



Airlines Passenger Satisfaction

Classification Project

By: Njood Alqahtani & Reem Alsugaih



Table of Content

✈ Introduction

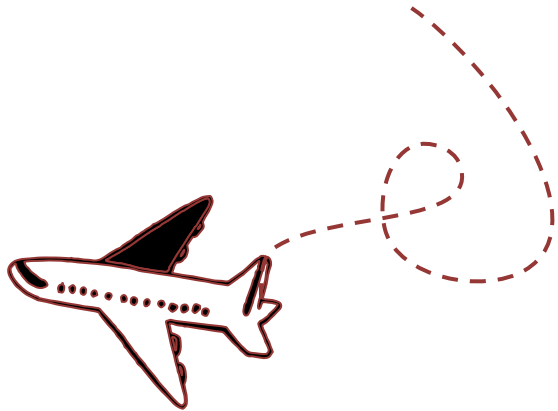
✈ Experiments

✈ Methodology

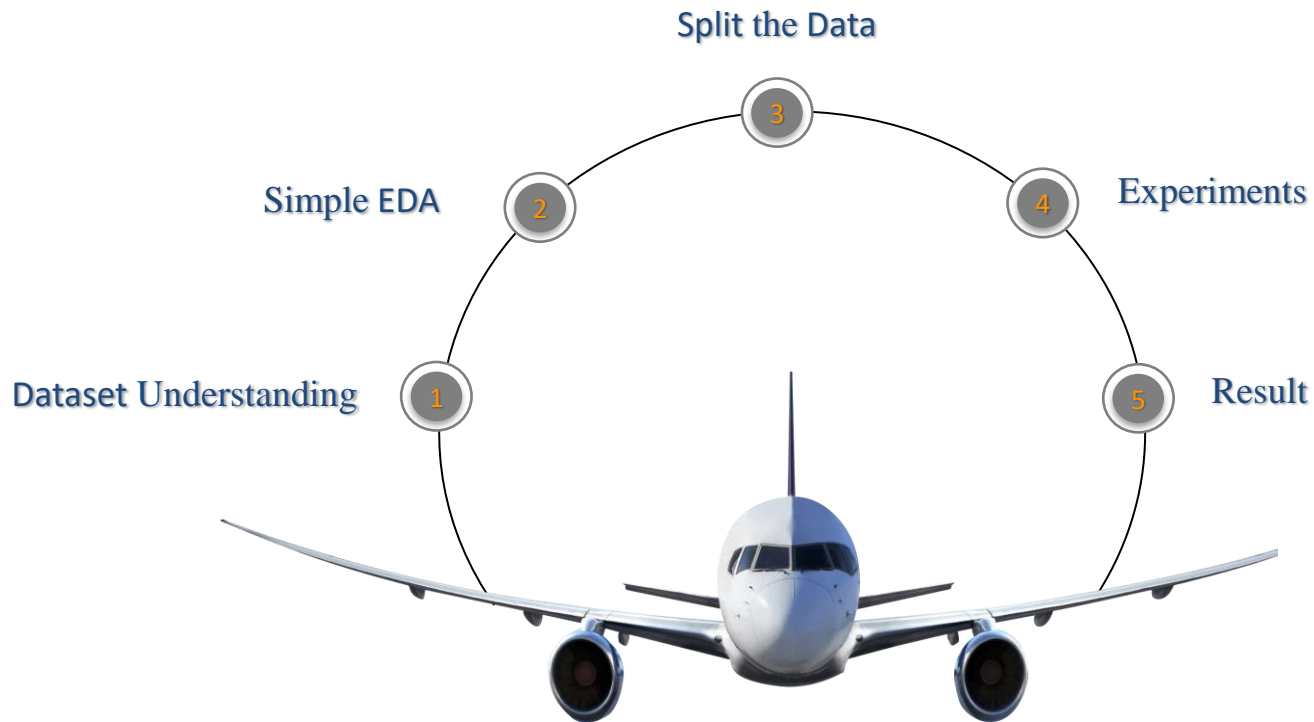
✈ Result

Introduction

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?



Methodology



Data Understanding

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

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)



Pre-processing the dataset & Simple EDA



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.



