

Airline Passenger Satisfaction

- **Abstract**

The goal of this project was to use classification models to predict the Satisfaction of airline passenger in order to help such airlines improve their services and to make it better so their passenger will be more satisfied and that will benefit them to be the one of the top 10 in the best airlines. We worked with the data the provided by <https://www.kaggle.com>. The data was token from a survey from the US airline. We will work with some models that are suitable for our classification model.

- **Design**

The data is provided by <https://www.kaggle.com>, and presents a two status of **(satisfaction, neutral or dissatisfaction)**.

Classifying statuses accurately via machine learning models would enable such airlines to take action to improve their services and planning to develop their flight

- **Data**

The dataset contains 129,880 passenger decisions with 24 features for each, 5 of which are categorical. The most of the numerical features include a number from 0-5 that represent how was the passenger satisfied. Also, we have another numerical features that represent the delay time in minute. Which we believe that an important thing for any passenger.

- **Algorithms**

1. Exploring the data and do some cleaning steps which we call it EDA
2. Converting categorical features to binary dummy variables
3. Split the data to training and testing data set

- *Models:*

k-nearest neighbors, Decision Tree, random forest and XGBoost

- *Model Evaluation and Selection:*

The entire training dataset of 129,880 records was split into 80/20 train vs. holdout, and all accuracy scores reported below only.

The official metric for data was classification rate (accuracy).

Final model's accuracy:

- **Tools**

- Numpy and Pandas for data manipulation.
- Scikit-learn for modelling.
- Matplotlib and Seaborn for plotting.

- **Communication**

In addition to the slides the whole code will be embedded on my personal website in the github.

