Capstone 2 Project Proposal

I am going to create a model to predict winners of NFL games against the Las Vegas spread. My client will be an NFL franchise. Hopefully the model can pick winners against the spread and offer insight as to why the spread was wrong. By doing this it could reveal under or overvalued traits of an NFL team or an NFL game.

I am going to use data from pro-football-reference.com. After downloading all the data individually from games, the data will be combined and broken down by play type, along with other features. By breaking down each team’s performance by play type (pass short left, pass short middle, pass short right, pass deep left, pass deep middle, pass deep right, run left end, run left tackle, run left guard, run center, run right guard, run right tackle, run right end, and middle plays) offensively, defensively, and their completion percentage (complete pass or run of 3 yards or greater), I hope to quantify advantages or disadvantages in this realm that others have not revealed.

For example, I believe that a team that is strong running the ball to the left should have a notable advantage against a team that is weak defensively against runs to the left.

Additionally, columns will be created that represent the difference between the offense’s completion percentages vs. the defense’s completion percentages by play type. Columns representing the home team’s time zone, the away team’s time zone, the time zone at the location of the game, the time after their previous game, and whether it is a division game or not will be added as well. Other columns will likely also be created.

Vegas spreads are notoriously accurate, and thus creating a model slightly above 50% could be considered useful. Hopefully using extensive play data along with other, more standard data, this model will be able to consistently outperform the Vegas spreads. I will test a random forest classifier, a logistic regression model, and a deep learning model to see which is most accurate.

I will include code for the data and the model, along with sample predictions. A presentation of how the data was collected, new columns created, the model, and how to use the model will be given.

One example of how to use the model:

The model finds that teams that are weak against runs to the left and often do not see them often do worse than expected (worse than the Vegas line would suggest) against a team that is strong running the ball to the left. When a team is playing another team that is weak against runs to the left in the following week, emphasis should be placed on running the ball to the left during practice or calling run plays to the left to exploit that team’s weakness and accentuate their own strength.