Akash Alaparthi, Neel Avancha, Aryan Kale,

Aneesh Ponduru, Druv Sarin

Dr. Clachar

Foundations of Data Science

16 October 2022

Phase 1: Project Proposal

Over the course of the last 75 years, the emergence and development of professional sports leagues has led to star athletes of their respective sport gaining and sometimes eclipsing celebrity status and fame. Now, with the emergence of social media, athletes are more famous now than ever in history. Currently, international soccer star Cristiano Ronaldo is the most famous Instagram account with roughly 484 million followers. The explosive emergence of the sports industry has also led to the creation of another massive industry, sports betting. This year, LegalSportsBetting.com conservatively estimated a whopping $165 billion was bet legally in the United States alone, netting the company a $44 billion profit. (Malone). Each bet a user can place has corresponding odds which are designed to net the betting company the most profit. In order for these odds to be most optimal, we need to be able to best predict the outcome of these events. Our employer, Caesars Entertainment, is recognized as one of the most prominent sports betting companies in the world, with an estimated current revenue of $3.47 billion as of 2021 (Porter). The NBA is one of the most popular sports to place bets on currently in the United States, with a plethora of various outcomes to predict. Caesars has tasked our team of five data scientists to construct an algorithm that allows the company to successfully predict the NBA award winners in the next year. These awards are made up of the MVP (Most Valuable Player), DPOY (Defensive Player of the Year), Sixth Man of the Year, and Most Improved Player.

With so much potential money on the line, our team must be able to identify relevant data, process it, and use it to create models to best predict the winners of the awards listed above. Failure to do so properly may result in a major net loss for the company. This is because we will be providing betters with odds for these awards. Betters can potentially earn a lot of money very easily if the odds for the player most likely to win are higher than they should be. Betters will take advantage of this opportunity and if these players do end up winning, a lot of money will have to be paid out. Sports betting services already offer many free promotions to attract users to their platform, spend millions of dollars every year on marketing, and lose close to 50% of their profits to tax, so they must ensure that their odds are as effective as possible so that they will not put themselves in a position to lose even more money (Kosman).

This topic is of interest to all of us because from a young age, we have all been involved in various forms of sports including playing, watching, and even attending sporting events. The one thing in common among us five is our love for basketball which is why we specifically chose that as the sport for this project. Furthermore, the relevance of sports betting has skyrocketed in the past century which has piqued our interest as it has become so heavily correlated with each basketball game. With our knowledge of computer science and recent knowledge of data science, we believed our passion for sports would be a perfect application for our project with the publication of new advanced statistics, allowing us to have an abundance of potential data available to be scrapped. We believe that it would be very interesting to attempt to create these models as well because this is essentially what goes on behind the scenes at the major sports betting companies and is how they determine their odds for the public.

Our goal on this project is to find which metrics and factors are the most important in determining which player gets the MVP, DPOY, Sixth Man of the Year and Most Improved

awards and use this information to create a predictive algorithm that can predict the top player for each of these awards. This will allow us to inform Caesar Entertainment of what the best odds would be to increase revenue and win more bets. First, we will analyze historical data of the players' statistics for the season that they received each of these awards to better understand what statistics they have in common that led to them receiving their respective awards. Then, after we understand the key factors needed for each of these awards, we will create a predictive model that uses the data from the current and previous NBA seasons and current player statistics to determine which player is most likely to win each award this season.

Works Cited

Kosman, Josh. “New York Mobile Sports Betting Operators Have Lost $200M, Analyst Claims.” *New York Post*, New York Post, 10 Mar. 2022, https://nypost.com/2022/02/23/new-york-mobile-sports-betting-operators-have-lost-200m-analyst/.

Malone, George. “How Large Is the Sports Gambling Industry in America?” *GOBankingRates*, GOBankingRates, 11 Nov. 2021, https://www.gobankingrates.com/net-worth/sports/how-large-is-sports-gambling-industry-in-america/#:~:text=Although%20the%20figure%20is%20considered%20conservative%2C%20LegalSportsBetting.com%20estimates,billion%20in%20gambling%20revenue%20for%202021%2C%20Forbes%20.

Porter, Melanie. “20 Largest Gambling Companies in the World (2022).” *GamblingNews*, 20 May 2022, https://www.gamblingnews.com/blog/best-gambling-companies/.