

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:

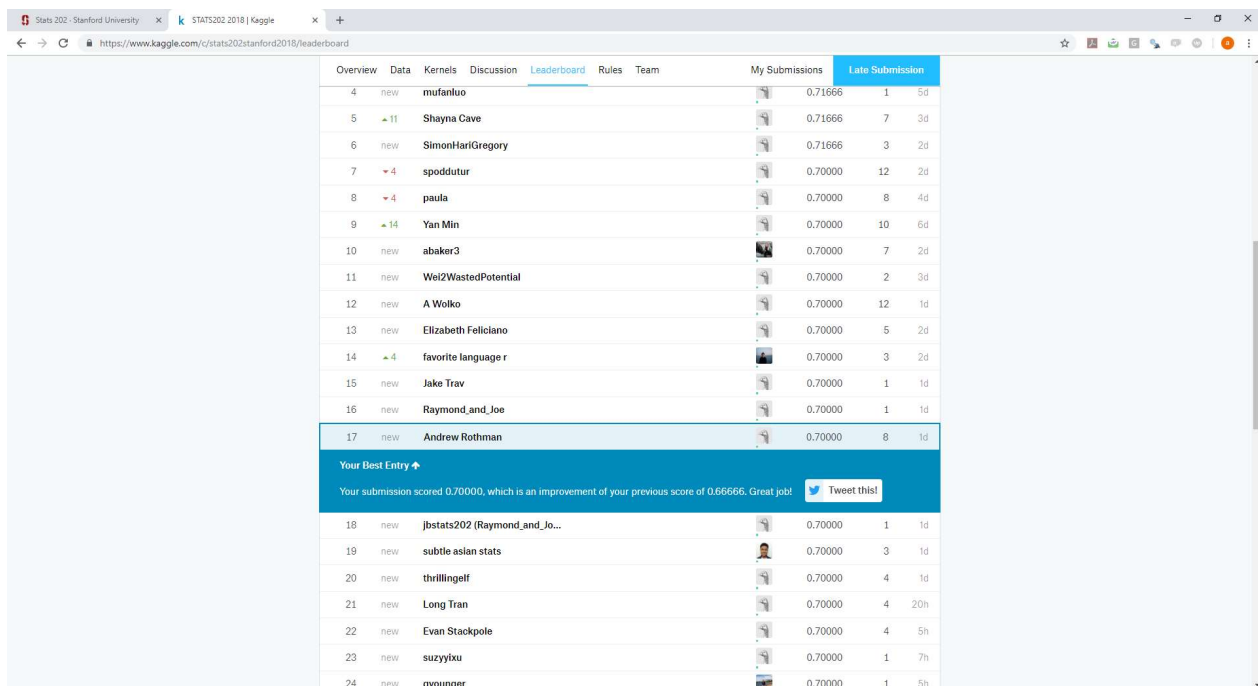
- Logistic Regression
- Regularized Logistic Regression (LASSO, Ridge, Elastic Net)
- Tree methods (Bagging, Random Forests, Boosting)
- Generalized Additive Models
- Support Vector Machines

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:



The screenshot shows the Kaggle Stats 202 leaderboard. The table lists submissions with their rank, status, author, score, and time. The author 'Andrew Rothman' is highlighted at rank 17 with a score of 0.70000. Below the table, a message states: 'Your submission scored 0.70000, which is an improvement of your previous score of 0.66666. Great job!'. The table continues with ranks 18 through 24.

| | Overview | Data | Kernels | Discussion | Leaderboard | Rules | Team | My Submissions | Late Submission |
|---|----------|-------------------------------|---------|------------|-------------|-------|------|----------------|-----------------|
| 4 | new | mufanluo | | | | | | 0.71666 | 1 5d |
| 5 | new | Shayna Cave | | | | | | 0.71666 | 7 3d |
| 6 | new | SimonHartGregory | | | | | | 0.71666 | 3 2d |
| 7 | new | spoddutur | | | | | | 0.70000 | 12 2d |
| 8 | new | paula | | | | | | 0.70000 | 8 4d |
| 9 | new | Yan Min | | | | | | 0.70000 | 10 6d |
| 10 | new | abaker3 | | | | | | 0.70000 | 7 2d |
| 11 | new | Wei2WastedPotential | | | | | | 0.70000 | 2 3d |
| 12 | new | A Wolko | | | | | | 0.70000 | 12 1d |
| 13 | new | Elizabeth Feliciano | | | | | | 0.70000 | 5 2d |
| 14 | new | favorite language r | | | | | | 0.70000 | 3 2d |
| 15 | new | Jake Trav | | | | | | 0.70000 | 1 1d |
| 16 | new | Raymond_and_Joe | | | | | | 0.70000 | 1 1d |
| 17 | new | Andrew Rothman | | | | | | 0.70000 | 8 1d |
| Your Best Entry ⬆ | | | | | | | | | |
| Your submission scored 0.70000, which is an improvement of your previous score of 0.66666. Great job! Tweet this! | | | | | | | | | |
| 18 | new | jbstats202 (Raymond_and_Jo... | | | | | | 0.70000 | 1 1d |
| 19 | new | subtle asian stats | | | | | | 0.70000 | 3 1d |
| 20 | new | thrillingelf | | | | | | 0.70000 | 4 1d |
| 21 | new | Long Tran | | | | | | 0.70000 | 4 20h |
| 22 | new | Evan Stackpole | | | | | | 0.70000 | 4 5h |
| 23 | new | suzyyixu | | | | | | 0.70000 | 1 7h |
| 24 | new | gyounger | | | | | | 0.70000 | 1 5h |