

Project Deliverable 1

For this assignment each team needs to provide 3 items: a written report in pdf format, predictions (response values) for the test data in a csv format and working code in python notebook format.

1. A report must include the following sections (no more than 4 pages in total):
 - a. Project description (details of the given data set, objective of the project, etc.)
 - b. Data visualization (descriptive analysis of data mostly consisting of visuals/plots)
 - c. Data pre-processing and data splitting techniques (provide proper justification/explanation for each technique utilized in your project)
 - d. Description of the model(s) used for analysis and provide justification for the following:
 - a. model selection
 - b. resampling method(s) utilized during training
 - c. performance metric used to gauge the prediction accuracy of the model
 - e. Visual representation of results: confusion matrix, ROC curve, accuracy and other plots
 - f. Conclusion: thoughts about why the selected model produces such results
2. Obtain the best model on the training data set and then apply that model on the test data set.
 - a. Report the predicted (\hat{y}) values for the samples in the test data set in a csv format.
3. A working code should be in python notebook format and satisfy the following criteria:
 - a. The code should run without any modification from the TA/Instructor.
 - b. The code should include comments describing each of the steps.