Writeup— Classification Module

Abidemi

Abstract

The purpose of this project is to build a classification model which will be used to identify features that result in a satisfied passenger. By focusing on these features, airlines can increase satisfaction, loyalty, and retention among passengers.

Design

A Kaggle dataset that has the results of an airline passenger satisfaction survey was used. The original dataset was created in June 2018. Interpreting the model is valuable to airlines because they have lost many business class passengers during the pandemic (due to corporate safety restrictions).

Data

The balanced dataset has over 100, 0000 unique records and twenty-three features including the target feature. Most features are rated on a scale from 1 to 5. The target feature was converted to binary values. Sample features: *Age, Class, Inflight wifi service, Gate location, Food and drink, Online boarding.*

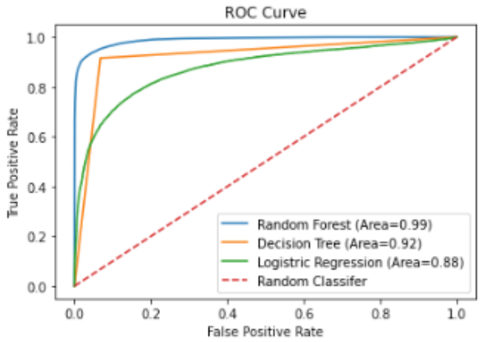
Algorithms

1. Data was split into training and validation sets. A separate test dataset was included.

2. Decision Tree, Logistic Regression, and Random Forest model were built using Sklearn. ROC AUC score was used to select best performing model.

3. Random Forest hyperparameters were tuned. RMSE and ROC AUC score were used to select best model. Computational cost was also used to determine final model.

4. SHAP values were used to interpret model feature importance. Subsets of data were used to gather further insight into the model.

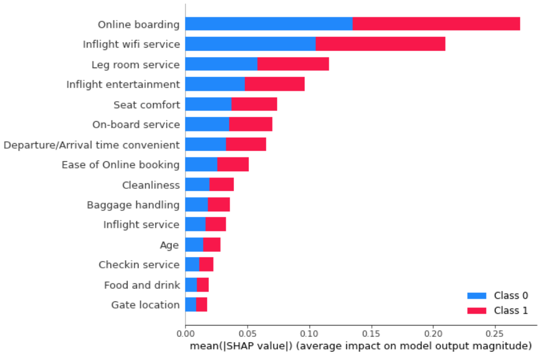


Tuned Random Forest Model on Validation Set:

Precision: 0.92 Recall: 0.91 F1-score: 0.91

Tuned Random Forest Model on Test Set:

Precision: 0.92 Recall: 0.90 F1-score: 0.91



Tools

Python and Python libraries (Numpy, Pandas, Sklearn, Seaborn, SHAP).

Communication: Slides and presentation.