

Implementasi Algoritma *K-Means Clustering* dalam Identifikasi Kelompok Popularitas Hotel di Jawa Barat Berdasarkan Rating dan Ulasan Pengguna

Adik Anindia Putri¹, Dhesta Maulina², Nabilla Dian Azizah Widodo³, Ade Fitria Lestari⁴

^{1,2,3,4} Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika Jakarta

e-mail: 19221159@bsi.ac.id, 19220895@bsi.ac.id, 19220469@bsi.ac.id,
ade.afr@bsi.ac.id

Abstrak

Pertumbuhan industri perhotelan di Jawa Barat membuat masyarakat menghadapi berbagai pilihan hotel dengan karakteristik yang bervariasi. Situasi ini menyulitkan konsumen untuk menentukan hotel berdasarkan tingkat popularitasnya. Rating dan jumlah ulasan di Google Maps dapat digunakan sebagai indikator untuk merepresentasikan tingkat popularitas hotel. Penelitian ini bertujuan untuk mengelompokkan popularitas hotel di Jawa Barat dengan menggunakan algoritma *K-Means* berdasarkan atribut rating dan jumlah ulasan. Data untuk penelitian dikumpulkan melalui proses *web scraping* di Google Maps dengan menggunakan platform Apify. Dataset yang didapat kemudian melalui tahap *preprocessing* hingga menghasilkan 1.147 data hotel yang siap untuk dianalisis. Penentuan jumlah *cluster* dilakukan dengan metode *Elbow*, sedangkan evaluasi kualitas hasil *clustering* menggunakan *Davies-Bouldin Index*. Hasil penelitian menyatakan bahwa jumlah *cluster* yang paling tepat adalah empat, terdiri dari *Cluster 0* sebanyak 44,73%, *Cluster 1* sebanyak 45,42%, *Cluster 2* sebanyak 6,10%, dan *Cluster 3* sebanyak 3,75%. Nilai *Davies-Bouldin Index* yang diperoleh adalah 0,636, yang menunjukkan bahwa proses pengelompokan memiliki kualitas yang baik. Berdasarkan karakteristik nilai centroid, *cluster* yang terbentuk diinterpretasikan menjadi empat kategori tingkat popularitas, yaitu Sangat Populer, Populer, Cukup Populer, dan Kurang Populer. Hasil penelitian menunjukkan bahwa algoritma *K-Means* mampu mengelompokkan tingkat popularitas hotel secara efektif sehingga dapat memberikan informasi yang lebih terstruktur bagi masyarakat dalam memilih hotel serta menjadi bahan pertimbangan bagi pengelola hotel dalam meningkatkan kualitas layanan.

Kata kunci: K-Means Clustering, Popularitas Hotel, Rating, Jumlah Ulasan, Google Maps

Abstract

The growth of the hotel industry in West Java has resulted in various hotel choices with different characteristics. This situation makes it difficult for consumers to determine hotels

based on their popularity level. Ratings and the number of reviews on Google Maps can be used as indicators to represent hotel popularity. This study aims to classify hotel popularity levels in West Java using the *K-Means* algorithm based on rating attributes and the number of reviews. Data collection was carried out through a web scraping process on Google Maps using the Apify platform. The collected dataset was then processed through a preprocessing stage, resulting in 1,147 hotel data records ready for analysis. The determination of the optimal number of clusters was performed using the *Elbow* method, while the quality evaluation of the clustering results was measured using the *Davies-Bouldin Index*. The results showed that the most appropriate number of clusters was four, consisting of *Cluster 0* with 44.73%, *Cluster 1* with 45.42%, *Cluster 2* with 6.10%, and *Cluster 3* with 3.75%. The obtained *Davies-Bouldin Index* value was 0.636, indicating that the clustering process had good quality. Based on the centroid characteristics, the formed clusters were interpreted into four popularity categories: Very Popular, Popular, Moderately Popular, and Less Popular. The results demonstrate that the *K-Means* algorithm can effectively classify hotel popularity levels, providing more structured information for consumers in selecting hotels and serving as a consideration for hotel managers to improve service quality.

