Bank Marketing (Campaign)

Name: Mustafa Fakhra
Location : Dubai, UAE
Project: Data Science
Date : 24-September-2022

Executive Summary

Problem Statement:

- 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).

ML Problem:

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

The highest model accuracy and precision were attained using the Random Forest model.

Project Steps

1. Understanding the case
2. Importing Required libraries and dataset
3. Understanding our data (data exploratory)
4. Data processing and transformation
5. Model Building
6. Model evaluation
7. Model Deployment

Data Processing

- File Used: bank-additional-full.csv
- Correlation between all variables and the predictor.
- Missing data or nulls exist and have been handled.
- Data wrangling transformation included normalizing data and standardize them.
    - This has increased the correlation between the features and the predictor variable.
- Dummy variables have been created (Categorical variables to 0 and 1).

Model Building

- The case is classification so we will be using three models:
- Logistic Regression, Ridge Regression, & Random Forest classifier.

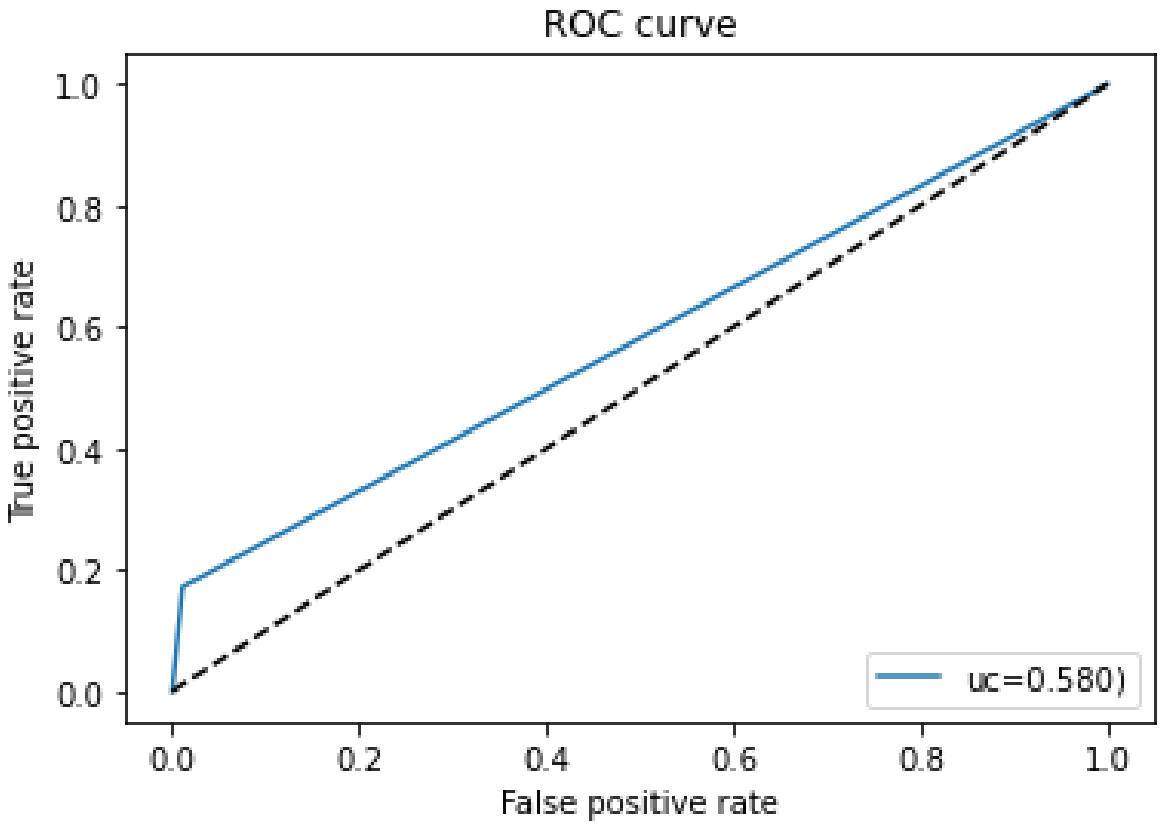
Logistic Regression Model Results

```
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
```



