

PERBANDINGAN ALGORITMA K-MEANS DAN HIERARCHICAL CLUSTERING DALAM IMPLEMENTASI SISTEM REKOMENDASI DESTINASI WISATA DI INDONESIA

Shabbina Amellia¹, Yusuf Kurnia^{2*}

^{1,2} Program Studi Informatika, Fakultas Sains dan Teknologi, Universitas Buddhi Dharma

*Corresponding Author, email: yusuf.kurnia@ubd.ac.id

ABSTRAK

Pariwisata merupakan sektor strategis yang terus berkembang di Indonesia dengan keberagaman destinasi yang tersedia. Untuk membantu wisatawan dalam menemukan tempat yang sesuai preferensi, dibutuhkan sistem rekomendasi yang mampu menyaring informasi secara relevan. Penelitian ini bertujuan untuk membandingkan dua algoritma clustering, yaitu K-Means dan Hierarchical Clustering, dalam penerapan sistem rekomendasi destinasi wisata. Data yang digunakan berasal dari kombinasi dataset destinasi wisata Indonesia dari Kaggle serta data rating dan komentar wisatawan dari Google Maps. Penelitian dilakukan melalui tahapan preprocessing data, penerapan kedua algoritma clustering, pembangunan sistem berbasis Python dan Flask, serta evaluasi performa menggunakan metrik Silhouette Coefficient dan efisiensi waktu proses. Hasil pengujian menunjukkan bahwa K-Means memberikan skor Silhouette lebih tinggi sebesar 0.65 dan waktu komputasi yang lebih cepat, dibandingkan Hierarchical Clustering dengan skor 0.61. Visualisasi dari Hierarchical Clustering tetap menunjukkan keunggulan dalam menampilkan struktur relasi antar data. Berdasarkan hasil tersebut, sistem rekomendasi dikembangkan menggunakan hasil kluster K-Means dan diimplementasikan dalam aplikasi web. Kesimpulan dari penelitian ini adalah bahwa pemilihan algoritma clustering sebaiknya disesuaikan dengan konteks sistem dan kebutuhan pengguna. Pendekatan hybrid juga berpotensi memberikan sistem rekomendasi yang lebih akurat dan adaptif terhadap data wisatawan.

Kata kunci: *Hierarchical Clustering*, K-Means, Sistem Rekomendasi, *Silhouette Coefficient*, Wisata Indonesia.

I. PENDAHULUAN

Pariwisata merupakan salah satu sektor strategis dalam pembangunan ekonomi Indonesia. Keberagaman budaya, keindahan alam, serta kekayaan sejarah menjadikan Indonesia sebagai destinasi unggulan baik untuk wisatawan domestik maupun mancanegara. Seiring berkembangnya era digital, informasi destinasi wisata kini dapat diakses dengan mudah melalui internet, sehingga perilaku wisatawan dalam memilih destinasi pun turut berubah. Menurut (Fahira Murzani & Arianto, 2023), sistem rekomendasi memiliki peranan penting dalam membantu wisatawan menyaring informasi yang relevan dengan preferensi mereka. Teknologi

sistem rekomendasi berbasis data menjadi salah satu solusi yang banyak digunakan untuk meningkatkan efisiensi pencarian informasi. Menurut (Cholil et al., 2023; Mutiasari et al., 2021), yang dapat menunjukkan bahwa sistem rekomendasi yang memanfaatkan algoritma clustering dapat meningkatkan relevansi rekomendasi berdasarkan riwayat dan rating pengguna. Salah satu metode yang umum digunakan adalah K-Means Clustering, yang mampu mengelompokkan objek berdasarkan atribut numerik seperti harga dan rating (Ambar Wasesha & Syafrianto., 2024; Ashari et al., 2022). Namun, penelitian (Ayadi et al., 2025; Indra et al., 2023) menekankan bahwa metode alternatif seperti Hierarchical Clustering juga memiliki potensi, terutama dalam memberikan visualisasi struktur data yang lebih informatif. (Cui, 2025) menyatakan bahwa penerapan algoritma clustering dalam sistem pariwisata dapat membantu pengambilan keputusan berbasis preferensi pengguna yang kompleks. Sementara itu, (Oti et al., 2021) menambahkan bahwa integrasi berbagai atribut, termasuk sentimen ulasan dan lokasi geografis, akan memperkaya hasil klasterisasi. (Damos et al., 2024) mengembangkan metode peningkatan performa K-Means melalui penggabungan algoritma genetika, menunjukkan bahwa personalisasi sistem rekomendasi sangat mungkin ditingkatkan. (Vardakas et al., 2024) juga memperkenalkan metrik alternatif untuk meningkatkan pemisahan antar klaster.

