**Assignment 2 on Machine Learning**

**Problem Statement*:***

**“Who will subscribe to the term deposit?”**

A Portuguese Banking Institution needs to understand which of their existing customers are most likely to invest in a term deposit. They launched a tele-marketing campaign to run an experiment on approx.. 41000 customers and analyse the differences between the ones who do opt for a TD and the ones who do not.

**Apply following classification algorithms:**

* Logistic Regression
* KNN with suitable K
* Decision Tree
* Random Forest
* AdaBoost
* Gradient Boost
* XGBoost

**Tasks:**

* Apply data cleaning and/pre-processing techniques, if necessary.
* Apply normalization technique, if needed, wherever applicable.
* Check if the dataset is imbalanced. If required, apply the appropriate technique.
* Use Hyperparameter tuning if required.
* Compute appropriate evaluation metrics.
* Display confusion matrix
* Draw ROC curve
* Compare the performance of the models.

**Documents for submission for evaluation:**

**Send files in html and ipynb format, detailed analysis should be written.**

**Only code and output not accepted**. File name should be **yourname\_bank\_ML.**

**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')

**# related with the last contact of the current campaign:**

8 - contact: contact communication type (categorical: 'cellular','telephone')  
9 - month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')  
10 - day\_of\_week: last contact day of the week (categorical: 'mon','tue','wed','thu','fri')

**# other attributes:**  
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')

**# social and economic context attributes**  
15 - emp.var.rate: employment variation rate - quarterly indicator (numeric)  
16 - cons.price.idx: consumer price index - monthly indicator (numeric)  
17 - cons.conf.idx: consumer confidence index - monthly indicator (numeric)  
18 - euribor3m: euribor 3 month rate - daily indicator (numeric)  
19 - nr.employed: number of employees - quarterly indicator (numeric)  
  
**Output variable (desired target):**  
**20 - y - has the client subscribed a term deposit? (binary: 'yes','no')**