**Exercise: Predicting Current Alcohol Consumption from Behavioral Scores**

**Dataset Description:**

These data were collected as part of an online survey related to drug and alcohol use and personality traits. Individuals answered standardized questions which were used to calculate continuous scores on personality traits. Individuals were also asked about consumption of alcohol and multiple drugs. Further information on this dataset can be found at [http://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29#](http://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29).

For the purpose of this exercise, the data has been subset to include only 7 features on personality traits and the variable which distinguishes those who reported current alcohol use (defined as alcohol use in the past month or more frequently) vs no current use. Data are stored in the csv file alcohol\_use.csv on the course site.

**Feature Information:** Below is a list of the 7 features and outcome variable within the dataset. In general, the higher value of the score, the greater the personality trait observed within the individual based on the questionnaire.

1. alc\_consumption: CurrentUse, NotCurrentUse
2. neurotocism\_score: Measure of Neuroticism
3. extroversion\_score: Measure of Extroversion
4. openness\_score: Measure of Openness to Experiences
5. agreeableness\_score: Measure of Agreeableness
6. conscientiousness\_score: Measure of Conscientiousness
7. impulsiveness\_score: Measure of Impulsivity
8. sens\_seeking\_score: Measure of Sensation-Seeking Behaviors.

Instructions for Exercise:

Goal: You want to predict current alcohol consumption but it is expensive and time-consuming to administer all of the behavioral testing that produces the personality scores.

1. Conduct a reproducible analysis to build and test a classification model using the approach of your choice.
2. You should create and compare at least two different models, but the choice of variables and algorithm to utilize are up to you. Evaluate the models using cross-validation and obtain final metrics for the selected model in the test set.
3. Produce a shareable report of your analysis and results using R Markdown.

Considerations:

1. What research questions could this analysis either a) directly address or b) indirectly help to address by providing information that could be used in subsequent analyses?