Generating Relative Draft Value in the NBA Draft and Predicting Success from College Basketball

A Major Qualifying Project Report:

submitted to the faculty of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

degree of Bachelor of Science

by

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Michael Krebs

­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Jake Scheide

Date:

Approved: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Professor Craig Wills, Major Advisor

# Abstract

# Acknowledgements

# Table of Contents

[Abstract 1](#_Toc885873)

[Acknowledgements 1](#_Toc885874)

[Table of Contents 1](#_Toc885875)

[Table of Figures 2](#_Toc885876)

[Executive Summary (*Jake, Later)* 2](#_Toc885877)

[1. Introduction *(Later)* 2](#_Toc885878)

[2. Background (*Mike & Jake)* 3](#_Toc885879)

[3. Methodology (*Jake)* 3](#_Toc885880)

[3.1 Analyze existing basketball player performance metrics 3](#_Toc885881)

[3.2 Feature engineer new player performance metrics addressing shortcomings with existing metrics 3](#_Toc885882)

[3.3 Find the highest value picks based on various measures of cost 3](#_Toc885883)

[3.4 Calculate the approximate value of every pick in the NBA Draft 3](#_Toc885884)

[3.5 Create a Jimmy Johnson-style NBA Draft value chart 3](#_Toc885885)

[4. Design (*Mike)* 3](#_Toc885886)

[5. Results (*Jake)* 3](#_Toc885887)

[5.1 Analyze existing basketball player performance metrics 4](#_Toc885888)

[5.2 Feature engineer new player performance metrics addressing shortcomings with existing metrics 5](#_Toc885889)

[5.3 Calculate the approximate value of every pick in the NBA Draft 5](#_Toc885890)

[6](#_Toc885891)

[5.4 Find the highest value picks based on various measure of cost 7](#_Toc885892)

[5.5 Create a Jimmy Johnson-style NBA Draft pick value chart 8](#_Toc885893)

[6. Methodology for NCAA (*Jake)* 10](#_Toc885894)

[6.1 Create a model which predicts various measures of NBA success based on NCAA DI statistics 10](#_Toc885895)

[7. Design for NCAA (*Jake)* 11](#_Toc885896)

[8. Results for NCAA (*Mike)* 11](#_Toc885897)

[9. Discussion and Future Work (*Mike)* 11](#_Toc885898)

# Table of Figures

# Executive Summary (*Jake, Later)*

# 1. Introduction *(Later)*

# 2. Background (*Mike & Jake)*

# 3. Methodology (*Jake)*

## 3.1 Analyze existing basketball player performance metrics

In professional sports, ‘value’ can be quantified in many ways. Some measures look purely at statistical output, whereas others take factors such as contract cost, minutes played, and team wins into account. To contextualize our entire project, which involves measuring the performance of basketball players, we analyzed the common metrics used to evaluate players. These four metrics were Player Efficiency Rating (PER), Win Shares (WS), Value over Replacement Player (VORP) and Fantasy Points (FP).

## 3.2 Feature engineer new player performance metrics addressing shortcomings with existing metrics

After analyzing the existing player performance metrics, we identified potential areas for improvement with different metrics that allowed for a more accurate comparison of players in the same season. These metrics were called Basic Percentile (BP) and Advanced Percentile (AP). Additionally, we created a metric which rewarded recognition rather than statistical output, called Cumulative Individual Accolades (CIA).

## 3.3 Find the highest value picks based on various measures of cost

One of the most important applications of talent evaluation is the NBA Draft. Each of the thirty teams are assigned two picks, generally in inverse order of team wins. A lottery is conducted for the first fourteen picks, to disincentivize intentional losing of games (commonly referred to as ‘tanking’) to obtain a highly talented player with the first pick. The NBA rookie salary scale provides an approximation of the talent level available at each pick, which we use with the performance metrics to find the draft picks which provide the highest output per dollar.

## 3.4 Calculate the approximate value of every pick in the NBA Draft

Another possibility in the NBA Draft is pick trading. Both before and during the draft, teams can swap picks for players or even high picks for multiple lower picks. As such, knowing the value of each position in the draft is critical to teams trying to improve their talent. We use the performance metrics to analyze the drop-off in talent at each pick in the draft.

## 3.5 Create a Jimmy Johnson-style NBA Draft value chart

Pick trading is far more common in the National Football League (NFL) where there are 224 picks between 32 teams. NFL Analyst Jimmy Johnson created a draft chart in the early 1990’s which seeks to quantitatively evaluate the talent available at each pick. We apply this to the NBA and create a value chart which accurately matches past draft pick trades in the NBA.