Keywords : K-Means Clustering, Hotel Popularity, Rating, Number of Reviews, Google Maps

LATAR BELAKANG

Perkembangan industri pariwisata di Indonesia mendorong peningkatan kebutuhan terhadap fasilitas pendukung, salah satunya layanan akomodasi seperti hotel. Jawa Barat menjadi salah satu provinsi dengan potensi wisata tinggi karena memiliki beragam destinasi unggulan serta wilayah yang strategis dalam aktivitas ekonomi dan perjalanan wisata. Sektor pariwisata terus dikembangkan karena memberikan kontribusi terhadap pendapatan daerah, dengan hotel sebagai salah satu komponen utama dalam mendukung aktivitas tersebut (Hakim, 2025). Peningkatan jumlah wisatawan juga diikuti oleh pertumbuhan akomodasi hotel di berbagai wilayah sehingga menyebabkan persaingan antarhotel semakin meningkat dalam menarik dan mempertahankan pelanggan.

Persaingan industri perhotelan membuat konsumen dihadapkan pada berbagai pilihan hotel dengan karakteristik yang berbeda, baik dari segi harga, fasilitas, maupun kualitas layanan. Tingkat popularitas menjadi salah satu faktor yang dapat digunakan sebagai pertimbangan dalam menentukan pilihan hotel. Popularitas hotel dapat direpresentasikan melalui *rating* dan jumlah ulasan pengguna pada platform digital seperti *Google Maps*. Ulasan pelanggan memiliki peran penting karena memberikan gambaran mengenai pengalaman pengguna terhadap kualitas layanan, fasilitas, dan kepuasan pelanggan. Informasi tersebut juga dapat memengaruhi keputusan konsumen, di mana ulasan positif mampu meningkatkan minat calon tamu, sedangkan ulasan negatif dapat mengurangi ketertarikan terhadap suatu hotel (Messias et al., 2025).

Jumlah informasi hotel yang tersedia pada platform digital menyebabkan konsumen mengalami kesulitan dalam melakukan perbandingan secara cepat dan objektif. Kondisi

tersebut menunjukkan perlunya metode pengolahan data yang mampu menyederhanakan informasi hotel berdasarkan karakteristik tertentu sehingga tingkat popularitas dapat dipahami dengan lebih mudah. Metode *K-Means Clustering* merupakan salah satu teknik dalam *data mining* yang dapat digunakan untuk melakukan pengelompokan data berdasarkan kemiripan karakteristik. Metode ini mampu mengolah data dalam jumlah besar secara efisien serta menghasilkan kelompok data dengan karakteristik yang serupa (Wulandari et al., 2024). Penggunaan atribut *rating* dan jumlah ulasan diharapkan mampu menghasilkan pengelompokan tingkat popularitas hotel yang lebih informatif.

Beberapa penelitian sebelumnya telah menerapkan algoritma *K-Means* pada bidang pariwisata dan perhotelan. Putra dan Dharma (2023) menerapkan metode *K-Means Clustering* untuk mengidentifikasi wisatawan potensial pada Bendesa Hotel dan menghasilkan empat kelompok karakteristik tamu berdasarkan kelas kamar. Hamdani dan Hartama (2024) melakukan analisis tingkat *rating* hotel di Bali menggunakan kombinasi algoritma *K-Means* dan *Decision Tree*, dengan hasil bahwa atribut layanan menjadi faktor yang berpengaruh terhadap tingkat *rating* hotel. Penelitian Yolandari et al. (2025) membandingkan algoritma *K-Means* dan DBSCAN pada pengelompokan data *travel review ratings* menggunakan evaluasi *Silhouette Index* dan *Davies-Bouldin Index*. Hasil penelitian tersebut menunjukkan bahwa algoritma *clustering* mampu digunakan untuk menemukan pola pada data wisatawan.

Penelitian lain oleh Obajha et al. (2023) menerapkan metode *collaborative filtering* untuk membangun sistem rekomendasi hotel dan wisma berbasis website. Penelitian tersebut menunjukkan bahwa pengolahan data pengguna dapat membantu proses pemilihan hotel, tetapi belum melakukan pengelompokan hotel berdasarkan tingkat popularitas secara langsung. Kajian terdahulu menunjukkan bahwa penelitian bidang perhotelan masih banyak berfokus pada segmentasi pelanggan, identifikasi wisatawan potensial, analisis *rating*, serta sistem rekomendasi. Penggunaan kombinasi atribut *rating* dan jumlah ulasan pengguna sebagai indikator utama dalam menentukan tingkat popularitas hotel masih memiliki ruang untuk dikembangkan.

