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 2019, an average of 15 million people—per finals game—watched Kawhi Leonard and the Toronto Raptors garner their first NBA championship[1]. The American-based league, at the age of 74, now sits in fourth place among all sports leagues in the world by revenue[2]. People from all over the world enjoy watching and cheering on their favourite teams, as well as purchasing jerseys and merchandise. But over the last few years, we have seen a rise in a separate—though related—industry: sports betting. With recent legalization efforts in a number of American states, the derivative industry has witnessed enormous support: over $21 billion in total US handle (amount of money wagered) since 2018[3]. It goes without saying, then, that the ability to accurately predict sports results—especially with a mainstream league like the NBA—is of high consequence. This paper will describe our efforts to use existing game and player data in order to predict the result of an arbitrary NBA game.

Related Work

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

Here are some examples of citations: (**?**; **?**; **?**).

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.