

## Segmentasi Karakteristik Pelanggan Hotel Menggunakan Algoritma K-Means Berdasarkan Pola Reservasi dan Preferensi Layanan

<sup>1</sup>Adit Nasruloh Rakasiwi, <sup>2</sup>Joshua Dwi Putra, <sup>3</sup>Muhamad Ari Permana

<sup>1,2,3</sup>Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

<sup>1</sup>[adit.nasruloh@gmail.com](mailto:adit.nasruloh@gmail.com), <sup>2</sup>[joshuaputra248@gmail.com](mailto:joshuaputra248@gmail.com), <sup>3</sup>[aripermana991@gmail.com](mailto:aripermana991@gmail.com)

### Abstract

*The modern hotel industry requires personalized marketing strategies because conventional approaches are ineffective at accommodating diverse guest preferences. This study aims to design an objective hotel customer segmentation model using the K-Means Clustering algorithm. Utilizing the hotel\_bookings.csv dataset, this research applies the Knowledge Discovery in Databases (KDD) framework, encompassing data collection, preprocessing, modeling, evaluation, and visualization. The model was built by selecting three main features: booking lead time (lead\_time), number of special requests (total\_of\_special\_requests), and average daily rate (adr). Through evaluation using the Elbow Method and Silhouette Score, the customer data is optimally partitioned into three clusters. Data proximity distance was calculated using Euclidean Distance. The clustering results are subsequently implemented into an interactive web-based dashboard using the Streamlit library. The integration of this system enables hotel management to identify the dominant characteristics of each customer segment, adaptively evaluate pricing, and formulate specific marketing campaigns to improve long-term guest satisfaction.*

**Keywords:** K-Means, Customer Segmentation, Hospitality, Data Mining, Streamlit.

### Abstrak

Industri perhotelan modern membutuhkan strategi pemasaran personal karena pendekatan konvensional kurang efektif dalam mengakomodasi beragamnya preferensi tamu. Penelitian ini bertujuan merancang model segmentasi pelanggan hotel secara objektif menggunakan algoritma K-Means Clustering. Memanfaatkan dataset hotel\_bookings.csv, penelitian ini menerapkan kerangka Knowledge Discovery in Databases (KDD) yang mencakup pengumpulan data, prapemrosesan, pemodelan, evaluasi, hingga visualisasi. Model dibangun dengan menyeleksi tiga fitur utama: durasi waktu tunggu pemesanan (lead\_time), jumlah permintaan khusus (total\_of\_special\_requests), dan rata-rata harga harian (adr). Melalui pengujian menggunakan Elbow Method dan Silhouette Score, data pelanggan dipartisi secara optimal ke dalam tiga klaster. Penghitungan jarak kedekatan data dilakukan menggunakan Euclidean Distance. Hasil klasterisasi tersebut diimplementasikan ke dalam sebuah dashboard interaktif berbasis web menggunakan pustaka Streamlit. Integrasi sistem ini memungkinkan manajemen hotel untuk mengidentifikasi karakteristik dominan dari setiap segmen pelanggan, mengevaluasi harga secara adaptif, dan merumuskan kampanye pemasaran yang spesifik untuk meningkatkan kepuasan tamu jangka panjang.

**Kata Kunci:** K-Means, Segmentasi Pelanggan, Perhotelan, Data Mining, Streamlit.

### A. PENDAHULUAN

Perkembangan industri pariwisata menuntut sektor perhotelan untuk mengoptimalkan data reservasi di tengah ketatnya persaingan digital. Saat ini, volume data historis pelanggan yang besar sering kali hanya menjadi tumpukan data pasif tanpa pemanfaatan strategis. Padahal, pemahaman mendalam mengenai perilaku konsumen adalah kunci utama dalam merumuskan strategi pemasaran yang efisien. Melalui pemetaan karakteristik pelanggan yang akurat, manajemen dapat menyediakan layanan personal dan mengalokasikan sumber daya secara optimal.

Oleh karena itu, analisis data transaksional menjadi instrumen krusial guna meningkatkan okupansi dan profitabilitas bisnis perhotelan.

Kendala utama yang dihadapi manajemen hotel saat ini adalah penerapan strategi pemasaran konvensional yang bersifat "pukul rata" (one-size-fits-all). Segmentasi pelanggan umumnya masih dilakukan secara manual hanya berdasarkan kelas kamar, seperti reguler, menengah, dan VIP. Pendekatan ini gagal menangkap kompleksitas perilaku konsumen, seperti sensitivitas harga, jarak waktu pemesanan (lead time), dan preferensi fasilitas tambahan

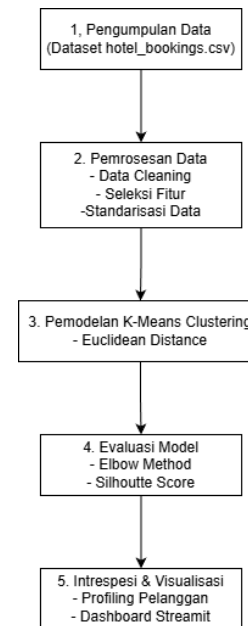
(special requests). Akibatnya, alokasi sumber daya pemasaran menjadi tidak tepat sasaran dan memicu inefisiensi biaya operasional. Jika dibiarkan, ketidakmampuan memetakan perilaku ini dapat menurunkan tingkat kepuasan dan daya saing hotel secara jangka panjang.

Untuk mengatasi kelemahan segmentasi manual, teknik data mining berbasis K-Means Clustering telah banyak diimplementasikan. K-Means terbukti efektif memetakan perilaku pelanggan hotel berdasarkan durasi menginap (Mahesa, Wijaya, & Dwiputranti, 2026), preferensi kamar (Wahyudi, Saragih, & Syahputra, 2023), hingga segmentasi geografis untuk target promosi (Klarita & Kaunang, 2020). Pendekatan ini juga andal dalam memproses dataset publik terkait pola konsumsi (Iqbal, Hidayat, Gevano, & Ilahi, 2025) dan pemetaan loyalitas melalui model RFM (Pramudiansyah & Munte, 2021). Guna menjamin stabilitas pemisahan kelompok, penentuan jumlah kluster optimal dalam algoritma ini lazim dievaluasi menggunakan Elbow Method dan Silhouette Score (Lashiyanti, Munthe, & Nasution, 2023).

Meskipun telah banyak diteliti, mayoritas klusterisasi perhotelan terdahulu masih menggunakan perangkat lunak statis sehingga sulit dieksekusi secara dinamis oleh manajemen. Selain itu, kombinasi variabel Average Daily Rate (ADR), lead time, dan special requests belum dieksplorasi secara simultan. Mengisi celah tersebut, penelitian ini menerapkan K-Means Clustering dengan optimasi Elbow dan Silhouette Score untuk mengelompokkan pelanggan berdasarkan ketiga variabel krusial tersebut. Sebagai kebaruan (novelty), hasil model ini diintegrasikan ke dalam rancangan Dashboard Interaktif berbasis Streamlit. Inovasi ini tidak hanya memvisualisasikan profil kluster secara real-time, tetapi juga menyajikan rekomendasi strategi bisnis otomatis agar hotel dapat mengambil keputusan berbasis data dengan cepat.

## B. METODE

Penelitian ini menerapkan tahapan sistematis penambangan data (data mining) untuk memetakan karakteristik pelanggan hotel. Alur penelitian dirancang secara terstruktur agar proses pengolahan data reservasi dapat berjalan secara teratur, seperti yang diilustrasikan pada Gambar 1. Tahapan utama ini meliputi pengumpulan data, pemrosesan data, pemodelan klusterisasi, evaluasi model, serta interpretasi dan visualisasi. Pendekatan terstruktur ini diadopsi untuk memastikan model segmentasi yang dihasilkan memiliki validitas dan akurasi yang tinggi.



Gambar 1. Alur Penelitian

Tahap pertama adalah pengumpulan data sekunder menggunakan dataset `hotel_bookings.csv` yang diperoleh dari repositori Kaggle. Selanjutnya masuk ke tahap pemrosesan data yang diawali dengan proses data cleaning. Proses ini meliputi penanganan nilai yang hilang (missing values) dan menghapus baris duplikat guna menghindari bias, seperti prosedur pembersihan data yang diimplementasikan oleh (Iqbal, Hidayat, Gevano, & Ilahi, 2025). Setelah bersih, dilakukan seleksi fitur dengan mengekstraksi atribut Average Daily Rate (`adr`), lead time, dan `total_of_special_requests` yang merepresentasikan nilai ekonomi, pola waktu, dan preferensi layanan.

Tahap pemrosesan diakhiri dengan standarisasi data menggunakan metode `StandardScaler` untuk menyamakan skala varians pada seluruh fitur numerik terpilih. Data yang telah distandarisasi kemudian masuk ke tahap pemodelan K-Means Clustering dengan inisialisasi `k-means++`. Pada tahap ini, pengukuran kemiripan karakteristik antar pelanggan dihitung menggunakan metrik Euclidean Distance guna meminimalkan varians dengan titik pusat kelompoknya (centroid), sejalan dengan penerapan algoritma (Mahesa, Wijaya, & Dwiputranti, 2026).

