Classification regression

(E-commerce Shipping)

Ebtesam Sultan - Nuha Aljohani - Hana Algarni

Abstract

The goal of our project to making a model that predicts the product shipment delivered on time or not,to make our customers more satisfied .

Data Description

- We extract our data set from Kaggle .com
- with 1999 rows and 12 columns .
- The Columns include (ID, Warehouse block, Mode of shipment, Customer care calls, Customer rating, Cost of the product, Prior purchases, Product importance, Gender, Discount offered, Weight in gms and reached on time) .
- Data Type :String -Integer .
- Data set hasn't null values .
-

Tools

- **Technologies**: Python, Jupyter Notebook, PowerPoint
- **Libraries**:Pandas,NumPy,Sklearn-Learn,sklearn-metrics, imblearn.under_sampling, Statsmodels, Warnings, Matplotlib and Seaborn.
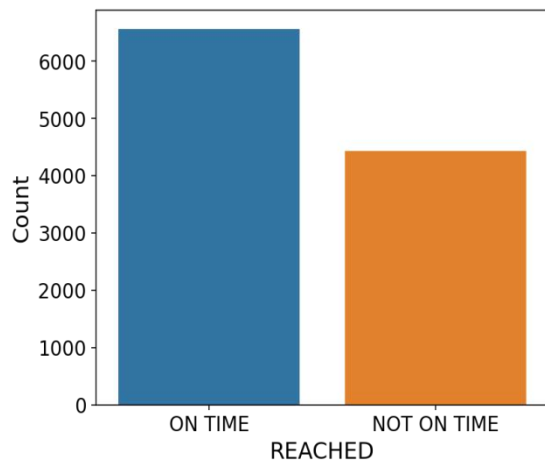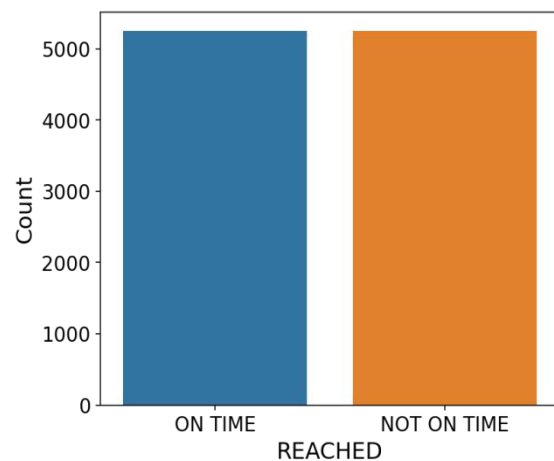
Communication:
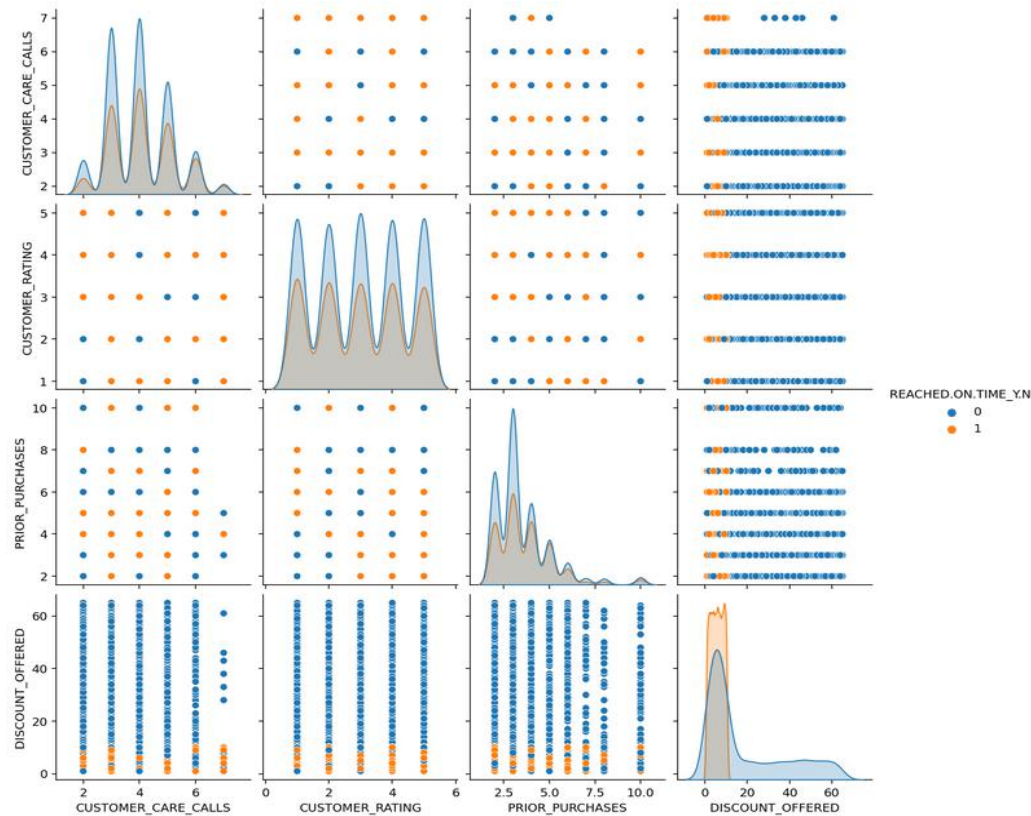
- ▪ The figure visualization the data that's mean our data nonlinear .
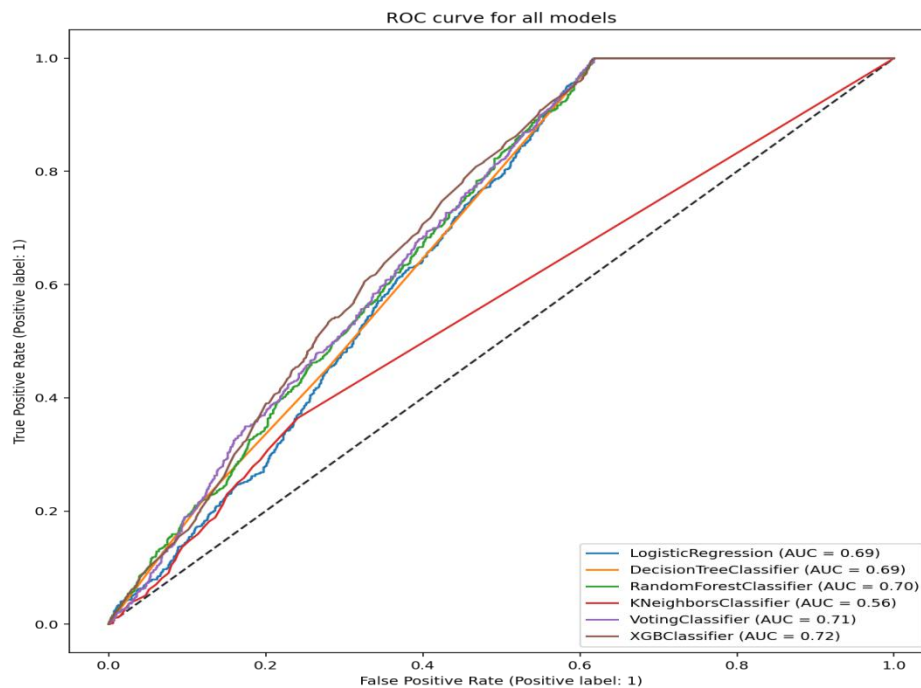

