Airline Passenger Satisfaction

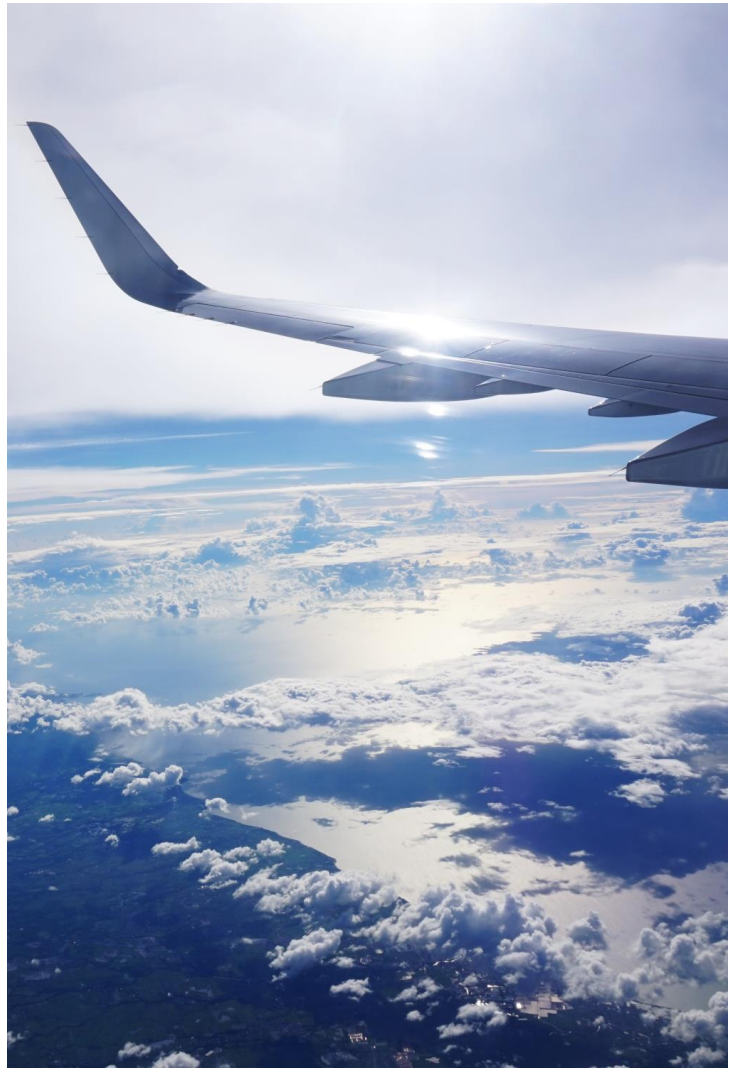
Presented By:
Wafa Albattah
Leena Alshwaihi
Taghreed Albaiz

Outline:

- Introduction
- Goal of Project
- Workflow plan
- Data description
- Modeling
- Conclusion