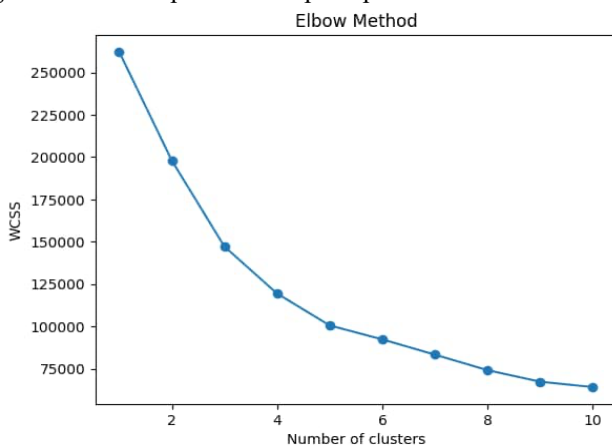
Untuk menjamin kualitas pemodelan, dilakukan tahap evaluasi model menggunakan kombinasi Elbow Method dan pengujian Silhouette Score. Elbow Method digunakan untuk menemukan jumlah kluster ( $K$ ) optimal dengan mengevaluasi nilai Within-Cluster Sum of Squares (WCSS), yang mengidentifikasi titik siku terbaik pada  $K = 3$ . Penggunaan optimasi titik siku ini didukung oleh penelitian Lashiyanti dkk. Selanjutnya, metrik Silhouette Score diterapkan untuk mengukur tingkat kepadatan objek di dalam kluster sekaligus pemisahannya dengan kluster lain secara kuantitatif, sebagaimana direkomendasikan (Fauzi, Wicaksono, & Kurnianingtyas, 2026).

Tahap terakhir dari metodologi ini adalah interpretasi dan visualisasi. Pada tahap ini, dilakukan profiling pelanggan untuk mendeskripsikan secara mendalam karakteristik unik dari masing-masing klaster. Sebagai bentuk solusi aplikatif, hasil segmentasi tersebut divisualisasikan dan diintegrasikan langsung ke dalam Dashboard Streamlit interaktif berbasis web. Sistem dashboard ini dirancang agar manajemen hotel dapat memantau distribusi segmen tamu secara dinamis sekaligus memperoleh rekomendasi strategi bisnis secara otomatis dan akurat.

### C. HASIL DAN PEMBAHASAN

#### 1. Evaluasi Model Klasterisasi

Berdasarkan tahapan pemodelan yang telah dilakukan, evaluasi penentuan jumlah kelompok divisualisasikan menggunakan Elbow Method pada Gambar 2. Grafik tersebut menunjukkan perbandingan antara jumlah klaster pada sumbu horizontal dan nilai Within-Cluster Sum of Squares (WCSS) pada sumbu vertikal. Terlihat bahwa kurva mengalami penurunan yang tajam hingga mencapai titik ketiga, lalu pergerakan garis mulai melandai secara perlahan. Berdasarkan pola tekukan bersudut tersebut, jumlah klaster optimal ditetapkan pada nilai  $K = 3$ .



Gambar 2. Elbow Method

Penetapan jumlah klaster ini dikonfirmasi melalui pengujian kuantitatif menggunakan Silhouette Score, yang diuraikan pada Tabel 1.

Tabel 1. Evaluasi Klasterisasi

Metrik Evaluasi	Nilai
Silhouette Score	0.322

Berdasarkan Tabel 1, model menghasilkan metrik sebesar 0,322. Nilai tersebut mengindikasikan bahwa kualitas pemisahan antar ketiga klaster sudah tergolong baik dan valid untuk ukuran dataset perilaku bisnis.

#### 2. Profiling Karakteristik Pelanggan

Data pelanggan hotel yang telah disegmentasikan ke dalam tiga kelompok utama menunjukkan pola perilaku konsumen yang berbeda secara signifikan. Hasil profiling data ini dirangkum secara kuantitatif pada Tabel 2.

