Data Science: Bank Marketing (Campaign) (Group Project)

May 2022

Prepared by Mustafa Fakhra

Team Members

Team Name:			*SAAN*		
SL. No	**Name**	**Email**	**Country**	**College / Company**	**Specialization**
1	Mustafa Fakhra	mostafafakhra@hotmail.com	UAE	Rasan	Data Science

Agenda

- Executive Summary
- Problem Statement
- Approach
- EDA
- Model Development
- Model Selection
- Model Evaluation
- Conclusion

Executive Summary

One of the challenge for all banking companies is to understand the whether the customers are eligible for deposit. To solve this problem ABC company approached an analytics company to automate this process of identification.

ML Problem:

With an objective to gather insights on the factors that are impacting the deposited, build a classification for the given dataset

Target Variable: Deposited

Problem Description

- ABC Bank wants to sell it's term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

Business Understanding

- Bank wants to use ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc) can focus only to those customers whose chances of buying the product is more.

- This will save resource and their time (which is directly involved in the cost (resource billing)).

- Develop model with Duration and without duration feature and report the performance of the model.

- Duration feature is not recommended as this will be difficult to explain the result to business and also it will

- be difficult for business to campaign based on duration.

Task

- Problem understanding
- Data Understanding
- Data Cleaning and Feature engineering
- Model Development
- Model Selection
- Model Evaluation
- Report the accuracy, precision and recall of both the class of target variable
- Report ROC-AUC as well
- Deploy the model
- Explain the challenges and model selection

Approaches taken

- Data was taken from github and analysed
- Problem understanding
- Data Understanding
- Data Cleaning and Feature engineering
- Model Development
- Model Selection
- Model Evaluation

Data Intake Report

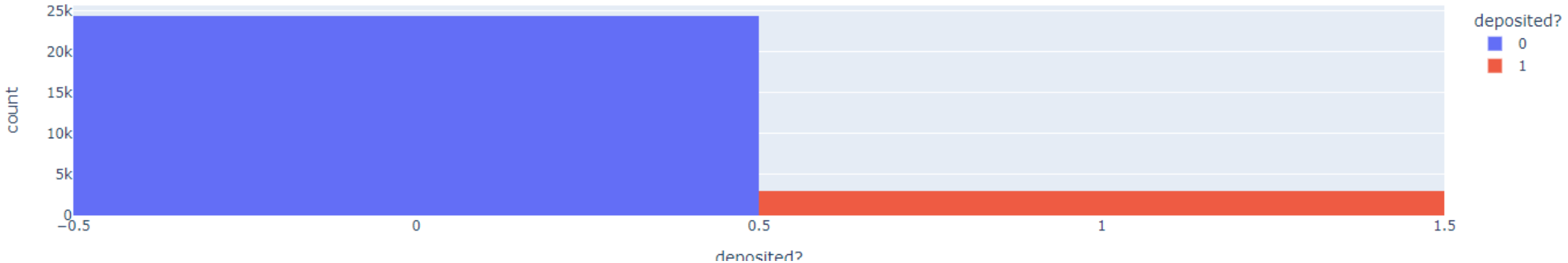
- Name: Bankuing – Data Science Report date: 24/9/2022
- Data storage location: https://github.com/mostafafakhra/DataGlacierInternship---30-July-to-30-October-2022
- Total number of files 1
- Base format of the file .csv

Missing Values

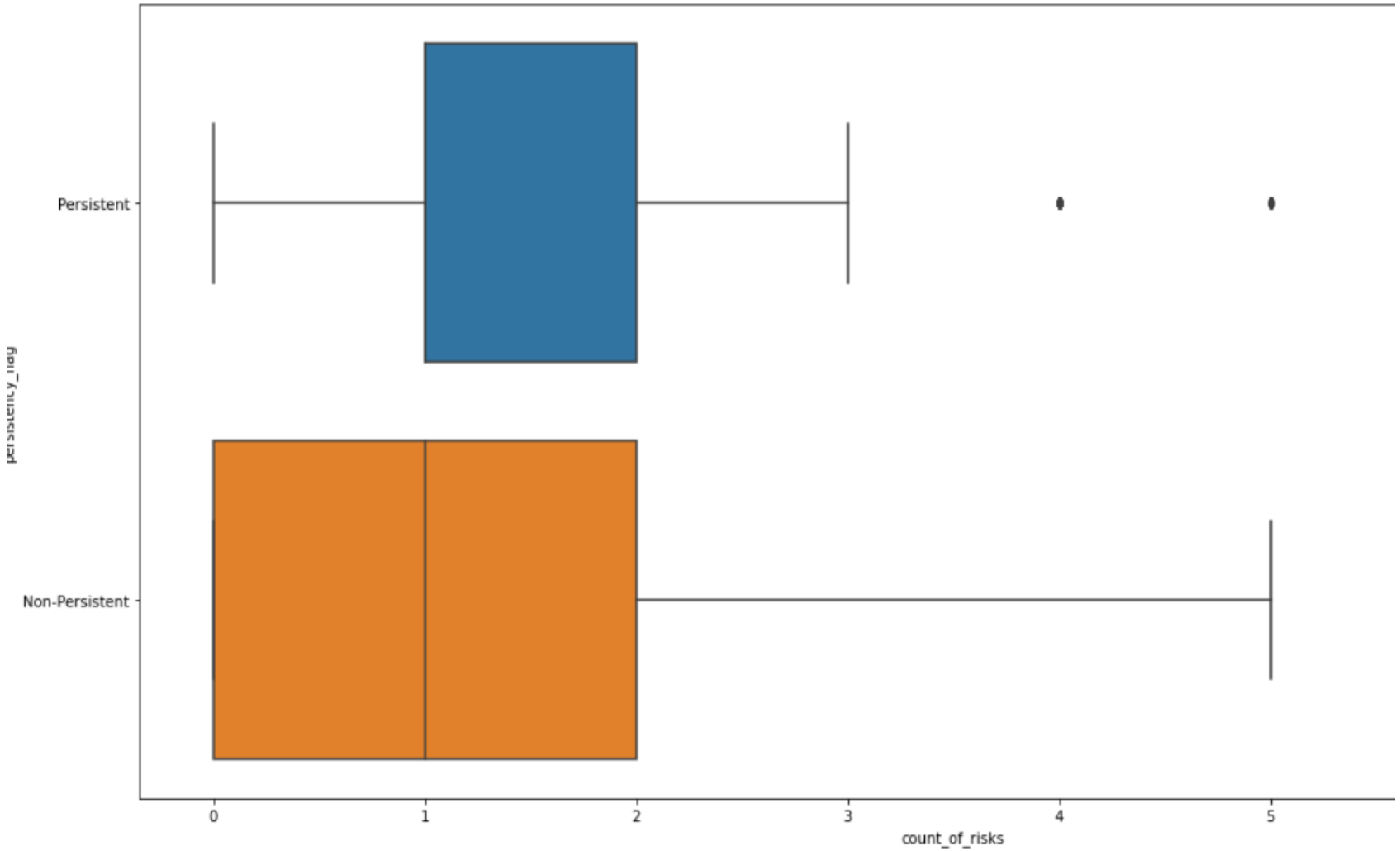
```
age                0
job              330
marital           80
education       1731
default         8597
housing          990
loan             990
contact            0
month              0
day_of_week        0
campaign           0
pdays              0
previous           0
poutcome           0
emp.var.rate       0
cons.price.idx     0
cons.conf.idx      0
euribor3m          0
nr.employed        0
deposited?         0
dtype: int64
```

