



Analisis Pendapatan dan Pola Pembatalan Reservasi Hotel Menggunakan *Exploratory Data Analysis* (EDA)

Suci Intania Indah¹

Universitas Singaperbangsa Karawang

Apriade Voutama²

Universitas Singaperbangsa Karawang

Alamat: Jl. H.S. Ronggowaluyo, Karawang, Jawa Barat

Korespondensi penulis: hi.suciintania@gmail.com

Abstract. *This study aims to analyze customer behavior between Resort Hotels and City Hotels. The method used was Exploratory Data Analysis (EDA) on 119,390 hotel booking data for the period of July 2015-August 2017. The results show that City Hotels experience more frequent last-minute bookings and cancellations, while Resort Hotels have high revenue fluctuations during the holiday season. These findings can help hotel management in formulating business strategies.*

Keywords: *Data Visualization; EDA; Hotel Management*

Abstrak. Penelitian ini bertujuan menganalisis perilaku pelanggan antara Resort Hotel dan City Hotel. Metode yang digunakan adalah *Exploratory Data Analysis* (EDA) pada 119.390 data pemesanan hotel periode Juli 2015-Agustus 2017. Hasilnya menunjukkan bahwa City Hotel lebih sering mengalami pemesanan dan pembatalan *last-minute*, sedangkan Resort Hotel memiliki fluktuasi pendapatan tinggi selama musim liburan. Temuan ini dapat membantu manajemen hotel dalam menyusun strategi bisnis.

Kata kunci: Visualisasi Data, EDA, Manajemen Hotel.

LATAR BELAKANG

Dalam era digital ini, industri perhotelan menghadapi tantangan yang cukup kompleks, terutama dalam mengelola reservasi. Pembatalan reservasi sangat berpengaruh pada pendapatan dan reputasi sebuah hotel (Azhar *et al.*, 2021). Strategi promosi yang efektif sangat penting bagi perusahaan untuk mempertahankan proporsi pasar mereka (Abdussalam Amrullah *et al.*, 2022). Maka dari itu, penting bagi hotel untuk memahami faktor-faktor yang mempengaruhi keputusan pemesanan dan pembatalan hotel.

Menurut teori perilaku konsumen, pembatalan reservasi hotel dilakukan dengan berbagai alasan, seperti adanya perubahan rencana, jadwal yang tidak sesuai dan lain sebagainya. Dalam industri perhotelan, pembatalan reservasi berpengaruh signifikan terhadap pendapatan dan reputasi kinerja bisnis sebuah tempat penginapan (Qani'ah *et al.*, 2023).

Received Maret 24, 2025; Revised Mei 24, 2025; Accepted Juni 4, 2025

*Suci Intania Indah, e hi.suciintania@gmail.com

Dengan tingkat pembatalan reservasi yang tinggi, banyak hotel yang mengalami kerugian akibat kamar yang tidak terisi. Oleh karena itu, proses *data preparation* dan *exploratory data analysis* (EDA) menjadi langkah awal untuk menggali informasi perilaku pelanggan dan menentukan faktor-faktor yang berkontribusi dalam keputusan pemesanan (Rizki *et al.*, 2024). Visualisasi data sebagai representasi grafis yang memudahkan para *stakeholder* untuk menganalisis informasi kompleks agar lebih efisien (Nisa *et al.*, 2023).

Namun, belum banyak penelitian yang membandingkan secara eksplisit antara tipe hotel, yaitu *Resort Hotel* dan *City Hotel*, dalam konteks pembatalan reservasi dan dampaknya terhadap pendapatan secara visual menggunakan EDA. Padahal, EDA mampu memberikan *insight* awal yang penting sebelum menerapkan metode *machine learning* untuk prediksi dan pengambilan keputusan yang lebih tepat.

Berdasarkan uraian di atas, berikut adalah rumusan masalah pada penelitian ini:

1. Bagaimana pola pembatalan reservasi pada *Resort Hotel* dan *City Hotel*?
2. Bagaimana perbedaan pola pendapatan antara *Resort Hotel* dan *City Hotel* sepanjang tahun?
3. Strategi apa yang dapat diterapkan untuk mengurangi pembatalan dan meningkatkan pendapatan pada kedua tipe hotel?

