What makes a Super Bowl Winner

IST718 Final Project

By Darrell Nelson,Taylor Moorman,Eric Yager,Sarena Zhao

**Overview**

There are a variety of different factors that go into determining a Superbowl winner, and the overall placing of every team in the playoffs. In addition, normal use cases of machine learning algorithms were rendered insufficient in this analysis because of the absence of an unlimited number of classes for each observation. Overall, it is very difficult to outright predict the placing of all teams, although it is much easier to predict the two teams who will be in the Superbowl.

Our recommendations from this analysis fall into two categories, for NFL teams and for gamblers. For NFL teams our model gives them an opportunity to projects where they may end up on the onset of the playoffs. They also may be able to potentially focus their gameplans to cover up statistical holes in their teams. For gamblers, our modeling gives them a reasonable guess as to who will appear in the Super Bowl and who may potentially win it.

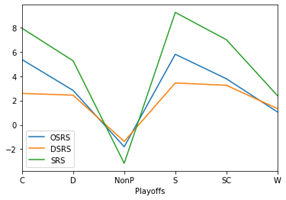
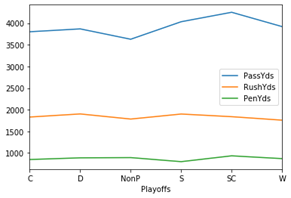
**Specification**

The NFL is inherently a very difficult game to predict. On average the better team wins, however, with so many game-changing plays and so little room for error, there can be huge swings in win probability at the drop of a hat. Furthermore, when zooming out to predict an entire season, so many things can go right or break wrong for a team. If one looks at overall season statistics there may be trends that highlight top-performing teams in distinct statistical areas, but which ones are the most important for getting to the end goal of the NFL, the Super Bowl? What combination of skills, performance, and coaching maximizes the chances of winning a Super bowl? What is the best winning formula? Our team of analysts will seek to crack the code of what constitutes a Super Bowl winner. Armed with this knowledge any team can go from worst to first and take home a bigger slice of the NFL revenue pie while basking in all the glory and immortality a Super bowl championship brings to their city.

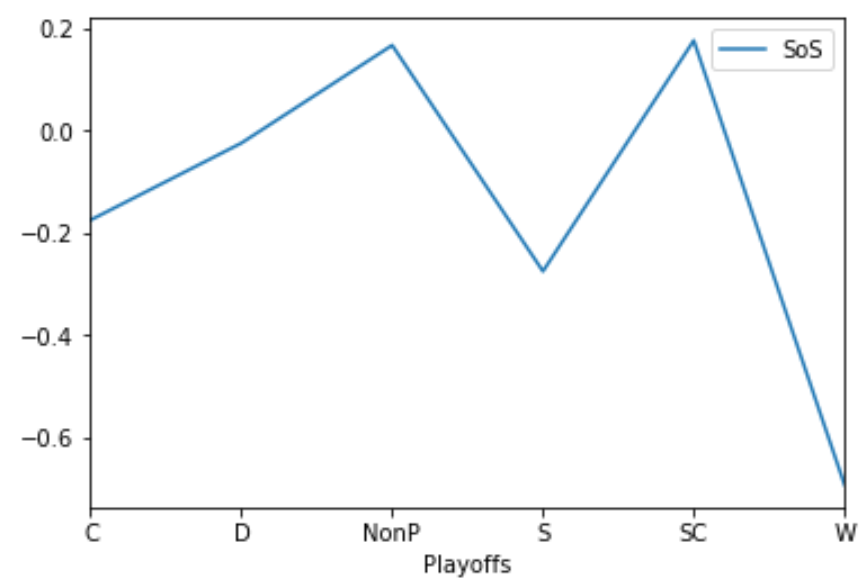
Our group has decided to investigate over 30 offensive and defensive metrics of all 32 NFL teams over the past 10 years to see if there is a winning formula for surviving the regular season, playoffs, and ultimately winning the big game. All of the offensive/defensive metrics will be taken from Pro Football Reference (www.pro-football-reference.com) and exported to a CSV. All final team rankings over the past 10 years will be taken from ESPN (www.espn.com) and using team names and year as a primary key will join the two datasets. Models will be developed on the premise of predicting the Super Bowl winner and all of the playoff results for every team based off of their regular season statistics. Our initial hypothesis would be that the team with the best overall statistics would be the Super Bowl favorite.

