	Buatan	4.6	55045	70000	85000	-78.062.344	1.103.967.977	<p>gambira loka adalah kebun binatang yang ber...

Gambar 2. Data Awal

Pada Gambar 2 menampilkan data awal yang diimpor dari Excel menggunakan pustaka pandas dalam bahasa pemrograman python. Proses impor data menjadi langkah pertama dalam proses pengolahan data yang dilakukan guna untuk membaca dan memuat data ke dalam struktur dataframe. Data diwakili dalam setiap baris, dan berbagai variabel diwakili dalam kolom-kolomnya

```
nama_wisata      0
kategori         0
rating           0
jumlah_voting    0
htm_weekday      0
htm_weekend      0
latitude         0
longitude        0
deskripsi        0
dtype: int64
```

Gambar 3. Hasil Data Missing Value

Pemeriksaan setiap kolom pada dataset untuk menemukan nilai yang kosong atau tidak ada. Tahap ini sangat penting dalam proses *preprocessing data* karena nilai yang kosong atau bernilai null dapat berdampak pada kualitas analisis. Dataset telah memiliki tingkat kelengkapan yang baik dan tidak memerlukan penanganan khusus seperti hapus data atau pengisian nilai. Kondisi ini menunjukkan bahwa proses penyimpanan data telah dilakukan dengan baik sehingga seluruh atribut memiliki informasi yang sangat lengkap. Gambar 3 menunjukkan bahwa tidak ada nilai yang kosong dalam setiap kolom dataset.

	nama_wisata	kategori	rating	jumlah_voting	htm_weekday	htm_weekend	latitude	longitude	deskripsi	harga_tiket
0	Candi Prambanan	Budaya_Dan_Sejarah	4.7	108803	50000	50000	-7.751.834.561	1.104.915.318	<p>candi prambanan adalah kompleks candi hindu...	50000
1	Candi Borobudur	Budaya_Dan_Sejarah	4.7	103839	150000	150000	-7.607.086.854	1.102.036.226	<p>candi yang pernah masuk sebagai salah satu ...	150000
2	Tebing Breksi	Alam	4.5	66707	10000	10000	-7.781.476.547	1.105.045.757	<p>tebing breksi merupakan tempat wisata yang ...	10000
3	Nol Kilometer Jl.Malioboro	Budaya_Dan_Sejarah	4.8	58129	0	0	-78.013.803	1.103.647.652	<p>titik pusat kota yogyakarta yang menjadi te...	0
4	Gembira Loka Zoo	Buatan	4.6	55045	70000	85000	-78.062.344	1.103.967.977	<p>gambira loka adalah kebun binatang yang ber...	77500

Gambar 4. Penambahan Kolom Harga Tiket

Pada tahap persiapan data, ditambahkan kolom harga_tiket yang digunakan sebagai salah satu variabel dalam penelitian. Kolom diperoleh dari perhitungan rata-rata antara htm_weekday dan htm_weekend. Perhitungan ini dilakukan untuk menghasilkan satu nilai harga tiket yang dapat mewakili biaya masuk destinasi wisata secara keseluruhan untuk memudahkan proses analisis dan pengelompokan menggunakan metode *K-Means Clustering* yang terdapat pada Gambar 4.

	nama_wisata	kategori	rating	jumlah_voting	htm_weekday	htm_weekend	latitude	longitude	deskripsi	harga_tiket
0	Candi Prambanan	Budaya_Dan_Sejarah	4.7	11.597303	50000	50000	-7.751.834.561	1.104.915.318	<p>candi prambanan adalah kompleks candi hindu...	10.819798
1	Candi Borobudur	Budaya_Dan_Sejarah	4.7	11.550607	150000	150000	-7.607.086.854	1.102.036.226	<p>candi yang pernah masuk sebagai salah satu ...	11.918397
2	Tebing Breksi	Alam	4.5	11.108080	10000	10000	-7.781.476.547	1.105.045.757	<p>tebing breksi merupakan tempat wisata yang ...	9.210440
3	Nol Kilometer Jl.Malioboro	Budaya_Dan_Sejarah	4.8	10.970437	0	0	-78.013.803	1.103.647.652	<p>titik pusat kota yogyakarta yang menjadi te...	0.000000
4	Gembira Loka Zoo	Buatan	4.6	10.915924	70000	85000	-78.062.344	1.103.967.977	<p>gambira loka adalah kebun binatang yang ber...	11.258046