No missing values were found

Histogram

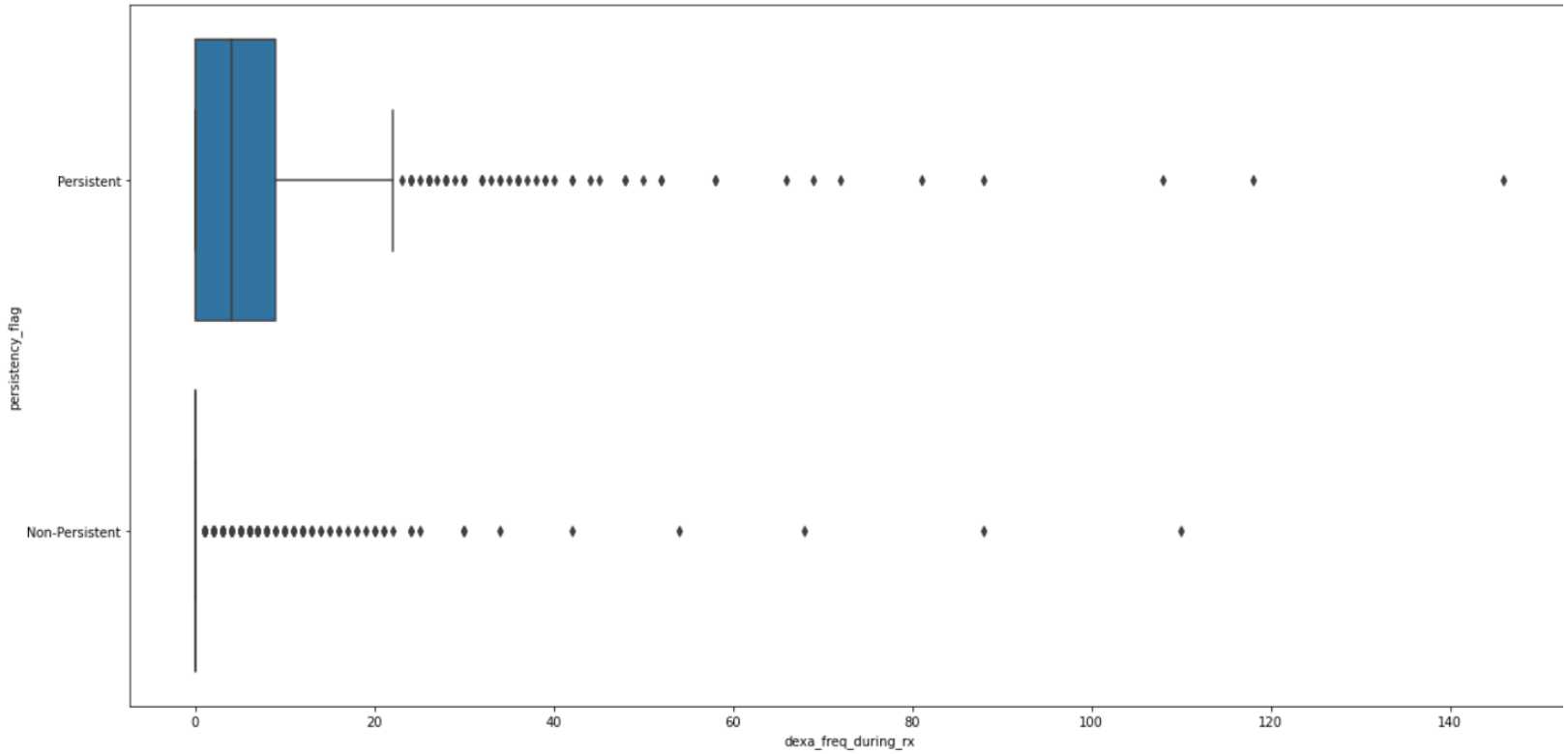


Analysis of Outliners



Visual analysis showing the outliners in one column by box plot Analysis

Analysis of Outliners



Box plot analysis showing the outliners

Analysis of Skewness and kurtosis

```
skweness: 0.8797905232898707
Kurtosis: 0.9004859968892842
```

