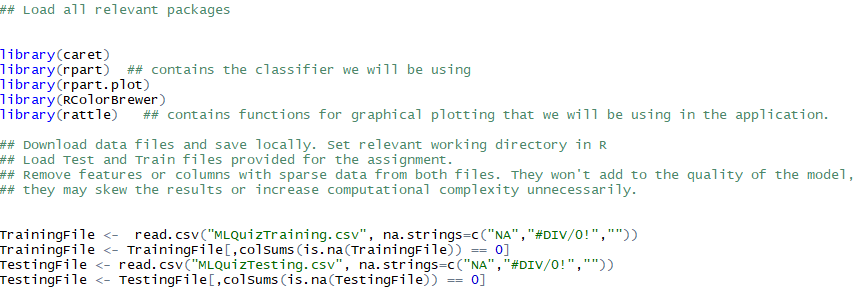
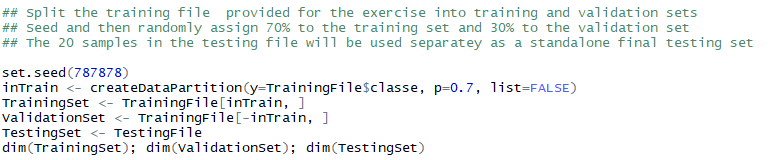
**Final Assignment for Practical Machine Learning course**

Load all relevant libraries and data files

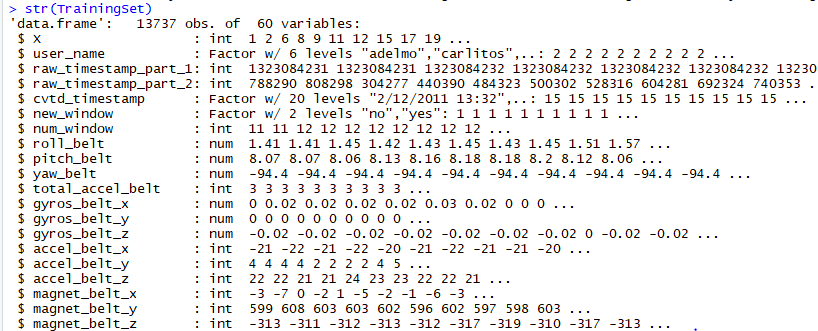


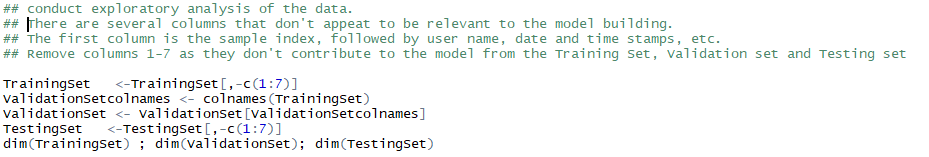
Split Training file into Training and Validation data sets





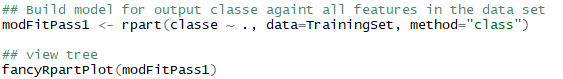
Remove unnecessary columns that are not reflective of features for the model building exercise. This will also prevent overfitting. Remove first 7 columns for all 3 data sets (Training, Validation and Testing)

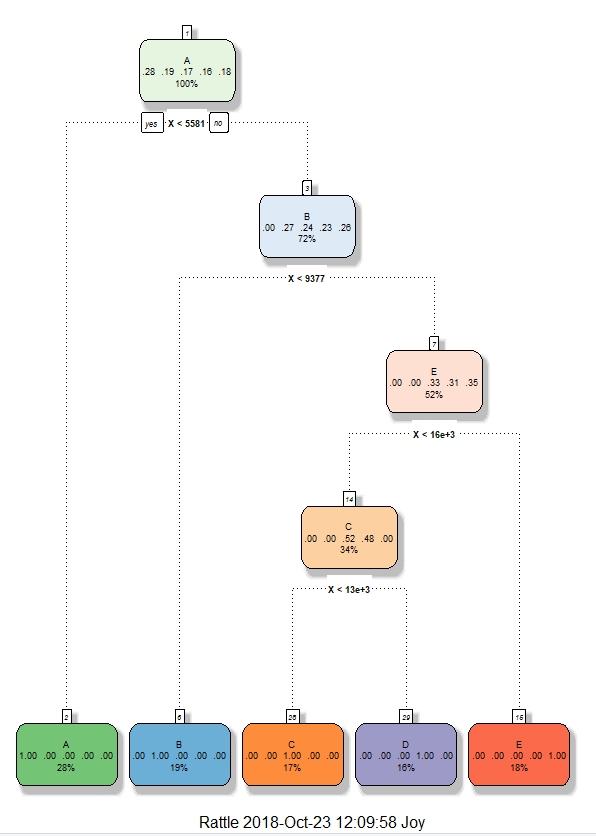






Build model and plot outcome

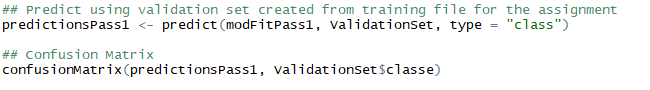




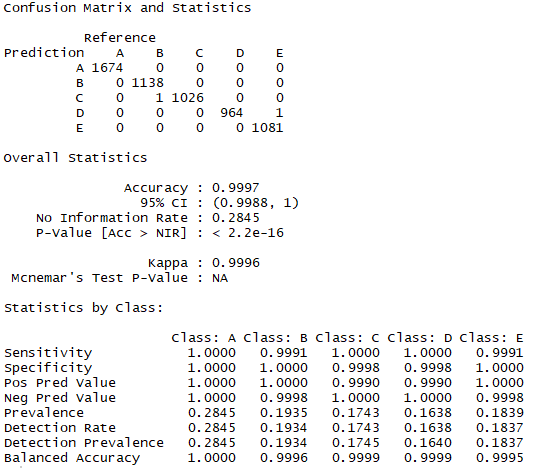
Clean tree structure with A-28%, B-19%, C-17%, D-16%, E18%

Given a strong model fit, with steps being taken to prevent overfitting this model is expected to perform well against the validation data set.

Run Prediction against validation set and analyze outcome via ConfusionMatrix



Very high accuracy outcome of the validation set right off the bat with only 2 items out of 5885 cases wrongly classified. No further models need to be test. Performance against any test set expected to be good.



Final Prediction