Penelitian ini menerapkan algoritma *K-Means Clustering* untuk mengelompokkan tingkat popularitas hotel di Jawa Barat berdasarkan atribut *rating* dan jumlah ulasan pengguna yang diperoleh dari *Google Maps*. Penentuan jumlah *cluster* optimal dilakukan menggunakan metode *Elbow*, sedangkan evaluasi kualitas hasil pengelompokan dilakukan menggunakan *Davies-Bouldin Index*. Penelitian ini bertujuan menghasilkan informasi tingkat popularitas hotel yang lebih terstruktur sehingga dapat membantu konsumen dalam menentukan pilihan akomodasi serta memberikan masukan bagi pengelola hotel dalam meningkatkan kualitas layanan.

METODE

Berisi Penelitian ini menggunakan pendekatan *data mining* dengan metode *K-Means Clustering* untuk mengelompokkan tingkat popularitas hotel di Jawa Barat berdasarkan atribut *rating* dan jumlah ulasan pengguna. *Data mining* merupakan proses penggalian informasi dari kumpulan data menggunakan teknik statistik, matematika, *artificial intelligence*, dan *machine learning* untuk memperoleh pola atau pengetahuan yang bermanfaat. Tahapan

dalam *data mining* meliputi proses seleksi data, pengolahan, transformasi, analisis data, hingga evaluasi hasil (Liyanto et al., 2024).

Data penelitian diperoleh dari platform *Google Maps* yang menyediakan informasi mengenai lokasi hotel, *rating*, serta jumlah ulasan pengguna. Pengumpulan data dilakukan menggunakan teknik *web scraping* melalui platform Apify. Teknik *web scraping* digunakan untuk memperoleh data dari internet secara otomatis sehingga proses pengumpulan data menjadi lebih cepat, efisien, dan mengurangi kesalahan dalam proses pencatatan manual (Hidayati et al., 2024). Dataset yang diperoleh terdiri dari atribut nama hotel, nilai *rating*, dan jumlah ulasan pengguna yang selanjutnya digunakan sebagai data penelitian (Aryadi et al., 2023).

Tahapan analisis data terdiri dari proses *preprocessing*, penentuan jumlah *cluster* optimal, implementasi algoritma *K-Means*, evaluasi hasil *clustering*, dan interpretasi hasil pengelompokan. Tahap *preprocessing* dilakukan melalui seleksi data, pembersihan data, transformasi data, serta normalisasi untuk memastikan data memiliki kualitas yang sesuai sebelum dilakukan proses pengelompokan.

Penentuan jumlah *cluster* optimal dilakukan menggunakan metode *Elbow*. Metode ini menggunakan nilai *Average Within Centroid Distance* (AWCD) untuk melihat tingkat kedekatan data terhadap pusat *cluster* pada beberapa nilai K. Nilai AWCD dihitung menggunakan persamaan berikut (Permata et al., 2025):

$$AWCD = \frac{\sum_{i=1}^n d(x_i, c)}{n}$$

Keterangan:

$d(x_i, c)$ = jarak data ke-i terhadap *centroid*

x_i = data ke-i

c = *centroid cluster*

n = jumlah data dalam *cluster*

Titik optimal jumlah *cluster* ditentukan berdasarkan titik siku (*elbow point*), yaitu ketika penurunan nilai AWCD mulai mengalami perlambatan sehingga penambahan jumlah *cluster* tidak memberikan peningkatan pengelompokan yang signifikan.