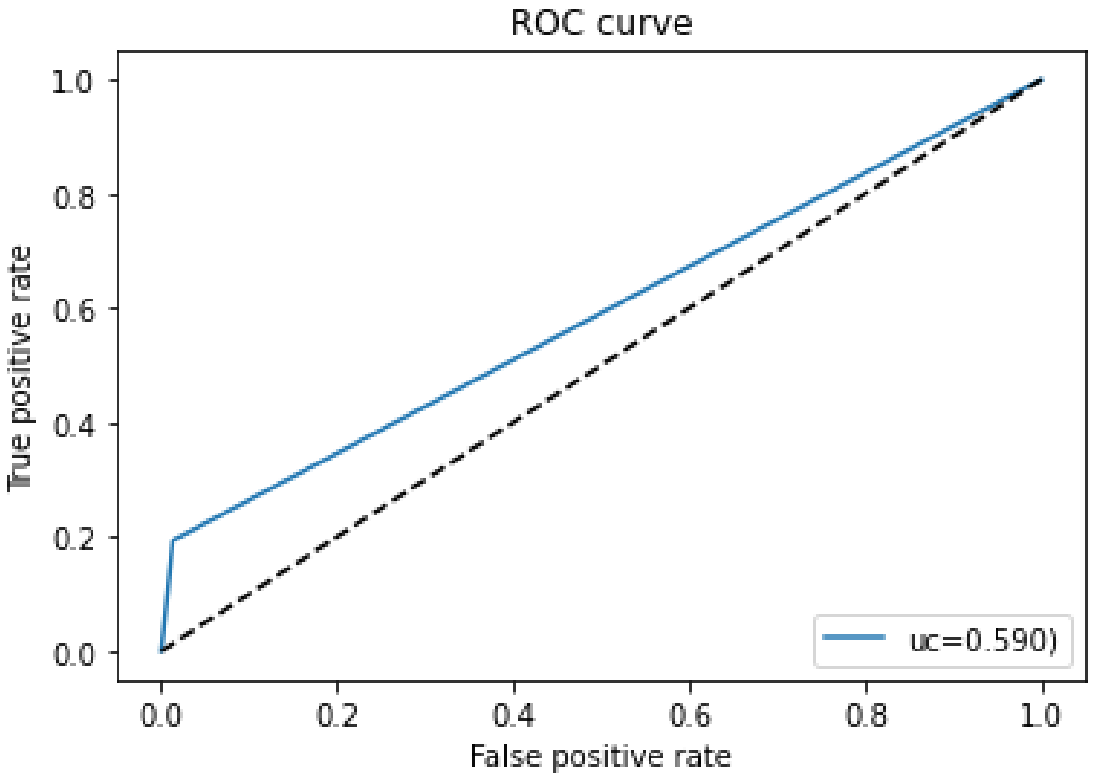
Ridge Regression Model Results

```
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
```



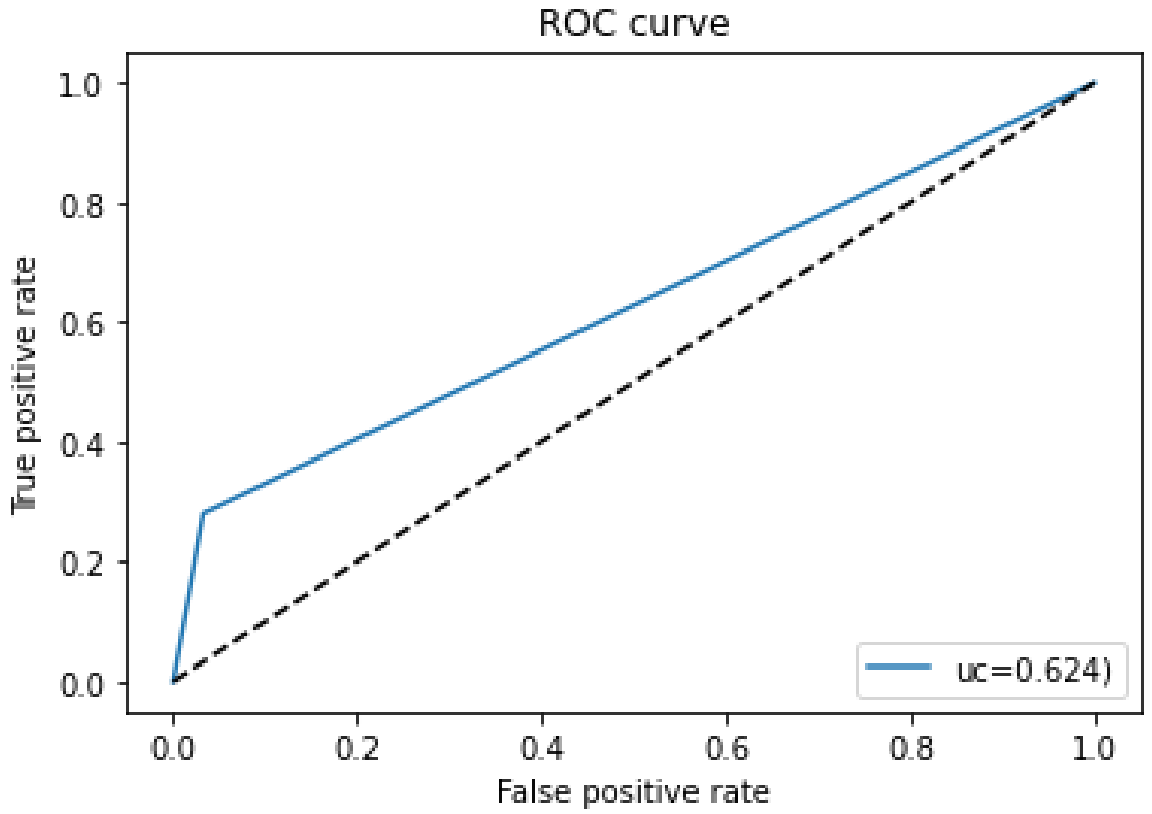
Random Forest Model Results

```
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
```



Random Forest Model

- Model Trade-offs:
  - Advantages:
    - Insensitive to Outliers.
    - Insensitive to Null values.
    - Less Prone to overfitting.
  - Disadvantages:
    - Losing Interpretability.
    - Difficult to diagnose and improve.
- Results obtained:
    - Accuracy: 87 – 89 %

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