



# Projecting 3-Point Shooting for NBA Draft

Roland Centeno, Hilary Sun, Jerold Yu

rcenteno, hsun3, jeroldyu@stanford.edu  
Stanford University, CS 229

## Introduction

With the rise of the three-point shot in NBA basketball, more emphasis has been placed on shooting and spacing the floor. Consequently, teams are starting to value players who are proficient three-point shooters. The goal of this project is to determine the best three-point shooters among top NCAA prospects by predicting 3-point NBA percentages from NCAA/NBA team and player statistics.

## Data

All data was scraped from Basketball Reference and Sports Reference. We selected players from nbadraft.net's Top 100 Prospects Big Board (which dates back to 2009) that met several three-point shooting benchmarks (e.g. player must have taken at least 30 threes in the NCAA).

## Features

We had 7 different features, such as the player's number of 3-point field goals made and attempted in the NCAA. We normalized the number of threes that an NBA team attempted with respect to all the teams in the NBA because if a player plays on a team that takes more threes, he may shoot more threes later as well. We projected our data into a four-dimensional feature space using principal component analysis. We predicted that some variables were correlated with each other, so this way we could reduce the dimensionality of our data.

## Models

We first split our data of 149 players into 80% train and 20% test. We then used k-fold validation with k = 10 to select our best model.

### Linear Regression

We used linear regression as a baseline.

$$h(x) = \theta^T x \quad x \in \mathbb{R}^{N \times 4}, \theta \in \mathbb{R}^4$$

### Weighted Linear Regression

We used weighted linear regression to account for features of more importance, using the reciprocal of the variances as the weights. We maximized:

$$\sum_i w^{(i)} (y^{(i)} - \theta^T x^{(i)})^2 \quad x \in \mathbb{R}^{N \times 4}, y \in \mathbb{R}^N, w \in \mathbb{R}^4, \theta \in \mathbb{R}^4$$

### Gradient Boosting Regression

We decided used gradient boosting decision trees to expose any nonlinearities in our data. It employs gradient descent to minimize for each weak model  $F$  at each stage:

$$\sum_i (y - F(x))^2$$

### Random Forest Regression

In addition to gradient boosting, we used a random forest regression. Each stage minimizes the mean-squared error:

$$\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

## Results

Model	Training MSE	Test MSE
Linear Regression	9.141	9.983
Weighted Linear Regression	9.202	9.947
Gradient Boost	10.061	13.056
Random Forest	5.276	11.792

Based on k-fold, we chose to run weighted linear regression on our full training data: our training MSE was 9.161 and our test MSE was 7.838. The decision trees did not work

as well because they may have tried to overfit our data, especially since our dataset was only 149 players, with 119 in the training and 30 in the test. Outliers could have affected our results, such as Stephen Curry. We also only chose top NCAA prospects as part of our dataset, so our model would probably not work so well for less proficient players. We found that it's difficult to predict a player's 3-point performance just based on college statistics since players change their games and adapt to the NBA environment.

## Future Work

As basketball expands globally, international athletes are joining the NBA from leagues outside the NCAA. One area of interest we have is finding a way to account for these players in our model. In addition, there are players who did not meet the college shooting volume requirement for our dataset that became shooting threats during their NBA career. Finding other parameters to predict their performance would be another interesting avenue for exploration (e.g. shooting ability from mid-range).

## References

- [1] Abrams, William, Barnes, John C., and Clement, Annie. "Relationship of Selected Pre-NBA Career Variables to NBA Players' Career Longevity," *The Sport Journal*, April 2008.
- [2] Coates, Dennis and Oguntimein, Babatunde. "The Length and Success of NBA Careers: Does College Production Predict Professional Outcomes?" August 2008.
- [3] Gajewski, Byron J. and Bishop, Todd. "Drafting a Career in Sports: Determining Underclassmen College Players' Stock in the NBA Draft," June 2004.
- [4] Harris, Jill and Berri, David J. Predicting the WNBA Draft: What Matters Most from College Performance?", *International Journal of Sport Finance*, vol. 10, no. 4, 2015.