**Motivation:**

The NBA draft is widely regarded as one of the most pivotal moments in the basketball world each year, if not the most crucial. Players from all corners of the globe converge at the draft, hoping to seize the opportunity to become a part of the prestigious National Basketball Association (NBA) in the United States.

Historically, the primary talent pool for the NBA has been the North American colleges. Many of the greatest players of all time have emerged from universities within the borders of the USA. However, this trend has started to shift in recent years. A growing number of talented and promising players from continents such as Europe and Asia have made their mark in the NBA draft and experienced notable success.

The question of who will be selected by which team in the current NBA draft evokes tremendous excitement among league followers and basketball enthusiasts alike. It isn't merely an intriguing query for fans; it represents a significant and potentially life-altering decision for NBA teams, as they strive to identify the most suitable young talents among the vast pool of thousands, if not tens of thousands, of basketball players. Prior to the draft night, NBA teams invest an unimaginable amount of effort and time into meticulous analysis and research, examining every potential candidate and uncovering details about their skills and abilities on the court. While intuition certainly plays a role in this process, teams also rely, to some extent, on statistical analysis because "numbers don't lie."

As a true basketball lover and someone deeply interested in the inner workings of the NBA and its teams, I have taken it upon myself to dive into player statistics and attempt to predict which individuals, based on their numbers, have the potential to be selected in the NBA draft. Currently, my analysis is focused solely on statistical data from college basketball, which means it doesn't cover all the draft picks, as some players come from other countries and leagues. For example, there are players who compete in the EuroLeague, widely regarded by some as the most challenging basketball championship in the world, surpassing even the NBA in terms of difficulty.

Looking ahead, it would be advantageous to expand the data sources and integrate statistics from players who are not affiliated with North American college basketball, but rather showcase their skills at a young age while playing for top European or Asian teams. This broader approach has the potential to provide more precise and accurate predictions in the future.

**Research questions:**

Building upon the motivations mentioned in the previous chapter, I aim to predict the players who will be selected in the NBA draft. This prediction task raises several intriguing questions that I intend to address throughout my research:

Are there any statistics that can be considered as a baseline model for the prediction task? Can a model be constructed that outperforms the baseline model? If so, what level of accuracy can be achieved by such a model? What kind of machine learning algorithm can be used for modelling the data?

I am eager to uncover satisfying answers to all the aforementioned questions as I progress with my research.

**Background:**

…

**Exploratory Data Analysis:**

References:

<https://sci-hub.se/https://doi.org/10.1007/s11123-010-0187-x>

<https://medium.com/@data.science.enthusiast/how-to-improve-logistic-regression-in-imbalanced-data-with-class-weights-1693719136aa>

<https://www.kaggle.com/datasets/adityak2003/college-basketball-players-20092021>

<https://www.basketball-reference.com/about/per.html>