**Abstract**

The hotels have been experiencing drastic changes in their online reservation channels' booking possibilities due to customers' behavior in cancelling the hotel reservations in vast numbers as it is made easier to do so free of charge or preferably at a low cost which benefits the hotel guests, but it is a revenue reducing factor for the hotel to deal with. So, this project aims to identify the main reasons behind hotel reservation cancellations and also develop a machine learning model which can predict the customer who might cancel the reservation based on the details provided. This project proposes to explore to answer the 4 research questions in order to understand the importance of dropping unwanted columns, one-hot encoding and its significance to the dataset, and algorithms performance on over-sampled and under-sampled data to deal with imbalanced data and to know what factors are crucial for customer canceling the hotel reservation. The project was conducted in four steps: data cleaning and preprocessing, exploratory data analysis, feature engineering, and model training and evaluation. The project found that over-sampled data using SMOTE technique performed better than under-sampled data. Random Forest with over-sampled data resulted in the highest performance with training accuracy of 99.46% and validation and test accuracy of 90.46% and 89.80%, respectively. This Project finds that the most important factors that impact hotel reservation cancellations are lead time and average price per room. These insights can be utilized by hotels to design methods to decrease cancellations.