PREDICTIVE MODEL TO IDENTIFY AIRLINE NO-SHOW RATE BASED ON SRILANKAN CONTEXT

MOHAMED SHABIR MOHAMED SAHAB NISTHAR

A dissertation submitted in partial fulfilment of the requirement for

Master of Science degree in Business Analytics

Department of Computing

Information Institute of Technology, Sri Lanka

In collaboration with

Robert Gordan University, UK

2022

Abstract

In this research we have examined the effect on No-Show of airline passengers with multiple variables, where we have given the introduction on to the issues related to No-Show and the airlines struggles in the 21st century while maintaining the competitive advantage staying competitive. We therefore not only primarily indicated the benefits and the efficient matter how an airline can operate if the No-Show predictions can be classified using ML.

This Research was mainly done to identify the No-Show by initially collecting the data via an online survey and then utilizing the primary data, which is then used to build machine learning via Colab. Where the personal features were examined to identify how these features explain or effect the NO-SHOW related incidents in the airline industry. Where these features were transformed and utilized to build the ML models of Random Forest, Decision Tree, SVM, Logistic Regression and Naïve Bayes.

It was identified, during this study that the logistic regression outperformed the rest of the algorithms used in this study with an accuracy of 78.13%. by indicating if we were to develop a model it should be done using logistic regression as per the indicated accuracy report.

The main struggle during this research would have been collecting higher number of samples, for this study it was just above the minimum requirements. This resulted in different methods of evaluation because we were unable to use the test train method event though the test train is used for reference in this study the main validation method utilized and appropriate for this study was cross validation. For the hypothesis testing the chi-square test was used to identify if there are any relationship between the variables. The test indicated there are some significance relationships between few features like age, gender, education and airline No-Show while other features at this study and sample indicate there are no significant relationship between them.

Keywords: ***Airline overbooking, Algorithms, Classification, No-Show forecasting, Predictive modeling, Revenu Management***