

Prediksi Harga Masuk dan Rating Wisata di Pulau Jawa dengan Metode K-Means & DBSCAN

Ahmad Zulqifar¹, Apip Wahyudi², Achmad Hanif³, Riza Samsinar⁴

^{1,2,3,4}Teknik Elektro, Universitas Muhammadiyah Jakarta

E-mail: 120210410200036@student.umj.ac.id 20210410200008@student.umj.ac.id,

324040270002@student.umj.ac.id [riza.samsinar@umj.ac.id*](mailto:riza.samsinar@umj.ac.id)

Abstrak – Penelitian ini bertujuan untuk memprediksi harga masuk dan rating wisata di Pulau Jawa dengan menggunakan metode clustering K-Means dan DBSCAN. Data yang digunakan terdiri dari informasi harga tiket masuk dan rating dari berbagai destinasi wisata. Metode K-Means dipilih karena kemampuannya dalam mengelompokkan data ke dalam jumlah cluster yang sudah ditentukan sebelumnya, sedangkan DBSCAN digunakan untuk mengidentifikasi pola berbasis kepadatan data tanpa memerlukan jumlah cluster awal. Analisis dilakukan dengan mengevaluasi performa kedua metode melalui metrik seperti Silhouette Score dan Davies-Bouldin Index, serta menghasilkan cluster berdasarkan statistik harga dan rating. Hasil penelitian menunjukkan bahwa K-Means memiliki performa lebih baik dengan Silhouette Score sebesar 0,42 dan Davies-Bouldin Index sebesar 0,70 dibandingkan DBSCAN yang hanya menghasilkan Silhouette Score 0,18 dan Davies-Bouldin Index sebesar 1,49. Statistik clustering menunjukkan bahwa wisata dengan harga tiket lebih tinggi cenderung memiliki rating yang lebih baik, sedangkan harga tiket yang lebih rendah ditemukan pada destinasi dengan variasi rating yang lebih beragam. Penelitian ini memberikan kontribusi penting dalam memahami pola harga dan preferensi wisata di Pulau Jawa. Hasil ini dapat dimanfaatkan oleh pengelola wisata untuk menentukan strategi segmentasi harga, meningkatkan kualitas layanan, dan menarik berbagai segmen pengunjung. Penelitian lanjutan disarankan untuk memasukkan variabel tambahan, seperti jenis wisata dan fasilitas, serta memperluas analisis ke wilayah lain di Indonesia.

Kata Kunci — prediksi harga masuk, rating wisata, Pulau Jawa, K-Means, DBSCAN

1. PENDAHULUAN

Indonesia merupakan negara yang kaya akan kekayaan alamnya, kekayaan budayanya hingga keindahan tempat wisatanya. Potensi pariwisata di indonesia sangat luar biasa sehingga menjadi favorit bagi wisatawan domestik serta wisatawan mancanegara. Pariwisata merupakan salah satu sektor unggulan yang berkontribusi besar terhadap perekonomian di Indonesia[1]. Pulau Jawa, sebagai pusat ekonomi dan budaya, memiliki beragam destinasi wisata yang menarik perhatian wisatawan domestik maupun mancanegara. Destinasi-destinasi ini mencakup objek wisata alam, budaya, kuliner, dan rekreasi modern yang tersebar di berbagai provinsi seperti Jawa Barat, Jawa Tengah, dan Jawa Timur. Berdasarkan data Badan Pusat Statistik (BPS), jumlah kunjungan wisatawan di Pulau Jawa terus meningkat setiap tahunnya.[2] Dengan beraneka ragamnya tempat wisata di pulau Jawa, tentunya para wisatawan asing maupun lokal yang akan berkunjung ke pulau Jawa membutuhkan waktu yang tidak sedikit untuk menjelajahi berbagai tempat wisata di pulau Jawa. Para wisatawan asing maupun lokal tentunya memiliki keterbatasan dalam hal biaya dan waktu untuk melakukan kunjungan wisata. Oleh karena itu, penelitian ini bertujuan untuk memudahkan para wisatawan dalam memilih tempat wisata berdasarkan dari rating dan harga tiket masuk destinasi [3].

