Problem Description

The following is the story of a challenge as faced by a reputed Portuguese banking institution. They aim to determine the relative importance of each parameter with regards to their contribution as to whether a client is going to subscribe to a term deposit. Provided is a random sample of 17323 clients who have used various services offered by the bank and have been part of the previous campaign run by the bank. The response of the clients to the previous campaign run by the bank along with the clients’ information is published in the csv file named ‘Training Data’.  These clients were later asked whether they are going to subscribe to a term deposit with the same banking institution. The response of the clients is recorded in the variable ‘Y’.

The objective of this exercise is to understand which parameters play an important role in swaying a client response towards the positive scale and make the client subscribe to a term deposit. You are provided test data of 11475 clients from the same banking institution. These details were collected in the same time period.  
You are expected to predict whether a client is going to subscribe to a term deposit or not given the rest of the details.

**Dataset**

The problem consists of 2 separate datasets: Train data & Test data. The train data has information related to clients and the performance of the clients in the previous campaign run by the bank. You are expected to treat both the datasets (Train and Test) as raw data and perform any necessary cleaning/validation steps as required.

**Model Evaluation:**

Each Model will get evaluated on the basis of accuracy. The model with the best accuracy will get ranked as the best model.

**Working with Data**

Data has been split into two groups and provided in the module:

* training set
* test set

The training set is used to build your machine learning model. For the training set, we provide the whether the client has subscribed to a term deposit (also known as the variable Y) for each client.

The test set should be used to see how well your model performs on unseen data. For the test set, it is your job to predict the outcome variable (Y) for each client.

This is an example of what your submission file should look like:

Data Dictionary

The data dictionary is attached here.