

Predicting Football Wins/Losses and Scores Based on Average Player Position Statistics

Members: Maxwell Smith, Joshua Draper, Coulson Rich, Michael Hawks

How much we all want to be part of the project: **10/10**

Description of the project

Our project will be to build a dataset of NFL football games and team statistical averages and train an MLP model on those historical games. We will take as input a list of team stats for team A and team B and predict the outcome of the game (whether it was a win or a loss for the home team) as well as produce a distribution of potential scores. The distribution of scores will be all numbers from 2 to 73 (the historical max of all NFL games). Our idea came from a project that Max did to predict the result of baseball games given the player statistics who played any game. This inspired us to apply machine learning to another sport, football. We knew that football has thousands of records of games played, which will help our model converge. A topic we discussed was whether or not to use player stats or team averages for training the network.

What features the data set might include

Because we are taking the average statistics over a season, getting many different statistics per team will be easy. Here are some stats that we thought of and might use when looking over example datasets on nfl.com:

Inputs:

Offense: Attempts, completion%, rushing, rush yards, yards/carry, receptions, yards, yards/reception, total touchdowns

Defense: sacks, tackles, interceptions, fumbles forced

Special Teams: field goal percentage, punt average

Other: whether team1 was home or away

Outputs: Win/Loss for Team 1, Team 1 score distribution, Team 2 score distribution

So, an entry would look something like the example table below. I took out a lot of the things listed above to fit on the page.

Team 1 Passing Yards	Team 1 Sacks	Team 2 Passing Yards	Team 2 Sacks	Team 1 Home or Away	Team 1 Win/Losses	Team 1 Score	Team 2 Score
353	2	234	3	1	1	24	17

How, from, and where would the data set be gathered and labeled

[Operations.nfl.com](https://www.operations.nfl.com), www.footballdb.com, and www.nfl.com have information on team statistics back to 1970. These stats will be used to train the model on each game that the team played that season and the outcome of the game. These averages will generalize teams and players so

they become a non-factor. We will use an ARFF or something similar to store all the game averages and the resulting wins and losses of the teams throughout their respective seasons. This ARFF file will also label each of our features. We will join the data of wins/points with the data of the average team features. All of this data combined will make up our training/testing data. We should then be able to predict the outcome of a game by using both the team's averages and the current state of the game to make an accurate prediction of the game's outcome.