Data shows a moderate positive skewed data on this column and fairly platykurtic so the data has little outliers

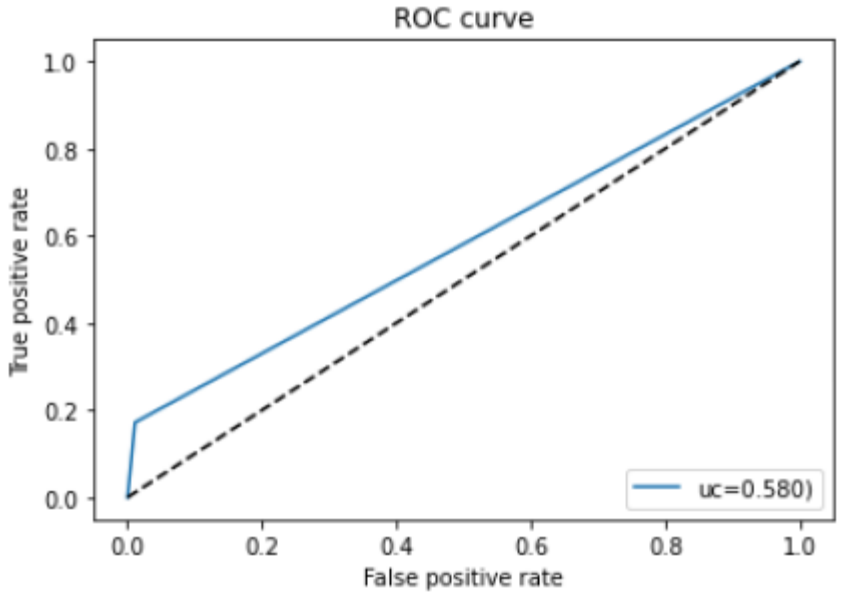
Model Creation-Logistic Regression

```
Accuracy : 0.8979418268412995
Precision : 0.6578947368421053
Recall : 0.1720183486238532
F1 Score : 0.2727272727272727
              precision    recall  f1-score   support

No Deposited       0.91      0.99      0.95     10450
   Deposited       0.66      0.17      0.27      1308

    accuracy                           0.90     11758
   macro avg       0.78      0.58      0.61     11758
weighted avg       0.88      0.90      0.87     11758

AUC : 0.5804110881875246
```



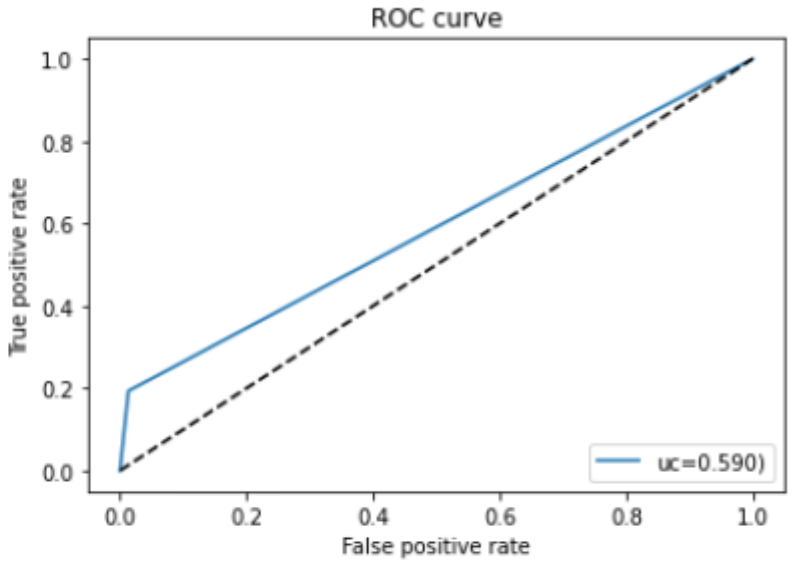
Model Creation- RidgeClassifier

```
Accuracy : 0.8985371661847253
Precision : 0.6470588235294118
Recall : 0.19342507645259938
F1 Score : 0.2978222483814008
              precision    recall  f1-score   support

No Deposited       0.91      0.99      0.95     10450
   Deposited       0.65      0.19      0.30      1308

    accuracy                           0.90     11758
   macro avg       0.78      0.59      0.62     11758
weighted avg       0.88      0.90      0.87     11758

AUC : 0.5901096674129026
```