Proses pengelompokan dilakukan menggunakan algoritma *K-Means Clustering*. Algoritma ini merupakan metode pengelompokan data nonhierarki yang membagi data berdasarkan tingkat kemiripan karakteristik antarobjek (Yani et al., 2023). Perhitungan jarak antar data dengan *centroid* menggunakan metode *Euclidean Distance* dengan persamaan berikut (Pansayta & Chansanam, 2024):

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

Keterangan:

$d(x, y)$ = jarak antara data dan *centroid*

x_i = nilai atribut ke-i dari data
 y_i = nilai atribut ke-i dari *centroid*
 n = jumlah atribut atau fitur

Setelah jarak setiap data terhadap *centroid* diperoleh, data dikelompokkan berdasarkan jarak terdekat. Selanjutnya dilakukan pembaruan nilai *centroid* menggunakan nilai rata-rata seluruh data dalam setiap *cluster* dengan persamaan berikut (Maori & Evanita, 2023):

$$C_i = \frac{\sum_{j=1}^n x_j}{n}$$

Keterangan:

C_i = centroid pada cluster ke-i
 x_j = nilai data anggota cluster
 n = jumlah anggota dalam cluster

Proses iterasi dilakukan hingga nilai *centroid* stabil dan tidak terdapat perubahan anggota *cluster*. Pengolahan data dan implementasi algoritma dilakukan menggunakan perangkat lunak RapidMiner yang mendukung proses *preprocessing*, penerapan algoritma *K-Means*, serta evaluasi hasil *clustering* (Sudrajat et al., 2022).

Evaluasi kualitas hasil pengelompokan dilakukan menggunakan *Davies-Bouldin Index* (DBI). DBI merupakan metode evaluasi internal yang mengukur tingkat kedekatan data dalam satu *cluster* (*cohesion*) dan tingkat pemisahan antar-*cluster* (*separation*). Nilai DBI yang semakin rendah menunjukkan kualitas hasil *clustering* yang semakin baik karena menghasilkan kelompok data yang lebih kompak dan memiliki pemisahan yang lebih jelas (Maiyana & Aulia, 2025).

HASIL DAN PEMBAHASAN

HASIL

Pengumpulan dan Preprocessing Data

Data penelitian diperoleh melalui proses *web scraping* pada platform Google Maps menggunakan Apify. Proses pengumpulan data menghasilkan sebanyak 1.260 data hotel di wilayah Jawa Barat. Data yang diperoleh terdiri dari beberapa atribut informasi hotel, seperti nama hotel, *rating*, dan jumlah ulasan pengguna.

No	title	totalScore	reviewCount	street	city	state	countryCode	website	phone	category	category	category	category	category	category	category	category	category	url	categoryName
1	Laut Biru Resort Hotel	4.4	7547	Jl. E. Jaga Laut Kab. Par	Jawa Barat	ID	https://www.lautbiru.com	+62 261 8000000	Hotel										https://www.lautbiru.com	Hotel
2	Hotel Horizon Palma P	4.4	4878	Jl. Pantar Barak Kab. Par	Jawa Barat	ID	https://www.horizonpalma.com		Hotel										https://www.horizonpalma.com	Hotel
3	The Allure Villas Panga	4.6	1909	Kawasan, Jl. G Kab. Par	Jawa Barat	ID	https://sahid.com	+62 81 2000000	Hotel R Villa	Hotel									https://www.sahid.com	Hotel Resor
1258	ALMIRA HOMESTAY			Dusun bojong Kabupaten Jawa Barat	Jawa Barat	ID		+62 831 7000000	Hotel										https://www.almira.com	Hotel
1259	Hotel			Jl. Gununggun, Kab. Tas Jawa Barat	Jawa Barat	ID		+62 871 7000000	Hotel										https://www.hotel.com	Hotel
1260	Hotel Anias			P591+H36 Kab. Tas Jawa Barat	Jawa Barat	ID			Hotel										https://www.hotel.com	Hotel

Sumber: (Hasil Penelitian 2026)

Gambar 1. Hasil Pengumpulan Data Hotel dari Google Maps

Tahapan *preprocessing* dilakukan untuk memastikan dataset memiliki kualitas yang sesuai sebelum digunakan dalam proses *clustering*. Tahapan ini meliputi seleksi atribut, penghapusan data duplikat, penghapusan data kosong (*missing value*), dan normalisasi data.

Atribut yang digunakan dalam penelitian terdiri dari *rating* dan jumlah ulasan pengguna. Atribut nama hotel hanya digunakan sebagai identitas data dan tidak digunakan dalam proses perhitungan algoritma *K-Means* karena tidak memiliki nilai numerik yang dapat dihitung menggunakan metode jarak. Penggunaan atribut *rating* dan jumlah ulasan sebagai variabel penelitian sesuai dengan penelitian Aryadi et al. (2023) yang memanfaatkan data ulasan hotel pada Google Maps untuk memperoleh informasi karakteristik hotel.

