CSCI 184 Final Project Proposal

***Group Members:***

Ryan Chappelle, Michael Nocerino, Esteban Samayoa Mayora, Dhwanil Mehta

***Project Title:***

Professional Basketball (NBA) Player Worth classifier/

Overpaid or underpaid NBA player classifier



***Dataset:***

Michael has a dataset containing the stats of individual NBA players (per game stats and advanced stats) from the 2022-2023 NBA season. The dataset involves two joined tables from basketballreference.com as well as a player salary column which has been added by Michael. The .csv for this dataset has been linked in the submission. It contains season stats for over 400 players and over 40 features describing the play of these basketball players. We hope to be able to perform some meaningful analysis and feature engineering to find new insightful ways to evaluate player value and worth using machine learning.

***Project Idea:***

Our project idea is to create and compare several models that classify NBA players as either undervalued or overvalued. The general process might follow these steps. First, we would determine what features in the dataset contribute to determining a player’s statistical value. Next we would need to determine what makes a player undervalued and overvalued. Knowing this, we can then begin to create several classification models such as random forest or naive bayes to classify players.

Now that we can classify players, we can apply this to many different aspects of the NBA. One idea is to determine if a player is overpaid or worthy of a max contract. Another idea would be to build an ideal team of players that are undervalued and could be acquired for cheap. A final idea is to recognize players on a current team that are overvalued, and determine which free agent player of the same position could replace the current player. Ultimately there are many different applications once the models are built, and we will be able to compare several models in order to achieve an optimal model.

***Software Needed***: Python, Pandas, Matplotlib, Seaborn, Sklearn, Tensorflow, Keras, Pickle

***Papers to Read***:

# Optimizing NBA Player Selection Strategies Based on Salary and Statistics Analysis: [Optimizing NBA Player Selection Strategies Based on Salary and Statistics Analysis | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/8328519)

# Estimating NBA players salary share according to their performance on court: A machine learning approach: [[2007.14694] Estimating NBA players salary share according to their performance on court: A machine learning approach (arxiv.org)](https://arxiv.org/abs/2007.14694)

# Are NBA Players’ Salaries in Accordance with Their Performance on Court?: [Are NBA Players’ Salaries in Accordance with Their Performance on Court? | SpringerLink](https://link.springer.com/chapter/10.1007/978-3-030-85254-2_25)