**Cross validations used for our boosting model:**

**KFold:** We have used k fold validation where, according to online documentation, the model is trained on each iteration at a proportion of the data set equal to (k-1)/k. There are k training iterations in the algorithm,

**Pro:** All observations are used for both training and validation. Each observation is used for validation exactly once. Therefore, there is proper representation of data in this technique.

**Con:** Expensive for large N, K since we train/test K models on N examples.

**ShuffleSplit:** Here, the model is trained at each iteration on a defined train\_size. The number of iterations is parameterized (n\_splits).

**Pro:** The process is quick. Training time is reasonable since ShuffleSplit can be configured to run with less demanding coverage

**Con:** Random splits do not guarantee that all folds will be different. Some observations may not be present in the test set. Therefore, the actual characteristics of the dataset in under-represented in this validation technique.