Penelitian oleh (Kusumastuti et al., 2024; Ramao, 2021), menekankan pentingnya pemanfaatan nilai lokal dalam pariwisata digital, yang dapat difasilitasi melalui pemodelan sistem berbasis Flask (Walingkas & Saian, 2023). (John et al., 2023; Paliwal, 2024) membuktikan bahwa pemilihan algoritma yang tepat sangat berpengaruh terhadap kecepatan respons dan kepuasan pengguna dalam aplikasi sistem rekomendasi. Dengan demikian, studi komparatif terhadap algoritma perlu dilakukan untuk memilih metode yang paling sesuai dengan konteks dan kebutuhan pengguna akhir. Penelitian ini bertujuan untuk membandingkan performa algoritma K-Means dan Hierarchical Clustering dalam membangun sistem rekomendasi destinasi wisata berbasis web di Indonesia. Sehingga penelitian ini bermanfaat untuk memberikan kontribusi dalam pengembangan sistem rekomendasi berbasis machine learning, khususnya dalam bidang pariwisata, serta menjadi referensi bagi pengembang yang ingin memilih algoritma clustering yang tepat. Penelitian ini

menggunakan data destinasi wisata Indonesia dari Kaggle (<https://www.kaggle.com/datasets/athreal/destinasi-wisata-dataset>) dan Google Maps, serta mengevaluasi algoritma berdasarkan metrik Silhouette Coefficient dan efisiensi waktu proses.

II. METODOLOGI

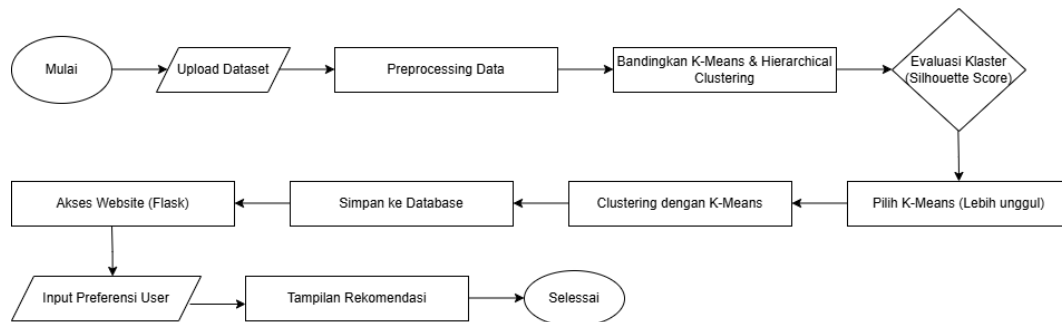
Penelitian ini dimulai dengan tahapan pengumpulan data destinasi wisata dari file CSV yang telah dikurasi, berisi informasi nama tempat, kota, kategori, harga tiket, dan rating. Data ini kemudian diunggah dan diproses melalui tahap *preprocessing* yang meliputi pembersihan data, standarisasi format numerik, dan transformasi fitur untuk keperluan pengelompokan. Menurut (Ashari et al., 2022), *preprocessing* data merupakan langkah penting untuk memastikan kualitas input dalam proses clustering.

