

Title: NBA Shot Prediction ('14-'15 Season)

Description: In the NBA, it's all about which team scores the most points at the end of the game. Being able to predict the result of every shot taken over the course of a game would be an extremely useful tool in predicting the overall outcome of a match. This project will focus on creating a model to predict whether a shot is successful or not based on the relevant statistics (True Shooting %, Field Goal %, 3-Point Field Goal %, etc) of the player taking the shot in tandem with the data of the selected shot (Distance of closest defender, distance from goal, etc).

Questions: The goal of the project is to determine whether or not we can accurately predict whether or not a shot was made given data we have on both the player and the shot itself. Of course, outcomes in sports are very hard to predict, but is there some kind of minimum accuracy we can achieve? How can we differentiate the results of a shot of players with similar stats taking similar shots? The data we're using for individual shots has information on how far the closest defender is when a shot is taken, but no details on the actual defender. How will this affect the accuracy of our model? Players certainly aren't all the same height, so it could be safe to assume that a 6'8" player 3 feet away would affect a shot more than a 6'3" player 3 feet away. Another question to consider is what kind of variable affects the ability for a player to make a shot the most?

As we move forward in this project, a thing to keep in mind would be to see what other insights could we gain from analyzing the data? Are there trends that develop within an individual season? Are certain shots attempted by star players more easy to predict than shots attempted by role/bench players?

Evaluation: Since this kind of prediction isn't exactly critical (we're not predicting the presence of cancer), we can simply use error rate to evaluate the performance of our model.

Baselines: We have discussed several possible baselines to use. Here are several:

1. Simply predict every shot as a miss. There aren't many players around the league that shoot over 50%, so starting with this baseline will most likely get us an accuracy of around 50% - 60%
2. Another baseline we can consider is using k-nearest neighbor classification. How many of a new data point's closest neighbors represent a made shot versus a missed shot?
3. A third one we considered is outputting the most popular class label. Given input data, how many similar shots taken by that specific player were a make? A miss? We'll take the one that happened the most as our prediction.

Data sets:

- <https://www.kaggle.com/dansbecker/nba-shot-logs>
- <https://www.kaggle.com/drgilermo/nba-players-stats-20142015>

Relevant research:

- <http://www.sloansportsconference.com/wp-content/uploads/2014/02/2014-SSAC-Quantifying-Shot-Quality-in-the-NBA.pdf>
- <https://arxiv.org/pdf/1609.04849.pdf>