Sid Iyer
Sparsh Jain
Wesley Graham
Roshan Srinivasan
Nabeil Kizilbash
10/30/19

Optimal Portfolio Theory in Daily Fantasy Sports

Motivation

The daily fantasy sports market is expected to grow to $1.5B USD by 2024. Many platforms (Fanduel, DraftKings) are opening up throughout the United States that allow individuals to assemble "the perfect daily fantasy sports team" - gambling on whether or not their assembled team can score the most points during the night's matches.

Our group seeks to generate predictive models that lead to optimal fantasy team selection - with hopes of "beating the market" on fantasy betting services.

Data

The nature of the data that is being used has to do with individual athlete's performance both during this season and historical, evaluated against similar players in the same position. The rationale behind this is that only a limited number of athletes can be chosen for any particular fantasy team, therefore it is most beneficial to have a group of athletes that will give you the most number of fantasy points. A lot of this data is readily available to consumers and fans via the various sports websites (NFL, NBA, etc), so it is very easy to collect and use this data in order to establish a ranking of players based on their expected fantasy points yield. Additional data to take into account is the opponent that a particular player is facing, the reason being that stronger opponents will likely pose a larger challenge, so their performance will likely be lower than against an easier opponent. In addition, information such as whether or not a particular player is injured or not playing on a particular day would also be vital in determining how many fantasy points they will get on a particular day.

Analytics Models

For this type of problem, it is important to consider the value of each feature in the model. As such, we will start by generating a linear regression model where our response variable is a value from 0 to 1 which is used to indicate the optimality of a team selection based on a series of factors. We will use VIF scores to identify key features in our model to be further explored in other analyses. For the purposes of this problem, we may use a higher VIF cutoff so we can further explore a larger set of features in further model training. From this initial model, we will then generate multiple other types of models such as logistic regression and CART and apply boosting and random forest methods with different numbers of trees and number of features selected to identify the best model accuracy. In these CART and logistic regression models that we apply ensemble methods to, our response variable is whether or not a certain team formation won a game or not (1 or 0). We will also apply bootstrapping to generate a confidence interval of our model accuracy to make sure that there is low variance.

Sid Iyer
Sparsh Jain
Wesley Graham
Roshan Srinivasan
Nabeil Kizilbash
10/30/19

From these models, we will identify which are the key features and key values that influence optimal team selection in fantasy sports and then run simulation tests using logistic regression on the subset of values that gave the highest optimality scores in our models to identify which values in the key features give the most optimal team for winning.

Impact

In a world where data collection and analysis are becoming increasingly important, our group realizes the impact that it has on the world of sports, specifically with regards to fantasy sports and sports betting. By determining an effective way to optimize fantasy team roster selections, fantasy and betting participants will be able to make more informed decisions that will result in increased accuracy and overall gains monetarily. In other words, creating a method to more accurately put together a fantasy roster or bet on sporting events will help take the guesswork out of a conjecture-heavy arena.

With the assumption that our project will solve this problem, this framework can be applied to events outside of sports, including possibly the medical field (prescribing medicine in a more accurate manner), the insurance field (determining what the best insurance package to purchase would be for a family given their income levels), or even the political sphere (determining the most effective use of campaign funds in an election in order to achieve success and efficiency). Regardless, all possible explorations involve the effective and efficient usage of money in order to solve a predicted outcome, making our project a universally relevant solution.

Goal:

- Look at salary over +/- ratio and see how our team performance compares to others based on aggregated total number of points per team
- Data:
    - Salary data per player
    - NBA data for points per game
    - Fantasy total number points