Setelah data siap, dilakukan penerapan dua algoritma clustering secara berurutan, yaitu K-Means dan Hierarchical Clustering. Masing-masing algoritma digunakan untuk mengelompokkan destinasi berdasarkan kemiripan atribut seperti harga dan rating. K-Means dipilih karena keunggulannya dalam efisiensi dan kesederhanaan proses iteratif (Ambar Wasesha & Syafrianto., 2024), sedangkan Hierarchical Clustering digunakan karena mampu memberikan representasi struktur relasi antar data yang lebih informatif (Indra et al., 2023). Masing-masing algoritma digunakan untuk mengelompokkan destinasi berdasarkan kemiripan harga dan rating. Hasil kluster dari kedua metode tersebut dievaluasi menggunakan Silhouette Coefficient, yang dirumuskan sebagai:

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

$a(i)$ adalah rata-rata jarak objek i terhadap anggota klasternya sendiri, dan $b(i)$ sebagai jarak rata-rata jarak terhadap kluster terdekat. Menurut (Vardakas et al., 2024), metrik ini sangat efektif dalam menilai pemisahan antar kluster dengan mengukur kesamaan internal dan perbedaan eksternal tiap data. Berdasarkan nilai rata-rata Silhouette tertinggi, algoritma dengan performa terbaik dipilih untuk implementasi sistem.

Selanjutnya, sistem melakukan clustering ulang dengan algoritma terbaik (K-Means) dan menyimpan hasil kluster ke dalam basis data SQLite. Hasil ini kemudian diintegrasikan dalam aplikasi web berbasis Flask. Framework Flask dipilih karena ringan dan mendukung pengembangan sistem interaktif, seperti dibuktikan dalam penelitian (Walingkas & Saian, 2023). Pengguna dapat memasukkan preferensi berupa kota, kategori, dan kisaran harga tiket. Sistem mencocokkan preferensi tersebut dengan hasil klusterisasi untuk menampilkan destinasi wisata yang paling relevan dalam bentuk daftar rekomendasi. Seluruh alur proses penelitian dan implementasi sistem dijelaskan secara sistematis dalam flowchart pada Gambar 1, mulai dari input data hingga sistem menampilkan rekomendasi kepada pengguna akhir.



Gambar 1. Flowchart Sistem Rekomendasi Destinasi Wisata Berbasis K-Means

Flowchart di atas menggambarkan alur proses penelitian dan implementasi sistem rekomendasi destinasi wisata. Tahapan dimulai dari pengumpulan dan preprocessing data, dilanjutkan dengan perbandingan dua algoritma clustering, yaitu K-Means dan Hierarchical Clustering. Setelah dilakukan evaluasi menggunakan Silhouette Coefficient, algoritma K-Means dipilih karena memberikan hasil kluster yang lebih optimal. Hasil kluster digunakan untuk membangun sistem rekomendasi berbasis web menggunakan framework Flask, di mana pengguna dapat memasukkan preferensi kota, kategori, dan harga tiket untuk mendapatkan rekomendasi destinasi yang relevan.

III. HASIL DAN PEMBAHASAN

Hasil pengujian menunjukkan bahwa algoritma K-Means memberikan hasil clustering yang lebih optimal dibandingkan Hierarchical Clustering dalam sistem

rekomendasi destinasi wisata. Nilai Silhouette Coefficient rata-rata untuk K-Means mencapai 0.65, sementara Hierarchical Clustering hanya sebesar 0.61. Hal ini menunjukkan bahwa K-Means mampu menghasilkan kelompok data yang lebih kompak dan terpisah dengan baik. Tabel berikut menampilkan hasil perhitungan evaluasi clustering dari kedua algoritma,

Tabel 1. Hasil Perhitungan Evaluasi Clustering

Algoritma	Silhouette Coefficient	Waktu Proses (detik)
K-Means	0.65	1.82
Hierarchical Clustering	0.61	4.27