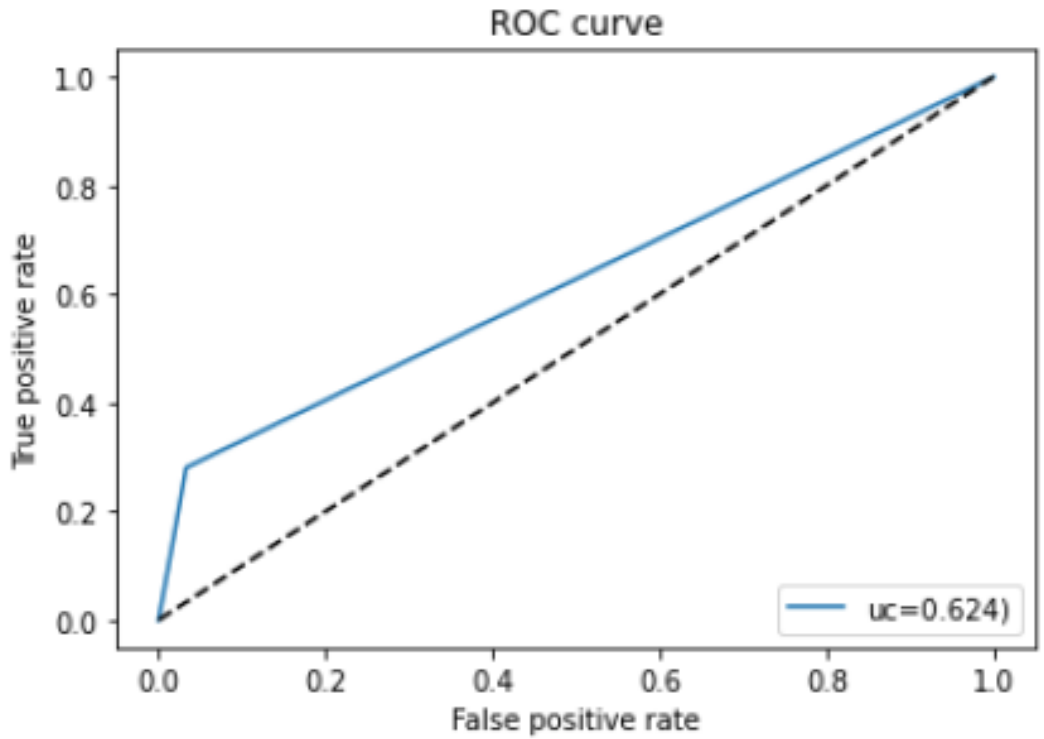
Random Forest Classifier

```
Accuracy : 0.8907127062425583
Precision : 0.5161744022503516
Recall : 0.2805810397553517
F1 Score : 0.36354631005448246
              precision    recall  f1-score   support

No Deposited       0.91      0.97      0.94     10450
   Deposited       0.52      0.28      0.36      1308

    accuracy                           0.89     11758
   macro avg       0.72      0.62      0.65     11758
weighted avg       0.87      0.89      0.88     11758

AUC : 0.6238311897341351
```



Conclusion

- Approximately all the classifiers have same result, but Random Forest was the best one.
- The model has around 89% Accuracy.
- Random Forest has 87% Precision, 89% Recall, & 88% F1 Score.
- We can also see the results for each classifier as well.

Thank You