ASSIGNMENT 3
ADVANCED METHODS AND ANALYSIS FOR THE LEARNING AND SOCIAL SCIENCES
PROFESSOR RYAN S.J.d. BAKER
BEHAVIOR DETECTION
DUE NOON, MONDAY FEBRUARY 13

The goal of this assignment is to build a behavior detector (classifier), using the data in Asgn3-dataset.csv

This data set involves a set of clips, generated from field observations synchronized with log files. You must build a detector of the behavior OffTask (e.g. a detector that can predict if the column OffTask is Y or N). You should make sure that your detector is not over-fit, paying particular attention to making sure that your detector does not use features that could not be used when applying the model to new data or new students. This can be done both by restricting the features used during model fitting, and setting up cross-validation in an appropriate fashion. (Hint: Try Batch Cross-Validation).

You must build the detector using an automated algorithm. You cannot simulate the algorithm in Excel. You can use any data mining package (e.g. SAS, R, Weka, KEEL) you want, but I strongly recommend using RapidMiner 4.6.

Please turn in:

- The data set you input into the data mining package, if different than the original data set
- The model built on the full data set
- Evidence of model goodness, when the model is applied to new students (see the Diagnostic Metrics lecture)
- All data mining code you used to generate the outputs
- A document explaining how you completed the assignment

You will be graded on completeness and comprehensibility of your hand-in, whether you correctly and validly apply the method you choose to this data, and whether the methods you chose fit the requirements of this assignment.

BONUS: The student who succeeds in producing the detector with the best A' and Kappa (averaged together) under appropriate cross-validation, gets the bonus. For this assignment, automated feature selection and creation is allowed, but creating new features by hand in Excel is not allowed.