Selain lebih akurat, K-Means juga menunjukkan efisiensi waktu komputasi yang lebih tinggi. Hal ini membuat K-Means lebih cocok digunakan dalam sistem berbasis web yang membutuhkan respons cepat. Visualisasi hasil clustering dari K-Means juga menunjukkan distribusi yang jelas dan mudah dipahami. Selanjutnya, sebanyak 34 destinasi wisata digunakan sebagai sampel untuk proses clustering. Destinasi ini mewakili berbagai kategori wisata seperti alam, budaya, dan kuliner. Berdasarkan hasil clustering menggunakan K-Means, destinasi terbagi menjadi dua kelompok besar. Kelompok pertama terdiri dari destinasi dengan rating tinggi dan komentar, seperti Dunia Fantasi, Candi Prambanan, dan Kebun Binatang Ragunan. Sedangkan kelompok kedua berisi destinasi dengan rating sedang hingga rendah dan ulasan yang bervariasi, seperti Pasar Seni dan Goa Pindul. Tabel berikut menampilkan hasil klasterisasi dari 34 destinasi wisata menggunakan algoritma K-Means,

Tabel 2. Hasil Akhir Pengelompokkan 34 Destinasi Wisata

No	Place Name	Cluster
1	Atlantis Water Adventure	1
2	Candi Prambanan	2
3	Curug Cipanas	2
4	Dunia Fantasi	2
5	Goa Pindul	1
6	Gua Belanda	1
7	Jalan Braga	2
8	Jembatan Kota Intan	2
9	Kawasan Malioboro	2
10	Kebun Binatang Ragunan	2
11	Keraton Yogyakarta	1
12	Kidzania	2
13	Kota Tua	2

No	Place_Name	Cluster
14	Monumen Bandung Lautan Api	1
15	Monumen Nasional	2
16	Monumen Palagan Ambarawa	1
17	Museum Bank Indonesia	1
18	Museum Fatahillah	1
19	Museum Gedung Sate	2
20	Museum Geologi Bandung	2
21	Museum Sepuluh Nopember Kota Surabaya	2
22	Ocean Ecopark	2
23	Pantai Parangtritis	2
24	Pasar Seni	1
25	Pelabuhan Marina	1
26	Pulau Bidadari	2
27	Pulau Pari	2
28	Pulau Pelangi	2
29	Pulau Pramuka	2
30	Pulau Tidung	2
31	Semarang Contemporary Art Gallery	2
32	Taman Impian Jaya Ancol	1
33	Taman Mini Indonesia Indah (TMII)	2
34	Trans Studio Bandung	1

Berdasarkan pembagian tersebut, sistem rekomendasi dapat menampilkan hasil yang lebih sesuai dengan preferensi pengguna. Pengguna yang lebih menyukai destinasi populer dan memiliki banyak ulasan positif akan diarahkan ke Cluster 1. Sementara itu, pengguna yang ingin mengeksplorasi tempat yang lebih tenang atau unik akan mendapat rekomendasi dari Cluster 2. Dari pembahasan tersebut, dapat disimpulkan bahwa K-Means lebih unggul secara umum, baik dari segi hasil clustering maupun efisiensi proses. Hierarchical Clustering masih dapat digunakan sebagai alat eksploratif dalam memahami struktur data, tetapi tidak seefisien K-Means untuk implementasi sistem rekomendasi real-time.

IV. SIMPULAN

Penelitian ini bertujuan untuk membandingkan performa algoritma K-Means dan Hierarchical Clustering dalam implementasi sistem rekomendasi destinasi wisata berbasis web. Berdasarkan hasil evaluasi menggunakan metrik Silhouette Coefficient dan waktu proses komputasi, algoritma K-Means menunjukkan performa yang lebih unggul dengan nilai koefisien sebesar 0.65 dan waktu proses

yang lebih cepat dibandingkan Hierarchical Clustering yang hanya mencapai 0.61. Penggunaan algoritma K-Means terbukti menghasilkan kluster yang lebih kompak dan terpisah dengan baik, serta lebih efisien untuk diterapkan dalam sistem rekomendasi real-time. Sistem rekomendasi yang dikembangkan juga berhasil menampilkan destinasi wisata yang sesuai dengan preferensi pengguna berdasarkan hasil klusterisasi. Dengan demikian, pemilihan algoritma clustering yang tepat sangat penting dalam pengembangan sistem rekomendasi berbasis data. K-Means menjadi pilihan yang efektif untuk konteks sistem yang membutuhkan efisiensi dan skalabilitas, sementara Hierarchical Clustering tetap relevan dalam eksplorasi struktur data secara visual.

