

# **Project report**

Name: bank marketing

Internship Batch : LISUM04

project by : Lazghan Yassine Country: Morroco

Email: volubilisx7@gmail.com Specialization: Data Science

## Bank Marketing (Campaign)

Problem Statement :

ABC Bank wants to sell their term deposit product to customers and before launching the product they want to develop a model that helps them understand whether or not a particular customer will buy their product (based on the customer's past interaction with bank or other financial institution).

### **Dataset**

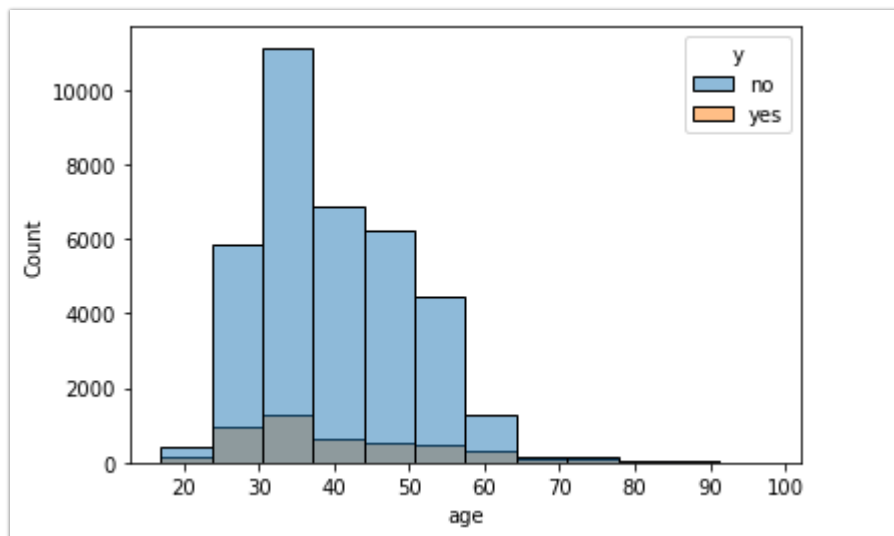
The data relates to direct marketing campaigns of a Portuguese banking institution. Marketing campaigns were based on phone calls. Often more than one contact with the same client was necessary, in order to access whether the product (term bank deposit) would be ('yes') or not ('no') subscribed.

The purpose of the classification is to predict whether the client will take out (yes / no) a term deposit (variable y).

The analysis was divided into 2 parts :

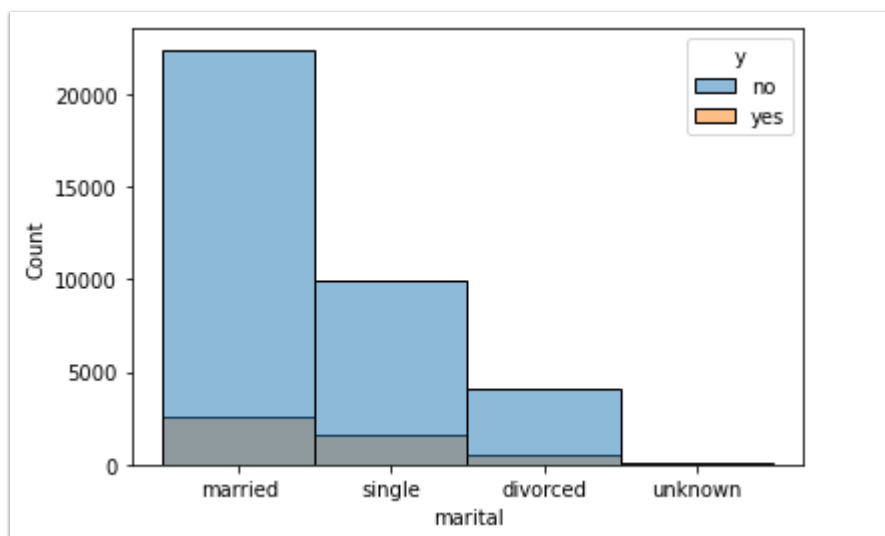
- EDA
- design of a classification model

