

Airlines Passenger Satisfaction

Abstract:

What factors lead to customer satisfaction for an Airline? This dataset contains an airline passenger satisfaction survey. What factors are highly correlated to a satisfied (or dissatisfied) passenger? Can you predict passenger satisfaction?

Design:

This project is one of the T5 Data Science Boot Camp requirements. Data provided by Kaggle. In this module we will be laying the foundation for our analysis by processing and exploring a large amount of data and classify it by using classification modules. The dataset contains an airline passenger satisfaction survey. *Get the data [here](#).*

Understanding the dataset:

Gender: Gender of the passengers (Female, Male)

Customer Type: The customer type (Loyal customer, disloyal customer)

Age: The actual age of the passengers

Type of Travel: Purpose of the flight of the passengers (Personal Travel, Business Travel)

Class: Travel class in the plane of the passengers (Business, Eco, Eco Plus)

Flight distance: The flight distance of this journey

Inflight Wi-Fi service: Satisfaction level of the inflight Wi-Fi service (0:Not Applicable;1-5)

Departure/Arrival time convenient: Satisfaction level of Departure/Arrival time convenient

Ease of Online booking: Satisfaction level of online booking

Gate location: Satisfaction level of Gate location

Food and drink: Satisfaction level of Food and drink

Online boarding: Satisfaction level of online boarding

Seat comfort: Satisfaction level of Seat comfort

Inflight entertainment: Satisfaction level of inflight entertainment

On-board service: Satisfaction level of On-board service

Leg room service: Satisfaction level of Leg room service

Baggage handling: Satisfaction level of baggage handling

Check-in service: Satisfaction level of Check-in service

Inflight service: Satisfaction level of inflight service

Cleanliness: Satisfaction level of Cleanliness

Departure Delay in Minutes: Minutes delayed when departure

Arrival Delay in Minutes: Minutes delayed when Arrival

Satisfaction: Airline satisfaction level (Satisfaction, neutral or dissatisfaction)

Algorithms

Use panda's library to analyze different features of the dataset, which includes:

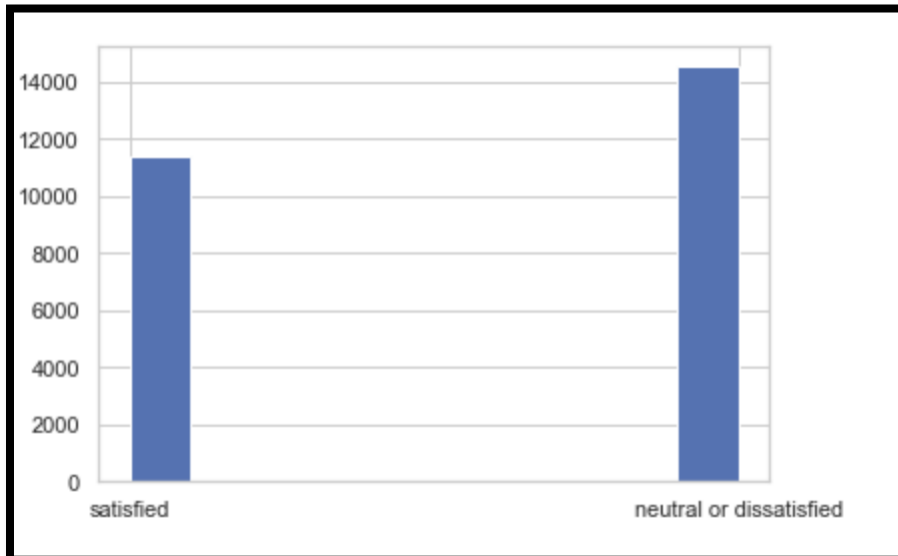
- Read the dataset.

```
df = pd.read_csv('Airline.csv')
```

- Cleaning the data and remove null values.

```
df.drop(['Gate location', 'Unnamed: 0', 'id'], axis= 'columns', inplace=True)
```

- Remove the outlier from dataset.
- Use seaborn and matplotlib library to visualize the given results.
- Plot graphs like bar graphs and pie chart to check the balanced data



- Use logistic regression & apply one of classification algorithm

```
tuned_lr Train score: 0.8150780163757145  
tuned_lr Test score: 0.8194315724436206
```

- Baseline Model.

```
0.7444240610215314  
0.755512647229194  
F(2): 0.733498457393647  
Precision : 0.6945073030240692  
Recall : 0.74394006170119
```

- Data Preparation.