Introductions

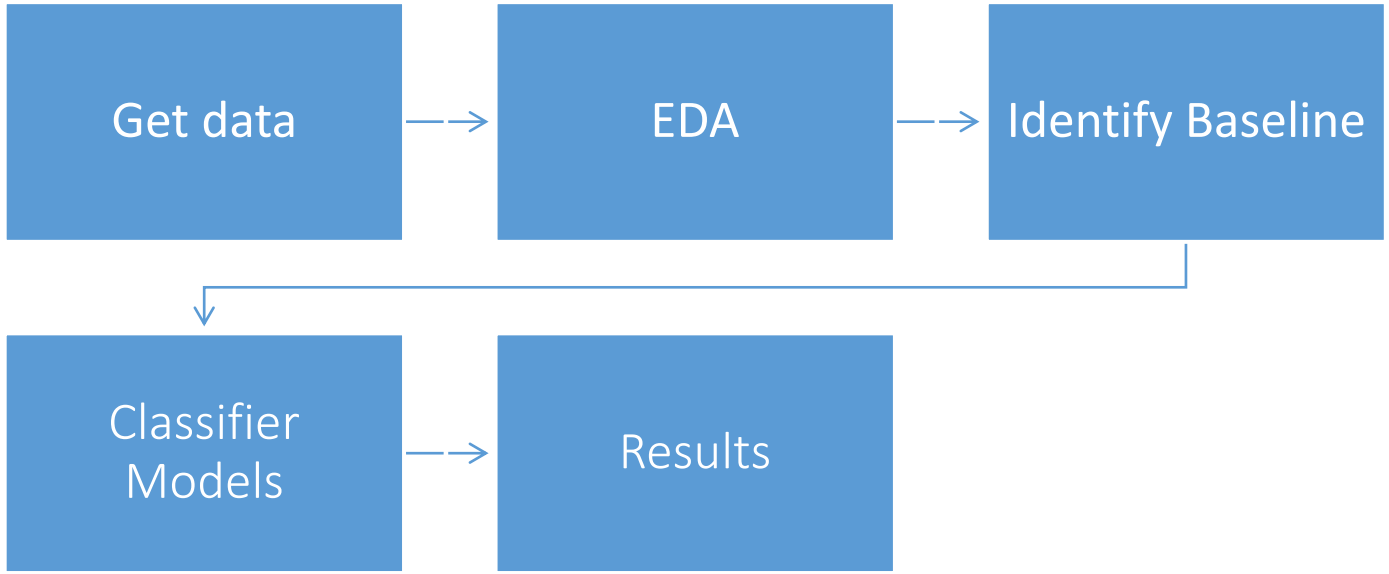
- This dataset contains an airline passenger satisfaction survey.
  It was created by observing the passengers from past flights of an airline company.
  It consists of factors that are supposed to affect the passengers' satisfaction during the journey.
  Most of them are personal information (eg. gender, age), flight information (eg. gate location, arrival delay) and ratings appointed by the passengers for flights services (eg. food, wi-fi).



The goal of project is:

- is to use the labeled training data to train our models and accurately predict, given a test dataset, which passengers were left satisfied

