Abstract

This project worked on a dataset that contains an airline passenger satisfaction survey. The project aimed to build classification models to predict passenger satisfaction and identify which model is outperformed based on accuracy. We have trained our data on six models which are: Decision Tree model, Logistic regression model, Random forest model, XGradient boosting model, LGBM model and Stacking model. The best accuracy among these algorithms is 0.93 of XGB and Stacking models.

Methodology

- Loading the dataset.
- EDA (cleaning and visualizing the data).
- Building different classification Models.
- choosing the model based on given best accuracy to prediction.

Data

We used a dataset of Airline Passenger Satisfaction survey from kaggle contain customers satisfaction (target variable) and various survey ratings, which contains 103904 rows and 25 features.

Classification Algorithms:

On this dataset we used different models of classification:

- Decision Tree model.
- Logistic Regression model.
- Random Forest model.
- LightGBM model.
- XGradient boosting model.
- Stacking model.

Tools

- Numpy and Pandas for data manipulation.
- Matplotlib and Seaborn for plotting.
- Sklearn for machine learning and statistical modeling.

Communication

The slides are provided here.