

## Introduction/Background

A couple of our team members are fans of the NBA and are interested in the idea of sports analytics. We as a group also wanted to create something that could potentially financially benefit us and from the combination of those two came our idea to look into different ways of beating the betting odds for NBA games.

## Problem Definition

We would like to find a way to correctly predict the outcome of NBA games, and specifically we would like to create a model that will determine the odds of a team beating the betting spread on those games. More broadly we would also like to look into finding ways to group good teams and bad ones using their statistical seasonal averages.

## Methods

### Unsupervised learning:

For unsupervised learning, we would want to cluster the teams or players that are likely to perform better and win. For this we would like to use GMM, as it will provide probability for each team, which is likely to go to playoffs or likely to perform better. We could also use K-means for unsupervised learning. But as K-means is a hard classifier, it might not be helpful while placing bets or looking for odds. Also, we would like to use hierarchical clustering to observe which teams and players come under an umbrella and how the model performs differently than K-means.

### Supervised Learning:

We noticed that Neural Networks is the most commonly used algorithm for sports betting and we would like to go in the same direction. As NN is the last topic in the class course, we want to explore other methods in the meantime. As our problem statement revolves around the odds or probability of a team winning (classification type), we would like to use SVM, Logistic Regression, or Naive Bayes approaches. Out of these approaches, we would like to use Logistic Regression as it will use the probability or Odds while classifying the output. Also, Naive Bayes approaches uses the probability of a team winning under the circumstances of events ( $x_1, x_2, x_3 \dots$ ). So, we would like to use these two methods and compare the outputs. SVM is a hard classifier, so it might be very helpful for our case.

## Potential results

Our primary focus is on beating betting spreads, so we would like results relating to the odds of teams beating that spread. Our unsupervised model has a different task and that is to group teams and players based on their statistical averages. The results for this will hopefully show groups of teams and players that end up performing better.

## Discussion

This project has a number of potential difficulties the first of which would be dealing with the data and the uncertain nature of sports betting. Sports are an area that is difficult to predict due to dealing with all of the variables that contribute to the success of a team and there are certain unpredictable aspects such as players getting injured. Not only is it hard to predict but the people creating these betting spreads usually use all the tools at their disposal to create even odds for both sides leaving it more difficult for us to predict one side or the other. Even with this all being the case this project could be very profitable for us should we succeed in creating a good model.

## References:

<https://towardsdatascience.com/machine-learning-for-sports-betting-not-a-basic-classification-problem-b42ae4900782>

<https://www.nbastuffer.com/analytics101/machine-learning-systems-in-sports/>

<https://blog.oursky.com/2019/11/26/machine-learning-applications-nba-predictions/>

<https://towardsdatascience.com/building-my-first-machine-learning-model-nba-prediction-algorithm-dee5c5bc4cc1>

<https://towardsdatascience.com/predict-nba-games-make-money-machine-learning-project-b222b33f70a3>

Use machine learning to bet on NBA playoff games (through spread, over/under, and/or pure team selection) based on predictors from the regular season.

Supervised - model knows results of playoff games (as well as regular season performance).

Unsupervised - model only trained with regular season data.-

From our study, we are hoping to learn how predictive an NBA team's regular season statistics are in forecasting its postseason performance. Additionally, we may uncover patterns in how

sports betting books tend to skew betting odds in ways that are inefficient. Ideally we hope that our model can identify betting lines that are mispriced and represent opportunities for profit.

We decided to narrow our study to only playoff games so that we would have a full season of data to train our model for each series of playoff games. However, we may still encounter difficulties if conditions change over the course of the postseason (roster changes, unexpectedly better or worse performance of players, etc.).