WORK FLOW PLAN



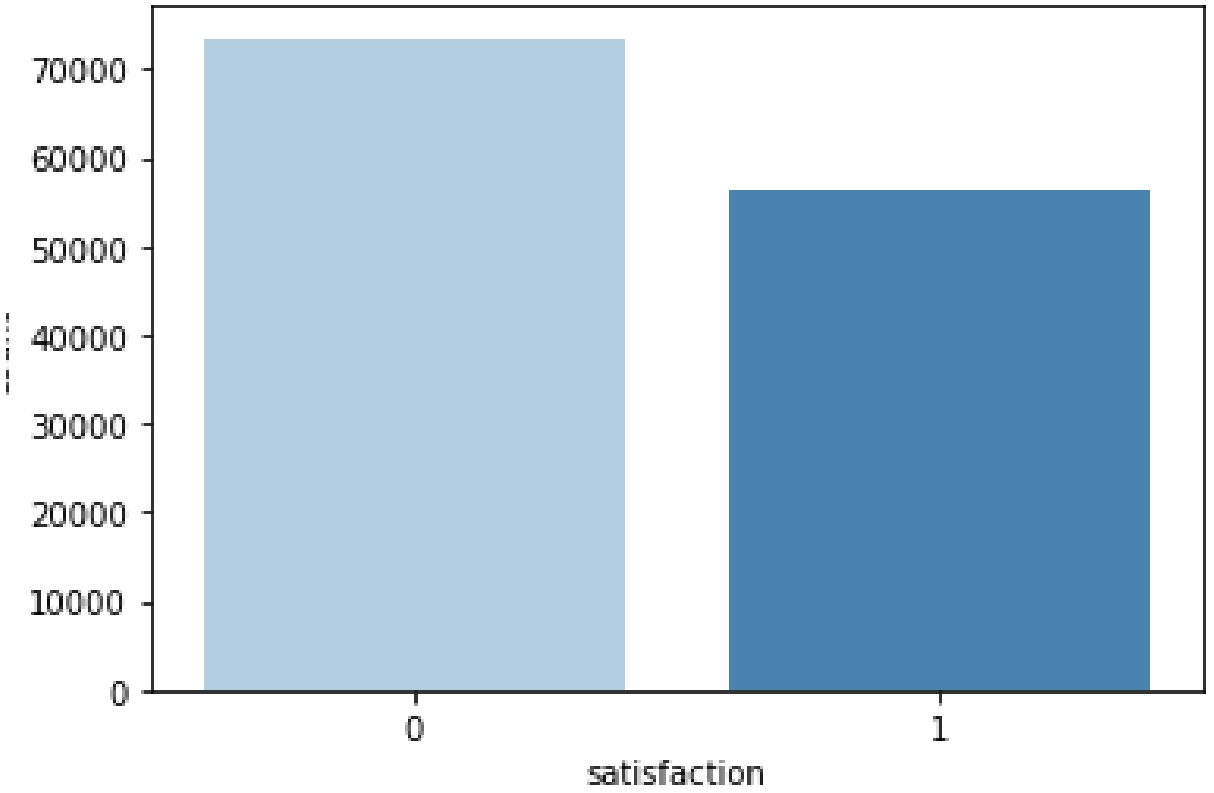
Data Description

- From Kaggle
- 129880 rows × 24 columns

Feature	Description
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	(Personal Travel, Business Travel)
Class	(Business, Eco, Eco Plus)
Cleanliness	Satisfaction level of Cleanliness