Salah satu tantangan dalam analisis pariwisata adalah mengelompokkan destinasi berdasarkan karakteristik dan popularitasnya (rating). Metode *clustering* atau pengelompokan data menjadi solusi yang efektif untuk mengatasi tantangan ini. Teknik *clustering* dapat membantu dalam mengidentifikasi pola kunjungan dan preferensi wisatawan berdasarkan data yang tersedia. Dalam penelitian ini, dua metode pengelompokan yang akan diterapkan adalah K-Means dan DBSCAN (Density-Based Spatial Clustering of Applications with Noise) [4].

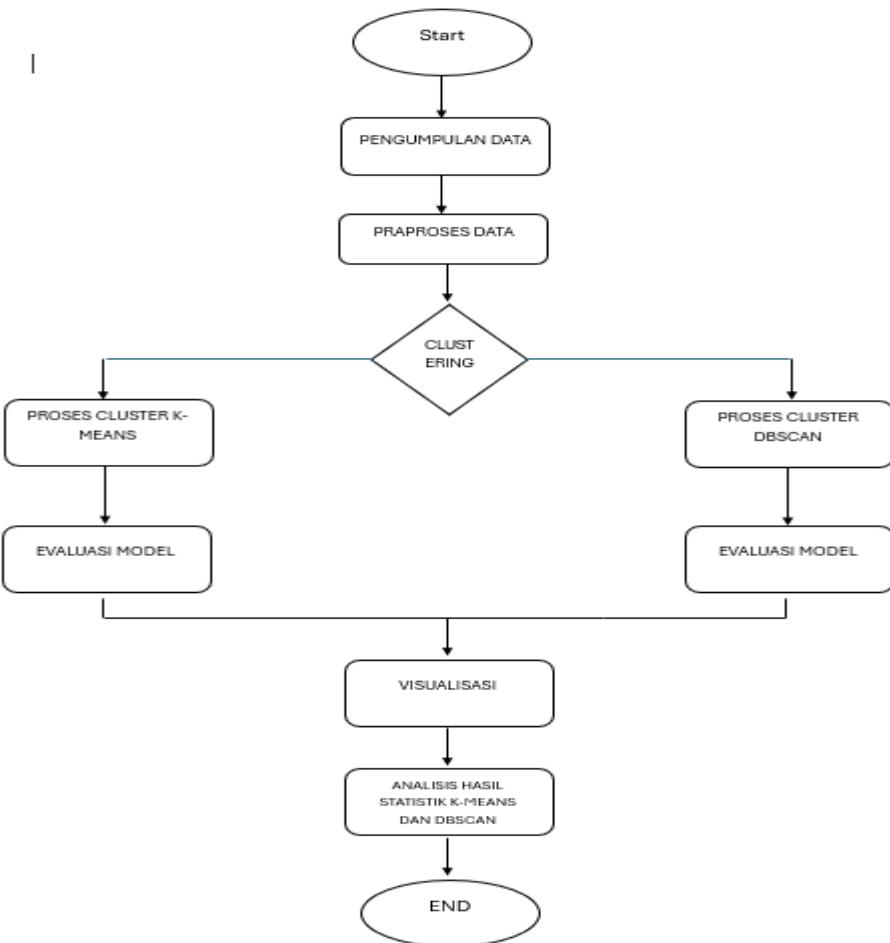
Algoritma K-Means merupakan salah satu metode clustering berbasis jarak yang digunakan dalam analisis data untuk mempartisi kumpulan data menjadi beberapa cluster atau kelompok yang berbeda. Metode K-Means mudah untuk dimplementasikan, memungkinkan proses yang cepat dan efisien, membandingkan kecepatan relatif pada komputasinya dengan metode – metode lainnya, dan metode ini cocok untuk kumpulan data yang besar. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) menyediakan metode yang sangat sederhana untuk mengumpulkan kumpulan data, DBSCAN dapat mengidentifikasi wilayah dengan kepadatan tinggi dan menggabungkannya dengan wilayah kepadatan rendah berdasarkan karakteristik kepadatan yang berbeda setiap cluster. Metode ini melakukan analisis yang efisien berdasarkan nilai ambang batas, unggul dalam mendeteksi klaster dengan bentuk tidak teratur serta menangani data outlier [5][6][7].