Penelitian ini bertujuan untuk menganalisis data reservasi hotel guna mengetahui pola perilaku pelanggan. Khususnya dalam membandingkan perbedaan pola pendapatan dan pembatalan antara *Resort Hotel* dan *City Hotel*, menggunakan teknik visualisasi data dalam EDA.

TINJAUAN TEORITIS

EDA *Exploratory Data Analysis* (EDA) adalah proses menganalisis dan menampilkan data bertujuan mendapatkan pemahaman yang lebih baik tentang wawasan dari data (Radhi *et al.*, 2022). Secara definitif, *Exploratory Data Analysis* mengacu pada proses kritis dalam melakukan investigasi awal pada data untuk menemukan pola, untuk menemukan anomali, untuk menguji hipotesis dan untuk memeriksa asumsi dengan bantuan statistik ringkasan dan representasi grafis (Mayasari *et al.*, 2023). Dengan melakukan EDA, kita dapat lebih memahami kondisi dataset yang kita miliki (Samosir *et al.*, 2021).

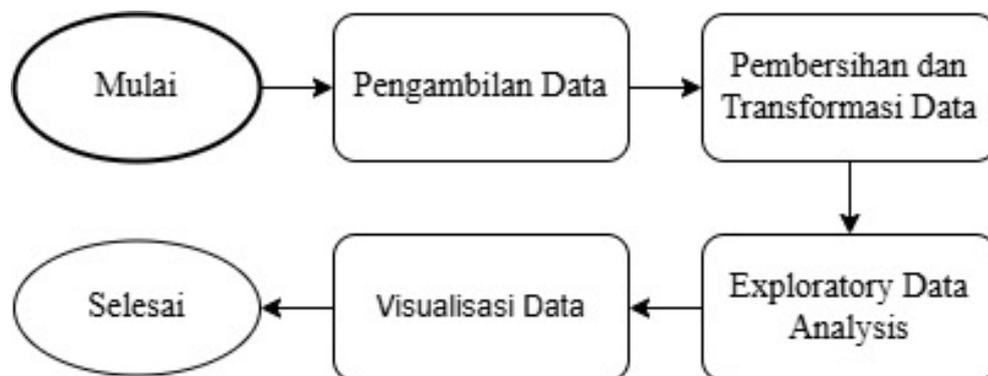
Tujuan utama dari visualisasi adalah untuk menyajikan informasi dengan cara yang jelas dan efektif (Prasetya & Susilowati, 2016). Dengan menyusun data dalam format

yang menarik dan mudah dipahami, pengambil keputusan dapat dengan cepat mendapatkan wawasan dari data yang kompleks dan merancang strategi yang lebih efektif untuk pertumbuhan bisnis (Susi *et al.*, 2023).

Google Colab (atau *Google Colaboratory*) adalah layanan *cloud computing* gratis yang disediakan oleh Google, yang digunakan untuk menulis, menjalankan, dan berbagi kode Python, serta melakukan pemrosesan data dan machine learning (Guntara, 2023). *Google Colab* adalah sebuah IDE untuk pemrograman Python dimana pemrosesan akan dilakukan oleh server Google yang memiliki perangkat keras dengan performa yang tinggi (Gelar Guntara, 2023).

METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis pola pemesanan dan faktor-faktor yang memengaruhi tingkat hunian hotel menggunakan visualisasi data dengan metode *Exploratory Data Analysis* (EDA). Data yang digunakan diperoleh dari modul Studi Independen, mencakup 119.390 data reservasi hotel dari periode 1 Juli 2015 hingga 31 Agustus 2017, yang terdiri dari pemesanan yang terealisasi dan dibatalkan.



Gambar 1. Alur Tahapan Penelitian

Metode EDA digunakan untuk mengeksplorasi data dan mengidentifikasi pola pembatalan reservasi, pemesanan last minute, serta perbedaan pendapatan antara Resort Hotel dan City Hotel. Beberapa variabel utama yang dianalisis antara lain:

1. *Is_canceled* (status pembayaran)
2. Hotel (tipe hotel)
3. *Lead_time* (waktu antara pemesanan dan check-in)
4. *Country* (asal negara tamu)
5. *Adr* (singkatan *Average Daily Rate* atau pendapatan harian rata-rata)
6. Reservation status date (tanggal status reservasi)

Berikut adalah langkah-langkah pada metode penelitian ini.

