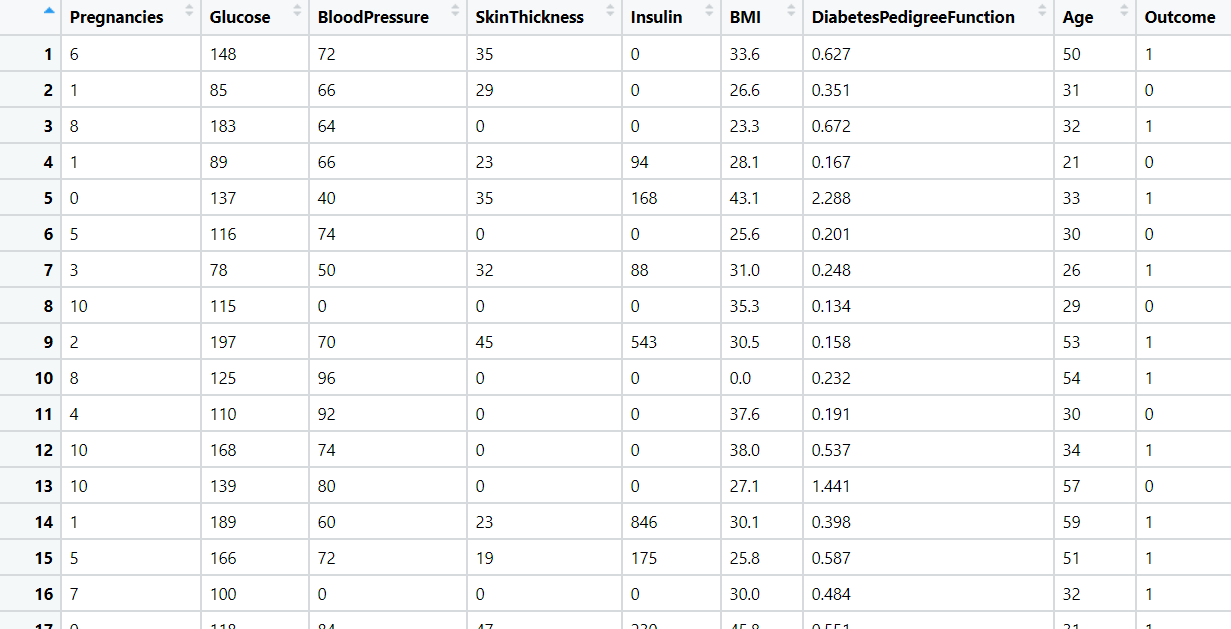
# Topic: AdaBoost- Extreme Gradient Boosting

Perform AdaBoost and Extreme Gradient Boosting for the following dataset

**1.) Diabetes\_RF.csv**



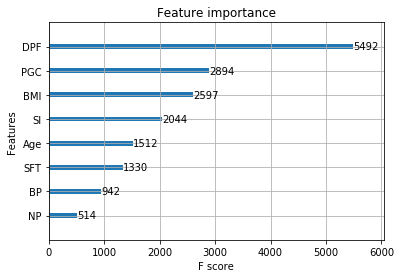
**Ans:**

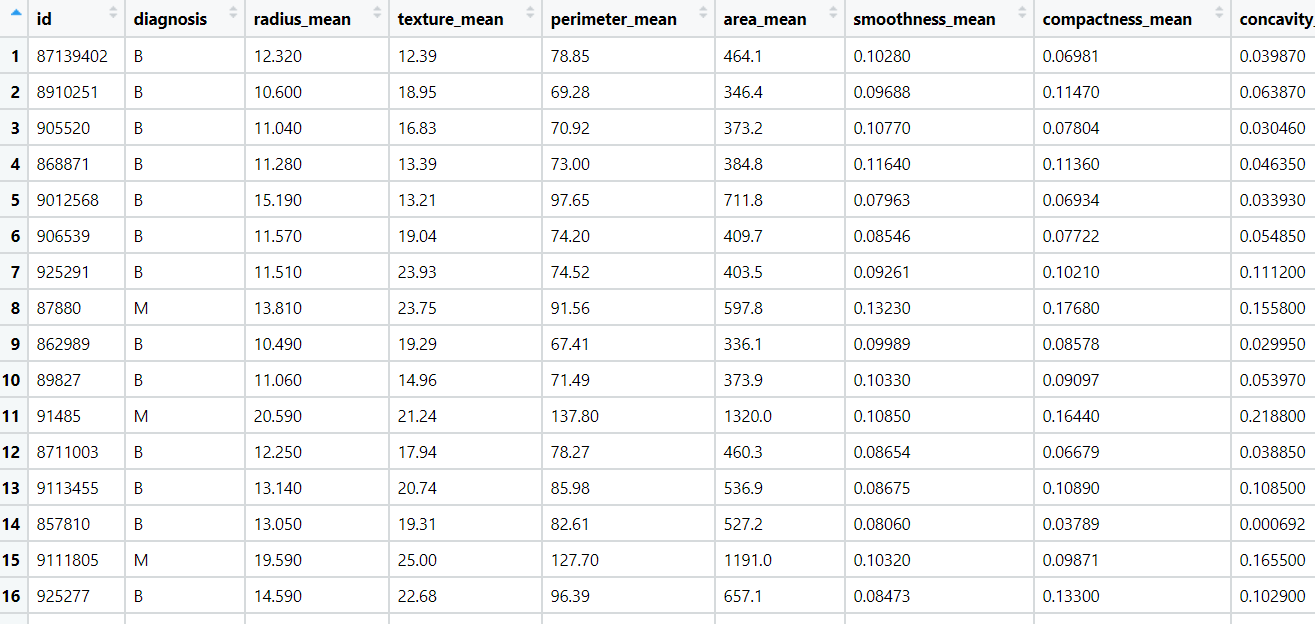
Data Analysis:

Output Variable = Class Variable

Input Variable = Other factors with influence the class variable

* Set.seed() function is used to keep the test and train data constant on over all analysis
* The data is divided into 80% train and 20% test
* The output variable is converted as factor
* Adaboost technique is applied on both test and train data to find the model accuracy
* The train and test accuracy of the model is 1.0 and 0.727 which show the model is overfit, need to do regularization method to make the model as right fit
* In python the train and test accuracy are 0.84 and 0.80 which shows the model is right fit
* Applying Xgboost technique to see the accuracy of the train and test
* Here both the test and train accuracy are 1 which still shows the model is not right fit need to do regularization methods to make the model right fit.



**2.) wbcd.csv**

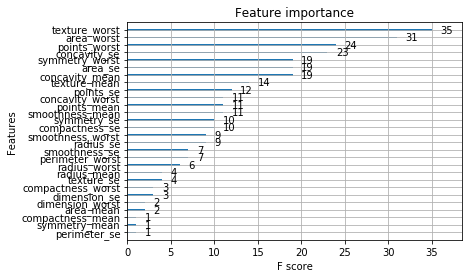
**Ans:**

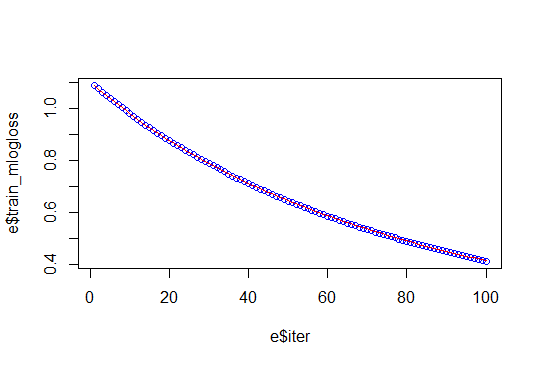
Data Analysis:

Output Variable = Diagnosis

Input Variable = Other factors with influence the class variable

* Removing unwanted columns from the dataset
* Set.seed() function is used to keep the test and train data constant on over all analysis
* The data is divided into 80% train and 20% test
* The output variable is converted as factor
* Adaboost technique is applied on both test and train data to find the model accuracy
* The train and test mean of the model is 0.97 and 0.97 which show the model is right fit
* By tweaking the gamma, eta, maxdepth, subsample the right fit model is achieved.
* In python the train and test accuracy are 1.0 and 0.96 which shows the model is right fit
* Applying Xgboost technique to see the accuracy of the train and test
* Here both the test and train accuracy are 1 which still shows the model is not right fit need to do regularization methods to make the model right fit.



****

**Hints:**

1. Business Problem
   1. Objective
   2. Constraints (if any)
2. Data Pre-processing

2.1 Data cleaning, Feature Engineering, EDA etc.

1. Model Building
   1. Partition the dataset
   2. Model(s) - Reasons to choose any algorithm
   3. Model(s) Improvement steps
   4. Model Evaluation
   5. Python and R codes
2. Deployment

4.1 Deploy solutions using R shiny and Python Flask.

1. Result Share the benefits/impact of the solution - how or in what way the business (client) gets benefit from the solution provided.

**Note:**

1. For each assignment the solution should be submitted in the format
2. Research and Perform all possible steps for improving the model(s) accuracy Ex: Feature Engineering, Hyper Parameter tuning, etc.
3. All the codes (executable programs) are running without errors
4. Documentation of the module should be submitted along with R & Python codes, elaborating on every step mentioned here