Hotel Cancellation Prediction

Karim AbouDaoud
900212779
Department of Mathematics and Actuarial Science
The American University in Cairo
Cairo, Egypt
karimaboudaoud@aucegypt.edu

Youssef Nakhla
900201430
Department of Mathematics and Actuarial Science
The American University in Cairo
Cairo, Egypt
youssef_n9212@aucegypt.edu

Introduction

In this report, we detail the process of implementing a Hotel Booking Cancellation Prediction System. The goal of this project is to develop a web application that allows users to input data related to hotel booking reservations, and predict whether a booking is likely to be canceled or not, along with the probability of cancellation. We used the Python library Streamlit to create the web interface, and a machine learning model trained on historical hotel booking data to make predictions.

Model Deployment

After training and evaluating our machine learning model, we saved it using the pickle library, enabling us to easily load the model for inference in our web application. The model encapsulates the knowledge learned from the historical hotel booking data and is capable of making predictions on new instances.

User Interface Desgin

Utilizing the Streamlit library, we designed a user-friendly web interface that allows users to interact with our predictive model. The interface presents users with input fields corresponding to various features of a hotel booking reservation. These features include lead time, arrival date, booking channel, and others.

Before making predictions, the input data is preprocessed to ensure it is in a suitable format for the model. As mentioned earlier, this preprocessing includes handling missing values and encoding categorical variables. Features that were previously one-hot encoded or label encoded during the preprocessing stage are processed accordingly to maintain consistency and compatibility with the model.

Making Predictions

Once the user submits the input data through the web interface, the application invokes the trained model to make predictions. The preprocessed input data is passed to the model, which then outputs the predicted probability of cancellation.

Displaying Results

The predicted outcome and probability of cancellation are displayed to the user on the web interface in a clear and understandable manner. Users can also see the characteristics of the model like the type of model used as well as the hyperparamters used of said model. Users can quickly interpret the prediction results and make informed decisions based on the model's evaluation tab that shows different performance metrics of the model.

Future Improvements

The designed utility application does not include a cloud feature and was rather done as a desktop application, hence a future improvement would be to convert it to a cloud based application as well as Gather feedback from users to identify pain points and areas for improvement in the user interface. Furthermore, it is vital to continuously monitor and evaluate the model's performance on real-world data to identify opportunities for improvement.