DAFTAR PUSTAKA

- Ambar Wasesha, D., & Syafrianto. (2024). *Penerapan Algoritma K-Means untuk Mengelompokkan Pekerjaan Baru*. 6(2), 181–186.
- Ashari, I. F. et al. (2022). Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies. *Journal of Applied Informatics and Computing*, 6(1), 07–15.
<https://doi.org/10.30871/jaic.v6i1.3485>
- Ayadi, A. et al. (2025). *Perbandingan tingkat performa metode k-means dan hierachical clustering pada sistem rekomendasi pemilihan kost*. 51–56.
- Cholil, S. R. et al. (2023). *Sistem rekomendasi tempat wisata di kota semarang menggunakan metode collaborative filtering*. 7(1), 118–125.
- Cui, X. (2025). Urban tourism management based on artificial neural networks analysis and data mining. *Scientific Reports*, 15(1), 1–17.
<https://doi.org/10.1038/s41598-025-01237-2>
- Damos, M. A. et al. (2024). *Enhancing the K-Means Algorithm through a Genetic Algorithm Based on Survey and Social Media Tourism Objectives for Tourism Path Recommendations*.
- Fahira Murzani, F., & Arianto, D. B. (2023). Implementasi Metode Collaborative Filtering pada Algoritma Sistem Rekomendasi Destinasi Wisata di Aceh. *Kopelma Darussalam, Kec. Syiah Kuala*, 8(3), 1. Retrieved from <https://jurnal.usk.ac.id/kitektro/article/viewFile/36168/20006>

- Indra, I. et al. (2023). Perbandingan K-Means dan Hierarchical Clustering dalam Pengelompokan Daerah Beresiko Stunting. *INOVTEK Polbeng - Seri Informatika*, 8(2), 356. <https://doi.org/10.35314/isi.v8i2.3612>
- John, J. M. et al. (2023). An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market. *Analytics*, 2(4), 809–823. <https://doi.org/10.3390/analytics2040042>
- Kusumastuti, H. et al. (2024). Leveraging Local Value in a Post-Smart Tourism Village to Encourage Sustainable Tourism. *Sustainability (Switzerland)*, 16(2). <https://doi.org/10.3390/su16020873>
- Mutiasari, H. et al. (2021). Sistem Rekomendasi Film Menggunakan Metode K-Means Clustering (Movie Recommendation System Using K-Means Clustering Method). *E-Proceeding of Engineering*, 8(5), 6755–6764. Retrieved from <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/16511>
- Oti, E. U. et al. (2021). Comprehensive Review of K-Means Clustering Algorithms. *International Journal of Advances in Scientific Research and Engineering*, 07(08), 64–69. <https://doi.org/10.31695/ijasre.2021.34050>
- Paliwal, P. (2024). International Journal of Research Publication and Reviews. *SSRN Electronic Journal*, 5(4), 3326–3333. <https://doi.org/10.2139/ssrn.4909110>
- Ramao, J. (2021). Tourism, Smart Specialization and Sustainable Development. In *Tourism, Smart Specialization and Sustainable Development*. <https://doi.org/10.3390/books978-3-03943-672-9>
- Vardakas, G. et al. (2024). *Deep Clustering Using the Soft Silhouette Score: Towards Compact and Well-Separated Clusters*. Retrieved from <http://arxiv.org/abs/2402.00608>
- Walingkas, H. L., & Saian, P. O. N. (2023). Penerapan Framework Flask pada Pembangunan Sistem Informasi Pemasok Barang. *Jurnal JTIC (Jurnal Teknologi Informasi Dan Komunikasi)*, 7(2), 227–234. <https://doi.org/10.35870/jtik.v7i2.729>