Proses pembersihan data (*data cleaning*) menghasilkan penghapusan sebanyak 18 data duplikat dari 1.260 data awal sehingga tersisa 1.242 data. Selanjutnya dilakukan penghapusan data yang memiliki nilai kosong pada atribut penelitian. Sebanyak 95 data dengan nilai kosong berhasil dihapus sehingga dataset akhir yang digunakan dalam penelitian berjumlah 1.147 data hotel.

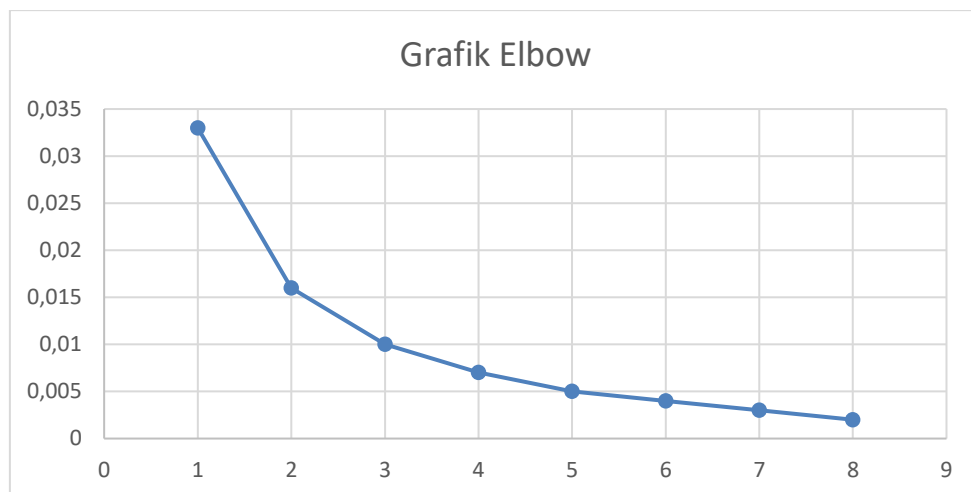
Normalisasi data dilakukan menggunakan metode Min-Max untuk mengubah nilai setiap atribut ke dalam rentang 0 sampai 1. Tahapan normalisasi dilakukan karena terdapat perbedaan rentang nilai yang cukup besar antara atribut *rating* dan jumlah ulasan, sehingga diperlukan penyetaraan skala agar proses pengelompokan tidak dipengaruhi oleh dominasi salah satu atribut.

Penentuan Jumlah *Cluster* Optimal

Penentuan jumlah *cluster* optimal dilakukan menggunakan metode *Elbow*. Metode ini digunakan untuk mengetahui jumlah kelompok terbaik berdasarkan perubahan nilai *Average Within Centroid Distance* (AWCD). Permata et al. (2025) menjelaskan bahwa metode *Elbow* dapat digunakan untuk menentukan jumlah *cluster* optimal dengan melihat titik perubahan penurunan nilai jarak dalam proses pengelompokan.

Tabel 1. Hasil Pengujian Metode *Elbow*

Jumlah Cluster (K)	Nilai AWCD
1	0,033
2	0,016
3	0,010
4	0,007
5	0,005
6	0,004
7	0,003
8	0,002



