Methodology

1) A description of the method you plan to employ to analyze your data

Conduct inferential statistics (i.e. chi-square test of independence) for each feature to see the trend and to have an insight about how each feature correlates with enrollment. Then incorporate all features to create a prediction model. By creating a decision tree model, we can, not only predict the possibility of a person enrollment decision, but also learn what features are more predictive than the others.

2) A list of assumption “gotchas” or other considerations you need to be aware of to use that method

Which type of decision tree model to use, and what is the order/hierarchy of the features are the main concerns I have when making the analysis.

3) How you plan to validate both model and constructs

I ‘m planning to use cross validation to validate the model, specifically group samples into several groups, and test the model on the testing group.

Predictive variables in the dataset include employee’s demographic information and working status. Target variable (enrollment) is used to represent staff’s motivation for online studying, due to the voluntary and self-paced nature of the online platform. Number of classes a staff is enrolled and completed can be used as a way to prove construct validity.