Penelitian ini bertujuan untuk:

1. Menganalisis karakteristik destinasi favorit di Pulau Jawa menggunakan metode K-Means dan DBSCAN.
2. Membandingkan performa kedua metode dalam mengelompokkan destinasi berdasarkan tingkat popularitas dan jenis wisata.

3. Memberikan rekomendasi destinasi wisata di pulau Jawa berdasarkan hasil pengelompokan.[8]

2. METODE PENELITIAN



Gambar 1. Alur Proses

Penelitian ini dilakukan melalui beberapa tahapan utama sebagai berikut:

A. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari dataset "kaggel.com" yang memuat informasi tentang destinasi wisata di Pulau Jawa. Dataset ini mencakup beberapa atribut seperti harga tiket masuk (Price), rating destinasi (Rating), dan kategori destinasi (Category). Data yang hilang atau tidak relevan akan dihapus untuk memastikan kualitas analisis [5].

B. Praproses Data

Praproses data dilakukan untuk meningkatkan kualitas data sebelum analisis. Tahapan praproses meliputi:

- Konversi Variabel Kategori: Atribut Category dikonversi menjadi variabel numerik menggunakan metode label encoding.
- Standarisasi Data: Fitur-fitur yang akan digunakan dalam clustering (Price, Rating, dan Category) distandarisasi menggunakan metode StandardScaler untuk menghilangkan perbedaan skala.

C. Clustering

Metode clustering yang digunakan adalah:

- K-Means: Metode ini diterapkan dengan jumlah cluster (k) yang telah ditentukan sebelumnya, yaitu 5 cluster. Algoritma ini diinisialisasi sebanyak 10 kali untuk memastikan hasil yang optimal [9][10].
- DBSCAN: Algoritma ini diterapkan dengan parameter epsilon (ϵ) sebesar 0.5 dan minimal sampel (min_samples) sebanyak 5, untuk mendeteksi cluster berdasarkan densitas data [9].

D. Evaluasi Model

Kualitas cluster dievaluasi menggunakan metrik berikut:

- Silhouette Score: Mengukur seberapa baik data dalam cluster tertentu dibandingkan dengan cluster lainnya. Nilai berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan kualitas clustering yang lebih baik [3][9].
- Davies-Bouldin Index: Mengukur seberapa baik pemisahan antar cluster. Nilai yang lebih rendah menunjukkan kualitas clustering yang lebih baik [5].

E. Visualisasi

Hasil clustering divisualisasikan menggunakan diagram pencar (scatter plot) untuk memetakan data berdasarkan fitur-fitur utama. Selain itu, bar chart digunakan untuk membandingkan hasil evaluasi kedua metode, dan heatmap digunakan untuk menampilkan statistik setiap cluster [2].

F. Analisis Statistik

Analisis statistik dilakukan pada setiap cluster untuk mengidentifikasi pola pada data, seperti rata-rata, nilai minimum, dan nilai maksimum untuk atribut Price dan Rating. Informasi ini membantu dalam memberikan interpretasi yang lebih mendalam terhadap hasil clustering.

Tahapan-tahapan tersebut diimplementasikan secara terstruktur untuk menjawab tujuan penelitian dan memberikan wawasan yang relevan bagi pengelola destinasi wisata.

3. HASIL DAN PEMBAHASAN

Penelitian ini bertujuan untuk menganalisis segmentasi harga dan rating destinasi wisata di Pulau Jawa dengan menggunakan metode K-Means dan DBSCAN. Berdasarkan hasil clustering, metode K-Means menghasilkan lima klaster utama, sedangkan metode DBSCAN menghasilkan enam klaster, termasuk satu klaster outlier (klaster -1).

A. Analisis dengan Metode K-Means

Tabel 1. Hasil Statistik K-Means

KMeans Cluster	Price Mean	Price Min	Price Max	Rating Mean	Rating Min	Rating Max
0	13265.43	0	150000	4.14	3.4	4.3
1	15390.51	0	125000	4.49	4.1	5.0
2	229166.67	150.000	500000	4.44	4.0	4.6
3	12825	0	150000	4.5	4.3	5.0
4	900000	900.000	900000	4.8	4.8	4.8

Metode K-Means membagi data wisata berdasarkan rata-rata harga dan rating dalam setiap klaster. Klaster dengan harga rata-rata tertinggi (Rp900.000) memiliki rating rata-rata 4,8, yang mengindikasikan destinasi eksklusif dengan kualitas dan tempat yang terbaik. Sebaliknya, klaster dengan harga rata-rata terendah (Rp13.265,43) memiliki rating rata-rata 4,13, yang menunjukkan destinasi dengan harga terjangkau namun tetap memiliki daya tarik tertentu.

