

Colton Storebo and Milo Libby

Gonzaga University

4/13/2022

Gina Sprint

Data Science Algorithms

Github link -> https://github.com/cstorebo/Data_Science_Algorithms_Final.git

Project Proposal

Data Description:

Source, format Contents:

Source:

<https://www.kaggle.com/datasets/willyiamyu/nba-2k-ratings-with-real-nba-stats?resource=download>

Format: csv

“This dataset contains the NBA 2K Ratings for each player (last column), as well as their corresponding real-life NBA stats for that season. Since NBA 2K generates ratings for players before the start of each season, we will join the player's previous season's stats with the game. For example, NBA 2K21 ratings are joined with player stats from the NBA 2019-20 season.

Data starts from NBA 2K16 (2014-15 season), up to NBA 2K21 (2019-20 season).”

With this dataset we are going to take an instance's team, age, GP(games played), W(wins), L, Losses and PTS(average points per game) to form the training set. Our goal is to try and predict players ranking based on the attributes stated.

Implementation/technical merit:

1. The dataset is in a csv file making it very easy to clean and manipulate the dataset as needed. We will load the dataset into a MyPyTable(). The dataset is large so we will discard any features with missing values.
2. One feature selection method that we can explore using is sklearn's univariate feature selection

Potential impact of the results:

A potential impact of the results could be showing some insight to the rating process in the video game. We could use the classifier to see how certain players ratings would be adjusted if the ratings were true to the dataset. (Ex. Does LeBron James deserve his 97 rating based on his

stats? Should Ja Morant be higher than 85? Does age prove to be a strong determining factor of rating?)

The stakeholders are the 2K game developers as we are analyzing their game and decision in the rating process. Also, one of the NBA players could be interested in these results.