

PROPOSAL TUGAS BESAR
DASAR KECERDASAN ARTIFISIAL
HOTEL BOOKING PREDICTION



Disusun oleh:

FARRAZ RAYA ABDILLAH	103012300149
AHMAD REFI WIDI KATIBIN	103012300231

**Program Studi S1 Informatika
Fakultas Informatika
Universitas Telkom
Bandung
2025**

1. Latar Belakang

Industri perhotelan menghadapi tantangan dalam mengelola tingkat hunian kamar dan meminimalkan pembatalan pemesanan. Pembatalan yang tidak terduga dapat mengganggu perencanaan operasional dan mempengaruhi pendapatan. Oleh karena itu, penting untuk mengembangkan sistem yang dapat memprediksi kemungkinan pembatalan pemesanan secara akurat.

Salah satu penelitian terdahulu yang relevan dilakukan oleh Martin Falk dan Markku Vieru (2018), yang bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi keputusan pembatalan pemesanan hotel. Penelitian ini menggunakan data pemesanan individu dari sembilan hotel di Finlandia, sehingga hasilnya memiliki keterbatasan dalam hal generalisasi ke konteks global. Metode yang digunakan adalah *probit model* untuk mengestimasi probabilitas pembatalan.

Dari total 233.000 data pemesanan, tingkat pembatalan keseluruhan tercatat sebesar 8%, dengan pembatalan secara offline mencapai 12% dan melalui agen perjalanan sekitar 4%. Hasil analisis menunjukkan bahwa probabilitas pembatalan meningkat secara signifikan pada pemesanan dengan *lead time* yang tinggi, pemesanan offline saat musim liburan, pemesanan tanpa anak-anak, serta pemesanan dari negara-negara tertentu seperti Tiongkok dan Rusia. Di antara seluruh variabel, *lead time* dan negara asal tamu menjadi faktor yang paling berpengaruh, terutama pada pemesanan online.

2. Tujuan

Menerapkan logika fuzzy untuk menangani ketidakpastian dalam data pemesanan hotel, lalu merancang sistem prediktif untuk mendeteksi potensi pembatalan berdasarkan variable-variable seperti durasi menginap, jumlah tamu, dan tipe hotel. Mengevaluasi kinerja model menggunakan metrik evaluasi yang relevan.

2.1. Logika fuzzy

Logika fuzzy merupakan pendekatan komputasional yang dirancang untuk menangani data yang tidak pasti dan tidak presisi dengan menggunakan nilai linguistik. Tidak seperti logika biner konvensional yang hanya mengenal dua nilai (benar dan salah), logika fuzzy memungkinkan representasi nilai dalam spektrum kontinu, seperti “low”, “medium”, dan “high”. Pendekatan ini sangat sesuai untuk permasalahan kompleks dan penuh ketidakpastian seperti prediksi pembatalan pemesanan hotel.

2.2. Dataset

Dataset yang digunakan adalah *Hotel Booking Demand* dari kaggle, yang memuat data pemesanan dari dua jenis hotel: resort dan city hotel, dengan total 119.390 data. Beberapa variable penting:

- Is_canceled: Menunjukkan status pemesana (0 = tidak dibatalkan, 1= dibatalkan)
- Lead_time: Jarak waktu antara pemesanan dan kedatangan
- Adults, children, babies: jumlah tamu dewasa, anak-anak, dan bayi.
- Hotel: jenis hotel (resort atau city)
- Arrival_date_month: Bulan kedatangan
- Meal: Jenis layanan makanan
- Country: Negara asal tamu

- Market_segment: segmen pasar
- Distribution_channel: kanal distribusi
- Previous_cancellations: Jumlah pembatalan sebelumnya
- Adr: tarif rata-rata harian
- Reservation_status: status akhir pemesanan

Pada dataset ini kami hanya akan menggunakan variable berikut :

- Lead_time
- Adr
- Total_of_special_request (di rename menjadi special_request)
- Booking_changes

2.3. Rencana implementasi

2.3.1. Pra-pemrosesan Data

- Menginput dataset
- Membersihkan data-data yang tidak ada nilainya / null
- Men-clip nilai data yang digunakan untuk menghindari outliers (nilai-nilai extreme), tidak melewati batasan membership function, dan membuat prediksi lebih stabil.

2.3.2. Penentuan Himpunan Fuzzy

Menetapkan himpunan fuzzy untuk variable input dan output:

Input:

- Lead_time: {Low, Medium, High}
- Special_request: {Low, Medium High}
- Adr: {Low, Medium, High}
- Booking_changes: {Low, Medium, High}

Output:

- Cancellation_risk: {Low, Medium, High}

2.3.3. Fuzzyfikasi

Menggunakan metode Mamdani untuk mengubah nilai input menjadi derajat keanggotaan fuzzy. Contoh aturan fuzzy:

- Jika lead_time high dan adr high, maka cancellation_risk akan tinggi.
- Jika booking_changes banyak, maka cancellation_risk tinggi.