Handling The Class Imbalance
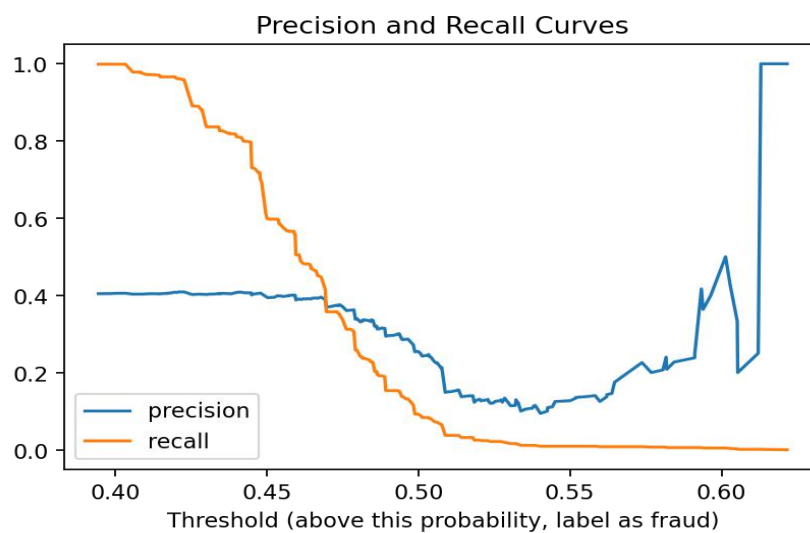
Data before under sampling

Data after under sampling

ROC curve for models :

ROC curve for all models

- This figure help us to find the best model .
- The best model is voting Classifier (AUC= 0.71)

- This figure show the precision and recall curve that give us the best threshold point .

Experiments :

Model	F1	Precision	Recall
Baseline Model	0.53	0.53	0.54
Dummy model	0.53	0.53	0.54
Scaling	0.53	0.53	0.54
Grid search	0.68	0.51	0.99
Decision tree	0.65	0.53	0.87
Random forest	0.63	0.54	0.77
Voting	**0.71**	**0.54**	**0.68**
Stacking classifier	0.42	0.51	0.36
XG-boots	0.67	0.53	0.89

We make a lot of Experiments and we noticed during the project that Voting Model gave us the best results among the models, and it can help the companies to predict the products has reached on time or does not reached on time.