Gambar 5. Log Transform Pada Variabel Tertentu

Log transform dilakukan pada variabel jumlah_voting dan harga_tiket untuk mengurangi rentang nilai yang terlalu besar. Kondisi ini berpotensi memengaruhi proses pengelompokan karena perbedaan skala dan jarak antar data. Adanya tahap ini distribusi data menjadi lebih seimbang dalam proses penggunaan metode *K-Means* untuk pengelompokan dan menghasilkan klaster yang optimal hasil *log transform* yang terdapat pada Gambar 5 menunjukkan bahwa sebaran data menjadi lebih terkonsentrasi dan tidak perlu dipengaruhi nilai ekstrem, Selain itu, transformasi log membantu mengurangi tingkat kemencengan (*skewness*) pada data sehingga pola yang terdapat dalam dataset lebih mudah dikenali oleh algoritma. Sehingga kualitas hasil pengelompokan menjadi lebih baik.

3.2 Seleksi Variabel

	rating	jumlah_voting	harga_tiket
0	4.7	11.597303	10.819798
1	4.7	11.550607	11.918397
2	4.5	11.108080	9.210440
3	4.8	10.970437	0.000000
4	4.6	10.915924	11.258046

Gambar 6. Hasil Seleksi Variabel

Seleksi variabel dilakukan dengan memilih atribut yang relevan terhadap tujuan penelitian, yaitu rating, jumlah_voting, dan harga_tiket. Karena tiga variabel ini dianggap memiliki pengaruh yang signifikan dalam menggambarkan kualitas, popularitas, serta biaya kunjungan. Ketiga variabel ini digunakan sebagai dasar untuk pengelompokan destinasi wisata, sementara variabel lain yang tidak digunakan dihilangkan untuk mengurangi kompleksitas data dan meningkatkan fokus analisis yang terdapat pada Gambar 6.

3.3 Normalisasi dengan Min-Max Scaler

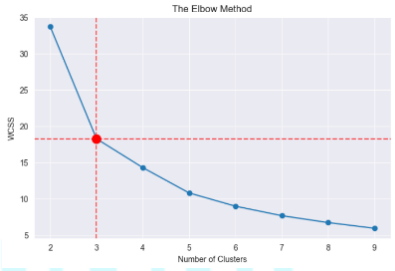
	nama_wisata	rating	jumlah_voting	harga_tiket
0	Candi Prambanan	0.925	1.000000	0.765321
1	Candi Borobudur	0.925	0.995718	0.843029
2	Tebing Breksi	0.875	0.955134	0.651486
3	Nol Kilometer Jl.Malioboro	0.950	0.942511	0.000000
4	Gembira Loka Zoo	0.900	0.937512	0.796320

Gambar 7. Hasil Normalisasi Data

Untuk menghindari dominasi variabel dengan nilai yang lebih besar, normalisasi data dilakukan dengan

menerapkan teknik *Min-Max Scaler* untuk menyesuiakan rentang nilai setiap variabel ke dalam skala yang seragam, yaitu antara 0 dan 1. Keseragaman skala antar variabel memungkinkan proses pengelompokan berlangsung secara lebih objektif dan menghasilkan klaster yang lebih akurat dalam merepresentasikan karakteristik destinasi wisata. Hasil normalisasi menunjukkan bahwa semua variabel telah memberikan kontribusi yang seimbang dalam proses pengelompokan, sehingga siap digunakan pada tahap *clustering* yang terdapat pada Gambar 7.

3.4 Penentuan Jumlah Cluster



Gambar 8. Hasil Elbow Method

Jumlah klaster yang optimal ditentukan dengan menggunakan metode *Elbow* melalui proses perhitungan nilai *Within Cluster Sum of Squares* (WCSS). Nilai WCSS menunjukkan tingkat homogenitas data dalam suatu klaster. Semakin kecil nilai WCSS menunjukkan bahwa anggota klaster memiliki karakteristik yang semakin mirip. Jumlah klaster terbaik ditentukan berdasarkan titik siku (elbow) yaitu ketika penurunan nilai WCSS mulai melambat secara signifikan. Berdasarkan grafik *Elbow Method* pada Gambar 8, perubahan kemiringan grafik mulai terjadi pada K=3 yang terdapat pada titik merah, yang menunjukkan penambahan klaster berikutnya hanya memberikan pengaruh yang kecil terhadap penurunan WCSS. Dengan mempertimbangkan hasil visualisasi grafik serta kemudahan interpretasi hasil pengelompokan, jumlah klaster yang digunakan dalam penelitian ini ditetapkan sebanyak 3 klaster.

Tabel 2. Hasil Silhouette Score

K=2	0.5595282332064421
K=3	0.5381089265596034
K=4	0.5182405138963271
K=5	0.4367627832452223

Hasil perhitungan *Silhouette Score* pada Tabel 2, nilai klaster tertinggi pada K=2 yaitu sebesar 0.5595. Namun, nilai *Silhouette Score* untuk K=3 masih tergolong baik yaitu sebesar 0.5381dan menunjukkan kualitas pengelompokan yang cukup kuat. Meskipun K=2 menghasilkan nilai *Silhouette Score* tertinggi, hasil analisis *Elbow Method* menunjukkan bahwa titik siku berada pada K=3, yang mengindikasikan bahwa penambahan klaster hingga tiga memberikan peningkatan kualitas pengelompokan yang signifikan sebelum laju penurunan WCSS mulai melambat. Selain itu ketika jumlah klaster ditingkatkan menjadi K=4 dan K=5, nilai *silhouette score* justru mengalami penurunan nilai. Hal ini menunjukkan

bahwa penambahan jumlah klaster tidak meningkatkan kualitas pemisahan antar klaster maupun kekompokan anggota dalam setiap klaster. Sehingga penambahan klaster setelah K=3 cenderung menyebabkan data terpecah menjadi kelompok yang lebih kecil tanpa memberikan peningkatan kualitas pengelompokan yang berarti. Dari sisi interpretasi penggunaan 3 klaster memberikan segmentasi data yang lebih rinci dan informatif dibandingkan dengan 2 klaster, yang memungkinkan interpretasi lebih baik dari karakteristik data wisata yang diteliti. Sehingga jumlah klaster yang dipilih untuk penelitian adalah 3 klaster karena keseimbangan kualitas klaster berdasarkan *Silhouette Score* dan *Elbow Method.*

Tabel 3. Data Ternormalisasi

No	Nama Wisata	Rating	Jumlah Voting	Harga Tiket
1	Candi Prambanan	0.925	1.000000	0.765321
2	Candi Borobudur	0.925	0.995718	0.843029
3	Tebing Breksi	0.875	0.955134	0.651486
4	Nol Kilometer Jl.Malioboro	0.950	0.942511	0.000000
5	Gembira Loka Zoo	0.900	0.937512	0.796320
6	Tugu	0.950	0.933627	0.000000
7	Taman Sari	0.900	0.919991	0.691066
8	Keraton Yogyakarta	0.925	0.906900	0.680163
…	…	…	…	…
450	Wisata Air Wanatirta Kencana	1.0	0.000000	0.667268

3.5 Penentuan Centroid

Tabel 4. Centroid Data Destinasi

Cluster	Rating	Jumlah Voting	Harga Tiket
C0	4.70	108803	50000
C1	3.70	154	20000
C2	4.80	58129	0

Berdasarkan Tabel 4 centroid awal dipilih dari tiga data destinasi wisata yang memiliki karakteristik berbeda. Centroid C0 memiliki rating sebesar 4,7, jumlah voting sebanyak 108803, dan harga tiket Rp50.000 sehingga merepresentasikan destinasi berbayar dengan tingkat partisipasi pengguna yang tinggi. Centroid C1 memiliki rating sebesar 3,7 jumlah voting sebanyak 154, dan harga tiket berbayar dengan tingkat partisipasi pengguna yang relatif rendah. Sementara itu, centroid C2 memiliki rating sebesar 4,8, jumlah voting sebanyak 58129, dan harga tiket gratis (Rp0), sehingga merepresentasikan destinasi tanpa biaya masuk dengan tingkat partisipasi pengguna cukup tinggi.

