

# Prediction of NBA 2K Player Ratings

Bhavik Jain, Aarsh Patel

21 March 2017

## 1 Problem

In order for 2K Sports to generate player ratings, they need to go through a tedious process. In order to expedite this process, we wanted to build a classifier that can easily predict a player's 2K rating based on their in season game stats from the past year. Being able to correctly rate and balance NBA players is tough, and thus we felt this problem was important for us to tackle.

## 2 Methodology

In order to build a classifier to predict 2K player ratings, we need to represent a player by some feature vector corresponding to their in-game season stats. These in-game stats would include features such as **points-per-game**, **assists-per-game**, **minutes-played-per-game**, etc. Thus a player would be represented by these features and the target would be the 2K rating for that player in that year. The goal would be to learn  $f : X \rightarrow Y$  that can predict a player's 2K rating (0-99) based on their stats. This would be considered a regression problem where we are trying to minimize the overall MSE. Initially, we want to use simple classifiers such as KNN and Linear Regression. After we can think about using more sophisticated methods such as Decision Trees, Random Forests, or SVR's. In order to apply these algorithms to our problem, we will be using python's scikit-learn library. We will target any hardware that supports python.

## 3 Data Sets

The data set we will be using for this project will need to be created. For each player there will be season stats, advanced season stats, as well as previous NBA 2K ratings. The data will go back from 2010 to present. This is an example of a data set we will be using for player stats: "[http://www.basketball-reference.com/leagues/NBA\\_2010\\_per\\_game.html](http://www.basketball-reference.com/leagues/NBA_2010_per_game.html)". This is an example of a data set we will be using for NBA 2K ratings: "<http://thereal2kinsider.blogspot.com/2009/09/nba-2k10-player-ratings.html>". To get the data from basketball-reference, we will be going on their website and just downloading the CSVs with the stats, then to get the 2K data, we will be writing a web crawler to parse all the data into a CSV. The next step will be then to combine all of this data into one CSV.

## 4 Experiments

- Use the baseline model to predict 2K player ratings based on all features
- Performing feature selection and finding which features contribute most to the overall 2k player rating
- Using the subset of optimal features found, use these features as input to the baseline Linear Regression model and predict 2K ratings and compute the MSE.
- If we see an improvement then we can use these features as input to our other models (RandomForestRegressor, Ensemble, etc.)
- If we don't see an improvement then we use all of the features for the other models (RandomForestRegressor, Ensemble, etc.)
- Try to come up with a linear combination of the features to represent a player's 2k rating
  - Based on the weights of these features, we can see exactly which stats are important
- Try to see if offensive stats or defensive stats contribute most to a player's 2k rating

## 5 Related work and Novelty

There has been another project done which uses NBA player stats to predict 2K ratings. Our project will use more advanced stats for these players to predict their ratings, basketball-references has a place for advanced players stats which show more detailed information about their performances throughout the year, which will help in predicting a more accurate rating. The next thing we will do is also set a minimum game limit on the players, so if they play less than 41 games in a season, their data/stats will not be used for the purposes of this data set, we will also be doing feature pre-processing. We may later try to determine a linear equation using stats and weights for the prediction of NBA 2K ratings.

## 6 Collaboration Plan

**Aarsh:**

- Collect NBA player/2k player ratings from 2013-2016.
- Create a train-validate-test split for the dataset
- Build a simple baseline model using Linear Regression
  - Using all of the features
  - No hyperparameter optimization
- Brainstorm and build an ensemble model for this data set
- Figure out which features contribute most to a player's 2k rating
  - Do offensive stats contribute most to a player's rating or do defensive stats contribute most

**Bhavik**

- Collect NBA player/2k player ratings from 2010-2013.
- Train and predict using random forest classifier for this data set
- Brainstorm and build an ensemble model for this data set
- Linear Equation for 2k player rating prediction
  - Find which features make a bigger difference on the prediction of the 2k player rating.

## References

- [1] "Basketball Statistics and History." Basketball-Reference.com. N.p., n.d. Web. 21 Mar. 2017.  
<http://www.basketball-reference.com>
- [2] "ESPN NBA Stats." ESPN. ESPN Internet Ventures, n.d. Web. 21 Mar. 2017.  
<http://www.espn.com/nba/statistics>
- [3] Khan, Ebran. "Advanced NBA Stats for Dummies: How to Understand the New Hoops Math." Bleacher Report. Bleacher Report, 25 Jan. 2017. Web. 21 Mar. 2017.  
<http://bleacherreport.com/articles/1813902-advanced-nba-stats-for-dummies-how-to-understand-the-new-hoops>
- [4] "NBA 2K Player Ratings Predictor." NBA 2K Player Ratings Predictor. N.p., n.d. Web. 21 Mar. 2017.  
<http://http://adamhe.me/2K-ML/>
- [5] Rappaport, Max. "Game Recognize Game." Complex. Complex, 20 Sept. 2016. Web. 21 Mar. 2017.  
<http://www.complex.com/sports/2015/09/how-nba-2k16-player-ratings-are-determined>