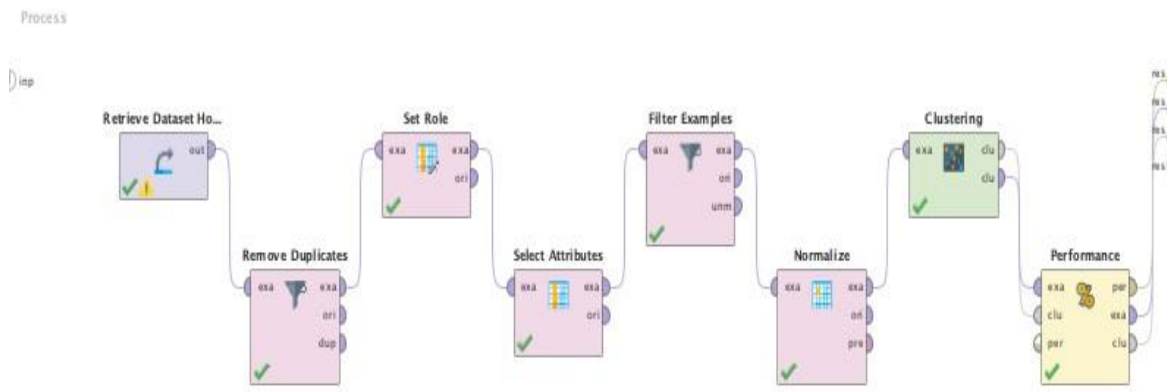
Gambar 2. Grafik Metode *Elbow*

Berdasarkan hasil pengujian, penurunan nilai AWCD terlihat signifikan hingga jumlah *cluster* $K=4$. Setelah nilai tersebut, perubahan nilai AWCD mengalami penurunan yang relatif kecil sehingga penambahan jumlah kelompok tidak memberikan peningkatan hasil pengelompokan secara signifikan. Oleh karena itu, jumlah *cluster* optimal yang digunakan dalam penelitian ini adalah sebanyak empat *cluster*.

Hasil Implementasi *K-Means Clustering*

Implementasi algoritma *K-Means Clustering* dilakukan menggunakan perangkat lunak RapidMiner dengan jumlah *cluster* sebanyak empat kelompok. RapidMiner digunakan untuk membantu proses pengolahan data, penerapan algoritma, dan evaluasi hasil *clustering*. Penggunaan RapidMiner dalam penelitian *clustering* juga diterapkan oleh Sudrajat et al. (2022) karena mampu mendukung proses analisis data secara sistematis.

Gambar 3. Proses Implementasi *K-Means Clustering* pada RapidMiner



Hasil pengelompokan terhadap 1.147 data hotel menghasilkan empat kelompok, yaitu *Cluster_0*, *Cluster_1*, *Cluster_2*, dan *Cluster_3*.

Tabel 2. Distribusi Data pada Setiap Cluster

Cluster	Jumlah Hotel	Persentase
Cluster_0	513	44,73%
Cluster_1	521	45,42%
Cluster_2	70	6,10%
Cluster_3	43	3,75%

Berdasarkan hasil tersebut, *Cluster_1* memiliki jumlah anggota terbesar yaitu 521 hotel, sedangkan *Cluster_3* memiliki jumlah anggota paling sedikit yaitu 43 hotel.

Evaluasi Hasil Clustering

Evaluasi kualitas hasil pengelompokan dilakukan menggunakan metode *Davies-Bouldin Index*. Metode ini digunakan untuk mengukur tingkat kemiripan antar data dalam satu kelompok dan tingkat pemisahan antar kelompok. Menurut Maiyana dan Aulia (2025), nilai *Davies-Bouldin Index* yang lebih rendah menunjukkan kualitas hasil *clustering* yang lebih baik karena kelompok yang terbentuk memiliki tingkat kekompakan dan pemisahan yang optimal.



Gambar 4. Hasil Evaluasi Davies-Bouldin Index

Hasil evaluasi menunjukkan nilai *Davies-Bouldin Index* sebesar 0,636. Nilai tersebut menunjukkan bahwa hasil pengelompokan memiliki kualitas yang baik karena menghasilkan kelompok data yang cukup kompak dengan tingkat pemisahan antar kelompok yang memadai.

PEMBAHASAN

Hasil penelitian menunjukkan bahwa algoritma *K-Means Clustering* mampu mengelompokkan tingkat popularitas hotel di Jawa Barat berdasarkan atribut *rating* dan jumlah ulasan pengguna. Penggunaan kedua atribut tersebut memberikan gambaran mengenai tingkat interaksi pengguna terhadap suatu hotel. Jumlah ulasan menunjukkan tingkat keterlibatan pengguna, sedangkan *rating* menggambarkan persepsi pengguna terhadap kualitas layanan hotel.

Hasil pembentukan empat *cluster* menunjukkan adanya perbedaan karakteristik tingkat popularitas antarhotel. Interpretasi kelompok dilakukan berdasarkan nilai *centroid* pada masing-masing *cluster*. Maori dan Evanita (2023) menjelaskan bahwa nilai *centroid* digunakan sebagai pusat kelompok yang menggambarkan karakteristik setiap *cluster* berdasarkan rata-rata nilai atribut anggota kelompok tersebut.

