Springboard Introduction to Data Science
Capstone Project
4-10-18

In American Football, many people have tried to predict the winners of football games versus the spread (the spread is a handicap given by oddsmakers in a particular match-up when one team is better than the other). Very few have had success consistently over the span of several years. It is especially difficult to do so without taking into account outside factors such as weather, injuries and team momentum. I have come up with a method of rating teams that might be able to calculate winners of games and want to test out the theory.

Being able to pick winners and losers in any sport is a very lucrative business. There are tv shows and websites devoted to helping people make gambling decisions. If this calculation is successful, I would say that the main stakeholder would be me, as I might be able to take this information and share it with others to help them make decisions on football games. I would possibly create a blog or YouTube channel to share this information for a price.

The data being used for this project will come from the following sources:

- Proprietary ratings data that I own. I will only be sharing the outcomes of the calculations and not the ratings calculations themselves.
- The data used to generate the ratings were obtained on www.cfbstats.com
- Matchups and outcomes for college football games since 2010 were obtained on www.espn.com
- Average scores for football teams over the course of each year was obtained from www.ncaasports.com

Our goal in this exercise is to be able to predict games with a spread at a rate of more than 52.4%. This is an industry benchmark considered “breakeven.” For the calculations, there will be two approaches: 1. Use the ratings themselves as a comparative for each team in a particular matchup. 2. Run a regression on the average rating over the 8 years of analysis versus the average score (offense and defense) and use the subsequent equation to turn the ratings into a potential score and see if that make a difference in predicting.

Deliverables will be: Code, results, analysis, presentation, and report detailing the exercise and results.