## Age wise analysis



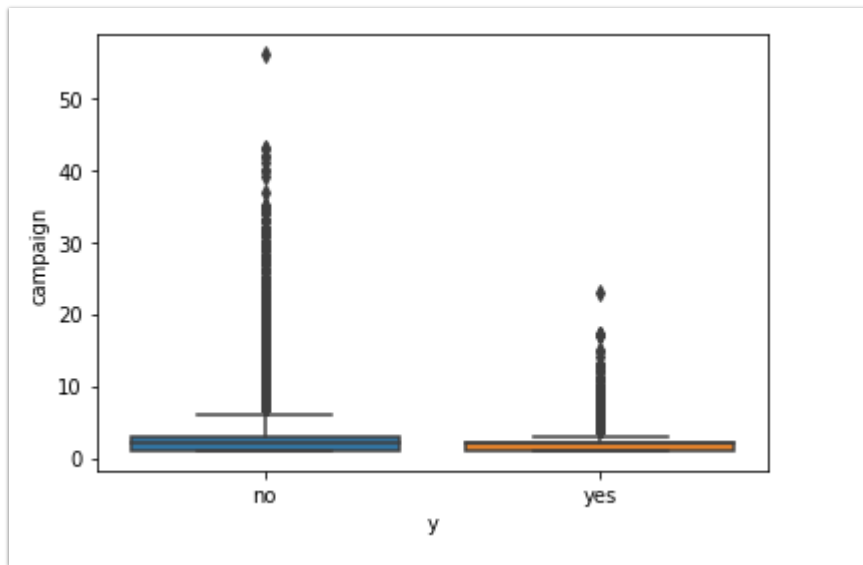
*1 number of calls made for every age section*

## status wise analysis



*2 number of calls made for every social status*

## Number of calls per campaign



*3 number of calls made in the last campaign*

## The modeling phase:

**Why the ML model:** The bank wants to use the ML model to pre-select those customers whose chances of buying the product are higher so that their marketing channel (telemarketing, SMS / email marketing, etc.) .

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

## Logisstic regression :

Logistic regression or logit model is a binomial regression model. As with all binomial regression models, it is a question of modeling at best a simple mathematical model with many real observations.

```
from sklearn.model_selection import cross_val_score

X = data.drop(['y'],axis=1).values
y = data['y'].values
X_train,X_test,y_train,y_test = train_test_split(X,y,random_state=0)
lr = LogisticRegression()
scores = cross_val_score(lr, X_train, y_train, cv=5)
scores.mean()
```

```
n_iter_i = _check_optimize_result(
```

```
Out[22]: 0.900715458132059
```

*4 cross-validation score*

## Random Forest :

This algorithm combines the concepts of random subspaces and bagging. The Decision Tree Forest algorithm trains across multiple decision trees trained on slightly different data subsets.

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score
clf = RandomForestClassifier(n_estimators=10,max_depth=None, min_samples_split=2, random_state=0)
scores = cross_val_score(clf, X_train, y_train, cv=5)
scores.mean()
```

0.8907448085813272

*RFC cross-validation score*

### **Adaboost classifier :**

An AdaBoost classifier is a meta-estimator that first fits a classifier on the original dataset, then fits additional copies of the classifier on the same dataset, but where the weights of misclassified instances are adjusted so that the following classifiers focus more on difficult cases.

```
from sklearn.ensemble import AdaBoostClassifier

cls = AdaBoostClassifier(n_estimators=100)
scores = cross_val_score(cls, X_train, y_train, cv=5)
scores.mean()
```

0.8999708753596899

*5 ABC cross-validation score*

Among the three classification models, Logistic regression performed best on this dataset.

## Generalisation error :

```
lr.fit(X_train, y_train)
lr.score(X_test, y_test)
```

```
C:\Users\Yassine\anaconda3\lib\site-packag
e (status=1):
STOP: TOTAL NO. of ITERATIONS REACHED LIM
```

```
Increase the number of iterations (max_ite
https://scikit-learn.org/stable/module
Please also refer to the documentation for
https://scikit-learn.org/stable/module
n_iter_i = _check_optimize_result(
```

```
0.8972516266873847
```