1. Pengambilan Data

Data diperoleh dari modul Studi Independen dan diekspor dalam format CSV.

2. Pembersihan dan Transformasi Data

Pertama, dilakukan identifikasi nilai yang hilang. Setelah itu, nilai yang hilang diisi dengan median, dan data yang terduplikasi dihapus. Untuk menangani outlier dilakukan dengan metode IQR (interquartile range), yaitu menghaous data yang berada di luar rentang $Q1-1.5iqr$ dan $Q3+1.5IQR$, khususnya pada kolom *lead_time*, *adr*, dan *stays_in_week_nights*. Setelah proses pembersihan selesai, data yang telah dibersihkan disimpan dalam format CSV untuk analisis lebih lanjut.

3. Exploratory Data Analysis

Data kemudian dianalisis menggunakan Python pada Google Collablatry. Tools yang digunakan dalam tahap ini mencakup:

1. Pandas untuk manipulasi data,
2. Matplotlib dan seaborn untuk visualisasi,
3. Numpy untuk perhitungan numerik dasar.

Teknik visualisasi antara lain:

1. Bar chart untuk membandingkan jumlah pembatalan dan pemesanan berdasarkan negara dan tipe hotel.
2. Time series untuk menganalisis tren pendapatan (adr) sepanjang tahun.

4. Visualisasi Data

Visualisasi yang dihasilkan akan dianalisis untuk memperoleh *insight* mengenai perilaku pelanggan hotel. *Insight* ini akan digunakan untuk merencanakan pengelolaan sumber daya di hotel.

