Prediction of Airline Passenger Satisfaction using Classification Models

Done By: Nisreen Alsayegh & Eman Alshehri

OUTLINE

- Introduction
- Tools
- Workflow
- Data
- EDA-Analysis
- Models
- Conclusion

Introduction

- In this project, we are going to work on a dataset contains an airline passenger satisfaction survey .

- The aim of this project is to build a classification models to predict passenger satisfaction and identify which model is outperformed based on accuracy.

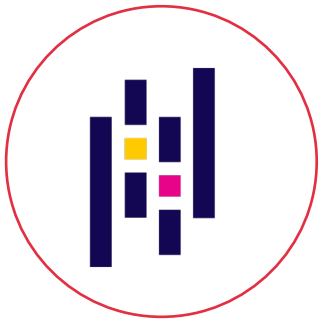
Goals

- Build a classification models to predict passenger satisfaction.
- Choose the model that give us the best predict.

TOOLS:

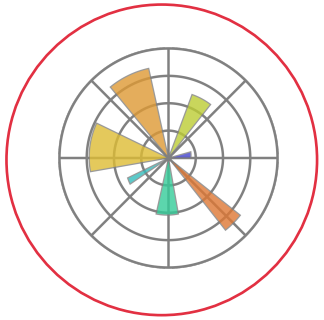
Jupyter

Seaborn



Pandas

mlxtend



Matplotlib

Scikit-Learn

Numpy

Workflow

Loading Data

EDA

Building Models

Choosing the best Model

Data Description:

•We used the dataset from https://www.kaggle.com

•The dataset contains 129880 rows and 25 columns

Cleaning Data :

By removing Null ,outliers and unwanted columns .

The data Becomes :116885 Rows , 23 columns

Convert some numerical data in to categorical.

Convert categorical data to dummies:

Before :116885 Rows , 23 columns

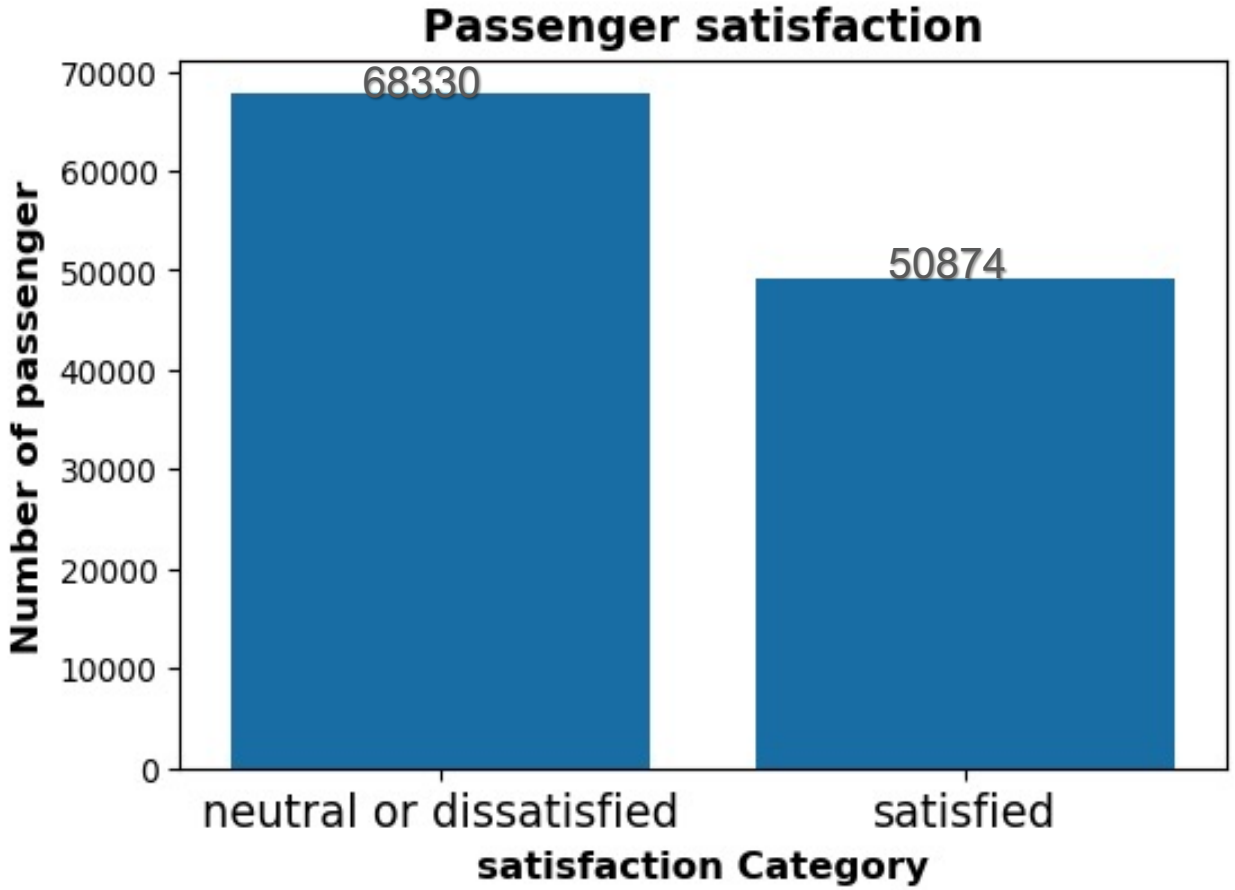
After :116885 Rows , 28 columns

Removing unwanted columns .

Before :116885 Rows , 28 columns

After :116885 Rows , 25 columns

EDA - Analysis



Models:

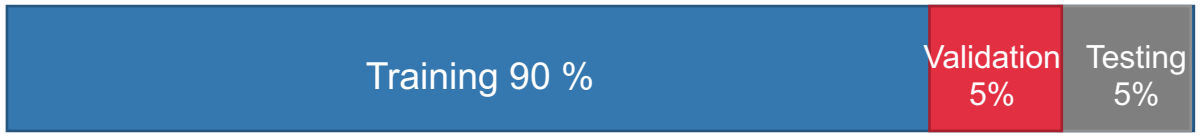
After Grid Search for the best hyperparameters

	KNN Model	Logistic Regression	Decision Tree	Random Forest
Training	100.00%	89.99%	95.46%	100.00%
validation	92.93%	89.07%	94.54%	95.83%

After tuning the hyperparameters manually

	XGBoosting	Naive Bayes	SVM Model	Gradient Boosting
Training	99.95%	88.24%	95.79%	100.00%
validation	95.83%	86.91%	94.94%	95.81%

Data Splitting :

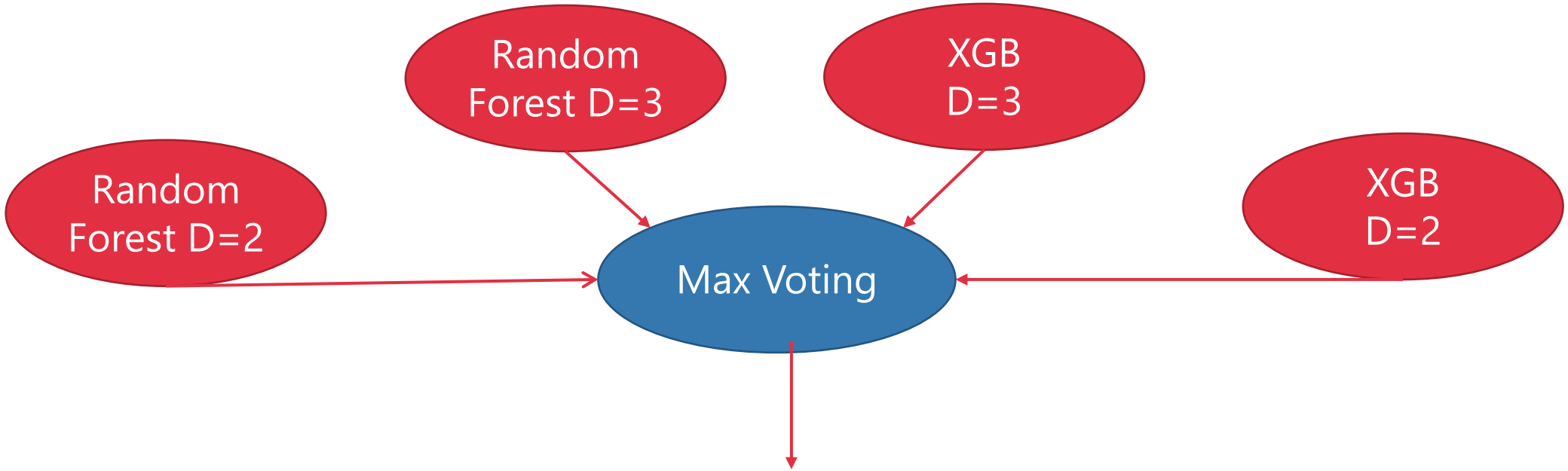


Models:

