

Final Project Data Memo

Daniel Pry, Rick Zheng, Nick Van Daelan

1/23/2022

Data Memo

An overview of your dataset

What does it include? Our data set includes ncaa basketball player statistics.

Where and how will you be obtaining it? Include the link and source.

We got the data from kaggle. (<https://www.kaggle.com/adityak2003/college-basketball-players-20092021>)

About how many observations? How many predictors? It has 25719 observations with 65 predictors.

What types of variables will you be working with? The variables are numerical.

Is there any missing data? About how much? Do you have an idea for how to handle it? There is some data missing but I think this is due to some players not having relevant data for that observation. For example one observation is Recruit Rank coming out of high school and there is a good amount of missing data but this is likely because not all the players were ranked. There are also some other observations with similar omissions of data for similar reasons. These missing data points will have to be taken into consideration when creating the model.

An overview of your research question(s)

What variable(s) are you interested in predicting? What question(s) are you interested in answering? We are trying to use the observations to predict which players will get drafted into the NBA.

Name your response/outcome variable(s) and briefly describe it/them. The response variable is whether or not the player gets drafted or not.

Will these questions be best answered with a classification or regression approach? Regression.

Which predictors do you think will be especially useful? I think points, blocks, assists, total rebounds, and recruit rank will be especially useful.

Your proposed project timeline and group work How is your group dividing up the work? We will meet up while working on it and depending on what we get done each time we meet up maybe assign out work each person should do before the next time we meet.

When do you plan on having your data set loaded, beginning your exploratory data analysis, etc? Later this week.

Any questions or concerns

Are there any problems or difficult aspects of the project you anticipate? None that I can think of right now but I imagine we will once we start working on it.

Any specific questions you have for me/the instructional team? None right now.