B. Analisis dengan Metode DBSCAN

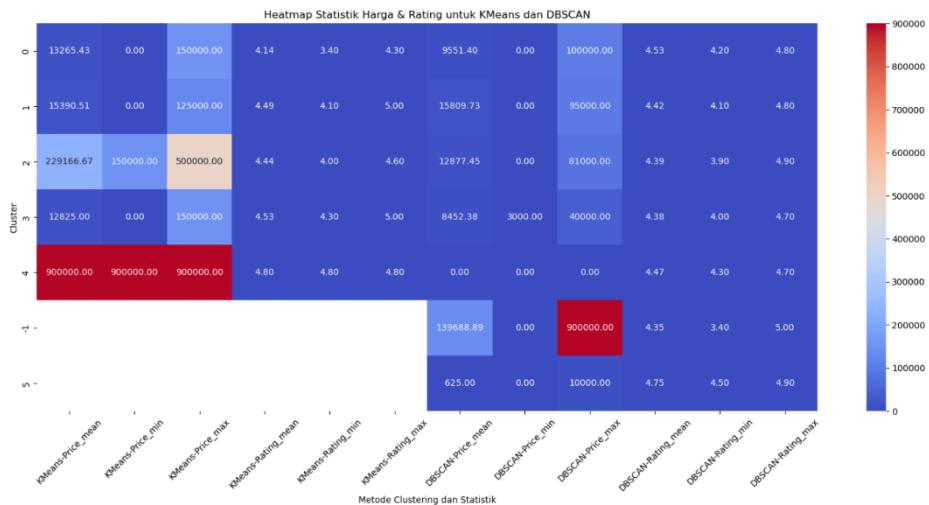
Tabel 2. Hasil Statistik DBSCAN

DBSCAN Cluster	Price Mean	Price Min	Price Max	Rating Mean	Rating Min	Rating Max
-1	139688.89	0	900000	4.35	3.4	5.0
0	9551.4	0	100000	4.53	4.2	4.8
1	15809.73	0	95000	4.41	4.1	4.8
2	12877.45	0	81000	4.39	3.9	4.9
3	8452.3809	3000	40000	4.38	4.0	4.7
	5					
4	0	0	0	4.46	4.3	4.7
5	625.00	0	10000	4.75	4.5	4.9

DBSCAN mampu mengidentifikasi klaster berbasis kepadatan, termasuk outlier yang terdeteksi dalam klaster -1. Klaster outlier memiliki rata-rata harga Rp139.688,89 dan mencakup destinasi dengan variasi harga yang sangat tinggi hingga Rp900.000. Klaster dengan harga rata-rata terendah

(Rp625) memiliki rating rata-rata 4,75, yang menunjukkan destinasi wisata murah dengan kualitas yang tetap baik.

C. Analisis Heatmap



Gambar 2. Heatmap Statistik Harga untuk Kmean dan DBSCAN

Heatmap yang ditampilkan memberikan gambaran statistik harga dan rating dari objek wisata di Pulau Jawa berdasarkan hasil analisis menggunakan metode K-Means dan DBSCAN. Visualisasi ini menampilkan rata-rata, nilai minimum, dan nilai maksimum untuk harga masuk serta rating di setiap cluster yang dihasilkan oleh kedua metode tersebut. Skala warna pada heatmap membantu dalam mengidentifikasi pola, di mana warna merah merepresentasikan nilai yang tinggi, sedangkan warna biru merepresentasikan nilai yang rendah.