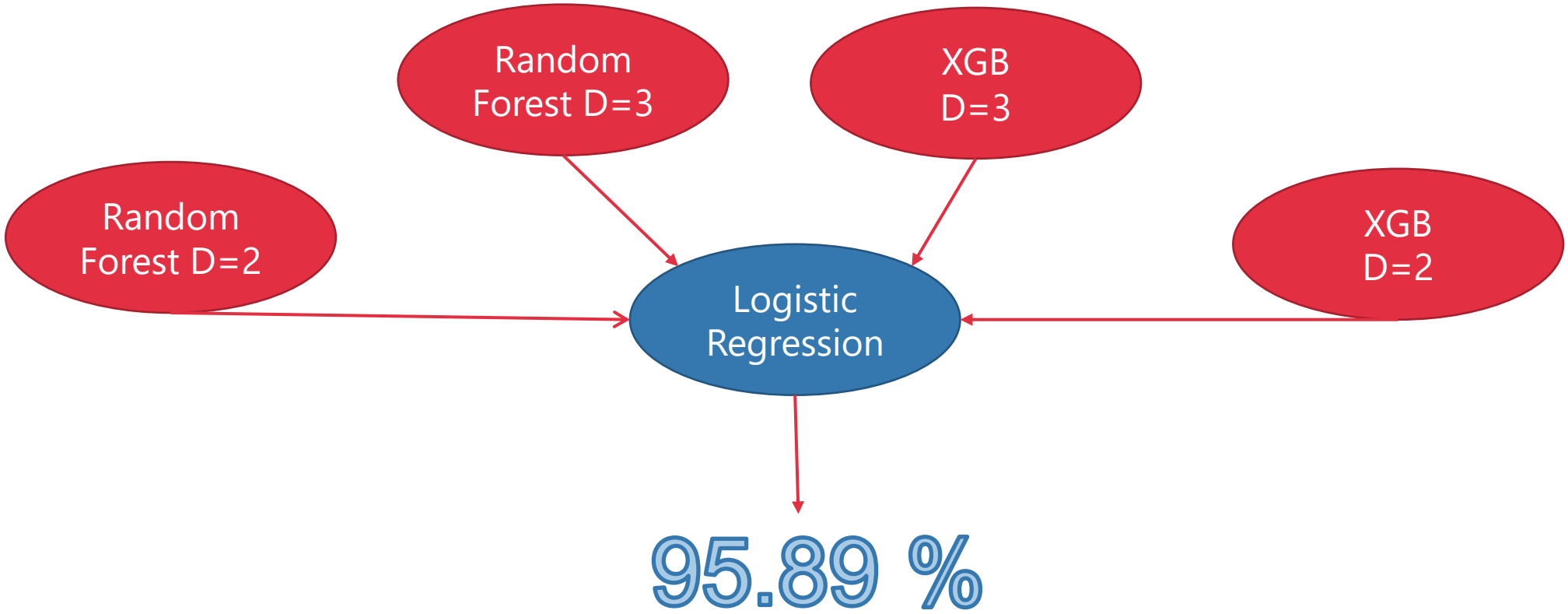
After feature engineering

	Polynomial Random Forest D=2	Polynomial Random Forest D=3	Polynomial XGBClassifier D=2	Polynomial XGBClassifier D=3
Training	100.00%	100.00%	100.00%	100.00%
validation	96.10%	95.91%	95.91%	95.89%