# 4. Design (*Mike)*

# 5. Results (*Jake)*

## 5.1 Analyze existing basketball player performance metrics

As discussed in the methods section, a crucial decision in evaluating player value is how ‘performance’ is quantified. The below table lists the top 20 players ranked using the four existing metrics, averaged out over the course of each player’s career.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Player | WS | PER | VORP | FP | AVG |
| LeBron James | 1 | 2 | 1 | 1 | 1.3 |
| Karl Malone | 2 | 4 | 2 | 2 | 2.5 |
| David Robinson | 4 | 1 | 4 | 5 | 3.5 |
| Tim Duncan | 8 | 6 | 9 | 4 | 6.8 |
| Chris Paul | 5 | 7 | 5 | 11 | 7.0 |
| Kevin Durant | 7 | 8 | 15 | 7 | 9.3 |
| Shaquille O’Neal | 14 | 3 | 18 | 3 | 9.5 |
| Michael Jordan | 3 | 11 | 3 | 21 | 9.5 |
| Charles Barkley | 10 | 5 | 6 | 21 | 10.5 |
| Russell Westbrook | 21 | 16 | 8 | 6 | 12.8 |
| Kevin Garnett | 17 | 17 | 12 | 8 | 13.5 |
| John Stockton | 6 | 13 | 19 | 16 | 13.5 |
| Hakeem Olajuwon | 21 | 10 | 17 | 10 | 14.5 |
| James Harden | 12 | 21 | 11 | 17 | 15.3 |
| Clyde Drexler | 16 | 19 | 7 | 21 | 15.8 |
| Stephen Curry | 15 | 21 | 10 | 19 | 16.3 |
| Kobe Bryant | 20 | 15 | 21 | 13 | 17.3 |
| Dirk Nowitzki | 13 | 18 | 21 | 18 | 17.5 |
| Magic Johnson | 9 | 21 | 21 | 21 | 18.0 |
| Dwight Howard | 18 | 21 | 21 | 12 | 18.0 |
| Yao Ming | 21 | 9 | 21 | 21 | 18.0 |
| Allen Iverson | 21 | 21 | 21 | 9 | 18.0 |
| Jason Kidd | 21 | 21 | 16 | 15 | 18.3 |
| Dwyane Wade | 21 | 12 | 20 | 21 | 18.5 |
| Reggie Miller | 11 | 21 | 21 | 21 | 18.5 |
| Scottie Pippen | 21 | 21 | 13 | 21 | 19.0 |
| Larry Bird | 21 | 21 | 14 | 21 | 19.3 |
| Anthony Davis | 21 | 14 | 21 | 21 | 19.3 |
| Gary Payton | 21 | 21 | 21 | 14 | 19.3 |
| Jeff Hornacek | 19 | 21 | 21 | 21 | 20.5 |
| Amare Stoudemire | 21 | 20 | 21 | 21 | 20.8 |
| Patrick Ewing | 21 | 21 | 21 | 20 | 20.8 |

Starting at the top, we can see that there’s a reasonable consensus among the top three players. Beyond that, the metrics begin disagreeing quite significantly. For example, Michael Jordan earns third place in Win Shares and VORP, but doesn’t feature in the top 20 for Fantasy Points. Because Win Shares distributes production by the number of wins the teams accrues, players on successful teams (such as the 90’s Bulls, arguably the greatest team ever) will feature strongly in the WS rankings. Similarly, Magic Johnson’s extremely strong Lakers teams boosts his WS rank to 9, which is the only time he features in these standings.

Extrapolating from this chart, if these metrics disagree so significantly for the absolute best players, it’s likely that mediocre players will also have large disparities in their statistical rankings by each metric.



To investigate just what these statistical disparities might be, we broke down each metric to its mathematical formula, to see their components.

Fantasy Points is the most basic metric – it multiplies each basic ‘counting stat’ by a coefficient and outputs a number representing the volume of statistical output by a player. The coefficients seek to equalize the value of assists, rebounds, and points. FP does not consider the player’s efficiency, or pace of play. Obviously, 20 points in a game ending 74-68 is more valuable than 25 points in a 135-123 game, but FP would rank the latter performance as stronger. By normalizing to pace, the metric would consider the amount of points the player scored per 100 possessions, allowing for a more accurate comparison.

In that case, let’s now move to PER, a stat which is normalized to pace, as well as minutes played. It multiplies counting stats by coefficients and analyzes the proportion of team field goals the player’s assists contribute towards. Additionally, PER subtracts what its creator, John Hollinger, calls “negative accomplishments” such as turnovers, personal fouls, and missed defensive rebounds. PER’s largest flaw is its greatest strength- minutes normalization. Because of limited sample size, the player with the all-time highest PER has only played a few minutes. Adding minimum games or minutes played removes these outliers, but on the other end, players who make significant contributions during their prime, only to decrease in efficiency in their career’s twilight are prone to having a low career average PER.

As such, there is no true ‘best metric’ for evaluating talent. Undoubtedly, every player on this list is a great player in their own right, but such significant difference in the ranking suggests there might be a better way to evaluate talent.

