CREATIVE ASSIGNMENT 1
CORE METHODS IN EDUCATIONAL DATA MINING
PROFESSOR RYAN BAKER
BEHAVIOR DETECTION
YOUR ASSIGNMENT IS DUE SEPTEMBER 18, 1159PM USA EASTERN
YOUR RESPONSE POSTS ARE DUE SEPTEMBER 21, 1159PM USA EASTERN

PART ONE: YOUR ASSIGNMENT

The goal of this assignment is to build a behavior detector (classifier), using the data in ca1-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.

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, RapidMiner, Python, Orange, Knime) you want.

Please post to the forum, in a new thread within the CA1 folder:

- Text explaining how you completed the assignment
- Evidence of model goodness, when the model is applied to new students (use Kappa at minimum)
- Along with files
    - The data set you input into the data mining package, if different than the original data set
    - All data mining code you used to generate the outputs
    - The model built on the full data set (can be in any format that I can read)

Solutions will be graded on completeness and comprehensibility, 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 Kappa, under appropriate cross-validation, gets the bonus.

PART TWO: YOUR RESPONSE POSTS

After completing your own assignment, you are expected to also provide substantive comments on at least four other students' submissions, as a response within that student's assignment thread. For these posts, there is no length requirement, but the posts must offer a critical and meaningful perspective on how that student did the assignment. (i.e. "Great job! You did really awesome!" and "Terrible! You totally messed up!" are insufficient)

This is not just for the benefit of the student whose solution you are commenting on. Seeing how other students did this assignment will be informative to you as well.

Although there is no requirement to do this, you are encouraged to give feedback to students who have received fewer feedback responses so far – i.e. I would like to avoid having one student get feedback from every classmate, and another student get feedback from no one. Thanks.