

# Progress Report I

Writer: Eric Foote

November 9, 2018

Supervisor: Dr. Stewart

Since the last progress report, I have finished up making all of the models that I stated were going to be applied. In total I ended up making two linear regression models using AIC variable selection, one forward selecting and one considering both forward selection and backward selection. Backward selection was also explored but it had the same AIC output as the both direction so no new interesting model was created. I made another linear model using the  $R^2$  values in the regression table of the full 27 variable model. Another linear model was constructed using the variables remaining from the forward selecting model and those with good  $R^2$  values. Then I took a look at building models based on ridge regression since I had more observations than variables rather than lasso and created two; one which I did not choose any lambda values and a second which I chose two lambda values based on the minimum error and one standard error of the minimum error. Following that I looked into applying logistic regression and had to change the problem rather than predicting points I looked into instead predicting creating a categorical variable which looked at whether or not a team passed the 45 win threshold. This value was chosen due to usually this is the threshold of a playoff team. I made two models for this new problem one with the lowest AIC value in the backward selection this time and a second using the variables from the forward selection in linear regression case. The reason for this was when forward selection or both were looked at both cases would default to  $y = 1$  so I decided to look at both. Finally I made four trees; one regression tree, one recursive partition tree, one gradient boosted decision tree and a random forest. Also during my exploration into trees; I learned why the predict function was not working for me in the linear regression section, it was due to when I was creating the models I would call the column in the dataframe using `$` and the predict function would search the new data for say `x$wins` instead of wins. I figured this out when reading some of the documentation on trees.

The remaining things to be done are to make any final modifications, tweaks to the models, write the rough draft and follow up on any suggestions before the final copy and the presentation