3.6 Perhitungan K-Means

K-Means Clustering digunakan dalam proses klasterisasi destinasi wisata berdasarkan karakteristik yang dimiliki masing-masing destinasi. Terdapat 450 data destinasi wisata Yogyakarta untuk dianalisis dengan penerapan teknik *K-Means Clustering.* Selain untuk pengelompokan destinasi wisata, hasil pengelompokan digunakan guna membantu pemilihan destinasi wisata yang dapat memenuhi kebutuhan dan minat pengunjung, sehingga proses pencarian dan pemilihan tujuan wisata menjadi lebih mudah. Data destinasi wisata yang sudah dilakukan proses normalisasi digunakan dalam penelitian ditunjukkan pada Tabel 3.

Proses pengelompokan diawali dengan penentuan *centroid* awal yang berasal dari beberapa data terpilih secara manual berdasarkan index data. Jarak antara setiap data ke pusat klaster dihitung menggunakan metode *Euclidean Distance.* Hasil perhitungan tersebut menghasilkan matriks jarak terdiri dari C0, C1, dan C2. Berikut merupakan perhitungan penentuan klaster pada iterasi pertama berdasarkan Centroid 0, Centroid 1, dan Centroid 2 dengan mengacu pada rumus (1). *Centroid* awal yang dipilih dari data terdapat pada Tabel 5.

Tabel 5. Centroid Awal

Cluster	Rating	Jumlah Voting	Harga Tiket
C0	0.925	1.000000	0.765321
C1	0.675	0.398956	0.700511
C2	0.950	0.942511	0.000000

Jarak pusat *cluster* pertama

$$d_0 = \sqrt{\{(0.925 - 0.925)^2 + (1.000000 - 1.000000)^2 + (0.765321 - 0.765321)^2\}}$$

= 0.00000

Jarak pusat *cluster* kedua

$$d_1 = \sqrt{\{(0.925 - 0.675)^2 + (1.000000 - 0.398156)^2 + (0.765321 - 0.700511)^2\}}$$

= 0.654182

Jarak pusat *cluster* ketiga

$$d_2 = \sqrt{\{(0.925 - 0.950)^2 + (1.000000 - 0.942511)^2 + (0.765321 - 0.000000)^2\}}$$

= 0.767884

Berdasarkan hasil perhitungan tersebut, diperoleh nilai jarak terhadap *cluster* 0 sebesar 0.00000 *cluster* 1 sebesar 0.654182 dan *cluster* ke 2 sebesar 0.767884. Hasil perhitungan selengkapnya ditunjukkan pada Tabel 5.

Hasil pengelompokan pada iterasi ke-1 yang ditunjukkan pada Tabel 5, menunjukkan bahwa *Cluster* 0 terdiri dari 69 data, *Cluster* 1 terdiri dari 294 data, dan *Cluster* 2 terdiri dari 87 data. Tahap

selanjutnya adalah memperbarui pusat klaster berdasarkan perhitungan nilai rata-rata atribut dari seluruh anggota pada masing-masing klaster. Nilai *centroid* baru diperoleh dari hasil pembagian total nilai setiap dimensi ke-k dengan jumlah anggota pada klaster yang bersangkutan. Berikut perhitungan menentukan centroid baru dari Tabel 6.

Tabel 6. Hasil Iterasi Ke-1

No	Nama Wisata	Cluster 0	Cluster 1	Cluster 2	Jarak Terpendek
1	Candi Prambanan	0.000000	0.654182	0.767884	0.000000
2	Candi Borobudur	0.077826	0.662522	0.845076	0.077826
3	Tebing Breksi	0.132179	0.593075	0.655910	0.132179
4	Nol Kilometer Jl.Malioboro	0.767884	0.928328	0.000000	0.000000
5	Gembira Loka Zoo	0.074099	0.591479	0.797904	0.074099
6	Tugu	0.768600	0.923154	0.008884	0.008884
7	Taman Sari	0.111983	0.567619	0.693238	0.111983
8	The Palace of Yogyakarta (Keraton Yogyakarta)	0.126172	0.566499	0.681553	0.126172
…	…	…	…	…	…
450	Wisata Air Wanatirta Kencana	1.007591	0.515651	1.155886	0.515651

