**Term Insurance Customer Prediction**

**Problem Statement**

The data is related with direct marketing campaigns (phone calls) of an insurance institution.

The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (term insurance) would be ('yes') or not ('no') subscribed.

You will learn how to apply Decision Tree, Random Forest/XGBoost/Adaboost to credibility of the customer. Also learn how to evaluate Random Forest/XGBoost/Adaboost model using various parameter like on Accuracy, Sensitivity, Specificity and area under the ROC curve.

Build Decision Tree, Random Forest/XGBoost/Adaboost models to predict if the client will buy a new insurance plan.

**Data Description**

* **Age:** Age of each customer
* **Job:** 0=Admin, 1=Blue-collar, 2=Entrepreneur, 3=Housemaid,

4=Management, 5=Retired, 6=Self-employed, 7=Services,

8=Student, 9=Technician, 10=Unemployed, 11=Unknown

* **Marital:** 0=Divorced, 1=Married, 2=Single, 3=Unknown; note: Divorced means divorced or widowed
* **Qualification:** 0=Basic, 1=High school, 2=Illiterate, 3=Professional course, 4=University degree, 5=Unknown
* **Default\_Premium:** has premium in default? (0=No, 1=Yes, 2=Unknown)
* **Health\_Insurance:** has health insurance? (0=No, 1=Yes, 2=Unknown)
* **General\_Insurance**: has general insurance? (0=No, 1=Yes, 2=Unknown)

**Related with Last Contact of The Current Campaign**

* **Contact:** Contact communication type (0=Cellular, 1=Telephone)
* **Month:** Last contact month of year (0=Apr, 1=Aug, 2=Dec, 3=Jul, 4=Jun, 5=Mar, 6=May, 7=Nov, 8=Oct, 9=Sep)
* **Last\_Contact\_Day:** Last contact day of the week (1=Mon, 2=Thu, 3=Tue, 4=Wed, 0=Fri)
* **Last\_Contact\_Duration:** Last contact duration, in seconds (numeric). Important note: this attribute highly affects the output target (e.g., if duration=0 then Client\_Subscribed='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.

**Other Attributes**

* **Contacts\_During\_Campaign:** Number of contacts performed during this campaign and for this client (numeric, includes last contact)
* **Previous\_Contact\_Days:** Number of days that passed by after the client was last contacted from a previous campaign (numeric; 999 means client was not previously contacted)
* **Contacts\_Before\_Campaign:** Number of contacts performed before this campaign and for this client
* **Previous\_Outcome:** Outcome of the previous marketing campaign (0:Failure, 2:Nonexistent, 3:Success)

**Social and Economic Context Attributes**

* **Employment\_Rates:** Employment variation rate - quarterly indicator
* **Price\_Variation:** Consumer price index - monthly indicator
* **Consumer\_Confidence\_Index:** Consumer confidence index - monthly indicator
* **Interest\_Rate:** Interest Rate - annual percentage rate
* **No\_Employees:** Number of employees - quarterly indicator

**Output Variable (desired target)**

Client\_Subscribed - has the client will buy a new insurance plan? (0:No, 1:Yes)

**Evaluation Parameters**

Evaluation will be based on:

* Data Preparation
* Model Comparison
* Model Selection

**Data Preparation**

Analyze the data statistically and treat the multicollinear variables.

**Model Comparison**

Apply Decision Tree, Random Forest/XGBoost/Adaboost algorithms for every change made in the datasets and compare results.

**Model Selection**

Select the best model. Model selection to be based on Accuracy, Sensitivity, Specificity, F1 score, and area under the ROC curve.

**Expected Outcome**

Higher AUC value and F1 Score in predicting the outcome using test data.