# Bank Marketing Campaign project description

**Name: Sarah Faisal Alzighaibi**

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

## Abstract

This project is about bank marketing data. In this project our goal is to predict the number of customers who will subscribe a term of deposit in a commercial bank. which help the marketing team with their next campaigns. To specify the clints, age range, perfect month for the campaign and many more other useful information. Using ANN, logistic regression and xgboost. The data used data provided by Kaggle include more than 41K rows with 20 features.

## Design

Using the available features of this data set, will allow us to explore more and find the most important success factors for the bank next marketing campaign. By using machine learning modules and measure their accuracy and f1 scores we can know which module give us more reliable answers to take the write decisions.

## Data

The data set includes 41188 rows and 21 columns whit categorical and numerical features. Thes e features related to customers personnel information and their history with the bank previous deposits. For more details analysis the relation between data features was explored, and calcula ting the mutual information score (MI) we dropped the columns with 0.0 MI. furthermore, we di vide the data to train and test data 30% and 70 % To be used in the models.

## Algorithms

### Feature Engineering

1. List the categorical and numerical data.
2. Deleting the duplicated rows.
3. balanced the data.
4. Converting categorical features to binary variables.
5. Create the correlation between the target value and the other features.
6. Find the relation between the target variable and other variables.

### Models

Logistic regression, artificial neural network, and XGBClassifier were used as models with this data set.

### Model Evaluation and Selection

The entire training dataset of 4118 records was split into 70/30 train vs. test. The classification rep ort and confusion matrix was used to measure the scores accuracy, precision, recall and  f1-score.

**Logistic regression:**

- Accuracy 0.93

**Xgboost:**

- Accuracy 0.949
- F1 score 0.95
- precision 0.95
- recall 0.95

**ANN:**

- Accuracy 0.937
- F1 score 0.94
- precision 0.95
- recall 0.92

## Tools

- Numpy and Pandas for data manipulation
- Scikit-learn (logistic regression), xgboost, ANN for modeling
- Matplotlib and Seaborn for plotting

## Communication

- Slide presentation.

- Python plots.