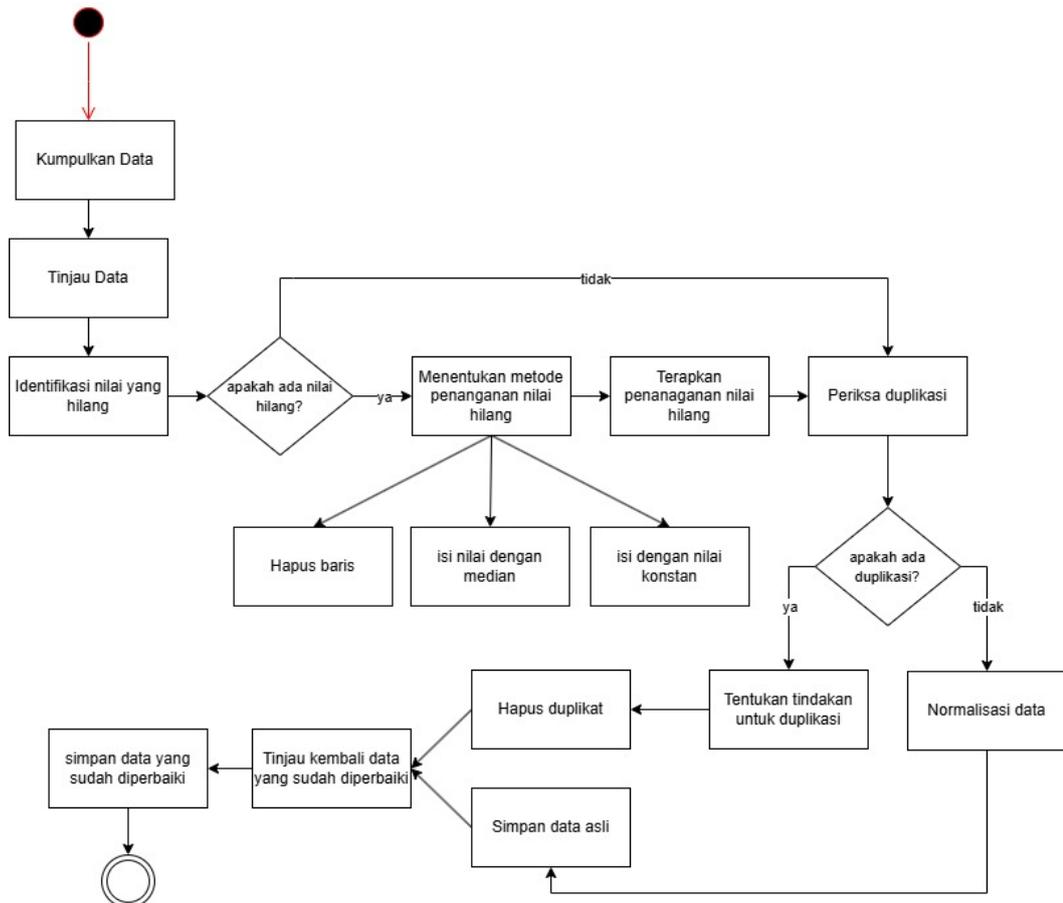
HASIL DAN PEMBAHASAN

Pada penelitian ini, dataset yang digunakan adalah data sebuah hotel yang disimpan dalam format CSV (*Comma-Separted Values*) yang berisi informasi tentang berbagai transaksi, Dataset ini terdiri dari 119390 baris dan 37 kolom yang diambil dari modul Studi Independen.

1. Pembersihan dan Transformasi Data

Pembersihan dataset dilakukan untuk memperbaiki data yang hilang, nilai yang terduplikasi, dan outlier atau nilai tidak wajar. Untuk menangani nilai yang hilang, kolom

numerik diisi dengan rata-rata, sedangkan kolom kategorikal diisi dengan modus. Outlier ditangani menggunakan metode IQR, terutama pada kolom `lead_time`, `adr`, dan `stays_in_week_nights`.



Gambar 2. Flow data preparation

Gambar 3 menampilkan tampilan dataset di *google colab* yang telah dibersihkan dan siap untuk dianalisis lebih lanjut.

```

df_clean.head()

```

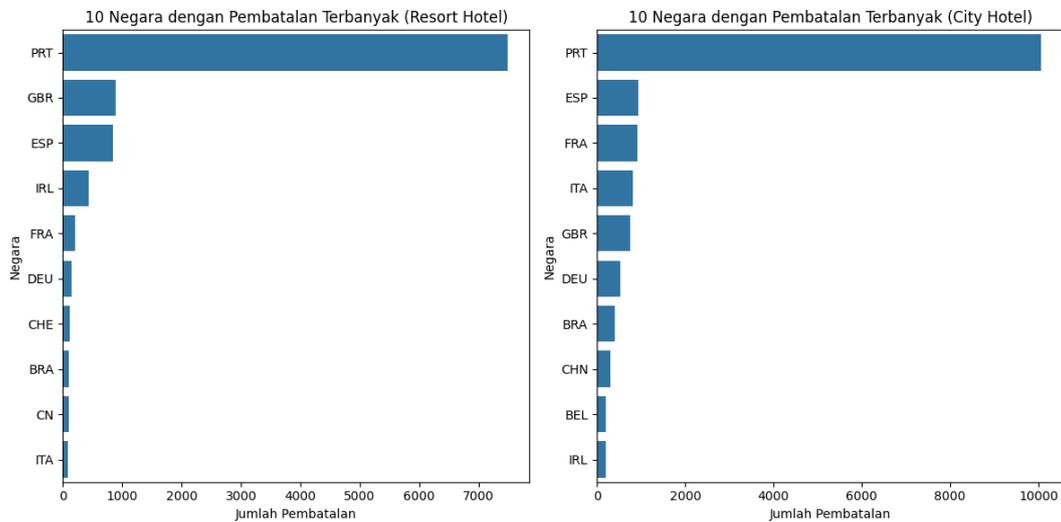
Unnamed: 0	hotel	is_canceled	lead_time	arrival_date_year	arrival_date_month	arrival_date_week_number	arrival_date_day_of_month
0	Resort Hotel	0	342.0	2015	July	27	1
1	Resort Hotel	0	737.0	2015	July	27	1
2	Resort Hotel	0	7.0	2015	July	27	1
3	Resort Hotel	0	13.0	2015	July	27	1
4	Resort Hotel	0	NaN	2015	July	27	1

Gambar 3. Dataset yang sudah siap dianalisis

Exploratory Data Analysis

1. Visualisasi tingkat pembatalan berdasarkan negara

Gambar 4 menunjukkan bar chart perbandingan antara Resort hotel dan City Hotel dari 10 negara dengan Tingkat pembatalan tertinggi.

