Kaggle Writeup:

For the Kaggle competition, I worked alone as a team of one. I considered several types of models. Models I experimented with and tuned included the following:
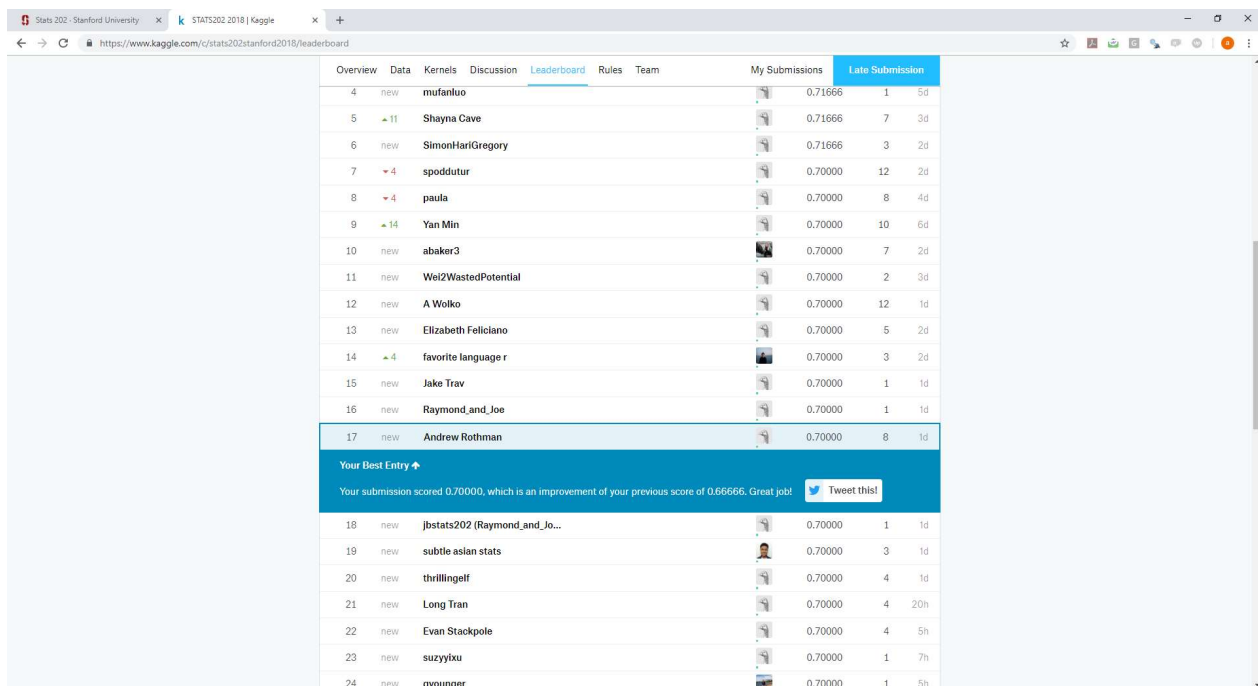
- Logistic Regression
- Regularized Logistic Regression (LASSO, Ridge, Elastic Net)
- Tree methods (Bagging, Random Forests, Boosting)
- Generalized Additive Models
- Support Vector Machines

I also took the initial 10 features available in the training dataset and expanded those potential features considerably. For each of the original 10 features, I created:

- second order polynomial terms
- Two-way interaction terms between every one of the 10 features

With the original 10 features included, this created a total of 57 possible features to choose from.

Using 10-fold cross-validation on the training set, from the array of different model types experimented with, I found the lowest cross-validated error I was able to achieve was with a Random Forest with 1000 trees and the "mtry" parameter set to 25. Therefore this was the best model I specified and submitted to the Kaggle Competition. With this model I achieved a score of 70%, and am currently ranked 17th on the public leaderboard. I have included a screenshot below: