NFL Game Predictions

Mary Manz\*

MATH637 - Math Techniques in Data Science

University of Delaware

Newark, DE 19716

May 21, 2021

1. Introduction

Most people know that when gambling, the house always wins. Each game at a casino is designed such that the odds are slightly in favor of the casino. And this slight advantage allows a casino to operate profitably. Does this phenomenon apply to sports betting as well? I do not think so. Every football season, I place bets on a parlay card through Delaware Lottery Games. I usually play the point spreads. These bets are setup in a way that if the Eagles are playing the Giants and are favored to win, the parlay card might look something like this: “Eagles -7.5” and “Giants +7.5.” In this instance, betting on the Eagles will win you money only if the Eagles win by 8 points or more. Betting on the Giants will win you money if the Giants win, or if the Giants lose by 7 points or less. I typically bet on three, which is the minimum, point spread games. And I hardly ever win, taking special note of the one game that fell within one point of the spread that I was on the wrong side of, causing me to lose my parlay card. And then I marvel and wonder how Vegas *knew* that the game was going to be that close?

However, according to one article, Vegas is within one point of spreads less than 10% of the time [1]. And Vegas is more than ten points off the spread 41% of the time [1]. Apparently, Vegas has a group of professional oddsmakers who are responsible for giving each game an “opening number” for a spread [1]. According to another article, spreads or “lines” are first determined in the beginning of the week by “a professional group of oddsmakers that convene to give each game an opening number” [2]. These oddsmakers will “use formulas, simulations or other statistics based programs to help them create these numbers, in addition to personal observations. However, the odds are not decided by a computer, but instead are primarily man made” [2]. Then, “sportsbooks,” which are basically a place that has sports betting, will adjust the spread throughout the week according to the supply and demand of the “NFL market,” consisting of professional and recreational gamblers. What “sportsbooks strive for is not *fair*odds, but attracting equal action on both sides” [3], profiting from what they charge betters on each wager [1]. By the Wednesday before the game, the closing line is usually determined for the parlay cards. In the case of my three-game bet on the parlay card, it seems the sportsbook makes its money by requiring a minimum of three bets. Assuming the NFL market is efficient and the line moves so that there is a 50/50 probability of covering a spread, the chances of being correct for all three bets is 12.5%. So, if the payout for a $5 bet on three spreads is $32.50 (according to one of my parlay cards), and there are 100 of these bets, the total revenue for the sportsbook would be $500 and its expected payout would be $406.25, leaving a $93.75 profit. The Delaware Lottery, run the by Delaware Government, then collects about 50% of the revenue of sportsbooks [4]. So just like gambling at a casino, if your odds are 8 to 1 and your potential payout is only 6.5 to 1, the house should always win! So maybe I give Vegas’ abilities to predict the future more credit than I should. Afterall, “even if closing lines were more predictive, the NFL market, and not “Vegas”, would be responsible” [1].

I understand that oddsmakers and professional gamblers are putting a ton of analysis into each game bet. However, I wondered if I, someone who does not know much about football, could apply the analysis of experts to a machine learning model to predict the outcome of a game, allowing me to win more parlay cards! I hypothesized that the ability of players are the most critical factors to the outcome of a game. As such, I figured that each player’s performance stats could be predictive of a game’s outcome. The Pro Football Focus (PFF) website grades each player for their performance during a given season, referred to as a PFF grade. Each NFL player is graded based on the performance for every play that they do. The player analysis that determines a player’s PFF grade goes beyond basic player stats, for example, if “the quarterback made a great throw, but it was dropped, the quarterback will receive a positive grade for that effort, even though the receiver let him down, earning a negative grade along the way” [4]. To start, rather than trying to predict a game’s spread, I thought I should try for something less complicated. I hypothesized that the PFF grades of both teams are predictive of the game score total. I considered, for example, that is both teams have offensive players with relatively high PFF grades and defensive players with relatively low PFF grades, that the game would be a high scoring game. In this report, I discuss the data I used, the models I created and their results, and my analysis supporting my results.

1. Data

One of the issues I faced when completing this project was getting a good data set. There were not any existing publicly available data sets that I could find that would allow me to achieve my objective. As a result, I created my own data set by copying and pasting into excel data from the following two websites:

<https://www.pff.com/news/nfl-roster-rankings-all-32-teams-ravens-first-jaguars-last>

<https://www.footballdb.com/games/index.html>

I then manually cleaned the data in excel and used excel formulas to combine the data, creating the data set I needed to complete my analysis.

1. Model results

The first model features I tried were the 2019 PFF grades, for both teams, of every single starting player for the 2020 season (48 PFF grades). The target values were the total scores of each game played in the 2020 season. I used seven different models with these parameters: Random Forest Regressor, Linear Regression, Ridge Regression, Lasso Regression, Support Vector Regression, Decision Tree Regressor, and Gradient Boosting Regressor. I tried several different models as I was not sure which would most appropriately fit my data.

I tried four different groups of features in case having less features might cause the models to perform better. I did not change anything for the first group of features and left all 48 features as is. For the second group of features, I combined both teams’ PFF grades for each position, resulting in 24 features. For the third group, I combined the PFF grades for all offensive players and for all defensive players separately for both teams, resulting in 4 features. For the fourth group, I took the difference between both team’s total offense PFF grades and total defense PFF grades, resulting in two features. Lastly, I created a dummy array of predictions equivalent to the mode of the target, which was a total game score of 44. Then, for each model and for each group of features, I got the following model metrics: mean absolute error, mean squared error, root mean squared error, and R squared (for testing predictions).

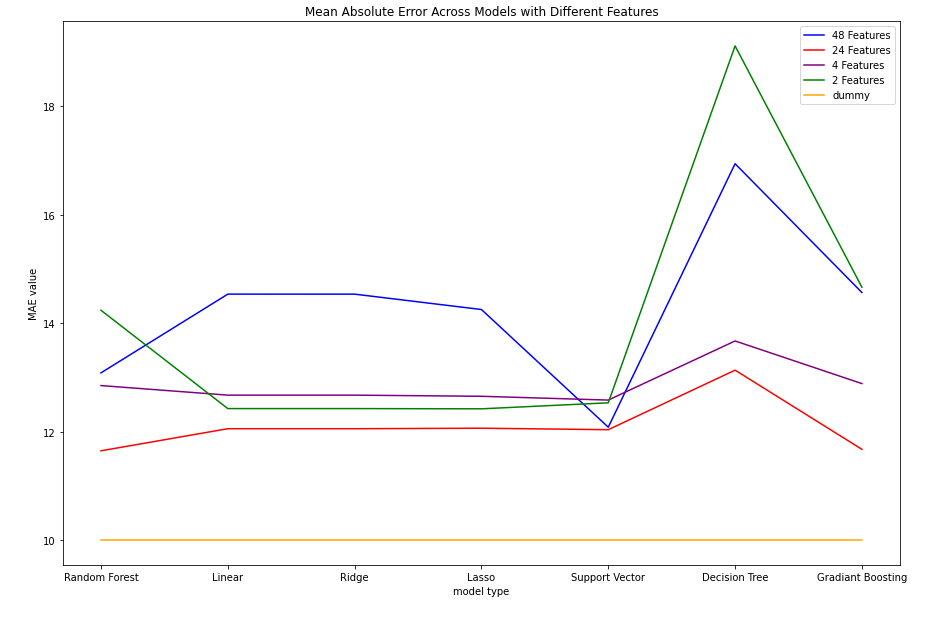


Figure 1 – All groups of features across all models have a mean absolute error higher than that of the dummy array.

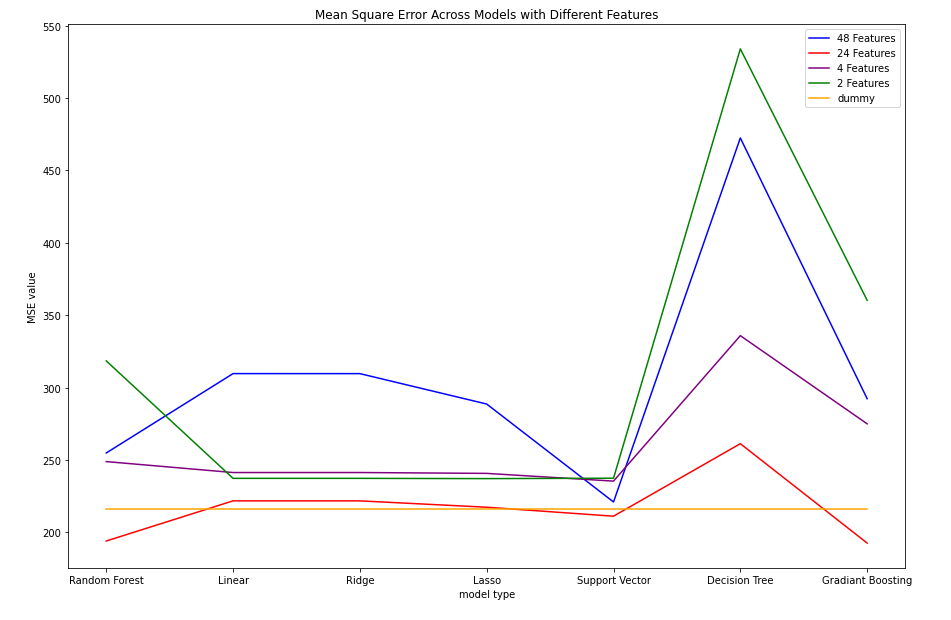


Figure 2 – All groups of features across all models have a mean square error higher than that of the dummy array, except for the 24 features used in the random forest, support vector, and gradient boosting models.

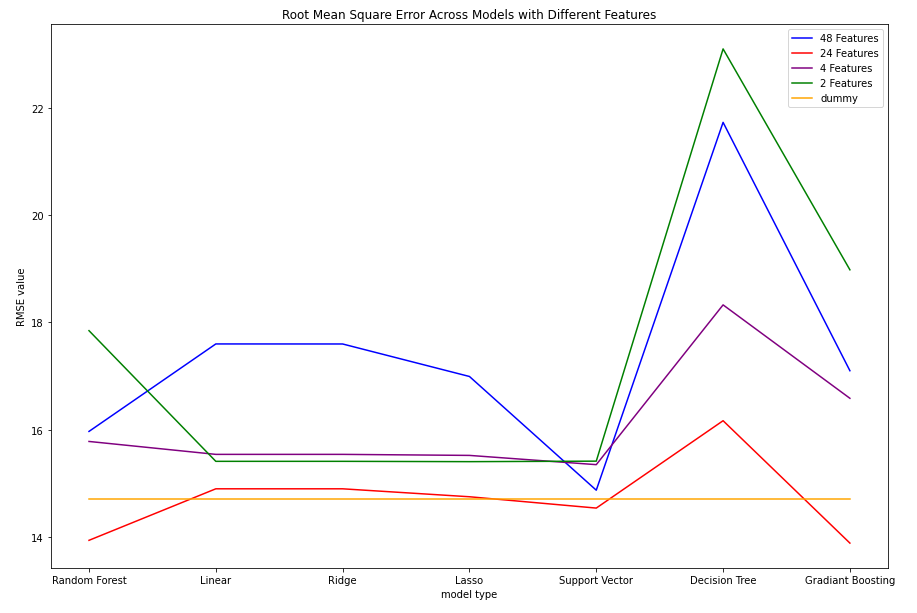


Figure 3 – All groups of features across all models have a root mean square error higher than that of the dummy array, except for the 24 features used in the random forest, support vector, and gradient boosting models.

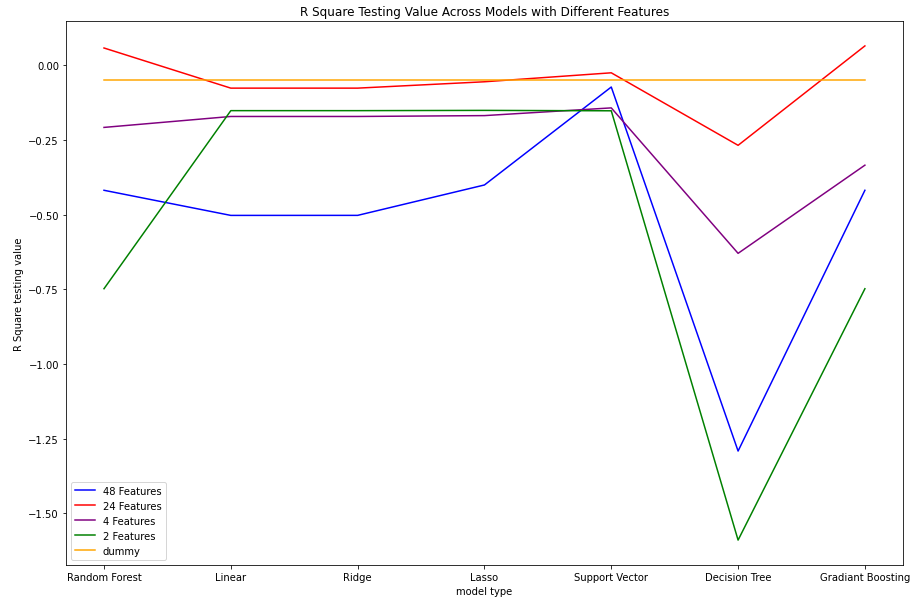


Figure 4 – All groups of features across all models have a R Square value lower than that of the dummy array, except for the 24 features used in the random forest, support vector, and gradient boosting models.

Based on the above figures, all groups of features except for the group of 24 features perform worse than the dummy array. Meaning that guessing for every game that the total score is 44 will typically serve as a better prediction than using the models I created. The group of 24 features performed only slightly better than guessing for some of the model metrics and for some of the models. This is not very promising either. For any model, tuning it can result in slightly better results; however, in this case it seems safe to conclude that there is a weakness in the relationship between PFF grades and total score as I originally hypothesized.