Cluster 0 ada 69 data:

C0=

$$\left(\frac{0.925+0.925+0.875+\ldots+0.900}{69}, \frac{1.000000+0.995718+0.955134+\ldots+0.667224}{69}, \frac{0.765321+0.843029+0.651486+\ldots+0.778217}{69}\right)$$

C0 = (0.88478261, 0.77556423, 0.67466851)

Cluster 1 ada 294 data:

C1=

$$\left(\frac{0.850+0.800+0.875+\ldots+1.000}{294}, \frac{0.681048+0.679033+0.678781+\ldots+0.000000}{294}, \frac{0.680163+0.691066+0.602464+\ldots+0.667268}{294}\right)$$

C1= (0.86522109 0.3076782 0.55858136)

Cluster 2 ada 87:

C2=

$$\left(\frac{0.950+0.950+0.950+\ldots+1.000}{87}, \frac{0.942511+0.933627+0.821836+\ldots+0.127134}{87}, \frac{0.000000+0.000000+0.000000+\ldots+0.000000}{87}\right)$$

C2= (0.85301205, 0.44032434, 0)

Setelah centroid ditentukan, perhitungan kembali diulang untuk menentukan keanggotaan klaster pada setiap data hingga menghasilkan hasil pada iterasi ke-2. Iterasi ke-2 terdapat pada Tabel 7.

Tabel 7. Hasil Iterasi Ke-2

No	Nama Wisata	Cluster 0	Cluster 1	Cluster 2	Jarak Terpendek
1	Candi Prambanan	0.245371	0.725000	0.959012	0.245371
2	Candi Borobudur	0.280054	0.746915	1.019696	0.280054
3	Tebing Breksi	0.181324	0.654161	0.839884	0.181324
4	Nol Kilometer Jl.Malioboro	0.698070	0.849832	0.524211	0.524211
5	Gembira Loka Zoo	0.203120	0.674107	0.947537	0.203120
6	Tugu	0.695999	0.843216	0.515446	0.515446
7	Taman Sari	0.146149	0.627446	0.850663	0.146149
8	The Palace of Yogyakarta (Keraton Yogyakarta)	0.137466	0.614347	0.835608	0.137466
…	…	…	…	…	…
450	Wisata Air Wanatirta Kencana	0.784111	0.353050	0.802550	0.353050

Hasil pengelompokan dari iterasi berikutnya dibandingkan dengan hasil dari iterasi sebelumnya, dan proses iterasi akan terus dilakukan hingga tidak terjadi lagi perubahan posisi data. Pada penelitian ini, proses perhitungan berhenti pada iterasi ke-8. Hasil iterasi ke-8 terdapat pada Tabel 8.

Tabel 8. Hasil Iterasi ke-8

No	Nama Wisata	Cluster 0	Cluster 1	Cluster 2	Jarak Terpendek
1	Candi Prambanan	0.385516	0.843350	1.033935	0.385516
2	Candi Borobudur	0.410096	0.851968	1.089993	0.410096
3	Tebing Breksi	0.321558	0.790282	0.919044	0.321558
4	Nol Kilometer Jl.Malioboro	0.731451	1.026325	0.640513	0.640513
5	Gembira Loka Zoo	0.335027	0.784555	1.016185	0.335027
6	Tugu	0.727745	1.019611	0.631698	0.631698
7	Taman Sari	0.289691	0.756338	0.924278	0.289691
8	The Palace of Yogyakarta (Keraton Yogyakarta)	0.279466	0.744636	0.908594	0.279466
…	…	…	…	…	…
450	Wisata Air Wanatirta Kencana	0.646844	0.214783	0.745744	0.214783