Gate location	Satisfaction level of Gate location
Food and drink	Satisfaction level of Food and drink
Seat comfort	Satisfaction level of Seat comfort
Satisfaction	Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

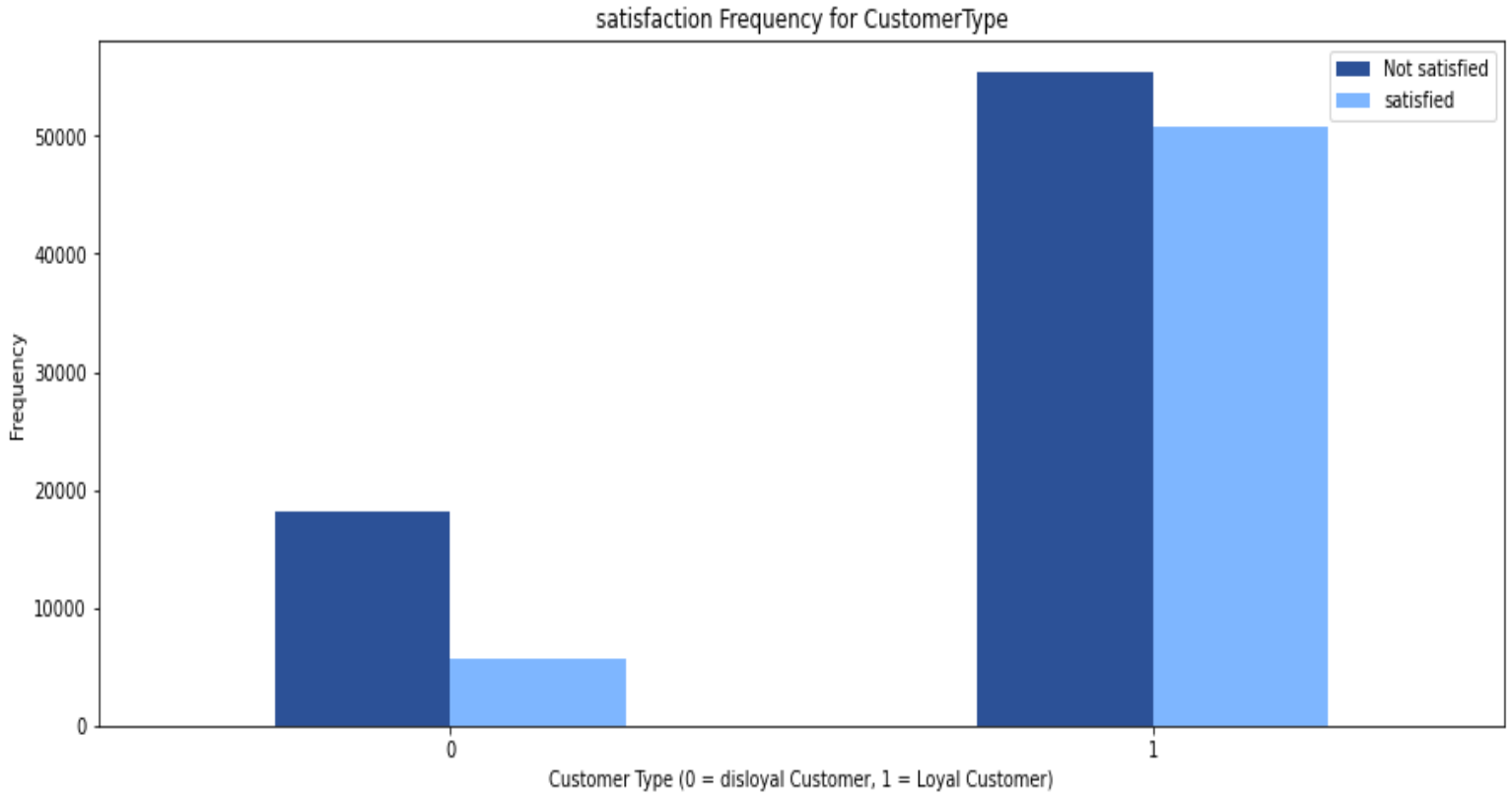
EDA:

The bar chart shows class imbalance



EDA:

The relationship between the customer type shows whether is satisfied with the services or not



Baseline Model :

- This is the confusion matrix of the baseline model, which was created Using Dummy Classifier. The accuracy of the baseline model is: **0.56**

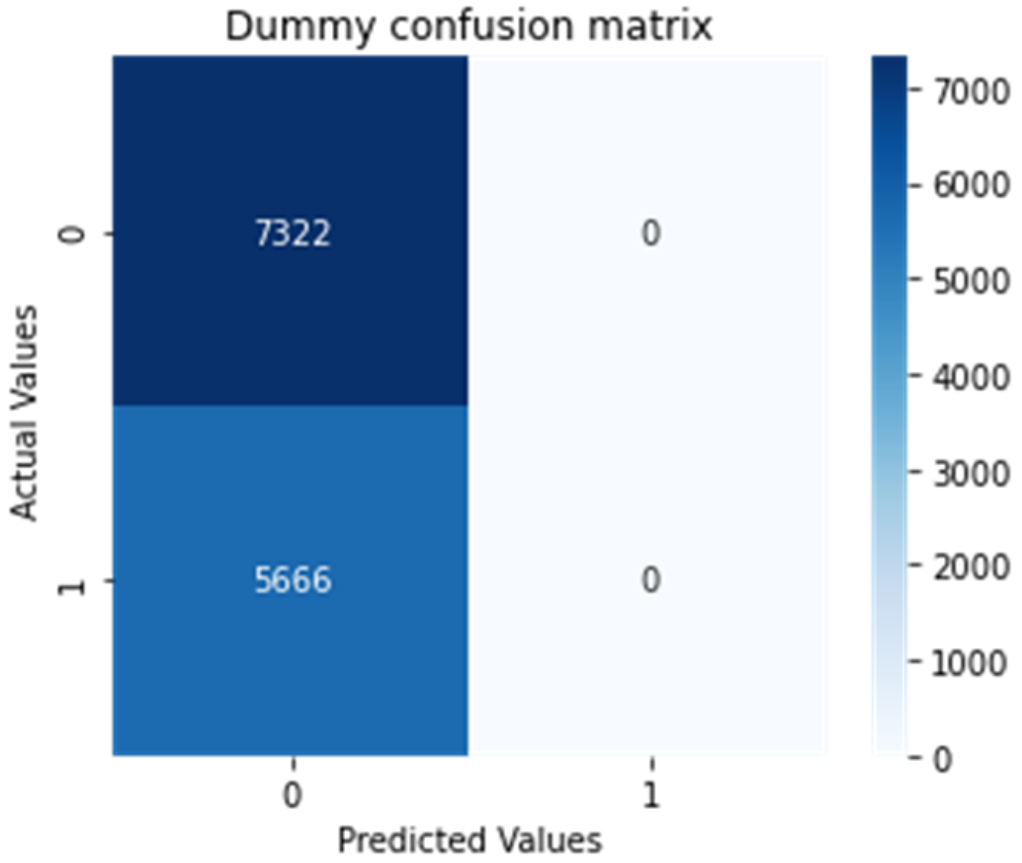


Table showing all confusion matrix values of the models:

