**NBA Draft Combine Measurements and Early Career Performance Model**

Kyle Ke

Data Science for Sports

**Objective:**

The objective of the project is to use raw NBA draft combine measurements to predict overall performance of a player during their rookie contract period (year 1 -year 4) and performance during rookie year only. It also serves to provide inference on which anthropometric and strength/agility measurements are most associated with player performance.

**Model or Analysis Type:**

* We are going to use regression predictive modeling to predict the average BPM of players during rookie contract period as well as BPM for rookie season only.
* The response variable is the averaged BPM or rookie season BPM of the players. The response variables are raw NBA draft combine measurements and other features deemed necessary.
* The players are going to be divided into three groups (guard, wing, big man). Three separate models are going to be constructed, one for each of the three groups.
* The inferential part of the project is going to be performed through linear regression coefficients or interpretable machine learning with SHAP scores.

**Data Requirements:**

* NBA draft combine raw measurements from 2000 to 2019. Source: [https://docs.google.com/spreadsheets/d/10KairkGJLeB\_FsDbZB\_om\_az8jwuZCQZEEEWfE-Gt4o/edit#gid=0&fvid=208595013](https://docs.google.com/spreadsheets/d/10KairkGJLeB_FsDbZB_om_az8jwuZCQZEEEWfE-Gt4o/edit%23gid=0&fvid=208595013)
* NBA player advanced statistics from 2000-2001 season to 2020-2021 season.

Source: <basketball-reference.com/>

**Projected Timeline:**

* September: Scraping, joining, and cleaning data
* October: EDA, feature engineering, and feature selection
* November: Model construction and selection
* December: Conclude which combine measurements are most associated with performance. Prepare and finalize report and presentation.