

Predicting Fantasy Points for NFL Players

Using regression techniques to build winning fantasy football lineups

Introduction

The fantasy football industry has seen a massive spike in both revenue and the number of fantasy players in the past five years. One of the main reasons for this spike is how the concept of daily fantasy sports (DFS) has revolutionized the industry. Daily fantasy players no longer need to deal with the risk of season-ending injuries to their star players, drafting player for a season that underperform, making trades that end up in favor of the other team-owner, etc. Also, daily fantasy players are able to win money and get paid immediately after the final contest in a slate has been completed, instead of having to wait until the end of the season to collect their winnings. Players have the ability to draft a new fantasy team each week, or every day for certain sports like baseball or basketball. Contests between players who participate in DFS may include anywhere from 2 participants (1 versus 1) all the way up to 500,000+. Entry Fees for participants may range from no cost (usually a promotional contest with a small reward for playing) up to \$10,000, and prizes may reach up to \$2,000,000 for winning a tournament with a high volume of participants. Daily fantasy sites, such as DraftKings and FanDuel are more popular and lucrative than ever. The motivation of this project is to explore ways to implement machine learning to predict points scored for players, which will help us build profitable daily fantasy sports lineups.

How it Works:

FanDuel and DraftKings both host DFS games in nearly every major sport, which include NFL, MLB, NBA, NHL, and NCAAF among others. For our purposes, we will be exploring professional football (NFL), and using FanDuel's format for building a team. For each NFL contest using FanDuel, the participant is allowed a \$60,000 salary cap to draft his or her players. For each team, the participant may choose one quarterback (QB), two running backs (RB), three wide receivers (WR), one tight-end (TE), one flex (a choice of an extra RB, WR, or TE), and one team defense/special teams (DST). The participant chooses players for his/her team, and pays a fee to enter a contest. The site, or host (FanDuel in this case), of the DFS contest automatically takes 10% of the participant's entry fee upon entering the contest. This is known as the "rake", and is how the site makes its profit. Once a participant has entered a contest, in order for his or her entry to win money, it must place higher than a certain percentage of entries in the contest. The amount of money a single entry wins depends on what type of contest it is, as well as how many entries are in it. For example, if someone entered a two-person contest, and paid \$1 to enter, that person's lineup must score higher than the opponent to win \$1.80, since the host keeps 10% of each entry. Another example is if someone entered a tournament that had 20,000 entries for \$5, the host keeps 10% of \$100,000, and the remaining \$90,000 would be up for grabs among the entries. The top prize might win \$40,000, second place might win \$15,000, third might win \$5,000, and the remaining \$30,000 is distributed in a decreasing manner among the remaining top 10-15% of entries in this contest.

Data

Attribute Selection

The attributes that will be used for this project will be mainly individual player statistics based on a player's overall history, and that player's recent history. We will also be using matchup attributes for an instance. Every position within a lineup is different, and thus different attributes will be important. WE need to take FanDuel's scoring system for players into account. For individual overall history attributes, we will be using features such as pass yards per game, rush yards per game, receiving yards per game, pass TD's per game, rush TD's per game, receiving TD's per game, receptions per game, targets per game, rush attempts per game, and pass attempts per game. For recent history attributes, recent FanDuel points per game (over last 5 games), and recent TD's (rushing, passing, receiving). For current matchup attributes, we will be using defensive points allowed per game, whether the match is home or away, defensive yards allowed per game (pass, rushing, receiving), and many others.

Obtaining the Data

For all of the attributes described, there are many sources from which to obtain the data. One source is pro-football-reference.com, which has a database that contains nearly every relevant football statistic for every NFL game played dating back to 1920, which includes weather, venue, and referee data. Another source for obtaining the needed data will be fantasydata.com, which has defensive and offensive rankings for every team in each contest, and data from sports books, which includes the totals for each game. Additionally, fantasydata.com has every relevant fantasy stat needed for individual players.