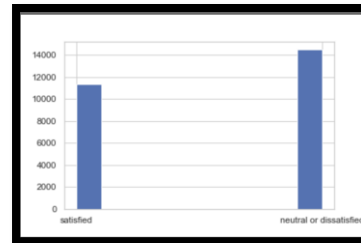
Missing values & outliers was observed across the dataset, so a good approach it to either remove it or correct it.



Split the data and do different features engineering on it.

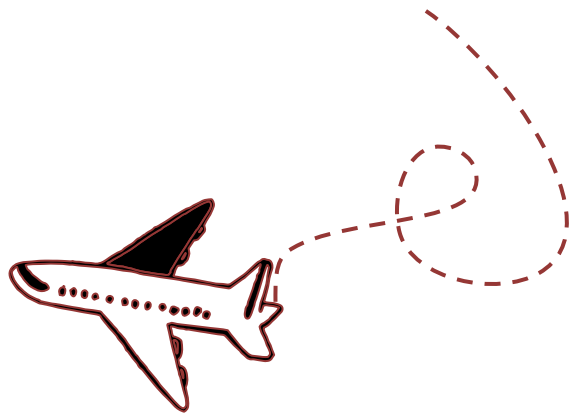


Check if our dataset is balance or not.





Experiments



Experiment 1 :Baseline Model

Train Score: 74.4 %

Test Score: 75.5 %



75%

It shows the score
percentage of
baseline experiment



Experiment 2: Logistic Regression

Train Score: 81.5 %

Test Score: 81.9 %



81%

It shows the score
percentage of
logistic experiment



Experiment 3: Standard Scale with the Grid search

Train Score: 87 %

Test Score: 87.1 %



87%

It shows the score percentage of the experiment with scale and grid search



Experiment 4: KNN Classifier

Train Score: 78 %

Test Score: 69.8 %



78%

It shows the score percentage of experiment with the KNN classifier



Experiment 5: Decision Tree Classifier

Best Score:
94.1 %



94%

It shows the score percentage of experiment with decision tree classifier



Experiment 6: Random Forest Classifier

Train Score: 97.35 %

Test Score: 94.74 %



94%

It shows the score percentage of the experiment with random forest classifier



Experiment 7: XGBoost

Best Score:

95.4 %



95%

It shows the score percentage of the experiment with XGBoost



Voting & Stacking

We do voting & stacking for all classifier, and we have these results :

