Case Study: Predictive Analytics in Hotel Booking Management

Objective

This case study aims to equip you with practical skills in data science, focusing on predicting customer behaviors and booking cancellations in the hotel industry. You will apply EDA, KNN, Decision Tree algorithms, and learn to handle class imbalances using SMOTE.

Data Overview

You are provided with a dataset from 'INN Hotels,' containing various features related to hotel bookings. Your task is to analyze this data to uncover insights and predict booking cancellations.

Tasks

1. **Data Cleaning and Preprocessing**

   - Inspect the dataset for anomalies and missing values.

   - Apply appropriate preprocessing techniques to prepare the data for analysis.

2. **Exploratory Data Analysis (EDA)**

   - Conduct a thorough analysis using statistical summaries and data visualizations.

   - Explore relationships between different variables and understand booking trends.

3. **Predictive Modeling**

   - Implement K-nearest neighbors (KNN) and Decision Tree models to predict booking cancellations.

   - Evaluate and fine-tune your models to achieve the best performance.

4. **Handling Class Imbalance with SMOTE**

   - Understand how class imbalance in your dataset can skew the results.

   - Apply SMOTE to balance the dataset and retrain your models.

   - Compare the performance of your models before and after applying SMOTE.

5. **Deliverables**

   - A comprehensive report containing your EDA, model implementations, results, and insights.

   - The report should not only consist a basic EDA and model implementation, it should also contain an extensive insight into the dataset. Generate at least 15 questions that you would ask from the data and answer them through visualizations.

Data Dictionary

You'll be working with the following features in the dataset:

- **Booking_ID**: unique identifier of each booking

- **no_of_adults**: Number of adults

- **no_of_children**: Number of Children

- **no_of_weekend_nights**: Number of weekend nights (Saturday or Sunday) the guest stayed or booked to stay at the hotel

- **no_of_week_nights**: Number of week nights (Monday to Friday) the guest stayed or booked to stay at the hotel

- **type_of_meal_plan**: Type of meal plan booked by the customer:

- **required_car_parking_space**: Does the customer require a car parking space? (0 - No, 1- Yes)

- **room_type_reserved**: Type of room reserved by the customer. The values are ciphered (encoded) by INN Hotels.

- **lead_time**: Number of days between the date of booking and the arrival date

- **arrival_year**: Year of arrival date

- **arrival_month**: Month of arrival date

- **arrival_date**: Date of the month

- **market_segment_type**: Market segment designation.

- **repeated_guest**: Is the customer a repeated guest? (0 - No, 1- Yes)

- **no_of_previous_cancellations**: Number of previous bookings that were canceled by the customer prior to the current booking

- **no_of_previous_bookings_not_canceled**: Number of previous bookings not canceled by the customer prior to the current booking

- **avg_price_per_room**: Average price per day of the reservation; prices of the rooms are dynamic. (in euros)

- **no_of_special_requests**: Total number of special requests made by the customer (e.g. high floor, view from the room, etc)

- **booking_status**: Flag indicating if the booking was canceled or not.

Learning Outcomes

- Gain hands-on experience in data preprocessing and EDA.

- Develop skills in building and evaluating predictive models.

- Learn to address real-world data challenges like class imbalance.

Evaluation Criteria

- Accuracy and thoroughness of EDA.

- Effectiveness of predictive models.

- Ability to handle class imbalance and articulate its impact on model performance.

- Quality and clarity of the final report.

This case study is designed to be both challenging and educational, providing you with a realistic scenario to apply your data science skills. We encourage creativity and critical thinking throughout your analysis.