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

# Table of Figures

# Executive Summary

# *Jake*

# 1. Introduction

# 2. Background

# *Mike & Jake*

# 3. Methodology

# *Jake*

## 3.1 Analyze existing basketball player performance metrics

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

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

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

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

# 4. Design

# *Mike*

# 5. Results

# *Jake*

# 6. Methodology for NCAA

# *Jake*

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

# 7. Design for NCAA

# *Jake*

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

Once we had a grasp on the value of a player and the expected value from a given draft pick we set out to predict NBA performance for NCAA Division I players. To do this, we first needed to gather statistics about al NCAA Division I players. Using the same methodology to pull data from Basketball-Refernce.com we were able to pull the college data from Sports-Reference.com. We were able to pull data from all NCAA division I teams from 2000 – 2018. But due to the need to lack of identifiers for an NCAA player (the ids used in sports reference are not the same as the ones used in basketball reference) we needed to manually enter where a player was drafted and so we focused on college players from 2010 to 2018. When we were evaluating NBA player performance only in game performance was accounted for, but since predicting NBA readiness and expected performance it is also necessary to consider physical attributes. Thus, we also made sure to collect height and weight measurements for all NCAA players. To further investigate how physical attributes play a role into the probability a college player will reach the NBA we also collected data from the NBA combines from 2010-2018.

After collecting all the data that we needed to we integrated Python with sklearn, a machine learning package, to predict whether or not a player would make the NBA. We defined making the NBA as playing in an official game during the NBA season. This excludes players who were drafted and never played a game, as well as those who signed contracts and were on NBA rosters but failed to play in a game. These distinctions echo the distinctions that are enforced on the sports reference page in order for a college player to be considered having gone on to play in the NBA. We created and ran a logistic regression, decision tree classifier, random forest classifier, MLP classifier, and Zero R model to see which model would be best at predicting whether a player would make the NBA. The Zero R model, predicting every player as never making the NBA, was going to be our baseline. Since the vast majority of NCAA DI players never make the NBA, a model that predicts no one will make the NBA is still correct over 99% of the time. But in order to tell a story worth listening to we needed to predict the players who did end up making the NBA.

A close up of a map

Description generated with high confidenceIn an attempt to improve the prediction ability of our model while also using realistic sub sections of NCAA DI players we broke up our dataset into the following categories to test our model with.

Freshmen only: We decided that it would be appropriate to only look at players who were in their freshmen year because the trend of freshmen being drafted, especially in lottery selections, has been increasing. From our previous work on NBA performance and the expected value of a pick it was appropriate to put an extra consideration on lottery picks. In the 2018 draft 11 of the 15 lottery picks were freshmen, the other four being international player at 3, junior at 10, sophomore at 12, an junior at 13. In the 2017 draft 11 of the 15 lottery picks were also freshmen. The other four being international at 8, sophomore at 12, sophomore at 13, and junior15

Last Year of College: We decided that including the last year a player played would be a good sub section of players to consider as well. This is because this subsection inherently captures a players best season or their

Guard:

Forward:

Center:

We also ran the model on all of the players in our dataset.

Note: due to an overwhelming amount of empty data points in the combine anthropology and agility datasets we didn’t these metrics. This is due to three main issues, the first is that players who attend the combine rarely perform all the tests, the second is often times the most notable college players rarely perform any of the tests if they attend at all (the vast majority of NCAA players also do not attend) and lastly the combine usually occurs only a month before the NBA draft and by then most scouts/ fans have already decided who they feel are most draft worthy. For these reasons we decided that adding the combine metrics to our machine learning models would negatively affect the model’s ability to predict NBA readiness.

# 8. Results for NCAA

# Form:

## 8.1 Predicting whether an NCAA DI player will play an NBA game

The linear regression model provided the highest f1 score when

A close up of a map

Description generated with high confidence

## 8.2 Predicting which 2018 NCAA DI players would play an NBA game

## 8.3 Predicting whether an NCAA DI player will be drafted

## 8.4 Predicting whether an NCAA DI player will be a lottery pick

## 8.5 Predicting whether an NCAA DI player will be a first round pick

## 8.6 Predicting whether an NCAA DI player will be a second round pick

## 8.7 Predicting whether an NCAA DI player will be a free agent

# 9. Discussion and Future Work

## 9.1 Predicting which current (2019) NCAA DI players will play an NBA game

# 4.1 Sourcing draft data

# 4.2 Experiment design

# 5. Results

# 5.1 Analyze existing basketball player performance metrics

# WS VORP PER WS FP

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

# CIA BP AP

# 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.

5.3.1 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.

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.

Next, we use Jimmy Johnson’s values of value for NFL picks.