Tabel 3. Interpretasi Tingkat Popularitas Hotel

<i>Cluster</i>	Rating	Jumlah Ulasan	Tingkat Popularitas
<i>Cluster_0</i>	0,753	0,013	Cukup Populer
<i>Cluster_1</i>	0,919	0,016	Populer
<i>Cluster_2</i>	0,363	0,001	Kurang Populer
<i>Cluster_3</i>	0,893	0,308	Sangat Populer

Berdasarkan hasil interpretasi, *Cluster_3* dikategorikan sebagai kelompok hotel Sangat Populer karena memiliki karakteristik jumlah ulasan yang tinggi dengan nilai *rating* yang baik. Kondisi tersebut menunjukkan bahwa hotel dalam kelompok ini memiliki tingkat perhatian dan keterlibatan pengguna yang lebih besar dibandingkan kelompok lainnya.

Hasil penelitian juga menunjukkan bahwa atribut jumlah ulasan memiliki kontribusi yang lebih dominan dibandingkan atribut *rating* dalam membedakan karakteristik antar-*cluster*. Hal tersebut menunjukkan bahwa tingkat popularitas hotel tidak hanya dipengaruhi oleh nilai penilaian pengguna, tetapi juga oleh jumlah pengguna yang memberikan pengalaman dan tanggapan terhadap hotel tersebut.

Temuan penelitian ini sejalan dengan penelitian Messias et al. (2025) yang menyatakan bahwa ulasan pelanggan memiliki peran penting dalam menggambarkan persepsi konsumen terhadap kualitas layanan hotel. Data ulasan dapat memberikan informasi tambahan yang tidak hanya berasal dari sistem penilaian numerik, tetapi juga menunjukkan tingkat keterlibatan pengguna terhadap suatu layanan.

Penelitian Putra dan Dharma (2023) juga menunjukkan bahwa metode *K-Means Clustering* dapat digunakan untuk menemukan pola karakteristik pada bidang perhotelan. Sementara itu, Hamdani dan Hartama (2024) menunjukkan bahwa analisis data hotel menggunakan algoritma *clustering* mampu membantu memahami karakteristik tingkat penilaian hotel.

Perbedaan penelitian ini dibandingkan penelitian sebelumnya terletak pada fokus pengelompokan tingkat popularitas hotel berdasarkan kombinasi atribut *rating* dan jumlah ulasan pengguna dari Google Maps. Pendekatan tersebut menghasilkan informasi kelompok hotel berdasarkan tingkat popularitas yang dapat digunakan sebagai bahan pertimbangan bagi konsumen maupun pengelola hotel.

SIMPULAN

Berdasarkan hasil penelitian, algoritma *K-Means* mampu mengelompokkan tingkat popularitas hotel di Jawa Barat berdasarkan atribut *rating* dan jumlah ulasan dari Google Maps. Hasil pengelompokan menghasilkan empat *cluster* dengan nilai *Davies-Bouldin Index* sebesar 0,636 yang menunjukkan kualitas pengelompokan yang baik. Keempat *cluster*

tersebut diinterpretasikan menjadi kategori Sangat Populer, Populer, Cukup Populer, dan Kurang Populer. Hasil penelitian menunjukkan bahwa jumlah ulasan memiliki pengaruh lebih dominan dibandingkan rating dalam membedakan karakteristik popularitas hotel.

