

Sports Betting Strategies Guided by Machine Learning Algorithms

Student: Eunice Park

Advisor: Ethan Meyers

Introduction and Background

Over these past few years, sports betting has been increasingly popularized in the United States. Through the assistance of online betting platforms such as FanDuel and DraftKings, over one-in-five U.S. adults engage in sports betting, and such popularity is expected to increase as more states legalize commercial sports betting.¹ (Pew Research 2022) By 2025, the cumulative value of the American sports betting market is expected to soar to \$8 billion.² This is the result from a May 2018 Supreme Court decision which ruled that individual states had the right to legalize sports betting, overturning a longstanding 1992 law.³ As sports betting is becoming more accessible, the availability of data and advancement of technology have driven new opportunities to utilize machine learning algorithms to predict optimal sports betting strategies.

Sports betting is an ideal problem for a machine learning algorithm, with large availability of historical data and a clear simplified objective of picking the winner.⁴ (Malafosse 2019) This thesis will examine the applications of machine learning to Daily Fantasy sports betting, particularly for gambling websites of DraftKings and FanDuel. These sites allow users to draft a fantasy sports team and have that team be entered into a pool of other teams. If that team's total number of points is higher than that of a large percentage of its competitors, the creator of that team receives a multiple of the buy-in as payout, with the potential to win much more if their team is in the top fraction of scorers in the pool.⁵ (FanDuel) Because of the predictive nature of selecting a high-scoring team and the large volume of data publicly available from previous sports seasons, the Daily Fantasy sports betting space is a prime candidate to apply machine

¹

<https://www.pewresearch.org/fact-tank/2022/09/14/as-more-states-legalize-the-practice-19-of-u-s-adults-say-they-have-bet-money-on-sports-in-the-past-year/>

² http://cs230.stanford.edu/projects_fall_2020/reports/55766293.pdf

³ <https://www.nytimes.com/2018/05/14/us/politics/supreme-court-sports-betting-new-jersey.html>

⁴ <https://towardsdatascience.com/machine-learning-for-sports-betting-not-a-basic-classification-problem-b42ae4900782?gi=01f58b2ead5a>

⁵ <https://www.fanduel.com/rules>

learning principles. As opposed to general game bets of an over/under calculation based on team outcomes, fantasy sports provide unique opportunities for potential betting profits with the application of machine learning. Fantasy sports allow one to curate a personalized team, and such personalization is suited for a machine learning model for fantasy value projection which can directly predict a player's fantasy score from a set of player features. In practice, a new fantasy lineup can be created for each slate of games scheduled for a particular day, following constraints of player positions and a salary cap of \$50,000 on the draft, with better players costing more. (Nguyen 2017) ⁶ In fantasy sports betting, the crux of the problem is in trying to determine and predict fantasy scores for all players such that the most efficient lineup can be constructed in terms of fantasy value per salary unit. These defined limitations of a salary cap and player positions create a unique opportunity for a machine learning model to best predict which individual players can construct the most efficiently generating team. Additionally, one can submit up to 150 fantasy lineups at any given daily fantasy time. Our model can test out an optimal level of "risk" a fantasy sports "lineup" should take, with "risk" being the distribution of different players with larger performance variance, and "lineup" potentially being a combination of x distinct lineups with varying risk levels.

More specifically, this thesis will examine the applications of machine learning for Daily Fantasy sports betting on the NBA.

Due to narrower distributions of team strengths and smaller home advantages, a contest in the NHL or MLB is much closer to a coin-flip than one in the NBA or NFL. (Lopez 2017) In comparison to other sports leagues, the NBA demonstrates the largest dispersion of talent and the largest home advantage. Yet even when accounting for the larger scale of these factors for outcomes, the NBA still boasts the largest between-week uncertainty, such that they are the least consistent league on a week-to-week basis. (Lopez 2017) Possible explanations for this include the prevalence of injuries, resting of starters, and in-season trades. An NBA team has fewer active players during play when compared to other common fantasy sport betting leagues, thus individual players have a larger impact on the outcome of a game in the NBA. These quick changing factors involving individuals players can significantly acutely shift the performance of

⁶ http://cs230.stanford.edu/projects_fall_2020/reports/55766293.pdf

a team. (Berri and Schmidt 2006) Therefore, due to increased week-to-week variability and the larger importance of individual players and game setting, more opportunities for potential fantasy sports betting profits exist in the NBA.

Problem Statement, Methods, Data

The problem can be split into two parts. First, the number of DraftKings/FanDuel points a player will score based on that player's previous performance needs to be predicted. This can potentially be framed through a regression problem based on box score statistics from that player in previous games and through a naive bayes classifier with discretized state space. After predicting the number of points for DraftKings/FanDuel players and fitting such into a machine learning model, a team can then be chosen based on the predicted player points for particular game nights. This part is framed through a constraint satisfaction problem, which includes the betting platform's spending limit to draft a team and the requirement for every position to be filled by a unique player. With this information, an optimal betting strategy can be predicted in order to select a fantasy sports team with the highest predicted score to be entered into the betting pool.

Utilizing python web scraping libraries (BeautifulSoup, etc), box score and team data from previous sports seasons can be collected from ESPN. Additionally, websites such as basketballreference.com provide consolidated statistics, scores, and historical performance for major sports league teams. DraftKings/FanDuel provide salary and position data for individual players, which can be downloaded daily.

Once these statistics are retrieved, features can be constructed to be the basis of the learning algorithm. These derived features can be fit to a model that creates a mapping from features to fantasy points. More research is needed to determine which features should be included, but a list of potential features include: season-to-date features of average player statistics, game characteristics of home/away games, and injury predictions (possibly identified beforehand through natural language processing of twitter news data). Such data will need to be extensively cleaned, especially if it is web scraped to reveal relevant information. With these features, a

machine learning model with a custom loss function can be implemented to predict optimal betting strategies for daily fantasy sports.

Deliverable and Timeline

The deliverable will be a thesis paper discussing background, implementation, and conclusions from my research.

January - Mid February: literature review, outline approaches to take to make predictions, collect relevant data, finish writing background/intro section of thesis

Mid February - Early March: clean relevant data

Early March - Early April: train and test various machine learning models with cleaned data (ideally by late March/early April some of the training/testing can be through the DraftKings/FanDuel platform)

Early - Mid April: draft conclusions and finalize thesis