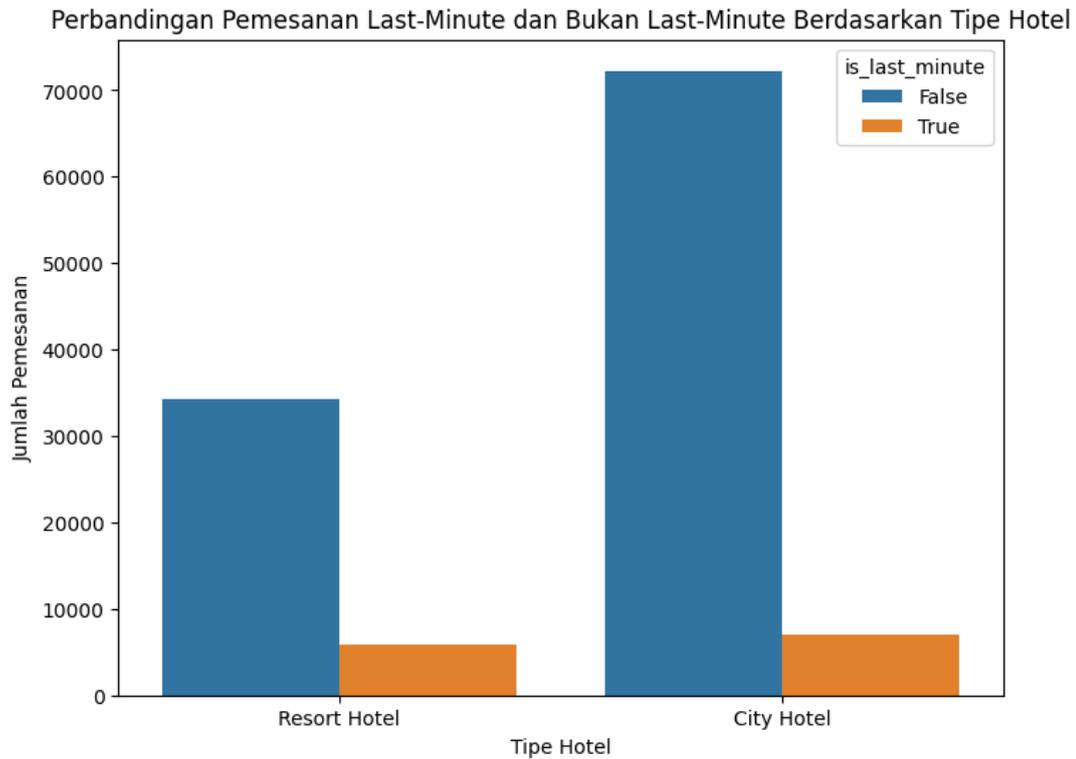


Gambar 4. Perbandingan pembatalan antara Resort dan City Hotel

PRT (Portugal) menempati jumlah pembatalan tertinggi untuk kedua kategori, baik Resort Hotel maupun City Hotel. Tingginya tingkat pembatalan di kedua kategori hotel sejalan dengan temuan (Qani'ah *et al.*, 2023) yang menyebutkan bahwa pembatalan dapat terjadi karena alasan seperti kurangnya komitmen dari pelanggan lokal yang melakukan pemesanan dalam negeri. Insight bagi industri hotel yang bisa diambil, untuk hotel dengan pembatalan tinggi sebaiknya mempertimbangkan sistem atau kebijakan yang lebih ketat untuk mengurangi kerugian.

2. Visualisasi pemesanan *last-minute* berdasarkan tipe hotel

Gambar 5 menampilkan bar chart jumlah pemesanan *last-minute* dan bukan *last-minute* berdasarkan tipe hotel. Terlihat bahwa city hotel memiliki proporsi pemesanan last minute yang lebih tinggi. Menurut (Sulistyaningtyas, 2023), pemesanan last-minute sering dikaitkan dengan kebutuhan yang tidak terencana, yang lebih sering terjadi di kawasan kota.

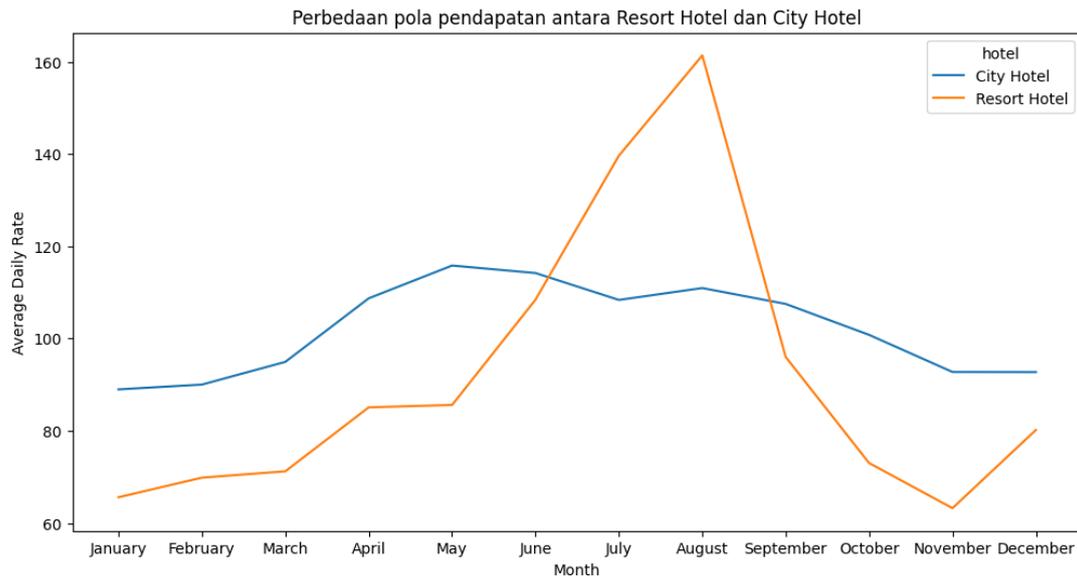


Gambar 5. Perbandingan pemesanan *last-minute* berdasarkan tipe hotel

Grafik ini menunjukkan jumlah pemesanan City Hotel jauh lebih tinggi dibandingkan dengan Resort Hotel, dengan pemesanan *last-minute* (true) yang signifikan, terutama pada City Hotel. Temuan ini mengindikasikan bahwa pelanggan lebih sering melakukan pembatalan di City Hotel. Sementara untuk Resort Hotel lebih dipilih untuk pemesanan jauh sebelumnya. Hal ini menunjukkan bahwa pentingnya manajemen hotel untuk menerapkan strategi seperti penalti pembatalan agar tingkat pembatalan bisa ditekan.

3. Visualisasi Perbedaan Pendapatan Antar Hotel Menggunakan *Time Series*

Gambar 6 menampilkan line chart perbedaan pola pendapatan antara Resort Hotel dan City Hotel sepanjang tahun. Diukur melalui *Average Daily Rate* (ADR).



Gambar 6. Perbedaan Pendapatan Antar Hotel Menggunakan *Time Series*

City Hotel menunjukkan pendapatan yang stabil dengan fluktuasi kecil, mencerminkan permintaan yang konsisten dari wisatawan. Sedangkan, Resort Hotel menunjukkan fluktuasi yang lebih mencolok, dengan lonjakan tinggi di bulan Agustus, kemungkinan berkaitan dengan tanggal merah atau musim liburan, sehingga meningkatkan jumlah wisatawan. Resort hotel dapat memanfaatkan momen liburan dengan strategi promosi dan penyesuaian harga. Sementara itu, City Hotel dapat menjaga pendapat stabil dengan program loyalitas untuk pelanggan.

KESIMPULAN

Penelitian ini bertujuan untuk menganalisis pola pemesanan dan faktor-faktor yang mempengaruhi tingkat hunian hotel, khususnya dalam konteks pembatalan reservasi, dengan menggunakan metode *Exploratory Data Analysis* (EDA). Hasil analisis menunjukkan bahwa PRT (Portugal) merupakan negara dengan jumlah pembatalan tinggi untuk kedua kategori, yaitu Resort Hotel dan City Hotel. Selain itu, City Hotel memiliki banyak pesanan *last-minute*, menunjukkan kecenderungan pembatalan yang tinggi dibandingkan Resort Hotel. Analisis *time series* menunjukkan pendapatan Resort Hotel yang fluktuatif, sedangkan City Hotel relatif stabil sepanjang tahun. Berdasarkan temuan ini, disarankan agar City Hotel menerapkan kebijakan pembatalan untuk mengurangi tingkat pembatalan, sementara Resort Hotel dapat memanfaatkan musim liburan untuk meningkatkan pendapatan.