DAFTAR PUSTAKA

- Aryadi, J. A., Aidil Basith, Y. A., Munawir, M., & Rimadhani Agustini, D. A. (2023). Analisis data review hotel di Google Maps melalui *text mining* (Studi kasus: Kabupaten Bandung). *JIKO (Jurnal Informatika dan Komputer)*, 7(2), 312. <https://doi.org/10.26798/jiko.v7i2.938>
- Hakim, D. P. S. (2025). Segmentasi pasar pariwisata Provinsi Jawa Barat menggunakan *K-Means* berdasarkan jumlah kunjungan, hotel, dan restoran. *Jurnal Informatika dan Teknologi Komputer*, 10–22.
- Hamdani, H., & Hartama, D. (2024). Implementation of data analysis hotel rating levels in Bali using the *K-Means* algorithm and *Decision Tree*. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 10(3), 467–474. <https://doi.org/10.33330/jurteks.v10i3.2894>
- Hidayati, L., Kusuma, L. P., Agustini, D., & Ardhana, V. Y. P. (2024). Implementasi *web scraping* untuk pengumpulan data media sosial lingkup Pemerintah Provinsi NTB. *Jurnal Sistem Informasi dan Informatika (SIMIKA)*, 7(1), 63–72. <https://doi.org/10.47080/simika.v7i1.3200>
- Liyanto, E. W., Homaidi, A., & Lutfi, A. (2024). Implementasi *K-Means clustering* menggunakan RapidMiner dalam pengelompokan data kunjungan wisatawan asing di Provinsi Jawa Timur. *Jurnal Teknik Elektro dan Informatika*, 19, 205–216.
- Maiyana, E., & Aulia, W. (2025). Optimization of stunting infant data clustering with *K-Means++* algorithm using DBI evaluation. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, 11(1), 162–170. <https://doi.org/10.33480/jitk.v11i1.7007>
- Majalaya, K., Azahrah, F. R., Afrinaldi, R., & Karawang, U. S. (2021). Keterlaksanaan pembelajaran bola voli secara daring pada SMA kelas X se-Kecamatan Majalaya. *Jurnal Ilmiah Wahana Pendidikan*, 7(4), 4. <https://doi.org/10.5281/zenodo.5209565>
- Maori, N. A., & Evanita, E. (2023). Metode *Elbow* dalam optimasi jumlah *cluster* pada *K-Means clustering*. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 14(2), 277–288. <https://doi.org/10.24176/simet.v14i2.9630>
- Messias, A., Rosa, G., Couto, P., Rodrigues, V., Silva, L. M., & Marques, J. (2025). A new hotel classification model combining guest reviews with official hotel classification systems: Bridging expert and consumer ratings. *Tourism and Hospitality Research*, 0(0), 1–19. <https://doi.org/10.1177/14673584251379305>
- Obajha, K., Sari, N. N. K., & Pranatawijaya, V. H. (2023). Implementasi metode *collaborative filtering* pada aplikasi rekomendasi hotel dan wisma di Kota Palangka Raya berbasis website. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, 3(2), 398–410.
- Pansayta, S., & Chansanam, W. (2024). Thai COVID-19 patient clustering for monitoring and prevention: Data mining techniques. *IAES International Journal of Artificial Intelligence*, 13(1), 256–265. <https://doi.org/10.11591/ijai.v13.i1.pp256-265>
- Permata, R. P., Alifah, A. N., & Sanjaya, I. M. W. A. (2025). Optimizing *K-Means clustering* through distance metric simulation for strategic enrollment segmentation in private

- universities. *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, 10(2), 616–629. <https://doi.org/10.18860/cauchy.v10i2.33089>
- Putra, I. G. K. K., & Dharma, I. G. W. S. (2023). Application of the *K-Means clustering* method to search for potential tourists of Bendesa Hotel. *TIERS Information Technology Journal*, 4(1), 8–15. <https://doi.org/10.38043/tiers.v4i1.4297>
- Sudrajat, W., Cholid, I., & Petrus, J. (2022). Penerapan algoritma *K-Means clustering* untuk pengelompokan UMKM menggunakan RapidMiner. *Jurnal JUPITER*, 14(1), 27–36.
- Wulandari, S., Astuti, R., & Basysyar, F. M. (2024). Implementasi teknik data mining pada data kunjungan wisatawan menggunakan *K-Means clustering* di hotel bintang Kota Cirebon. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 1142–1149.
- Yani, A., Azmi, Z., & Suherdi, D. (2023). Implementasi data mining menganalisa data penjualan menggunakan algoritma *K-Means clustering*. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 2(2), 315. <https://doi.org/10.53513/jursi.v2i2.6357>
- Yolandari, Y. N. A. B. L. E. R. G. C. H. Z. M. F. A. R. (2025). Analisis perbandingan *K-Means* dan DBSCAN dalam pengelompokan data *travel review ratings* menggunakan evaluasi *Silhouette Index* dan *Davies-Bouldin Index*. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3). <https://doi.org/10.23960/jitet.v13i3.6884>