Destinasi wisata Yogyakarta dikelompokkan ke dalam tiga klaster berdasarkan popularitas dan harga tiket. *Cluster* 0 memiliki kriteria wisata berbayar dengan popularitas tinggi terdiri dari 188 destinasi

wisata mencakup Candi Prambanan, Candi Borobudur, Tebing Breksi, dan Gembira Loka Zoo dengan nilai rating, jumlah voting yang tinggi dan harga relatif sedang hingga tinggi berdasarkan hasil pengelompokan. *Cluster* 1 memiliki kriteria wisata berbayar dengan popularitas rendah terdiri dari 131 destinasi wisata mencakup Pakem Sari Water Park, Ecotourism Jatisari Seropan 3, Watu Tekek, dan Goa Sentono dengan nilai rating cukup tinggi tetapi jumlah voting yang cukup rendah dan harga relatif sedang hingga tinggi. *Cluster* 2 memiliki kriteria wisata gratis dengan rating tinggi dan jumlah voting menengah terdiri dari 131 destinasi wisata mencakup Nol Kilometer Jl.Malioboro, Tugu, Masjid Gedhe Kauman, dan Embung Tambakboyo dengan harga tiket yang gratis.

3.7 Centroid Akhir

Setelah proses iterasi *K-Means* mencapai kondisi konvergen, diperoleh nilai centroid akhir untuk masing-masing klaster. Nilai centroid merupakan nilai rata-rata dari setiap variabel pada anggota klaster yang digunakan untuk menggambarkan karakteristik masing-masing kelompok destinasi wisata. Nilai centroid selanjutnya digunakan sebagai dasar dalam menginterpretasikan karakteristik setiap klaster berdasarkan variabel rating, jumlah voting, dan harga tiket. Nilai centroid akhir terdapat pada Tabel 9.

Tabel 9. Centroid Akhir

Cluster	Rating	Jumlah Voting	Harga Tiket
C0	0.870479	0.633680	0.658271
C1	0.862595	0.165054	0.664276
C2	0.870802	0.306914	0.000000

Berdasarkan Tabel 9, terlihat bahwa setiap klaster memiliki karakteristik yang berbeda berdasarkan nilai centroid dari masing-masing variabel. Nilai centroid digunakan sebagai acuan untuk membandingkan karakteristik antar klaster, sehingga dapat diketahui variabel mana yang memiliki nilai relatif tinggi maupun lebih rendah pada setiap klaster.

Cluster 0 memiliki nilai centroid rating sebesar 0.870479, jumlah voting 0.633680, dan harga tiket sebesar 0.658271 . Nilai jumlah voting pada klaster ini merupakan yang tertinggi dibandingkan klaster lainya, sehingga menunjukkan bahwa destinasi wisata pada Cluster 0 memiliki tingkat partisipasi pengguna yang paling tinggi. Selain itu, nilai harga tiket berada pada kategori berbayar dengan tingkat harga yang relatif sedang hingga tinggi. Berdasarkan karakteristik tersebut, Cluster 0 dapat diinterpretasikan sebagai kelompok destinasi berbayar dengan popularitas tinggi.

Cluster 1 memiliki centroid rating sebesar 0.862595, jumlah voting sebesar 0.165054, dan harga tiket sebesar 0.664276. Meskipun memiliki nilai harga tiket yang sedikit lebih tinggi dibandingkan Cluster 0, jumlah voting pada cluster ini relatif rendah. Hal ini menunjukkan bahwa destinasi wisata pada Cluster 1 memperoleh penilaian yang baik dari pengunjung, namun tingkat partisipasi masyarakat

dalam memberikan penilaian masih rendah. Selain itu, Cluster 1 memiliki nilai centroid harga tiket tertinggi sehingga didominasi oleh wisata berbayar. Cluster 1 dapat diinterpretasikan sebagai kelompok destinasi berbayar dengan popularitas rendah.

Cluster 2 memiliki nilai centroid rating sebesar 0.870802, jumlah voting sebesar 0.306914, dan harga tiket sebesar 0.000000. Nilai harga tiket sebesar 0 menunjukkan bahwa klaster ini didominasi oleh destinasi wisata tanpa biaya masuk (gratis). Meskipun jumlah voting lebih rendah dibandingkan Cluster 0, nilai rating pada klaster ini tetap tinggi. Hal ini menunjukkan bahwa destinasi wisata gratis juga mampu memperoleh tingkat kepuasan pengunjung yang baik. Cluster 2 dapat diinterpretasikan sebagai destinasi wisata gratis dengan popularitas sedang.