## 5.2 Feature engineer new player performance metrics addressing shortcomings with existing metrics

## 5.3 Calculate the approximate value of every pick in the NBA Draft

First, we summed up the total value of each metric of each draft pick.

This graph is oversensitive to extremely good players who come up at particular positions, which makes the graph jagged. In order to provide a more accurate curve, we cluster the draft picks into groups. These groups are 1-3, 4-7, 8-14, 15-30, 31-45, and 46-60. We felt these clusters fall in line with how picks are generally compared to one another

This graph provides a much clearer picture of the values of each metric. Also featured in this graph is the NBA Rookie Salary scale. As there is no mandatory salary for second round picks, we use the league minimum salary. We also display the number of players calculated in each cluster, for context.

Using trendlines, we were able to construct mathematical equations for each metric’s value.

## 5.4 Find the highest value picks based on various measure of cost

First, we use the obvious measure of cost, salary, to divide the pick values by. This shows where the best ‘bang-for-the-buck’ can be found in the NBA draft. We again use the clustering technique to clearly visualize the curves.

As shown, the metrics disagree greatly in where the highest value can be found. Advanced Percentile suggests the early second round has the best value players, but VORP values the top three picks as the superior selections.

## 5.5 Create a Jimmy Johnson-style NBA Draft pick value chart

We created draft value charts for each pick. NFL Analyst Rich Hill used Jimmy Johnson’s chart as a baseline to evaluate draft-pick only trades to create a new draft value chart. With this in mind, we found an assortment of draft-pick only trades in the NBA to evaluate each of the draft charts and select a ‘best’ chart.

|  |  |
| --- | --- |
| Metric | Mean Abs Error |
| VORP | 0.045443858 |
| WS | 0.070661068 |
| FP | 0.081395181 |
| RS | 0.096852287 |
| AVG | 0.112119185 |
| PER | 0.149346209 |
| BP | 0.167310009 |
| AP | 0.198652996 |

Clearly, VORP is the most accurate chart.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| DraftPos | VORP | DraftPos | VORP | DraftPos | VORP | DraftPos | VORP |
| 1 | 3000 | 16 | 1082 | 31 | 390 | 46 | 141 |
| 2 | 2803 | 17 | 1011 | 32 | 364 | 47 | 131 |
| 3 | 2619 | 18 | 944 | 33 | 340 | 48 | 123 |
| 4 | 2446 | 19 | 882 | 34 | 318 | 49 | 115 |
| 5 | 2286 | 20 | 824 | 35 | 297 | 50 | 107 |
| 6 | 2135 | 21 | 770 | 36 | 278 | 51 | 100 |
| 7 | 1995 | 22 | 719 | 37 | 259 | 52 | 94 |
| 8 | 1864 | 23 | 672 | 38 | 242 | 53 | 87 |
| 9 | 1741 | 24 | 628 | 39 | 226 | 54 | 82 |
| 10 | 1627 | 25 | 587 | 40 | 212 | 55 | 76 |
| 11 | 1520 | 26 | 548 | 41 | 198 | 56 | 71 |
| 12 | 1420 | 27 | 512 | 42 | 185 | 57 | 67 |
| 13 | 1327 | 28 | 478 | 43 | 172 | 58 | 62 |
| 14 | 1239 | 29 | 447 | 44 | 161 | 59 | 58 |
| 15 | 1158 | 30 | 418 | 45 | 151 | 60 | 54 |

Compared to the NFL, the NBA follows a different level of apparent talent drop-off.

For the first 20 picks, NBA talent is relatively better than the same draft pick in the NFL. However, after that, the NBA talent continues to decline quickly while the NFL flatlines.

# 6. Methodology for NCAA (*Jake)*

## 6.1 Create a model which predicts various measures of NBA success based on NCAA DI statistics

We designed a series of experiments to examine the ability of machine learning models to predict various success criteria in the NBA. These criteria are as follows:

- Was drafted

- Made NBA (played a regular-season game)

- First round pick

- Lottery pick

All these target values are binary, so this is a classification problem. The models we experimented with were:

- Logistic Regression

- Decision Tree

- Random Forest

- Multi-layer Perceptron (Neural Networks)

# 7. Design for NCAA (*Jake)*

8.5 Create a model which predicts various measures of NBA success based on NCAA DI statistics

Mike talk about how you joined the data together for NCAA stuff

Once we had a clean dataset, we used stratified sampling to split the data proportionally based on class value. We also normalized the non-target attributes, to make sure no attribute was being artificially weighed more than another.

We tinkered with the parameters for each of the models, until we found the best performing set of parameters for each model. At that point, we ran our experiments on each of the target classes, which were: madeNBA, wasDrafted, firstRound, secondRound, and lotteryPick. We then used sk-learn’s classification\_report to print the resulting precision, recall, accuracy, and f1 score for each of the classes.

# 8. Results for NCAA (*Mike)*

# 9. Discussion and Future Work (*Mike)*