To get started, we need to import some useful libraries that will help us import the dataset into our python environment, manipulate and analyze the same and later help us to visualize it.

Missing values & outliers was observed across the dataset, so a good approach it to either remove it or correct it .

Split the data and do different features engineering on it .

Check if our dataset is balance or not .

- Experimentations:

Decision Tree Classifier

```
Best score: 0.9410754314254788
```

KNN Classifier

```
knn_best Train score: 0.7815026520418147  
knn_best Test score: 0.6983317886932344
```

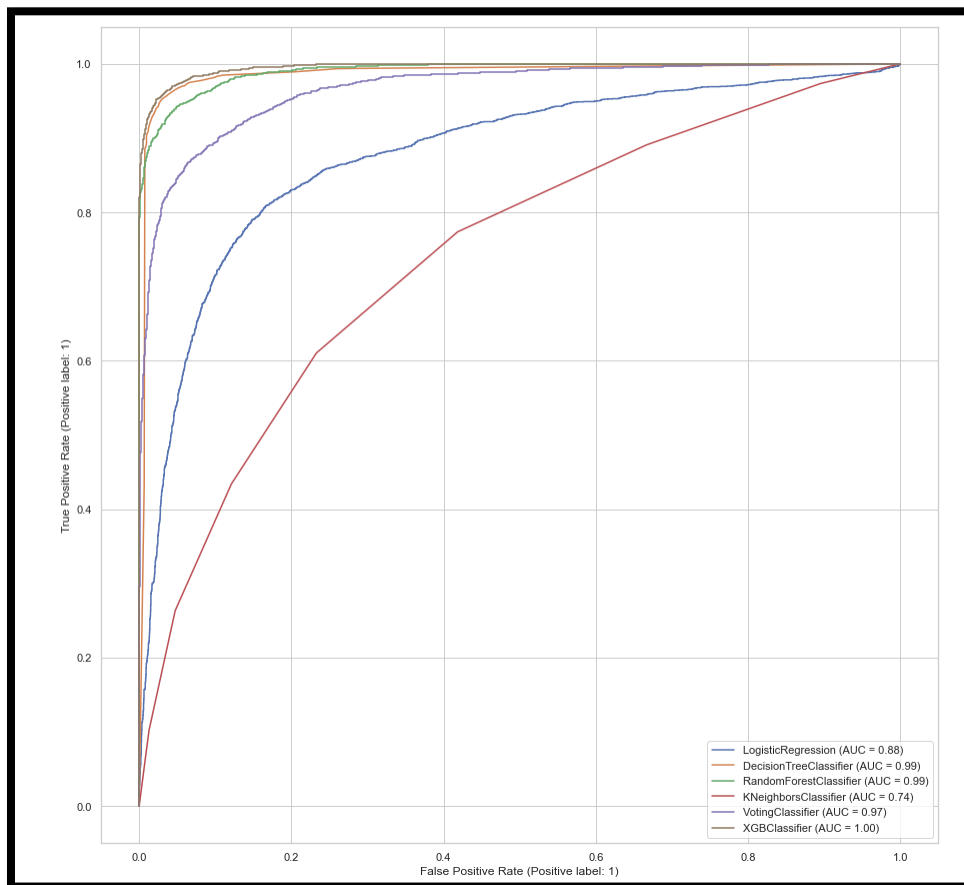
Random Forest Classifier

```
rf_best Train score: 0.9735310778103918  
rf_best Test score: 0.9474822366388631
```

