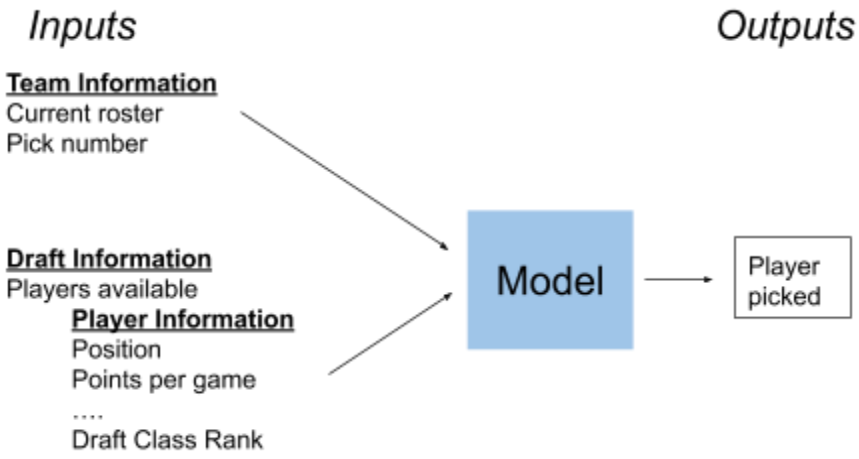
Emma McLean and Alex Goodman
Profs. Sadigh and Charikar
CS 221 Spring 2019
3 May 2019

Project Proposal

With the NBA season coming to an end, it's almost time to start gearing up for the next season. With the draft lottery coming up in just over a week and the draft following immediately after school is over, sixty new prospects will join the best basketball league in the world. Fans from non-playoff teams will be watching the lottery in the hopes that their teams is selected to pick first. Fans of all thirty teams will be watching the draft to see who gets the top prospects first and who gets snubbed without warning.

We want to predict how the draft will go. Around 200 players enter the draft each year (1), and we want to model how NBA teams choose their picks when they get their time on the clock. To do this, we would like to be able to feed in a list of rosters for all of the teams as of draft day along with a draft order dictating which team picks when, and receive as output the order in which players will be drafted. Another input feature we may add is the player's rank within his draft class by a subjective analysis like the ones ESPN does. At each step of the draft, we want to predict which player the team will take based on player statistics, including their rank, and team statistics, like which positions they need to fill. An example input and output is shown below to make this clear:



To make the above picture concrete, consider the example of a likely scenario from this upcoming draft:

Inputs:

Team: New York Knicks, pick no. 1

Draft: Players available

Zion Williamson, Duke, Rank 1, 22.6 PPG, …

….

Ja Morant, Murray State, Rank 2, 24.5 PPG, …

Output: Zion Williamson

This seems to be a relatively hard problem to solve, especially using our success metric of percentage of picks predicted correctly. We would also like to take a different approach than other people who have tried in the past. We were able to find a few reflex-based models (machine learning classifiers) that achieved just below 80% accuracy on the final test set, but just for the first round of the draft (2). Second-round accuracy drops off precipitously, which we found in our baseline model as well. Our baseline just predicted based off of draft class ranking and achieved significantly lower accuracy than our oracle of aggregated opinions from professional sportscasters. This gap increased slightly in the second round, but both the baseline and the oracle became much worse in the second round of the draft. This is likely because the first half of the draft is usually full of players who were consensus draft picks. In other words, most of the first 30 picks knew they would be picked in the draft, whereas in the second round most players are not positive about securing a spot on an NBA team.

We think that a search task might be better suited to the NBA than a reflex-based method because of the domain knowledge and changing state of the draft (for example, teams tend to have ranked lists of whom they want to draft, but what happens if their favorite player gets picked earlier and their second- and third-favorite players would not get along well?). Another article we found used unsupervised data mining techniques but again was not very successful, especially in the second round of the draft (3).

(1) https://watchstadium.com/news/up-to-date-list-of-college-basketball-players-who-declared-for-2019-nba-draft-04-01-2019/
(2) https://cornerthree.net/2018/06/12/applying-machine-learning-techniques-to-the-nba-draft/
(3) https://towardsdatascience.com/dissecting-the-nba-draft-part-2-79b6bd486a8d