Metode K-Means menghasilkan lima cluster utama, yaitu cluster 0 hingga cluster 4. Pada cluster 0 hingga 3, rentang harga rata-rata berada di kisaran Rp12.000,00 hingga Rp229.000,00, dengan harga maksimum berkisar antara Rp125.000,00 hingga Rp500.000,00. Ini menunjukkan bahwa sebagian besar objek wisata di Pulau Jawa memiliki harga masuk yang terjangkau. Sementara itu, cluster 4 secara khusus mencerminkan kategori wisata premium dengan harga yang jauh lebih tinggi, yaitu Rp900.000,00 secara rata-rata, minimum, dan maksimum. Cluster ini menunjukkan wisata eksklusif yang tidak hanya memiliki harga tinggi, tetapi juga mendapatkan rating yang cukup baik dan konsisten, yakni 4,80.

Berbeda dengan K-Means, metode DBSCAN menunjukkan kemampuan untuk menangkap variasi data yang lebih rinci, terutama dalam mendeteksi data yang dianggap sebagai outlier. Dalam heatmap, cluster dengan label -1 merepresentasikan data yang tidak masuk ke dalam kelompok manapun, yang bisa diartikan sebagai wisata yang memiliki karakteristik unik atau jauh berbeda dari kebanyakan data lainnya. Selain itu, DBSCAN juga mampu memisahkan cluster yang mencakup wisata dengan harga sangat murah, bahkan hingga Rp625,00, yang kemungkinan besar mencerminkan wisata gratis atau objek wisata dengan tiket masuk yang sangat terjangkau. Di sisi lain, cluster dengan harga rata-rata mencapai Rp139.688,89 menunjukkan keberadaan wisata yang relatif lebih mahal, tetapi tidak masuk dalam kategori wisata premium.

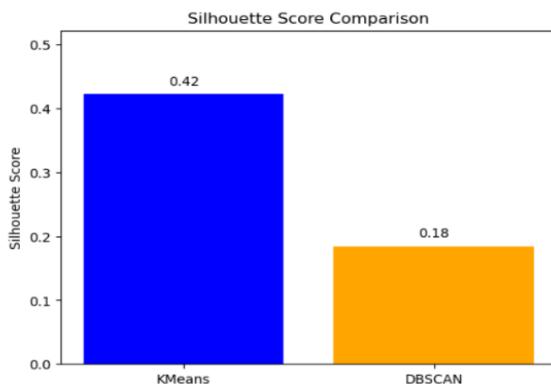
Perbandingan antara K-Means dan DBSCAN menunjukkan perbedaan dalam cara kedua metode tersebut mengelompokkan data. K-Means cenderung membagi data menjadi kelompok-kelompok yang lebih homogen tanpa mempertimbangkan outlier, sehingga cocok digunakan untuk mendapatkan gambaran umum dari distribusi data. Sebaliknya, DBSCAN lebih sensitif terhadap anomali dan variasi harga, yang terlihat dari keberadaan cluster -1 yang menunjukkan adanya data yang secara signifikan berbeda dari kelompok lainnya. Selain itu, DBSCAN juga menunjukkan kemampuan untuk mengidentifikasi wisata murah dan premium dengan lebih jelas.

Secara keseluruhan, hasil analisis ini memberikan wawasan yang penting bagi pengelola wisata di Pulau Jawa untuk memahami karakteristik pasar. Wisata dengan harga premium dan rating tinggi, seperti yang terlihat pada gambar.

D. Evaluasi Clustering

Evaluasi kualitas clustering dilakukan dengan menggunakan dua metrik utama, yaitu Silhouette Score dan Davies-Bouldin Index.

1. Silhouette Score

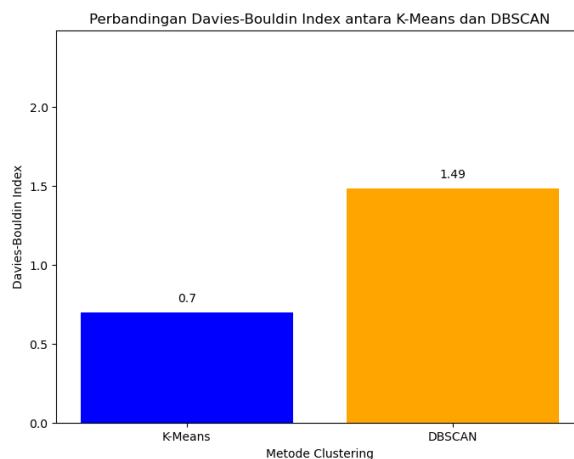


Gambar 3. Silhouette Score Kmean dan DBSCAN

Silhouette Score mengukur seberapa baik objek dalam klaster dikelompokkan. Nilai ini berkisar antara -1 hingga 1, dimana nilai mendekati 1 menunjukkan bahwa objek berada dalam klaster yang tepat. Berikut adalah hasil evaluasi:

- a. K-Means: Nilai Silhouette Score sebesar 0,42 menunjukkan bahwa kualitas clustering cukup baik. Data pada masing-masing klaster terpisah dengan jelas, sehingga memberikan segmentasi yang optimal untuk analisis destinasi wisata.
- b. DBSCAN: Nilai Silhouette Score sebesar 0,18 menunjukkan kualitas clustering yang kurang optimal. Overlap antar klaster masih cukup besar, terutama karena metode ini lebih sensitif terhadap parameter seperti epsilon dan jumlah minimum poin.

2. Davies-Bouldin Index

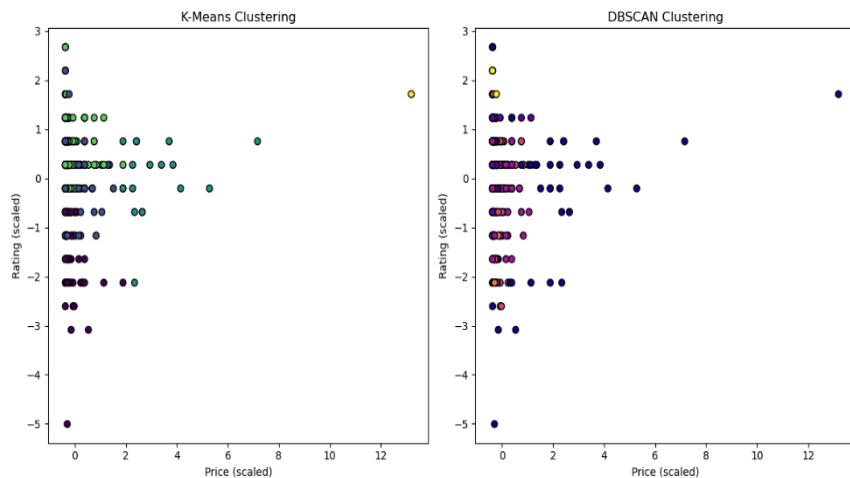


Gambar 4. Davies-Boulding Index antara Kmean dan DBSCAN

Davies-Bouldin Index mengukur rata-rata kemiripan antar klaster dengan mempertimbangkan jarak antara klaster dan lebar klaster itu sendiri. Semakin kecil nilai indeks, semakin baik kualitas clustering. Berikut adalah hasil evaluasinya:

- a. K-Means: Nilai sebesar 0,7 menunjukkan bahwa klaster yang dihasilkan terdefinisi dengan baik. Tingkat overlap antar klaster rendah, sehingga membuatnya lebih unggul dalam mendefinisikan segmentasi destinasi wisata.
- b. DBSCAN: Nilai sebesar 1,49 menunjukkan bahwa klaster yang dihasilkan memiliki tingkat overlap yang lebih tinggi dibandingkan K-Means. Hal ini disebabkan oleh sifat DBSCAN yang bergantung pada kepadatan data, yang mungkin kurang cocok untuk dataset dengan variasi harga dan rating yang cukup besar.

E. Interpretasi dan Prediksi



Gambar 5. Grafik Visualisasi Kmean dan DBSCAN

Hasil clustering menunjukkan bahwa destinasi wisata di Pulau Jawa memiliki segmentasi harga dan rating yang beragam, mulai dari destinasi murah hingga eksklusif. K-Means lebih efektif untuk mengelompokkan destinasi berdasarkan perbedaan harga dan rating yang jelas, sedangkan DBSCAN lebih baik dalam mendeteksi outlier atau destinasi dengan karakteristik unik.

Dengan hasil ini, prediksi harga dan rating untuk destinasi wisata dapat dilakukan lebih akurat berdasarkan klaster yang dihasilkan. Informasi ini dapat dimanfaatkan oleh pengelola wisata untuk menentukan strategi harga dan meningkatkan kualitas layanan guna menarik lebih banyak pengunjung.