**Observation**

Over the course of our analysis we found a myriad of interesting statistical anomalies which go against our original hypothesis and conventional wisdom. The most interesting facts have centered around the conclusion that the eventual Super Bowl champions are not the most statistically dominant in the regular season. On the average, the most dominant teams statistically lose in the Super Bowl. In the following graphs, it is clear that whatever metric is used in comparison, the Super Bowl Champions (SC) are always not at statistically dominate at the Super Bowl losers (S).



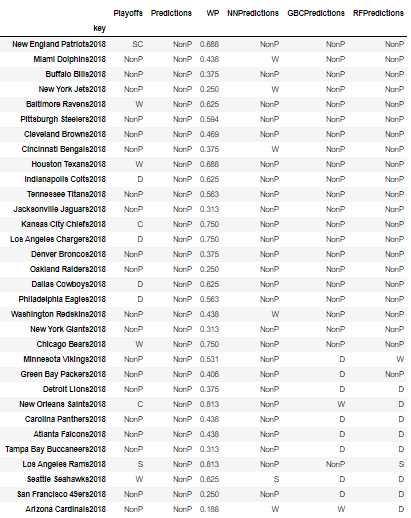
In addition, our team found one specific category the Super Bowl champions did lead in the regular season, Strength of Schedule (SoS). As shown in the below graph, the eventual Super Bowl Champions usually have the most difficult regular season schedule. The only teams that are close are teams that don’t make the playoffs, which makes sense, due to the fact that they played more games against playoff teams therefore having a higher strength of schedule.



This exploratory data analysis begins to paint an interesting picture of how teams will finish in the playoffs, and that while their stats may be a good indicator of their success, it will be a much murkier prediction than previously thought.

**Analysis**

For predicting seasons and playoff success we utilized a number of ML models. In order to try a large number of models and find the best predictor, our team utilized a Support Vector Machine, a Neural Network MLP Classifier, a Gradient Boosting Classifier, and a Random Forest. The results of these models can be seen below:



What one can glean from this analysis is that every model ended giving terrible results. While some were certainly better than others, overall they were highly inaccurate and unrealistic. This is the first time we had been faced with this type of dilemma, in normal classification problems there is nothing wrong with a model predicting every observation to the same class, but for this analysis, there were a finite number of each class. Only one team could be predicted as the Super Bowl winner, only one runner up, and so on and so forth. This limitation created the need to think of prediction in a different way.

For the final modeling, the GBC and MLP neural net models were selected due to them having the best performance in the initial modeling phase. Next, a loop was constructed to iterate through every team in the 2018 season otherwise called the test set, and predict the probabilities of each team receiving each classification. If the classification had not already been taken by another team, the team was given that prediction, otherwise the teams 2nd highest class probability was used. If the class of their 2nd highest prediction probability was already taken, their 3rd highest probability was used and so on and so forth.

What this iterative process gave us was a model that would at least classify teams in a way that was in line with how the actual NFL playoffs would unfold. Not only were these models more realistic in that fashion, they were also far more accurate.



As shown in the above table, the results given by the GBC and MLP models done in this iterative fashion were 100% realistic and much more accurate. 19 of the 20 non-playoff teams were correctly classified. While only a few of the playoff teams were correctly placed, the classifications were never more than +/- 1 round of the playoffs off from the real thing. The only exception were the New England Patriots who statistically were not a tremendous team in 2018 and were predicted to lose in the Wild Card, but in actuality won the Super Bowl. The most difficult teams to place were between the Wild Card and Divisional rounds of the playoffs. This makes sense logically because the teams losing in the first two rounds are likely very similar statistically.

**Recommendation**

As stated previously, armed with the analysis that we performed, our recommendations fall into two distinct categories. There are the NFL teams whose final playoff performance was classified here, and then there are the gamblers who bet on NFL games. For the NFL teams, we would recommend using our models to project their playoff success. Teams certainly need to believe that they can win every single game, but by using this model to predict how the playoffs will unfold can be beneficial in determining their most difficult matchups. Teams can also look at their statistical profile and where the model predicts they will finish, and create various points of emphasis in their gameplans to help overcome their deficiencies.

For the gamblers, the ability to pick and choose Super Bowl winners and NFL playoff brackets would be very beneficial when it comes to betting money. It would be huge for them to be able to look at a team’s regular season stats and make accurate predictions on who will win it all. With such insane predictions coming to fruition, this would hopefully result in large monetary wins for them.

**References**

* Data provided by [www.pro-football-reference.com](http://www.pro-football-reference.com)
* Final team standings provided by <https://www.espn.com/>