4. PENUTUP

Penelitian ini menerapkan teknik *K-Means* untuk pengelompokan destinasi wisata Yogyakarta berdasarkan popularitas dan harga tiket masuk. Beberapa tahapan pengolahan dilakukan untuk mengelompokkan data, termasuk pembersihan data, normalisasi atribut, penentuan jumlah klaster, dan penerapan teknik *K-Means* dalam tahapan pengelompokan. Hasil evaluasi dihasilkan dengan membandingkan beberapa nilai K untuk menghasilkan hasil yang baik dan maksimal. Berdasarkan hasil analisis yang telah diperoleh dengan penerapan metode *K-Means Clustering* pada data destinasi wisata di Yogyakarta, dapat disimpulkan bahwa metode *K-Means Clustering* berhasil diterapkan untuk mengelompokkan destinasi wisata berdasarkan karakteristik popularitas dan harga tiket.

Berdasarkan hasil evaluasi, K=2 menghasilkan nilai tertinggi. Namun, penelitian ini menggunakan K=3 karena menghasilkan pengelompokan yang lebih rinci dan mampu menggambarkan karakteristik destinasi wisata dengan lebih informatif. Hasil *clustering* menghasilkan tiga kelompok destinasi wisata, yaitu kategori wisata berbayar dengan popularitas tinggi sebanyak 188 destinasi atau 41,8%, wisata berbayar dengan popularitas rendah sebanyak 131 destinasi atau 29,1%, dan wisata gratis dengan rating tinggi sebanyak 131 destinasi atau 29,1%. Pengelompokan tersebut menunjukkan adanya perbedaan karakteristik antar destinasi berdasarkan tingkat popularitas dan harga tiket. Selain itu, hasil pengelompokan yang diperoleh dapat dimanfaatkan sebagai informasi pendukung dalam memilih tujuan destinasi wisata yang dapat memenuhi kebutuhan dan minat pengunjung.

DAFTAR PUSTAKA

Aldi, M. A., Fatah, Z., Informasi, T., Ibrahimy, U., Timur, S. J., Informasi, S., Ibrahimy, U., & Timur, S. J. (2025). Implementasi K-Means Clustering Dalam Pengelompokan Data Kunjungan. 2(1), 13–19.

Amellia, S., Kurnia, Y., & Rekomendasi, S. (2025). Perbandingan Algoritma K-Means dan Hierarchical Clustering dalam Implementasi Sistem Rekomendasi Wisata di Indonesia. 1(2), 308–315.

Apriyanto, J. (2024). Penerapan Clustering Pada Rata-Rata Lama Sekolah di Provinsi Indonesia Menggunakan Algoritma K-Means dan Fuzzy C-Means. 19(1), 36–42.

Dwididanti, S., Anggoro, D. A., & Sutanto, M. H. (2019). Analisis Perbandingan algoritme Bisecting

K-Means dan Fuzzy C-Means pada Data Pengguna Kartu Kredit. https://doi.org/10.23917/emitor.v22i2.15677

Efendi, M. R., Maulana, F., Ardiansyah, F. S., Rama, S., Rajib, M., Masdi, B., Alfian, D. R., Agung, M., Bazzury, T. A., Lestari, M., Studi, P., Informatika, T., Gedong, K., Rebo, P., & Timur, J. (2025). Analisis preferensi wisatawan terhadap objek wisata di kalimantan menggunakan clustering. 05(01), 81–89.

Fadilla, H. (2024). Pengembangan Sektor Pariwisata untuk meningkatkan Pendapata Daerah Di Indonesia. 2(1), 36–43. https://doi.org/10.31080/BENEFIT.

Hartono, B., & Lusiana, V. (2026). Analisis Metode Elbow SSE , Silhouette Score , dan Jaccard Stability dalam Pemilihan Jumlah Klaster Data yang Optimal TIN : Terapan Informatika Nusantara. 6(8), 1521–1532. https://doi.org/10.47065/tin.v6i8.9271

Maori, N. A. (2023). Metode Elbow Dalam Optimasi Jumlah Cluster Pada K-Means Clustering. 14(2), 277–287.

Murzani, F. F., & Arianto, D. B. (2023). Implementasi Metode Collaborative Filtering pada Algoritma Sistem Rekomendasi Destinasi Wisata di Aceh. 8(3), 111–116.

Mutawalli, L., & Fadli, S. (2023). Komparasi Metode Perhitungan Jarak K-Means Paling Baik Terhadap Pembentukan Pola Kunjungan Wisatawan Mancanegara. 5(1), 159–166. https://doi.org/10.47065/josh.v5i1.4377

Nahjan, M. R., Heryana, N., Voutama, A., Komputer, F. I., Karawang, U. S., & Miner, R. (2023). Implementasi Rapdiminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko OJ Cell. 7(1), 101–104.

Palanivinayagam, A., & Damaševi˘, R. (2023). Effective Handling of Missing Values in Datasets for Classification Using Machine Learning Methods. 1–15.

Permana, M. D. (2023). Klasterisasi Data Jamaah Umrah pada Tanurmutmainah Tour Menggunakan Algoritma K-Means. 10, 15–20. https://doi.org/10.35134/komtekinfo.v10i1.332

Purnomo, B. S., Prasetyaningrum, P. T., Mercu, U., Yogyakarta, B., Mining, D., & K-means, A. (2021). Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan di Kota Yogyakarta Menggunakan Metode K-Means. 1(1), 27–32.

Rahmiyati, D., & Siswantoro, E. B. (2025). Penerapan Metode K-Means untuk Pemetaan Objek Wisata Sebagai Rekomendasi Prioritas Pengembangan Pariwisata. 6(1), 17–24.

Ramadhan, B. F. (2025). Analisis Performa Clustering : K-Means dan Similarity Matrix dalam Evaluasi Silhouette, DBI, CHI, dan Dunn Index. 9, 22961–22967.

Romero-castro, N., & Miramontes-viña, V. (2022). Understanding the Antecedents of Entrepreneurship and Renewable Energies to Promote the Development of Community Renewable Energy in Rural Areas.

Salsabilla, A. (2021). Penerapan Algoritma K-Means Dan C4.5 Untuk Clustering Jurusan Siswa Baru Pada Sekolah Menengah Kejuruan (Studi Kasus: SMK Negeri 1 Paron).

Setiawan, D. (2015). Perancangan Aplikasi K-Means Sebagai Penentu Konsentrasi Bagi Mahasiswa Iinformatika UMS.

Shutaywi, M. (2021). Silhouette Analysis for Performance Evaluation in Machine. 1–17.

Sibuea, G. H. (2025). Strategi Percepatan Implementasi Teknologi Informasi untuk Transformasi Digital Pariwisata di Indonesia. 9(2), 1–8.

Sudrajat, R., Hadiana, A. I., Informatika, J. T., Jenderal, U., Yani, A., Artikel, I., Alam, B., Method, E., & Barat, J. (2025). Evaluasi Kualitas Klaster Wilayah Rawan Bencana Menggunakan K- Means

dengan Silhouette dan Elbow Method. 127–139. https://doi.org/10.33364/algoritma/v.22-2.2379

Tjortjis, P. K. and C. (2025). Data Preprocessing and Feature Engineering for Data Mining :

Utama, A., Sabilla, W. I., & Wakhidah, R. (2024). Sistem Rekomendasi Tempat Wisata di Malang Raya Menggunakan Metode K-Means Clustering. 16(1), 1–15.

Wang, T., & Azizurrohman, M. (2025). Segmenting International Tourists in Indonesia : A Cluster Analysis of Preferences , Motivations , and Behaviors. 9(1). https://doi.org/10.18196/jerss.v9i1.23582

Wasesha, D. A., & Syafrianto. (2025). Penerapan Algorima K-Means Untuk Mengelompokkan Kunjungan Wisatawan Pada Dua Puluh Tempat Wisata Di Jakarta. 16(1), 32–39.

Widyawati, F., Dawod, A. Y., & Santoso, H. A. (2025). K-Means Clustering Optimization of Toddler Malnutrition Status Using Elbow Method. 4(2).

Yetri, M. (2024). Metode K-Means Clustering untuk Pengolahan Data Pengunjung Wisata Kebun Binatang R-Zoo Serdang Bedagai. 3, 383–390.