# Bank Marketing (Campaign)

## Table of Contents



Prepared by: Fatimah Asiri

# 1. Abstract/Project Summary

Bank Marketing (Campaign)

For this project, ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which helps in understanding whether a particular customer will buy its product or not (based on the customer's past interaction with the bank or other Financial institutions).

# 2. Introduction

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

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

Develop a model with Duration and without duration features and report the performance of the model.

Based on PCA for the features, I choosed just 27 columns from dataset (job, contact, loan, default,age, bousing, balance, marital, and education).

## 3. Dataset description

Attribute Information:

Input variables:- bank client data:

1 - age (numeric)

2 - job : type of job (categorical: 'admin.','blue-collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')

3 - marital : marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)

4 - education (categorical: 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown')

5 - default: has credit in default? (categorical: 'no','yes','unknown')

6 - housing: has housing loan? (categorical: 'no','yes','unknown')

7 - loan: has personal loan? (categorical: 'no','yes','unknown')

8 - contact: contact communication type (categorical: 'cellular','telephone')related with the last contact of the current campaign

9 - month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')

10 - duration: last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then y='no'). Yet, the duration is not known before a call is performed. Also, after the end of the call y is obviously known. Thus, this input should only be included for benchmark purposes and should be discarded if the intention is to have a realistic predictive model.

11 - campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)

12 - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)

13 - previous: number of contacts performed before this campaign and for this client (numeric)

14 - poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')

15 - y: The classification goal is to predict if the client will subscribe (yes/no) a term deposit (variable y).

## 4. Methods and Algorithms.

In this project, I used several methods and Algorithms

## 5. Detailed Analysis.

The next step involves defining all the approaches, tools, and techniques used in the project. In Bank Marketing (Campaign) project report,

- First: Prepossessing step:

    Check for null and missing values

    Check for duplicate rows

    One hot encoding for ( job, marital,  education, contact)

    Label encoding for ( month, poutcome, loan, default, housing, y)

- Second:PCA

    Split Data to x and y

    Fit & transform data

    Apply PCA

Varinace Ratio

Scree Plot (Line plot chart of increasing variances)

- Third:

Build Multiple models for every model have

1. Test df  with Actual_Result and Predict_Result
2. Count values that Actual_Result == Predict_Result
3. Confusion Matrix in plot
4. Calculate Accuracy, Precision, and Recall

Logistic Regression: Accuracy: 0.95, Precision: 0.6, Recall: 0.33

Random Forest Classifier: Accuracy: 0.94, Precision: 0.6, Recall: 0.33

Decision Tree: Accuracy: 0.91, Precision: 0.6, Recall: 0.33

The Best model is (Logistic Regression) based on Accuracy, Precision, Recall

# 6. Final results

The model can predict a 0.95 form result so, In the future, I want to improve the model to increase the value and the bank in another campaign.

Employees should take the most important from clients to get the right data such as job, education, age, whether the clinet has a housing loan or not, whether the client has a loan before or not, whether the client has a credit card or not, balance,  and what is the marital status for the client.

## 7. Conclusion

After doing EDA, Feature Engineering, PCA, and Test Multiple models.

In the Future, I will be testing more models to improve the result to help the banking sector in future campaigns.

## 8. References

Data set UCI link:  https://archive.ics.uci.edu/ml/datasets/Bank+Marketing