Names: Reid Whitson and Luc Latiolait

Project Name: NBA Rookie Points Predictor

**Dataset description:**

Our Rookie Points Predictor will take all of the players in the NBA drafted in the first round from 1970-2018 and their associated stats and then given a new player who is supposed to go in the first round, we will predict how many points they will score. We are using a dataset from basketball-reference.com that has all of these data points. This dataset is very extensive and we will be taking a few of the columns which we think have the greatest impact on the score. As is, the dataset size is 1271 x 27; we hope to narrow down the columns to statistics that we see as having more direct impacts on players points per game (PPG), and possibly add data such as physical attributes and more college stats.

**Implementation/technical merit:**

Narrowing down the data will be a large part of this project, and is part of the pre-processing phase. I anticipate that we may run into issues when deciding what is most important with which attributes to keep and which to get rid of. Also, we may want to add additional data such as physical attributes like height and weight, and certain stats from the players’ college careers. Which attributes we choose to include in our classifier will be very important as there is much debate about what measures can indicate a player will transition well from college to the NBA. As a result, if we choose bad attributes, we could get very inaccurate results. In order to help select attributes, we may run regression analysis on the various attributes to find which have the highest correlation. We are planning on using a KNN classifier, since most of the relevant attributes are continuous, and then comparing it to the results of sci-kit learn and discussing the differences. The reason we choose to compare our results to sci-kit learn is because we both wish to learn more about it and believe this is a good way for us to dive into the library and analyze the results between it and a homegrown tests.

**Potential impact of the results:**

In recent years, NBA teams have increasingly focused on the draft as a vital opportunity for building and maintaining a successful roster, with some teams going as far as purposefully losing to improve their draft position. Draft picks have become a such a valuable commodity for NBA teams both because they allow small market teams to recruit potential star players who would likely not have chosen to play for them otherwise, and because players obtained through the draft can be paid significantly less than NBA veterans, meaning that teams that draft impactful players will more financial resources to improve the rest of their team. However, wasting a draft pick on a player who is not become an impactful NBA player can be very damaging because the limited roster spots teams have and the opportunity cost of trading that pick before the draft for other assets. For these reasons, drafting a player is a high stakes gamble that often has long term impacts, positive and negative, on the success of NBA franchises. With our project, we will attempt to predict college players’ rookie year impact (as measured by PPG) using data available to NBA scouts on draft night. The stake holders for this project would be NBA scouts and executives, who would stand to gain a significant advantage over their competition by consistently maximizing the value of their draft picks.