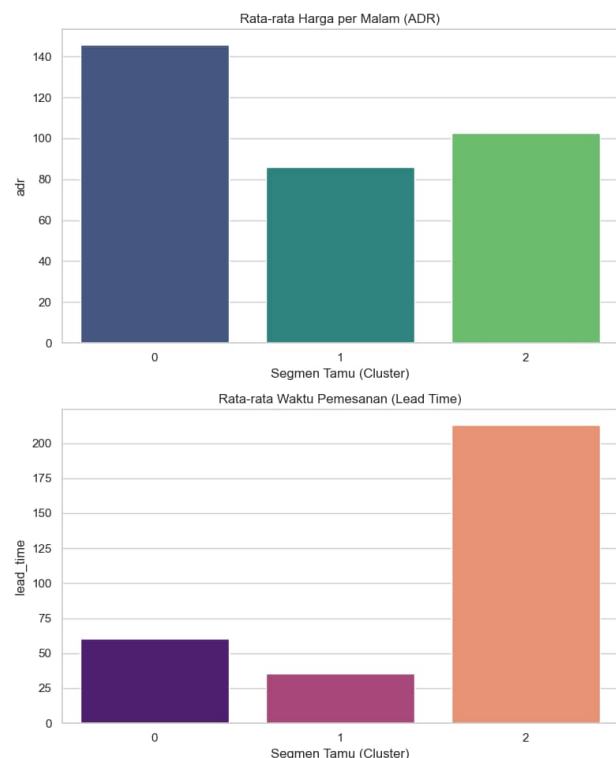
Tabel 2. Karakteristik Rata-rata Profil Pelanggan per Klaster

Klaster	Segmen Pelanggan	Average Daily Rate (ADR)	Lead Time (Hari)	Special Requests
Cluster 0	VIP / Premium	145.85	60.21	1.57
Cluster 1	Budget / Reguler	85.94	35.69	0.28
Cluster 2	Early-Bird / High-Risk	102.56	213.19	0.52

Berdasarkan Tabel 2, Klaster 0 diidentifikasi sebagai segmen VIP atau Premium, dengan rata-rata nilai ekonomi harian tertinggi (ADR 145,85) dan tingkat permintaan khusus terbanyak (1,57). Pelanggan pada kelompok ini cenderung melakukan reservasi rata-rata 60 hari sebelum kedatangan. Sebaliknya, Klaster 1 merepresentasikan segmen tamu Budget dengan tingkat harga terendah (ADR 85,94). Segmen ini didominasi oleh pemesanan waktu singkat (lead time 35 hari) dan minimnya permintaan layanan tambahan.

#### 3. Visualisasi Karakteristik Klaster

Untuk memperjelas perbedaan karakteristik antar segmen, hasil klasterisasi divisualisasikan menggunakan diagram batang (barplot) pada Gambar 3.



Gambar 3. Barplot Karakteristik

Melalui visualisasi pada Gambar 3, ketimpangan antara profil pelanggan terlihat dengan sangat jelas, khususnya pada perbandingan tingkat harga dan jeda waktu pemesanan. Klaster 2, meskipun menghasilkan nilai ADR tingkat menengah (ADR 102,56), secara visual memiliki lead time yang jauh melampaui kedua klaster lainnya. Kelompok ini secara bisnis tergolong sangat rawan terhadap risiko pembatalan (cancellation risk). Temuan

empiris dari visualisasi ini membuktikan bahwa pelanggan memiliki profil kebutuhan yang sangat spesifik.

#### 4. Implementasi Dashboard Rekomendasi

Sebagai bentuk implementasi teknis, hasil klusterisasi ini diintegrasikan ke dalam rancangan Dashboard Interaktif berbasis framework Streamlit. Sistem ini dirancang untuk memberikan rekomendasi strategi bisnis (Business Recommendations & Risk Analysis) secara otomatis berdasarkan profil risiko dan prediksi nilai transaksi dari masing-masing klaster tamu.



Gambar 4. Implementasi Model Klusterisasi pada Dashboard Interaktif Berbasis Streamlit

Berdasarkan dashboard tersebut, sistem memberikan rekomendasi paket bundling yang spesifik. Untuk segmen VIP (Klaster 0), sistem merekomendasikan peningkatan layanan melalui paket Room + Spa & Dinner. Bagi segmen Reguler (Klaster 1), strategi pemasaran diarahkan pada penawaran Room + Breakfast Promo. Sementara itu, segmen Menengah (Klaster 2) terdeteksi sebagai pelanggan berisiko tinggi (churn) dengan waktu tunggu mencapai 213 hari. Untuk memitigasi hal tersebut, sistem merekomendasikan paket Room + Cancellation Insurance. Selain itu, sistem juga secara otomatis mengeluarkan peringatan strategi utama berupa kewajiban pemberlakuan kebijakan Deposit Non-Refundable untuk segmen Menengah guna mengamankan pendapatan hotel dari risiko pembatalan sepihak.

