Predicting the Results of NBA Games Using Machine Learning

Sam Macpherson, Konstantin Topaloglou-Mundy, Lazar Vukoje

{skmacphe, ketopalo, lvukoje}@uwaterloo.ca

University of Waterloo

Waterloo, ON, Canada

Introduction

In recent years, the NBA has been extending its reach and attracting an increasingly larger and diverse audience. NBA games are viewable in many nations across the globe, and in the 2014-2015 season, the league featured 92 international players, hailing from 39 different countries (Jones 2016). In 2019, an average of 15 million people—per finals game—watched Kawhi Leonard and the Toronto Raptors garner their first NBA championship (Gough 2019). The American-based league, at the age of 74, now sits in fourth place among all sports leagues in the world by revenue (Amoros 2016). People from all over the world enjoy watching and cheering on their favourite teams, and as the NBA increases its global presence, we see an equal increase in the revenue the league pulls, not just from broadcasts, but from fans purchasing jerseys and other merchandise as well. Over the last few years, we have seen a rise in a separate—though related—industry: sports betting. On May 14, 2018, the federal ban that was in place in the United States which prohibited betting on the outcomes of sporting events was lifted, leaving states free to decide to pass legislation to legalize sports betting (Licata 2019). Since the ruling, 11 states have legalized sports betting—among them Nevada, Oregon, New York—and an additional 24 states are in the process of passing legislation (Licata 2019). With the recent legalization efforts, the derivative industry has witnessed enormous support: over $21 billion in total US handle (amount of money wagered) since 2018 (Legal Sports Report 2020). It goes without saying, then, that the ability to accurately predict sports results—especially with a mainstream league like the NBA—can be highly lucrative. This paper will describe our efforts to use existing game and player data in order to predict the result of future, arbitrary NBA games. We will focus on training a machine learning model using a season's worth of data as the input features. With high variance in team composition and individual player ability within the NBA from year to year, we believe it prudent to limit the input features to a single season, and use a model trained by such features to predict games from the playoffs for that season. We will take into account team statistics such as points per game, field goal percentage and turnovers. It is impor-

tant to note that many other—sometimes abstract—outside factors can impact the result of a game, such as the "home court advantage". To limit the scope of this project, we will limit our analysis to concrete statistics.

Related Work

The *Related Work* section should cite and summarize 3-5 papers related to the problem you are tackling.

It is not sufficient to summarize each prior work independently. You need to describe a coherent story that incorporates all of these prior work together. Compare and contrast them. Try to describe a story of how researchers have explored the topic in the past and what progress they have made so far.

Explain how your work is going to build on these prior work, i.e. how your work is similar to or different from the techniques that have been used.

The length of this section is typically 1/2 to 1 page.

Sources so far:

1. Predicting the Outcome of NBA Playoffs Based on the Maximum Entropy Principle
2. https://www.statista.com/statistics/240377/nba-finals-tv-viewership-in-the-united-states/
3. https://howmuch.net/articles/sports-leagues-by-revenue
4. https://www.legalsportsreport.com/sports-betting/revenue/

Methodology

In any given NBA matchup, numerous factors come into play to influence the game and its end result. To account for all these factors, we will be obtaining team and player data from website (Basketball Reference). This data will then be used to train our models. *Which algorithms are we using to train the models? Need 2. Why are we using these two? What is the performance of them?* Finally, we will compare the accuracy of the two techniques using a subset of past NBA games not used for training.

Results

Based on the papers weve read online, it seems at though none of the models have been able to predict NBA outcomes with an accuracy of much more than 75%. We do not anticipate to be able to predict outcomes with super high accuracy, but what we do hope to discover is if there is a certain type of algorithm that can train a model that is significantly better at predicting NBA outcomes than others based on the same input data.

References

[Amoros 2016] Amoros, R. 2016. Top professional sports leagues by revenue.

[Gough 2019] Gough, C. 2019. Nba finals average us tv viewership 2002-2019.

[Jones 2016] Jones, E. S. 2016. *Predicting outcomes of NBA basketball games*. Ph.D. Dissertation, North Dakota State University.

[Legal Sports Report 2020] 2020. Sports betting revenue tracker us betting revenue and handle by state.

[Licata 2019] Licata, A. 2019. 42 states have or are moving towards legalizing sports betting here are the states where sports betting is legal.