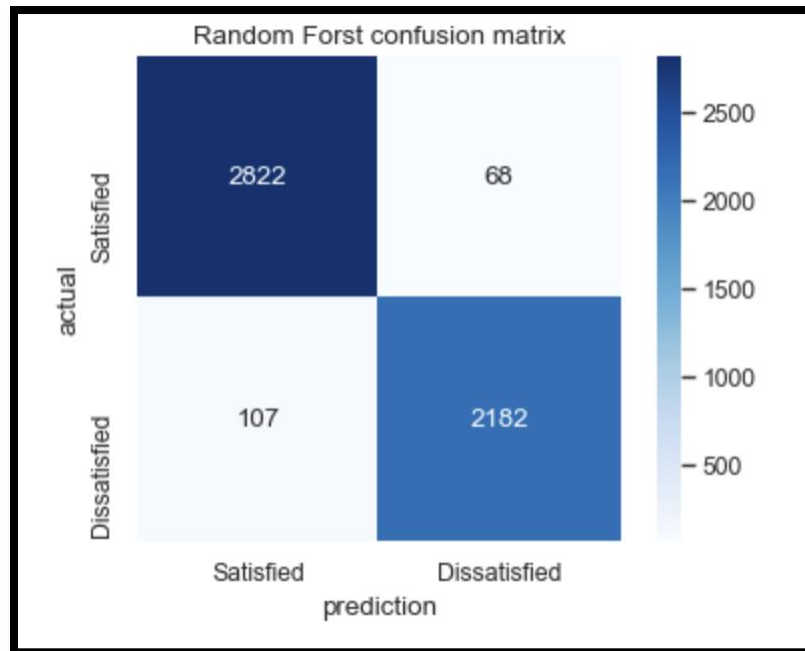
XGboost

```
0.9544313574049044
```

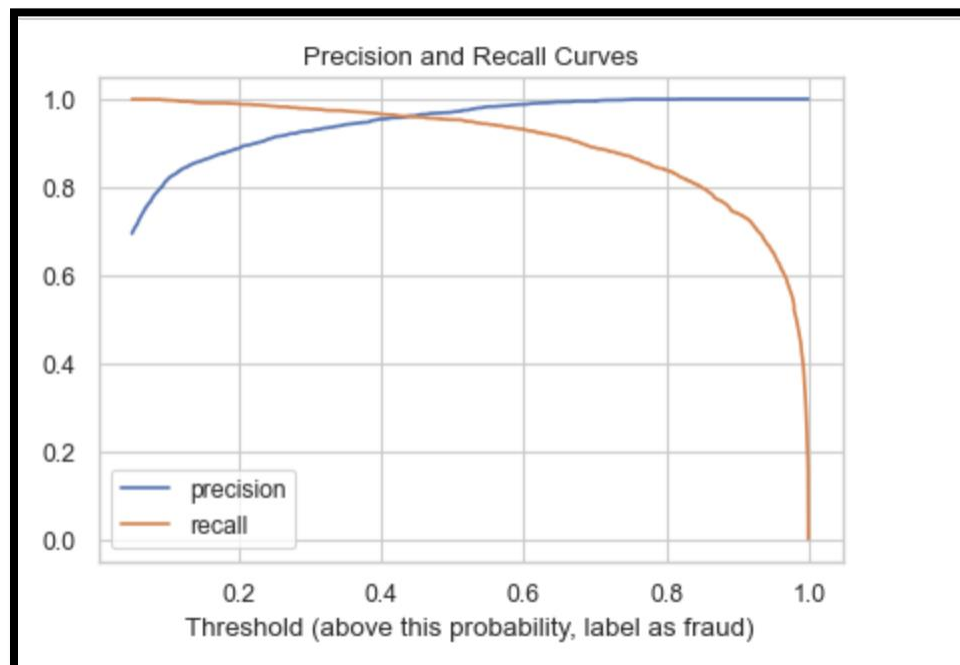
Plot the ROC curve:



The Confusion Matrix of Hight model is:



The precision and Recall curve



Tools

- Pandas for data manipulation
- IQR for discover outliers
- Remove Duplicate or unnecessary data
- Matplotlib for plotting
- Seaborn for visualizations
- Sklearn logistic Regression library

Communication:

- The slides will be provided here, feel free to any pull requests besides details are provided at the readme of the project.