#### D. PENUTUP

##### Kesimpulan

Penelitian ini telah berhasil mengimplementasikan algoritma penambangan data menggunakan metode K-Means Clustering untuk mensegmentasi karakteristik pelanggan hotel. Berdasarkan pengujian Elbow Method dan evaluasi Silhouette Score sebesar 0,322, data pelanggan terbukti dapat dipetakan secara optimal ke dalam tiga kelompok utama. Klaster 0 (VIP) memiliki nilai ekonomi tertinggi sehingga membutuhkan layanan eksklusif, Klaster 1 (Reguler) merupakan segmen spontan yang responsif terhadap promosi harga, sedangkan Klaster 2 (Menengah) teridentifikasi sebagai segmen dengan risiko pembatalan (cancellation risk) tertinggi akibat jeda waktu pemesanan yang sangat panjang.

Selain itu, penelitian ini tidak hanya berhenti pada analisis pemodelan, tetapi juga berhasil mengintegrasikan model tersebut ke dalam sebuah Dashboard Interaktif berbasis framework Streamlit. Sistem ini secara otomatis mampu menerjemahkan profil data menjadi rekomendasi strategi bisnis nyata, seperti penawaran paket bundling spesifik dan penerapan kebijakan Deposit Non-Refundable, guna

membantu pihak manajemen hotel dalam memaksimalkan pendapatan dan memitigasi risiko kerugian finansial.

#### Saran

Untuk pengembangan penelitian selanjutnya, disarankan untuk melakukan uji komparasi menggunakan algoritma klusterisasi lain, seperti DBSCAN atau Hierarchical Clustering, guna membandingkan efektivitas dan kualitas pemisahan model. Penelitian mendatang juga dapat memperkaya dimensi analisis dengan menambahkan variabel data eksternal, seperti analisis sentimen ulasan pelanggan atau faktor musim liburan (seasonality). Dari sisi implementasi sistem, dashboard yang telah dibangun dapat dikembangkan lebih lanjut menjadi sistem pendukung keputusan (SPK) yang terhubung langsung dengan database manajemen reservasi hotel secara real-time.

#### E. DAFTAR PUSTAKA

- Fauzi, M. R., Wicaksono, S. A., & Kurnianingtyas, D. (2026). Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer. *Analisis Segmentasi Pelanggan Pada Industri Laundry Menggunakan Metode K-Means Clustering Untuk Rekomendasi Strategi Pemasaran*, 1-9.
- Klarita, R. I., & Kaunang, F. J. (2020). Jurnal Teknologi Informasi dan Komunikasi (TelKa). *Analisis Target Pemasaran Menggunakan Algoritma K-Means Clustering (Studi Kasus: Hotel Sutan Raja Soreang)*, 107-115.
- Lashiyanti, A. R., Munthe, I. R., & Nasution, F. A. (2023). Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI). *Optimisasi Klusterisasi Nilai Ujian Nasional dengan Pendekatan Algoritma K-Means, Elbow, dan Silhouette*, 14-20.
- Mahesa, P. B., Wijaya, I. N., & Dwiputranti, M. I. (2026). JATI (Jurnal Mahasiswa Teknik Informatika). *PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENGANALISIS POLA RESERVASI PELANGGAN DI VILLA SAYONG HOUSE UBUD*, 2596-2601.
- Setiawan, M. W., & Syahputra, E. R. (2025). JURNAL SISTEM INFORMASI TGD. *Pengembangan Sistem CRM Berbasis Data mining: Segmentasi Nasabah*, 1346-1358.
- Iqbal, Hidayat, N., Gevano, D. P., & Ilahi, A. P. (2025). JUTIF (Jurnal Teknik Informatika). *Segmentasi Pelanggan Menggunakan K-Means Clustering Berdasarkan Data Kepribadian dan Pola Konsumsi*, 3914-3924.

- Pramudiansyah, A., & Munte, H. (2021). Jurnal Ilmiah Ilmu Komputer. *SEGMENTASI PELANGGAN MENGGUNAKAN ALGORITMA K-MEANS BERDASARKAN MODEL RECENCY FREQUENCY MONETARY*, 6-19.
- Wahyudi, A., Saragih, R., & Syahputra, S. (2023 ). Jurnal Nasional Teknologi Komputer. *Pengelompokan Tamu Hotel Dengan Menggunakan Metode K-Means Clustering*, 268-277.
- Wulandari, S., Astuti, R., & Basysyar, F. M. (2024). JATI (Jurnal Mahasiswa Teknik Informatika). *IMPLEMENTASI TEKNIK DATA MINING PADA DATA KUNJUNGANWISATAWAN MENGGUNAKAN K-MEANS CLUSTERING DI HOTEL BINTANG KOTACIREBON*, 1142-1149.