

Gian Batayola & Brett Geunes

Hamid Karimian & Hui Liu

CSE 482

13 April 2023

PREDICTING NFL POSTSEASON WINS USING MACHINE LEARNING

Report 3:

Analyzing the Data:

With the lower success of fantasy data, we ended up working with individual player statistics to train our model. We immediately found success in this niche, and it allowed us to predict wins and explore the roles individuals play in such a team-oriented sport. In analyzing our data, we found many key details we had not considered. One such example was we had to drop a lot of the week 17 data as it was not contributing well to our answers. Upon further research into why this may be, we found that many of the star players on playoff teams are rested week 17 to prepare for the playoffs. Given the fact that we aggregated individual performances for our investigation, these performances served as outliers to our goal.

Our feature selection provided some strong insight as well. We found through using ANOVA F-test that all our strongest features were defense and running statistics. It was interesting to see that in such a passing-oriented era of the NFL, the other aspects of the game tended to push one team over the other in the end.

Conclusion:

With our model, we found a 73% success rate predicting postseason wins using a linear support vector machine on the combined statistics of players on each team. The features we focused on for this model were passing interceptions, passing sacks, passing sack yards, sacks, sack yards, passes defended, interceptions, interception return yards, rushing attempts and rushing yards.

While achieving our goal to make a strong predictor of wins, we also found those features to be very important individual player metrics to a full team's success. This information can be used to determine what to focus on in both training and scouting players, as these individuals can make the difference above the normal scoring methods that contribute to success.

Future Work:

Given the opportunity to explore this problem more, we believe a natural next step would be to try to predict wins completely blindly with regards to the game we are predicting. While the nature of this model was just as much about what contributes to the win as predicting it, it would be an interesting next step to pair it with a model that predicts a game using only the data that has led up to the game itself.