1. Supporting analysis

To support my conclusion that there does not seem to be a relationship between PFF grades and total game score, I created some scatter plots to show this non-relationship visually.

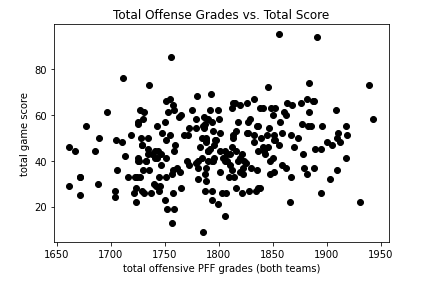


Figure 5 – Total offensive PFF grades for both teams combined vs. total game score. It does not appear as though the two variables are correlated.

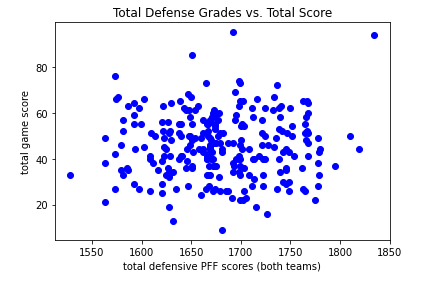


Figure 6 – Total defensive PFF grades for both teams combined vs. total game score. It does not appear as though the two variables are correlated.

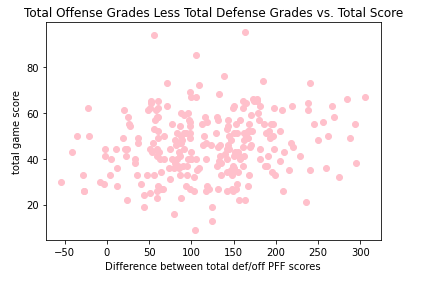


Figure 7 – Total offensive PFF grades for both teams combined less total defensive PFF grades for both teams combined vs. total game score. It does not appear as though the two variables are correlated.

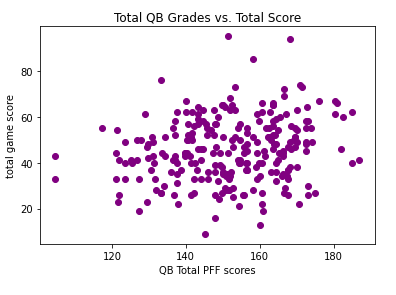


Figure 8 – Total quarterback PFF grades for both teams combined vs. total game score. It does not appear as though the two variables are correlated.

The above scatter plots further suggest that there is not a relationship between PFF grades and total game score.

1. Conclusion

In summary, according to my analysis, PFF grades (alone) are not predictive of NFL game score totals. However, one assumption that I made during my analysis is that the starting lineup for each team at the beginning of the season was the same throughout the season. Of course, in football there can be changes made to the roster due to injury, poor player performance, etc. Perhaps adjusting the features data to account for these changes could lead to better model performance, but I would not expect the performance difference to be substantial. Instead, for next steps, I am interested in building classification models that predict whether a game is high or low scoring, and to see if there is a relationship between weather and game score totals. I also wonder if a team performance’s is relatively similar between two seasons given that although several factors about a team change, most of the factors stay the same between two seasons. Could each team’s scoring average from last season be predictive of the game score total? Lastly, I am interested in seeing how often Vegas’ opening lines are better than the closing lines or how often they are better for instances when a large difference exists.

References

[1] Walsh, P. “No, Vegas Doesn’t Really Know the Outcome of Future Sporting Events.” *Action Network*, 27 Apr. 2020, [www.actionnetwork.com/how-to-bet-on-sports/general/las-vegas-odds-not-predictive-sporting-event-outcomes](http://www.actionnetwork.com/how-to-bet-on-sports/general/las-vegas-odds-not-predictive-sporting-event-outcomes).

[2] Ops, Dev. “NFL Lines.” *Online Football Betting,* 3 Aug. 2013, onlinefootballbetting.net/strategy/nfl-lines.

[3] Lorenzo, Frank, and Geoff Johnson. “Why Betting Lines Move at Sportbooks - Vegas Line Movement.” *My Top Sportsbooks*, 29 Apr. 2021, [www.mytopsportsbooks.com/guide/why-betting-lines-move](http://www.mytopsportsbooks.com/guide/why-betting-lines-move).

[4] Report, Betting, et al. “Legal Sports Betting in Delaware | Where to Bet on Sports.” *USBettingReport.Com*, 1 Mar. 2021, usbettingreport.com/sports-betting/delaware.

[5] “Player Grades.” *PFF*, 2021, [www.pff.com/grades](http://www.pff.com/grades).