Penelitian ini berkontribusi dalam memberikan gambaran visual perilaku pelanggan berdasarkan tipe hotel, yang dapat menjadi acuan awal dalam pengambilan keputusan bisnis. Keterbatasan penelitian ini adalah belum adanya analisis prediktif atau uji statistik lebih lanjut. Penelitian selanjutnya disarankan untuk menggunakan mode machine learning guna memprediksi pembatalan dan meningkatkan akurasi strategi marketing.

REFERENSI

- Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, & Apriade Voutama. (2022). Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang). *Jurnal Informatika Dan Rekayasa Elektronik*, 5(2), 244–252. <https://doi.org/10.36595/jire.v5i2.701>
- Azhar, Y., Mahesa, G. A., & Mustaqim, M. C. (2021). Prediction of hotel bookings cancellation using hyperparameter optimization on Random Forest algorithm. *Jurnal Teknologi Dan Sistem Komputer*, 9(1), 15–21. <https://doi.org/10.14710/jtsiskom.2020.13790>
- Gelar Guntara, R. (2023). Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(1), 55–60. <https://doi.org/10.47233/jteksis.v5i1.750>
- Mayasari, R., Nugraha, B., Juwita, A. R., & Heryana, N. (2023). Analisis Produktifitas Padi di Pulau Sumatera menggunakan Exploratory Data Analysis (EDA). *Jurnal Elektronik Sistem Informasi Unsika*, 1(1), 17–24.
- Nisa, N., Firdaus, D., & Aprilia, R. (2023). Implementasi Business Intelligence Untuk Menganalisis Jumlah Guru SD SMP SMA SMK Di Jawa Barat. *Simpatik: Jurnal Sistem Informasi Dan Informatika*, 3(1), 11–16. <https://doi.org/10.31294/simpatik.v3i1.1725>
- Prasetya, H. P., & Susilowati, M. (2016). Visualisasi Informasi Data Perguruan Tinggi Dengan Data Warehouse Dan Dashboard System. *Jurnal Teknik Informatika Dan Sistem Informasi*, 2(3). <https://doi.org/10.28932/jutisi.v2i3.504>
- Qani'ah, F. H., Ramadhan, R., & ... (2023). Prediksi Pembatalan Reservasi Hotel Menggunakan Algoritma Naive Bayes. *... of Informatics and ...*, 4(1), 76–80. <https://journal.univpancasila.ac.id/index.php/jiac/article/view/5499%0Ahttps://journal.univpancasila.ac.id/index.php/jiac/article/download/5499/2514>
- Radhi, M., Amalia, A., Sitompul, D. R. H., Sinurat, S. H., & Indra, E. (2022). Analisis Big Data Dengan Metode Exploratory Data Analysis (Eda) Dan Metode Visualisasi Menggunakan Jupyter Notebook. *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, 4(2), 23–27. <https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v4i2.2475>

- Rangga Gelar Guntara. (2023). Pelatihan Sains Data Bagi Pelaku UMKM di Kota Tasikmalaya Menggunakan Google Colab. *Joong-Ki: Jurnal Pengabdian Masyarakat*, 2(2), 245–251. <https://doi.org/10.56799/joongki.v2i2.1572>
- Rizki, I. N., Prayoga, D., Puspita, M. L., & Huda, M. Q. (2024). Implementasi Exploratory Data Analysis Untuk Analisis Dan Visualisasi Data Penderita Stroke Kalimantan Selatan Menggunakan Platform Tableau. *Jurnal Informatika Dan Teknik Elektro Terapan*, 12(1). <https://doi.org/10.23960/jitet.v12i1.3856>
- Samosir, F. V. P., Mustamu, L. P., Anggara, E. D., Wiyogo, A. I., & Widjaja, A. (2021). Exploratory Data Analysis terhadap Kepadatan Penumpang Kereta Rel Listrik. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(2), 449–467. <https://doi.org/10.28932/jutisi.v7i2.3700>
- Sulistyaningtyas, Y. (2023). Analisis Kontribusi Online Segment Terhadap Tingkat Hunian Di Swiss-Belhotel Balikpapan. *IKONIK: Jurnal Seni Dan Desain*, 5(2), 83–86. <https://doi.org/10.51804/ijsd.v5i2.14325>
- Susi, N., Sugiana, S., & Musty, B. (2023). Analisis Data Sistem Informasi Monitoring Marketing; Tools Pengambilan Keputusan Strategic. *Jutisi: Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 12(2), 696–708.