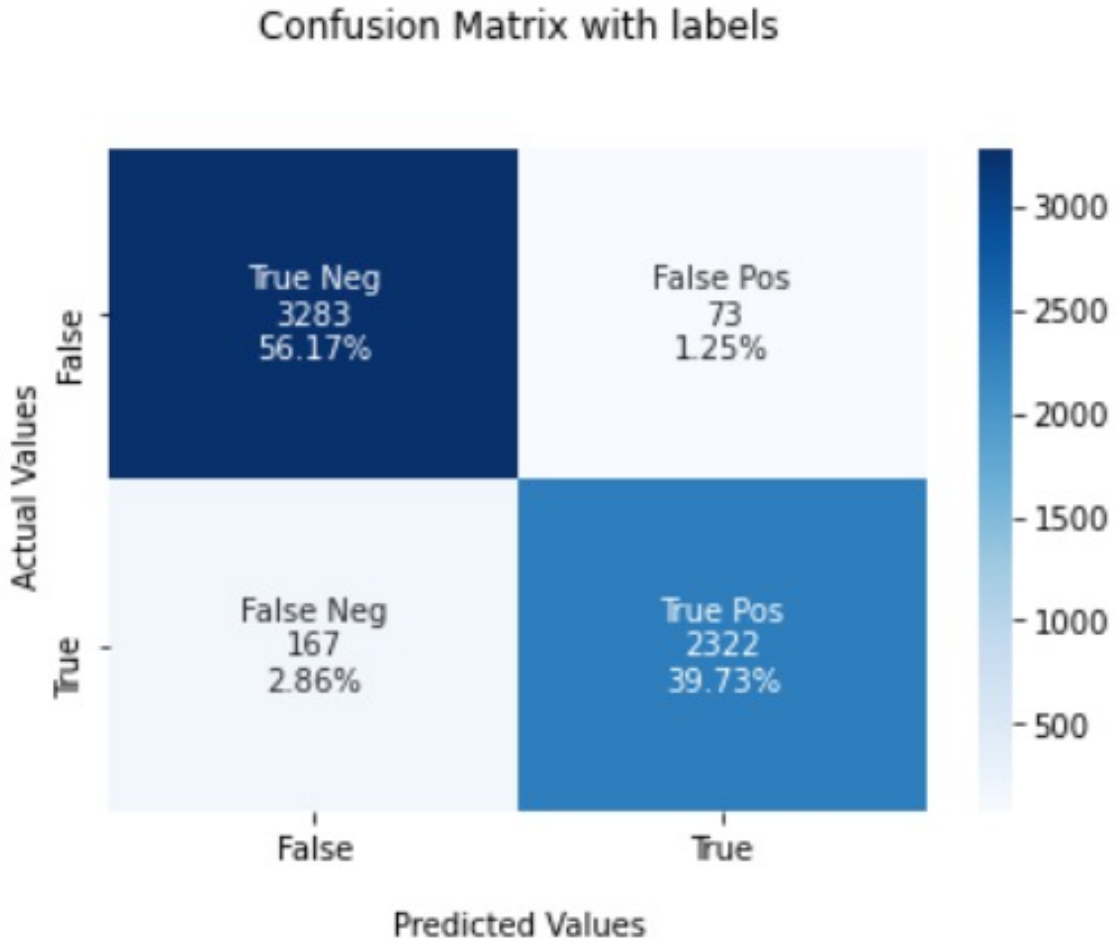
Ensemble:


Random Forest D=3
XGB D=3
Random Forest D=2
XGB D=2
Max Voting
95.84 %

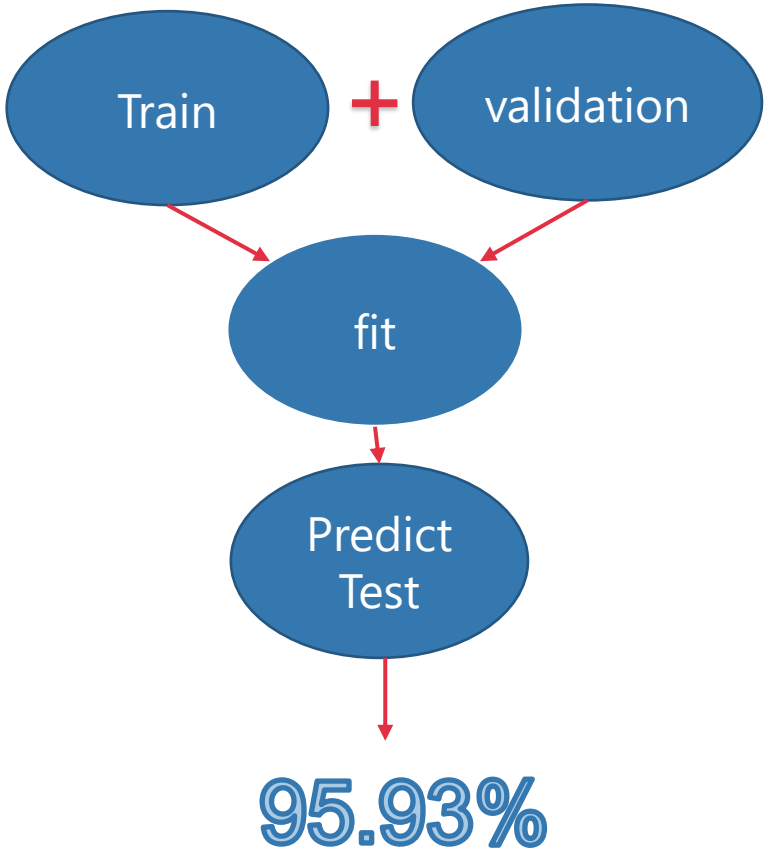
Stacking Classifier :



Testing our best model:



Random forest classifier D=2



95.93%

Conclusion

After building different models, Polynomial Random Forest degree=2 has the highest score in validation stage.

Future work:

- **Collect more data**
- **Explore different models**

THANK YOU