Outline

Intro

* Table 1 is general summary of data, showing percent of people who get addicted
* Table to show statistics on difference sampling techniques plus prob cutpoint technique
* Plot of ROC for various methods

Methods

* Temporal split training and test
* Keep it simple for clinical colleagues with regression
* Options involve stepwise selection, can change to not have p value cutoff but AIC or clinically meaningful vs using default p value
* If doing sampling technique and standard stepwise selection how that compares
* Not getting too complex but not just going with defaults
* Sampling: up, down and smote (for each what are the number of variables left in model, not specific coefficients)
* Look into changing defaults on sampling for smote and down

Results

Discussion

Conclusions

Where to publish: BMC?, BMC bioinformatics, Artificial intelligence in medicine