Abstract

The main goal of this project is to create a classification model on airline passenger satisfaction to predict if passenger satisfied or not?

Design

The data provided by kaggle has the main information about airline passenger satisfaction then we cleaned the data and preform EDA to visually present the data, preprocessing then modeling and evaluating the best model with F1 score as our metric.

Data

In this project we used the data from Kaggle https://www.kaggle.com/teejmahal20/airline-passenger-satisfaction it contains
 103904 rows with 25columns.

Algorithms

we've done the data cleaning and removing the nulls , also we transformed the categorical features into label encoder then we split our data and created our baseline model witch was KNN then we wanted to improve the model, we used decision trees and random forest and xgboosted classifier and ensemble stacking and we found that random forest has the best F1 score since it's our chosen metric because we want to find the tradeoff between recall and precession, then we deployed our model using streamlist

Tools

- Python and Jupyter Notebook
- Numpy and Pandas for data manipulation
- Matplotlib, Seaborn and Tableau for visuializations
- Sklearn for ML algorithms

Communication

In addition to the slides and visuals presented, we will share our work on our github accounts

https://github.com/emaanhajji

https://github.com/bushra3e/classification-project-Airline-Passenger-Satisfaction-info-Bushra-Eman