

Hotel Room Cancellation Prediction Project

Project Overview

Welcome to the Hotel Room Cancellation Prediction project! In this project, you will be working on a dataset related to hotel bookings and your task is to build a predictive model to predict whether a hotel room booking is likely to be cancelled or not.

Dataset

You can find the dataset here -

<https://github.com/anaspmachinelearning/DataScienceBootCamp/tree/main/Project>

The dataset contains information about hotel room bookings, including features such as room count, booking price, the type of room, number of booked nights, and more. The target variable is `cancelled`, indicating whether the booking was cancelled (1) or not (0).

Here are the initial rows of the dataset:

...

BookingID,roomCount,type,BookingPrice,numberOfBookedNights,numberOfReviews,hasFreeCancellation,cancelled

1,1,HOTEL,104,3,220,0,1

2,1,HOTEL,107,8,47,1,1

3,1,APART_HOTEL,96,2,50,1,1

4,1,HOTEL,114,6,62,1,1

5,2,APART_HOTEL,105,1,71,1,1

6,1,APART_HOTEL,118,4,22,1,1

7,1,HOTEL,105,5,103,0,0

8,1,HOTEL,91,1,22,0,0

...

Project Tasks

1. Exploratory Data Analysis (EDA) and Feature Engineering

Perform thorough Exploratory Data Analysis to understand the characteristics of the dataset. This includes:

- Summary statistics, data types, and missing values.
- Visualization of distributions and relationships between variables.
- Feature engineering: Create new relevant features if needed.

2. Model Evaluation

Evaluate the performance of 5 or 6 classification models to predict hotel room cancellations. Some suggested models include:

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine (SVM)
- Gradient Boosting

For each model, perform the following:

- Data splitting: Split the dataset into training and testing sets.
- Model training: Train the model on the training set.
- Model evaluation: Evaluate the model's performance on the testing set using appropriate metrics.

3. Performance Metric Selection

Provide reasoning for the selection of the performance metric. Consider metrics such as accuracy, precision, recall, F1-score, and ROC-AUC, and explain why you chose a particular metric.

4. Flask-based UI for Prediction

Create a Flask-based user interface (UI) to allow users to input data and obtain predictions from the best-performing model. The UI should include:

- A form for users to input relevant information about a hotel booking.
- A button to trigger the prediction.
- Display of the prediction result.

More about flask here : <https://flask.palletsprojects.com/en/3.0.x/quickstart/#a-minimal-application>

5. Project Documentation

Prepare documentation summarizing the entire project. Include details on:

- Project objectives and background.
- Data preprocessing steps.
- EDA findings and insights.
- Model selection and evaluation.
- Performance metric explanation.
- Flask UI development and deployment.

6. Final Presentation

Prepare a final presentation for the project, covering:

- Introduction and background.
- EDA findings and insights
- Feature Engineering
- Model selection and evaluation.
- Flask UI development.
- Future improvements and recommendations.

Deliverables

1. Jupyter Notebook containing all code, EDA, feature engineering, and model evaluation.
2. Flask application files for the UI.
3. Project documentation in PDF or Markdown format.

Submission Guidelines

Submit a zip file containing all the deliverables via mlwithap.com. Ensure that the code is well-commented, and the documentation is clear and concise.

Project Timeline

- Week 1: EDA, feature engineering, and initial model evaluation.
- Week 2: Finalize models, choose the best-performing model, and start Flask UI development.
- Week 3: Complete Flask UI development, conduct final testing, and prepare documentation.
- Week 4: Submission of the project.

Feel free to reach out for assistance or clarification during the project. Good luck!