Model	Accuracy	Recall	Precision	F1 - Score
KNN	0.93	0.89	0.96	0.92
Logistic Regression	0.87	0.84	0.87	0.85
Gaussian Naive Bayes	0.86	0.82	0.86	0.84
Random Forests	0.963	0.94	0.98	0.96
AdaBoost	0.95	0.94	0.96	0.95
XGBoost	0.9659	0.95	0.97	0.96
Bagging	0.9655	0.94	0.98	0.96
Voting	0.9654	0.94	0.98	0.96
Stacking	0.964	0.94	0.98	0.96

ROC Curves :

AUC of **Dummy** model: **0.5**

AUC of **KNN** model: **0.97**

AUC of **LogReg** model: **0.9945**

AUC of **Random forest** model: **0.924**

AUC of **AdaBoost** model: **0.993**

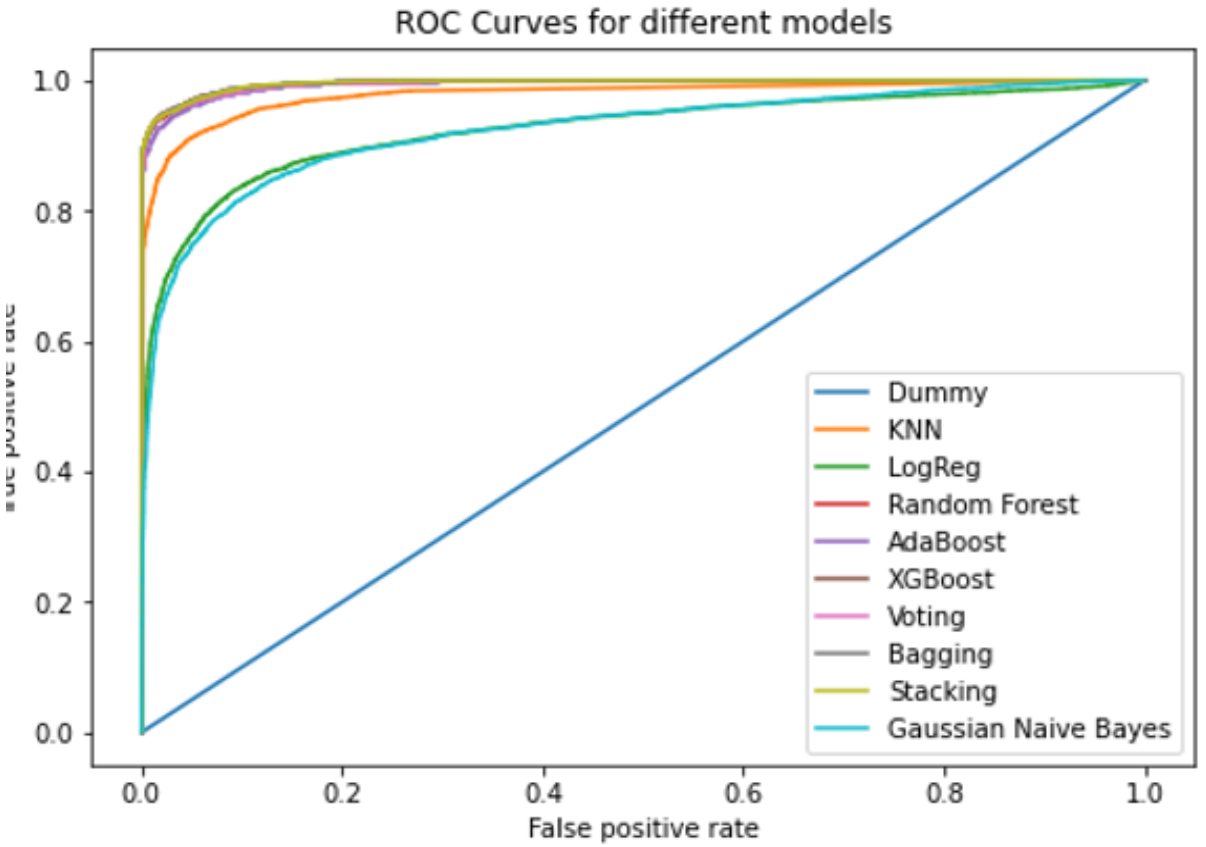
AUC of **XGBoost** model: 0.99547

AUC of **Voting** model: **0.9944**

AUC of **Bagging** model: 0.9945

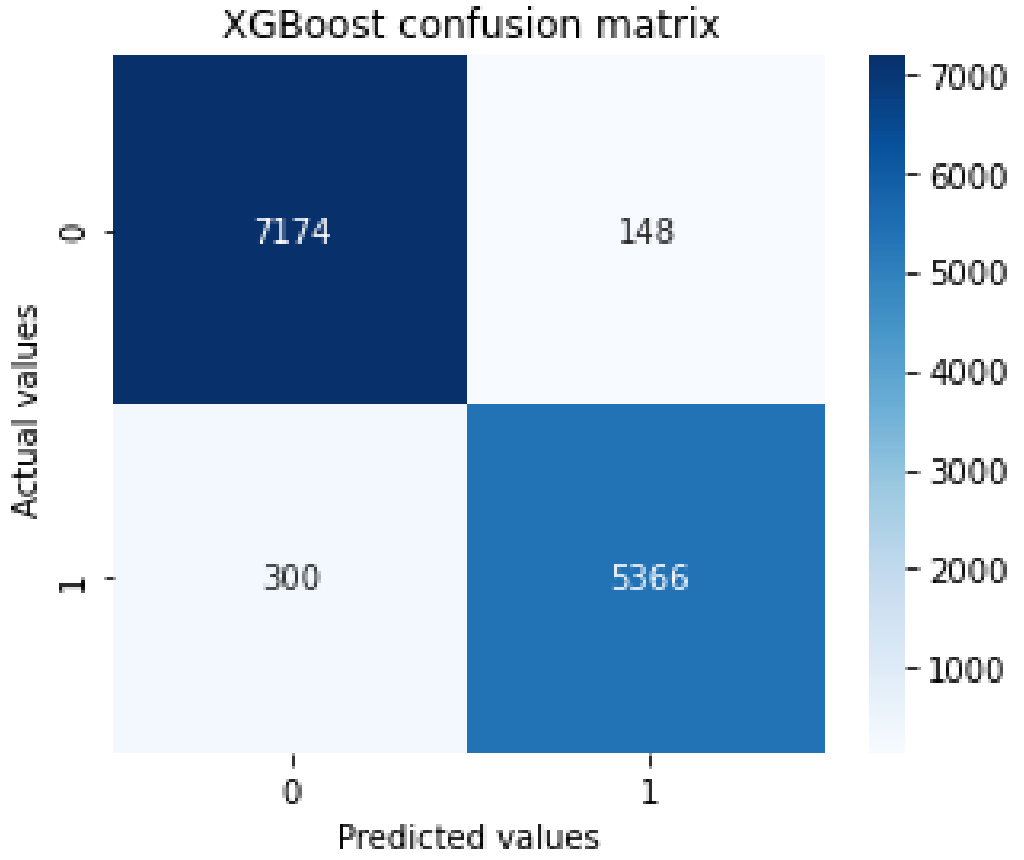
AUC of **Stacking** model: 0.9954

AUC of **Gaussian Naive** model: **0.922**



Best Model for our Project : (XGboost)

- XGboost Accuracy = 0.96
- XGboost F1 Score = 0.96



Thank you!