**Ensemble Techniques, Nov 2018 (DSE-Bangalore)**

**Faculty: Mahesh Anand. S**

1. Load the data (Parkinson.csv) file and perform the following task,
2. Refer to this link, https://archive.ics.uci.edu/ml/datasets/parkinsons and read the description of the dataset and attach suitable headers.
3. Status is the outcome variable, with ‘0’ – healthy and ‘1’ – Parkinson Disease. Perform EDA and apply suitable algorithm (pick among Logistic Regression, Naive’s Bayes, KNN), which is giving best score with 70/30 validation split with precision and recall score, ROC curve and AUC value. Finally, choose the best model out of 3.
4. Optimize the model (best model picked in ‘b’) with boosting technique and observe the validation score.
5. Comment on (c), whether boosting is improving the model performance or not.
6. Optimize the model (best model picked in ‘b’) with bagging technique and observe the validation score.
7. Comment on (e), whether bagging is improving the model performance or not.
8. Compare and Contrast both bagging and booting technique over the best model selected during the step ‘b’ with Bias and Variance Error using 5-fold Cross Validation Score.
9. Give your final inference, which optimization technique (Bagging/Boosting) would be chosen while the best model picked in step ‘b’, is going for production system.