Aturan-aturan ini dibangun berdasarkan analisis data serta pengetahuan domain.

2.3.4. Evaluasi Model

Evaluasi dilakukan menggunakan sejumlah metrik sebagai berikut:

- Akurasi
- Presisi
- Recall
- F1-Score
- Mean Absolute Percentage Error (MAPE)

Metrik tersebut akan membantu menilai seberapa efektif model dalam memprediksi pembatalan pemesanan.

2.4. Harapan Hasil

- Menghasilkan sistem prediksi pembatalan yang akurat dan efisien

- Mengidentifikasi variabel paling berpengaruh terhadap keputusan pembatalan
- Mendukung pengambilan keputusan strategis di sektor perhotelan
- Menambah kontribusi dalam penelitian mengenai penerapan logika fuzzy di industri perhotelan

3. Contoh hasil prediksi

3.1. Contoh Hasil Prediksi

- **Prediksi Pembatalan Reservasi**

Salah satu aplikasi utama dari dataset ini adalah memprediksi apakah suatu reservasi akan dibatalkan atau tidak. Model klasifikasi seperti Random Forest, Logistic Regression, dan Support Vector Machine (SVM) telah digunakan untuk tujuan ini. Dalam sebuah proyek, model Random Forest berhasil mencapai akurasi hingga 99,5% dalam memprediksi pembatalan reservasi. Fitur-fitur yang paling berpengaruh dalam prediksi ini meliputi:

- **Lead Time:** Jumlah hari antara tanggal pemesanan dan tanggal kedatangan.
- **Jumlah Permintaan Khusus:** Seperti permintaan tempat tidur tambahan atau makanan khusus.
- **Tipe Deposit:** Apakah pelanggan membayar deposit atau tidak.
- **Tipe Pelanggan:** Pelanggan baru atau berulang.

Model ini membantu hotel dalam mengidentifikasi reservasi yang berisiko tinggi untuk dibatalkan, sehingga mereka dapat mengambil tindakan proaktif seperti menghubungi pelanggan untuk konfirmasi atau menyesuaikan strategi overbooking.

3.2. Prediksi Permintaan Kamar Menggunakan Time Series

Analisis deret waktu telah diterapkan untuk memprediksi permintaan kamar hotel di masa mendatang. Model seperti Prophet, SARIMAX, dan Random Forest digunakan untuk memprediksi jumlah malam kamar yang dipesan hingga empat minggu ke depan.

Dalam sebuah studi, model Random Forest menunjukkan kinerja terbaik dengan Mean Absolute Percentage Error (MAPE) yang paling rendah, meningkatkan akurasi prediksi sebesar 53% dibandingkan model baseline. Faktor-faktor yang paling berpengaruh dalam prediksi ini termasuk:

- **Data Pemesanan Sebelumnya:** Tren historis pemesanan kamar.
- **Volume Penumpang Udara:** Jumlah penumpang yang tiba di kota tersebut.
- **Data Pencarian Penerbangan:** Frekuensi pencarian penerbangan ke destinasi tersebut.
- **Hari Libur dan Musiman:** Pengaruh hari libur nasional dan musim terhadap permintaan.

Prediksi ini memungkinkan hotel untuk merencanakan kapasitas dan strategi harga dengan lebih efektif.

3.3. Prediksi Pendapatan Harian Rata-rata (ADR)

Model regresi telah digunakan untuk memprediksi ADR, yaitu pendapatan rata-rata yang diperoleh per kamar per malam. Dalam sebuah proyek, model jaringan saraf regresi menunjukkan kinerja terbaik dengan Mean Absolute Error (MAE) sebesar 29, dibandingkan dengan rata-rata ADR sebesar 105 pada data uji. Fitur-fitur yang paling berpengaruh dalam prediksi ADR meliputi:

- **Durasi menginap:** Jumlah malam yang dipesan.
- **Tipe Kamar:** Kategori kamar yang dipesan.
- **Jumlah Tamu:** Jumlah orang dewasa dan anak-anak dalam reservasi.
- **Lead Time:** Waktu antara pemesanan dan kedatangan.

Prediksi ADR membantu hotel dalam menetapkan strategi harga yang kompetitif dan mengoptimalkan pendapatan.

4. Referensi

<https://www.emerald.com/insight/content/doi/10.1108/ijchm-08-2017-0509/full/html>

<https://www.sciencedirect.com/science/article/abs/pii/S0278431920300980>