# Bank Marketing Campaign Project Proposal

**Name: Sarah Faisal Alzighaibi**

**Email: Sarah.alnuzha@gmail.com**

Overview:

In this project our goal is to predict the number of customers who will subscribe a term of deposit in a bank. which help the marketing team with their next campaigns. Using ANN, logistic regression and xgboost. The data used data provided by Kaggel include more than 41K rows with 20 features. Trying to fulfil the below needs.

Question/need:

* In this project we planned to predict the number of clints who will subscribe for a term deposit in a bank.
* To know how many users willing to subscribe and provide the marketing team with the number of clints who will and will not subscribe to help to increase the sales
* For bank marking team and management to increase the sales and make new offers for unsubscribe customers.

Data Description**:**

* The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution from May 2008 to November 2010. The goal of this project is to predict if the number of clients will subscribe a term deposit (variable y)
* The data obtained from kaggle
* The data sent include 41188 rows with 20 features.
* The features of the data are:
  + age (numeric)
  + job: type of job (categorical: 'admin.','blue collar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')
  + marital: marital status (categorical: 'divorced','married','single','unknown'; note: 'divorced' means divorced or widowed)
  + education (categorical: 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown')
  + default: has credit in default? (categorical: 'no','yes','unknown')
  + housing: has housing loan? (categorical: 'no','yes','unknown')
  + loan: has personal loan? (categorical: 'no','yes','unknown')
  + contact: contact communication type (categorical: 'cellular','telephone')
  + month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
  + day\_of\_week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')
  + 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.
  + campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
  + 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)
  + previous: number of contacts performed before this campaign and for this client (numeric)
  + poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')
  + emp.var.rate: employment variation rate - quarterly indicator (numeric)
  + cons.price.idx: consumer price index - monthly indicator (numeric)
  + cons.conf.idx: consumer confidence index - monthly indicator (numeric)
  + euribor3m: euribor 3 month rate - daily indicator (numeric)
  + nr.employed: number of employees - quarterly indicator (numeric)
  + y - has the client subscribed a term deposit? (binary: 'yes','no')
* The data will be cleaned and check for duplication nulls and outlier, check the correlation between different features. And select the important feature to deal with.
* Apply different prediction models: ANN, logistic regression and xgboost. And check what is the best model to be used.

Tools:

In this project we planned to use

* NumPy and Pandas for data manipulation
* Scikit-learn, TensorFlow (Fast numerical computing), xgboost for modeling
* Matplotlib and Seaborn for plotting.

MVP Goal:

In MVP we are planning to provide a data set clean and optimized with the coloration between the features is provided as well as some useful plots.