

Deviations from Preliminary Report

The only minor deviations from the original plan are a change in the final dataset and a change in the set of statistical methods that will be implemented for the assignment.

Originally, the dataset was to have multiple response options that we would then pick and choose between different statistical methods. Only the broadest-most option allowed for an acceptable level of prediction power. As a result, the other response options were removed from the dataset and the accompanying help file.

There was also a change in the number of levels of our response variable. The level "Treatment" was removed from our Broad Category response as there were not enough records in this group to provide predictive power for these observations. This was done by removing the observations in the dataset that correspond to that level of response, reducing the total number of records from 293 to 285.

The change in statistical methods was that we decided against using multinomial logistic regression for this data as it proved to be fairly difficult to implement. We saw significant success in some of the other learning methods. One of the benefits of logistic regression (interpretability) is also lost in this context, as LSA makes the results very hard to interpret. The plan of using Neural Networks was decided against for the same reason: we were unconvinced that we would see significantly better results from this method, and the neural networks would be very hard to implement in our given time constraint.