Your MSc Project title (Module Code and Title):
7COM1039-0109-2022 - Advanced Computer Science Masters Project
Student ID Number: 20067093
Student Name: Nikhil Reddy Marella

Project Title:

Prediction of hotel reservation cancellation of a customer using reservation details.

Aim of the project: This project aims to analyze and observe the patterns and relations between the columns to identify the reasons behind the cancellations of hotel reservations and develop a machine-learning algorithm that can predict the customer who might cancel the reservation based on the customer details provided.

Research question/ hypothesis: Is the variables like lead_time, market_segment_type, repeated_guest, no_of_previous_cancellations, no_of_previous_bookings_not_canceled and avg_price_per_room responsible for reservation cancellation?

Objectives:

1) Figure out the variables responsible for reservation cancellations.
2) Perform EDA using statistics to find the data's relations and other findings.
3) Deal with the outliers and imbalanced data to generalize well on the data.
4) Training the model using data split methods.
5) Find out the best parameters for the model using cross-validation techniques to predict with great accuracy on the train and test set.
6) Make sure to collect every reference I have visited.
7) Detailed Project report with results achieved.

Short description of your idea:

The idea of this project is to develop a machine-learning model which can predict the customer who might cancel the hotel room reservation they made in advance. Dealing with the imbalanced data and identifying the relationship between the independent and dependent variables is crucial to developing the well generalizing model.

Specify how you plan to conduct your research:

I would like to investigate this research question using the following steps:

Dataset Link:

https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset

1) Data Variables Understanding, Data Cleaning, and Solving Data Imbalanced problems.
2) Exploratory Data Analysis and Visualization to understand the pattern of the columns.
3) Understanding the relations between the independent and dependent variables using statistics.
4) Find out a suitable data split ratio based on data and find algorithms suitable for the research.
5) Experiment with all the algorithms using cross-validation for best parameters and track the performance with metrics.
6) Find out the best-trained model and test it on unseen data and re-iterate the training process till the required performance is achieved.
7) Project report in detail with all references.