Final Step 1

Brian Reppeto

2023-07-24

Markdown Basics

Introduction

Fantasy football is a popular online game where participants act as team managers and draft real players from the National Football League (NFL) to form fantasy teams. The performance of the actual stats of the players in real-life games determines the fantasy team's score. Participants compete against one another on a weekly head to head basis and these scores determine a weekly winner. One of the most important aspects of fantasy football is the drafting phase, which takes place before the NFL regular season starts. Participants select players based on their projected performance and rankings. I will use the redraft type of league along with a snake draft type of draft. Additionally, I will use the standard scoring system, which is 1 point per 25 passing yards, 4 points per passing touchdown, 1 point per 10 rushing or receiving yards, 6 points per rushing or receiving touchdown, -2 points per fumble lost or interception.

The issue I want to research is being able to help predict fantasy football drafting rankings. Specifically, drafting quarterbacks, running backs, wide receivers, and tight ends. The goal is to develop a model or methodology that can help accurately predict the ending rankings of NFL players for fantasy football drafts. This project would be of particular interest to fantasy football participants, as it can help them better draft a fantasy teams for their season. By leveraging data science techniques, I can analyze historical player performance and other relevant data to make informed predictions for the upcoming season's draft rankings.

Research questions

1. What statistical features of NFL players historical performances are most indicative of their fantasy football draft rankings?
2. Can we identify any patterns or trends in player performance over the past few seasons that might help predict future draft rankings (rookies will be assigned a value based on draft position, and actual position (i.e. scoring role or none scoring role) ,and also team needs)?
3. How does a player's injury history affect their draft ranking, and can we account for injury risk in our predictions?
4. Are there specific positions (quarterback, running back, wide receiver, etc.) that are more valuable to draft in a fantasy league before others?
5. How does player performance vary across different teams and playing conditions, and how might this impact their draft rankings (i.e. does a better team give a player a better ranking)?

Approach

1. Data Collection: Gather historical player performance data, fantasy football draft rankings data, injury records, team data, and any other relevant datasets.
2. Conduct data analysis to understand the data, identify any missing values or outliers, and gain insights into player performance and rankings trends.
3. Select or create relevant features that may impact draft rankings (i.e. player statistics from previous seasons, team performance metrics (Strength of schedule), players age, injury history, etc.)

4. Cluster players using K-means to group players by past performances, which can help group the players into rankings.
5. Use Bayesian regression to generate predictions, which can help provide point estimates and also confidence intervals for the draft rankings.

6. Provide a recommendation for the most effective method to help predict fantasy football drafting rankings.

How approach addressed the problem

By combining the multiple approaches listed above I will be able to compare individual performances along with other factors to help provide valuable insights into a ranking and value list for fantasy football managers. This will all them to build competitive teams for the fantasy season. This will also allow me to group the players into clusters.

This project will not be fully operational as all cuts for the NFL season will not have happened by the time this project is complete.

##Datsets

Season stats:
https://www.pro-football-reference.com Strength of schedule:
https://www.sharpfootballanalysis.com/analysis/2023-nfl-strength-of-schedule/ Scoring stats: https://www.footballdb.com/statistics/nfl/player-stats/scoring/2022/regular-season Fantasy scoring points by two types of systems: https://www.footballdb.com/fantasy-football/index.html

Required Packages

'tidyverse' 'ggplot2' 'cran' 'factoextra' 'dplyr' 'ggfortify'

Plots and tables needed

1. Line plots to visualize trends in player performance over time.
2. Bar plots to compare the average fantasy points scored by players in different positions.
3. Scatter plots to explore relationships between player attributes and fantasy points.
4. Tables to display the top-performing players based on predicted fantasy points.
5. Feature importance plot to understand the model's top predictors.

Question for future steps

1. How can I incorporate additional data, such as team dynamics (coaching issues, player issues), weather conditions, or player off-field behavior, to improve the model's predictions?
2. Can I build an ensemble model that combines the strengths of multiple algorithms to achieve better prediction accuracy?
3. How can I continuously update the model during the NFL season to account for emerging trends and player performance changes?
4. Can I use this research for in-season predictors?
5. Understanding Feature Importance to help identify the best attributes for drafting success.