NBA Prospect Projections Proposal

Micaleb Johnson, Morgan Williams, Elias Cheleuitte

1. Specific Aims

1.1 Importance

The NBA, and sports in general, have a unifying effect on society bringing people who otherwise have nothing in common together. This has led the NBA to become a billion dollar industry where the primary goal of each team is to win. Each year the NBA holds a draft where each team is able to select the most promising prospects to add to their roster in hopes of developing these players into ones who contribute to a winning team. Given the limited number of draft picks, NBA teams spend a lot of time and money, each year, on scouting prospective players for the upcoming NBA Draft. Despite the large amount of resources that are put into finding out which players to draft, it is still an imperfect science. Oftentimes players who scouts deem as a future franchise player, fall out of the league within a few years and other prospects who fly under the radar end up becoming some of the league's top stars.

1.3 Aim 1

Our first aim is to determine the relative value of each of the first round draft selections. For instance, finding out whether a top 5 pick really is worth significantly more than one of the last 5 picks. This will allow us to observe when a team picks someone that outplays or underplays their draft position. We will also be able to assess trades involving teams' draft picks, determining which team received the most value in a deal.

1.3 Aim 2

Our second aim is to accurately predict the success of an NBA draft prospects career, by using player data that is available prior to them being drafted and their draft position. We hope to determine which factors are most and least influential in predicting NBA success and find the ideal athlete profiles for each draft position to result in the most successful NBA career.

1.3 Aim 3

Our third aim is to analyze the success of teams in the drafts of the past 25 years. Based on this analysis we will look at how this success or lack thereof has translated to the teams' overall on-court performance.

2. Research Strategy

2.1 Conceptual Framework and Hypotheses

It is vital to an NBA team's success to select good players in the draft. All drafted players have team friendly "rookie" contracts that are inexpensive and allow for the team to have the rights to retain players for up to 9 seasons before they are able to become unrestricted free agents. The importance to select star level players in the draft holds especially true for the league's small market teams, whose only realistic chance to build a championship caliber team is through the draft as the league's top stars almost never consider playing for these teams when they become free agents.

Our objective is to be able to consistently and successfully predict which prospects have the best chance of becoming good NBA players as well as determine how different pick positions should be valued. We will use data on all NBA players drafted in the first round between 1995-2015 to perform initial exploratory data analysis prior to training and testing our models with this data. This will allow us to assess each factor included in our dataset before using them in our prediction models. Upon finalizing our player value prediction model, we want to use the model to predict the value of the players who have been drafted from 2016-2020. For determining how well teams are at drafting we will use team data over the same time frame (1995-2015), we will then explore the relationship between the success of the team, indicated by a team's win percentage, and the success of their drafts, gauged by the value their drafted players have achieved relative to the projected value of the pick where they were taken.

We will define a player's value, as an estimation of a player's individual success. This will be a continuous variable made up of the player's cumulative statistics including win shares, box plus minus (BPM), and value over replacement player as well as non cumulative statistics such as win shares/48 minutes. In relation to Aim 1, we will define draft pick value based on the expected player value each pick should yield.

A player's win shares attempts to divide up a team's total wins among its players, based on the impact each individual player has on a win. This statistic can be positive or negative, while a player with positive win shares is seen as adding wins to his team and a player with negative win shares can be seen as a player who is taking away wins that his teammates otherwise would have generated. Win shares/48 minutes, is the same thing except the win shares are standardized based on how many win shares a player generates for every 48 minutes (full game) they play. BPM estimates each player's contribution in points added above league average per 100 possessions played. The league average is a 0, meaning below average players will have a negative BPM and, in general, a superstar player at their peak will record a BPM of above 8. The Value Over Replacement Player stat measures a player's overall contribution to their team vs. the contribution that replacement player (A low level player who makes the minimum salary) would be expected to make.

Hypothesis 1:

The draft position will have a negative effect on player value, as the position gets higher (later in the draft) the average player value will be lower.

Hypothesis 2:

We expect teams that have more success in drafting players to win more games than teams who draft poorly; this implies that good drafting teams will pick later on in the draft as draft order is largely determined by a team's win-loss record.

2.2 Data Collection

We will collect data from a variety of sources. Two datasets from data.world, containing the college and NBA stats of drafted players, will be merged together. Some of the data in these datasets are not up to date so we will scrape Basketball-reference.com to get updated NBA statistics for the players and merge this with the other data.world datasets. We will also possibly scrape NBA combine data from NBA.com and add this to our previous dataset. Team data will also need to be collected from Basketball-reference.com so that we can determine the relationship between how well an NBA team drafts and how successful the team is on the court.

Our dataset contains 623 players drafted in the date range 1995-2015. Each player's statistics include their physical measurements (age, height, weight, position), NCAA career statistics, draft year and pick, and their NBA statistics.

2.3 Statistical Analysis

Method 1:

We will use a parametric regression model to predict which factors from our dataset of draft prospects are the most influential when trying to predict player value of NBA draft prospects. We will define success through various advanced statistics from our NBA data.

Method 2:

We will use non-parametric regression or a regression tree to model the value of players drafted into the NBA. Therefore allowing us to analyze both the value of each draft position and player's probable value.

3 References

[1] https://www.basketball-reference.com/draft/

[2] https://data.world/bgp12/nbancaacomparisons/workspace/file?filename=players.csv

[3] https://data.world/gmoney/nba-drafts-2016-1989/workspace/file?filename=NBA+Drafts.xlsx

[4] https://data.world/bgp12/nbancaacomparisons/workspace/file?filename=players.csv

[5] https://data.world/gmoney/nba-drafts-2016-1989/workspace/file?filename=NBA+Drafts.xlsx

[6]https://github.com/SidTheKid007/NBARookieAnalysis/blob/master/CollegeRookieStatLog4.csvhttps://stats.nba.com/draft/combine-strength-agility/?SeasonYear=2002-03&sort=LANE_AGILITY_TIME&dir=-1