4. SIMPULAN

Penelitian ini berhasil memprediksi harga masuk dan rating wisata di Pulau Jawa dengan menggunakan metode clustering K-Means dan DBSCAN. Berdasarkan analisis dari 437 data, metode K-Means menunjukkan performa yang lebih baik dengan Silhouette Score 0.42 dan Davies-Bouldin Index 0.70 dibandingkan DBSCAN yang memperoleh Silhouette Score 0.18 dan Davies-Bouldin Index 1.49. Cluster yang dihasilkan menunjukkan bahwa destinasi wisata dengan harga tiket masuk lebih tinggi cenderung memiliki rating yang lebih baik, sementara destinasi dengan harga lebih rendah menunjukkan variasi pada rating. Hasil ini memberikan gambaran pola harga dan kualitas destinasi wisata di Pulau Jawa, yang dapat menjadi dasar untuk pengelompokan destinasi wisata berdasarkan segmentasi harga dan kualitas.

5. SARAN

Penelitian ini dapat dikembangkan lebih lanjut dengan mempertimbangkan variabel tambahan seperti jenis wisata, fasilitas yang disediakan, dan jumlah pengunjung untuk memberikan analisis yang lebih komprehensif. Selain itu, pengelola wisata dapat manfaatkan hasil penelitian ini untuk menentukan strategi harga yang sesuai dengan segmentasi pasar serta meningkatkan layanan guna meningkatkan rating destinasi. Metode clustering alternatif, seperti Agglomerative Clustering atau Gaussian Mixture Model, juga disarankan untuk digunakan dalam penelitian mendatang guna membandingkan hasil dengan metode yang telah digunakan. Studi serupa juga dapat diperluas ke wilayah lain di Indonesia untuk memahami pola harga dan rating wisata secara nasional.

DAFTAR PUSTAKA

- [1] B. M. Al-Fahmi, E. Rahmawati, and T. Sagirani, “Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 1–9, Aug. 2023.
- [2] A. A. Mahardika, E. N. Kencana, I. K. G. Sukarsa, K. Jayanegara, I. L. Wijayakusumaa, and I. W. Sumarjaya, “Klasterisasi Karakteristik Wisatawan Mancanegara Menggunakan Metode K-MEANS Clustering,” *E-Jurnal Mat.*, vol. 12, no. 2, pp. 1–9, May 2023.
- [3] A. Utama, W. Imama Sabilla, R. Wakhidah, P. Studi Teknik Informatika, P. Negeri Malang, and J. Soekarno Hatta No, “Sistem Rekomendasi Tempat Wisata di Malang Raya Menggunakan Metode K-Means Clustering,” *J. Inform. Multimed.*, vol. 16, no. 1, pp. 1–15, 2024.
- [4] S. Astiti and R. Harman, “Pengelompokan Destinasi Wisata di Batam Berdasarkan Daya Tarik dan Fasilitas Menggunakan Metode K-Means Clustering,” *J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol.

- 5, no. 4, pp. 1–8, 2022.
- [5] A. Saputra and R. Yusuf, “Perbandingan Algoritma DBSCAN dan K-MEANS dalam Segmentasi Pelanggan Pengguna Transportasi Publik Transjakarta Menggunakan Metode RFM,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 4, pp. 1–16, 2024.
- [6] A. Kristianto, “Analisa Performa K-Means dan DBSCAN dalam Clustering Minat Penggunaan Transportasi Umum,” *jurnal.stekom*, vol. 14, no. 2, pp. 1–5, 2021.
- [7] Qadrini, L. Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017. *J Statistika: Jurnal Ilmiah Teori dan Aplikasi Statistika*, 13(2), 5-11, 2020.
- [8] A. Yudhistira and R. Andika, “Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering,” *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 1–9, 2023.
- [9] R. Adha, N. Nurhaliza, and U. Soleha, “Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia,” *J. Sains, Teknol. dan Ind.*, vol. 18, no. 2, pp. 1–6, 2021.
- [10] Ediyanto, M. N. Mara, and N. Satyahadewi, “Pengklasifikasian karakteristik dengan metode k-means cluster analysis,” *Bul. Ilm. Mat. Stat. dan Ter.*, vol. 02, no. 2, pp. 1–4, 2013.