Voting



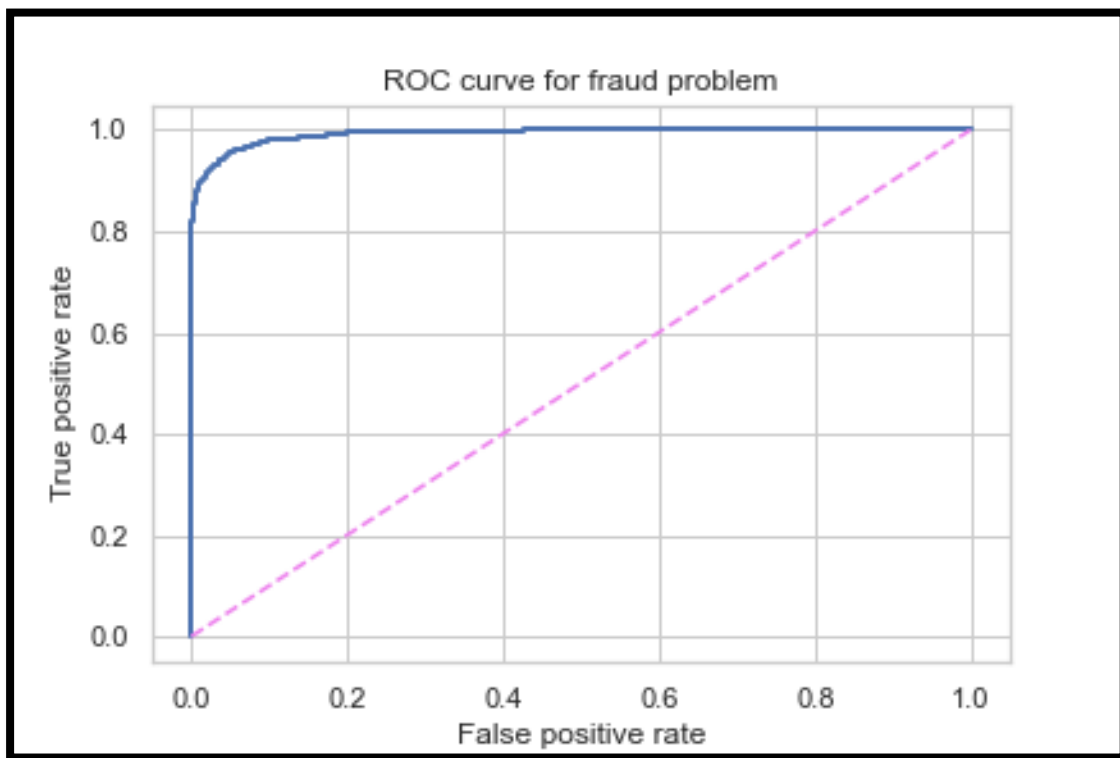
Stacking



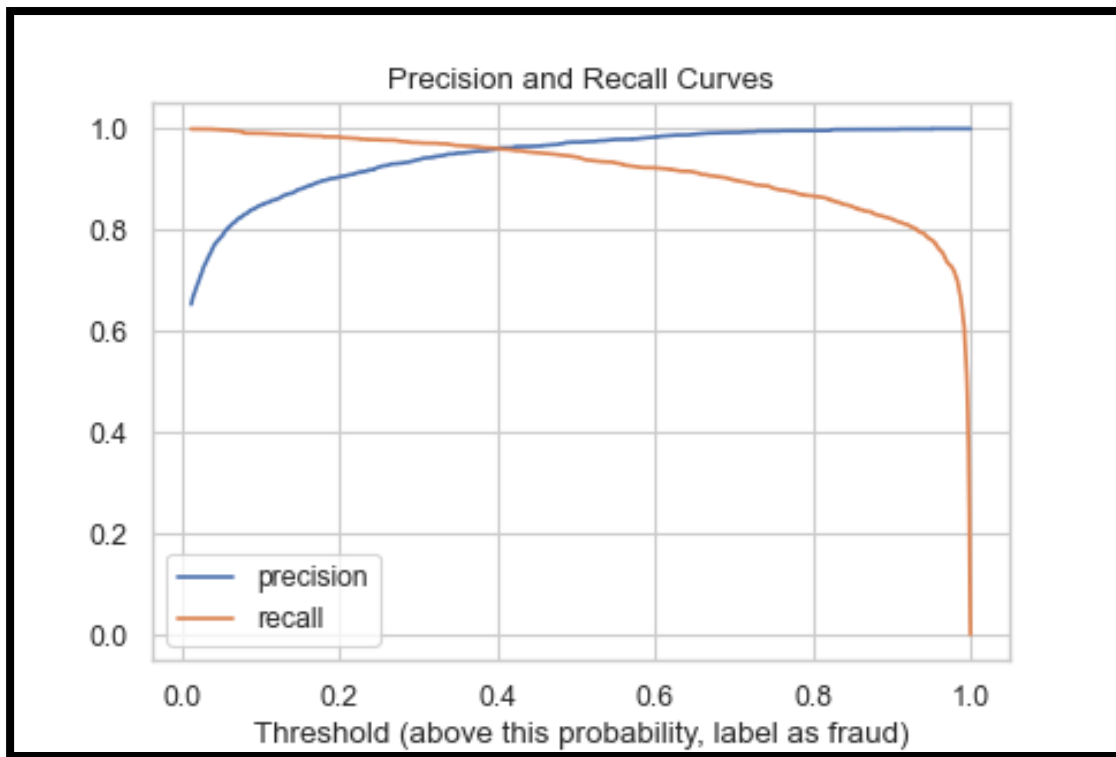
94 %



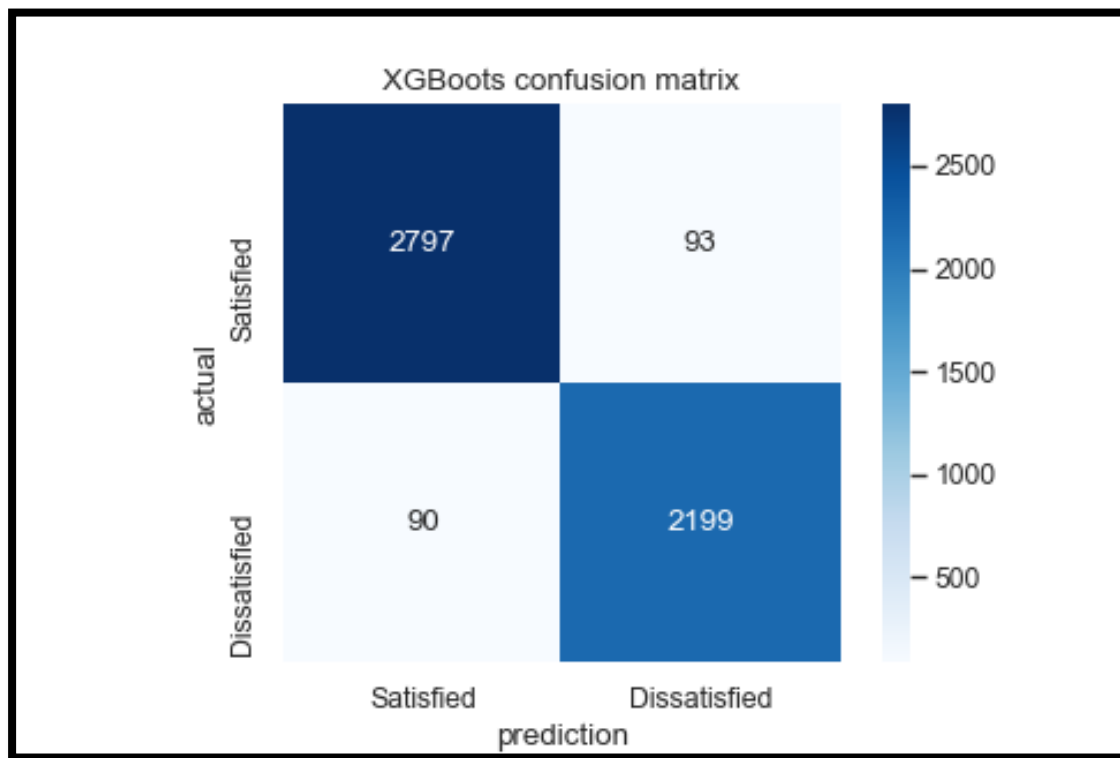
ROC Curve



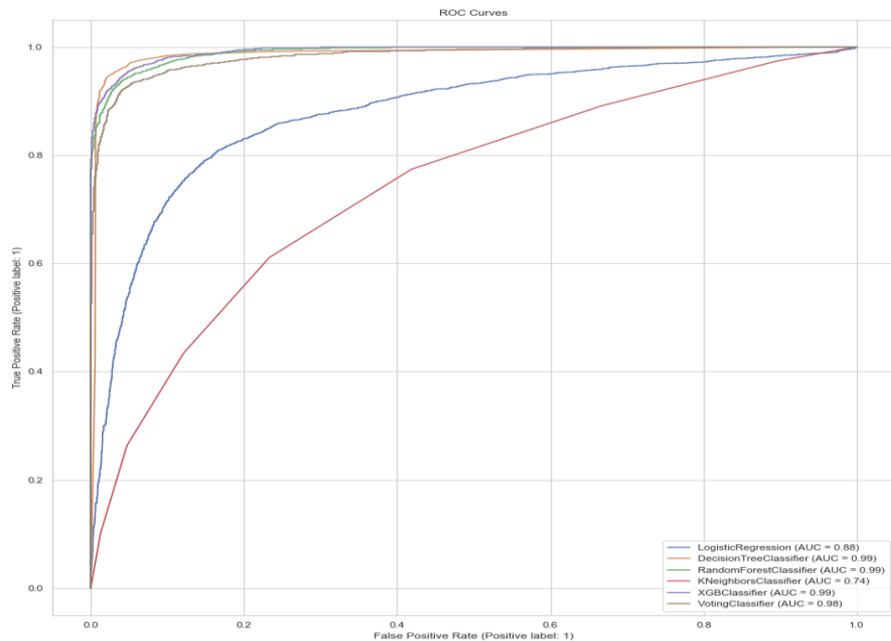
Precision & Recall For Best Model



Confusion Matric For Best Model



Conclusion



After doing the previous experiments, it shows that XGBoost scores the highest score, and the best threshold is 0.4



Thank you