

Project Phase 1

Marion Geary

2/25/2022

Why NFL Data?

For my Machine Learning project, I plan to analyze data about NFL players. In my Software Development class, I am currently working with a team to build an application for Fantasy Football players to help them use analytics to build their team. One important part of our project is helping our users make data-driven decisions when playing Fantasy Football. Most of our application will simply entail displaying data and letting the user make decisions. However, for this project, I would like to take the analytics a step further and use machine learning techniques to predict the performance of players, specifically as it relates to Fantasy Football scoring.

Research Questions

The main goal of my research is to predict the amount of fantasy points that a player will earn in their next game. This will entail using historic data about the player as well as data about players in the same position or on the same team to make predictions. Since fantasy points is a function of a player's stats, this project will be a way of predicting the player's performance in their next game.

How many fantasy points will this player score in their next game?

Another goal of my research is to be able to predict which of any two (or more) selected players will perform better in terms of fantasy points.

Which player will score the most fantasy points in the next game?

Data Source

My data will come from the `nflverse` group of packages in R. Much like the `tidyverse`, the `nflverse` is a group of packages designed to make working with NFL data easy and accessible (check out the documentation and repository). The package sources data from the NFL API and other resources and updates regularly when the NFL is in season. I plan to mainly use player statics. I will likely focus on the 2021 season, but I may make comparisons across seasons as well. The `nflverse` has data from every game since 1999.

```
library(nflverse)
# This will be my main data set. It includes all the stats
# from each players performance in every game since 1999.
player_stats <- load_player_stats()
#kable(head(player_stats, 3), "pipe")
names(player_stats)
```

```
## [1] "player_id"          "player_name"
## [3] "recent_team"        "season"
```

```
## [5] "week" "season_type"
## [7] "completions" "attempts"
## [9] "passing_yards" "passing_tds"
## [11] "interceptions" "sacks"
## [13] "sack_yards" "sack_fumbles"
## [15] "sack_fumbles_lost" "passing_air_yards"
## [17] "passing_yards_after_catch" "passing_first_downs"
## [19] "passing_epa" "passing_2pt_conversions"
## [21] "dakota" "carries"
## [23] "rushing_yards" "rushing_tds"
## [25] "rushing_fumbles" "rushing_fumbles_lost"
## [27] "rushing_first_downs" "rushing_epa"
## [29] "rushing_2pt_conversions" "receptions"
## [31] "targets" "receiving_yards"
## [33] "receiving_tds" "receiving_fumbles"
## [35] "receiving_fumbles_lost" "receiving_air_yards"
## [37] "receiving_yards_after_catch" "receiving_first_downs"
## [39] "receiving_epa" "receiving_2pt_conversions"
## [41] "special_teams_tds" "fantasy_points"
## [43] "fantasy_points_ppr" "pacr"
## [45] "racr" "target_share"
## [47] "air_yards_share" "wopr"
```

```
# I will use some data from the roster as well to get more
# demographic information about the players
roster <- load_rosters()
#kable(head(roster, 3), "pipe")
names(roster)
```

```
## [1] "season" "team" "position"
## [4] "depth_chart_position" "jersey_number" "status"
## [7] "full_name" "first_name" "last_name"
## [10] "birth_date" "height" "weight"
## [13] "college" "high_school" "gsis_id"
## [16] "espn_id" "sportradar_id" "yahoo_id"
## [19] "rotowire_id" "pff_id" "pfr_id"
## [22] "fantasy_data_id" "sleeper_id" "years_exp"
## [25] "headshot_url"
```