Identifying and Explaining the Factors that Drive Future Success in the NBA

Frank Baring
Columbia University
fb2493@columbia.edu

May 9, 2023

Abstract

This paper aims to identify and explain the factors that contribute to an elite college basketball player maintaining an elite level of play as a professional. Elite play is defined in this analysis as having an offensive or defensive "Robust Algorithm (using) Player Tracking (and) On/Off Ratings"(RAPTOR) rating in the top 90th percentile, for NBA players, and having a "Points Over Replacement Per Adjusted Game" (PORPAG or DPORPAG) in the top 90th percentile, for college players. This paper also considers the possible drivers that make a non-elite player at the college level develop into a top-tier professional, using random forrest as the central supervised learning model of choice for both parts of the analysis. The data used consists of individual statistics on players who were drafted after 2013 in order to restrict the scope of the analysis to the most modern manifestation of the game without confining the available data to the point where statistically significant insights are unachievable.

Introduction

What is the secret ingredient that makes a top level college athlete into an elite professional? Similarly, what causes a great college player to fall out of or break into the top tier within their cohort of drafted student athletes? This paper aims to answer these questions by employing the descriptive capabilities of the random forest classification model, tuned with grid search. Given the inherent complexity of the interconnected nuances that contribute to player performance, the analysis starts with first principles and then progressively builds in sophistication throughout. As such, the first stage when analyzing improvement and drop off is to quantify the 'elite' category at both the college and professional level, for both offense and defense.

Metrics

Our starting data sets for both college and professional player metrics are filtered such that we only observe players who were drafted from college to the NBA after 2013.

College players[1]:

Figure 1

	player_id	player_name	GP	Min_per	Ortg	usg	eFG	TS_per	ORB_per	DRB_per	...	blk_per	ast/tov	stl_per
0	gordoaa01	Aaron Gordon	1.125553	0.761066	-0.575066	0.063514	-0.304815	-1.342360	0.944002	0.679584	...	0.079651	0.325053	-0.297497
1	holidaa01	Aaron Holiday	0.287546	0.865068	-0.420670	0.171041	-0.058184	0.106144	-1.163201	-1.360117	...	-0.858478	0.654602	0.207023
2	nesmiaa01	Aaron Nesmith	-1.775241	-0.642967	0.496057	-0.326271	0.752178	0.891295	-0.763559	-0.030155	...	-0.285177	-0.678947	-0.234432
3	naderab01	Abdel Nader	-0.614924	-0.706343	-2.526247	1.199267	-1.590824	-1.826456	-0.472910	0.247986	...	-0.024585	-1.031457	0.017828
4	schofad01	Admiral Schofield	0.110275	-0.438212	-0.106088	-0.238905	-0.163883	-0.262691	-0.145930	0.363079	...	-0.563141	-0.284829	-0.770485
5	paynead01	Adreian Payne	0.352008	-1.192229	-0.331892	-0.400195	0.263319	0.273014	0.835009	0.919361	...	1.182822	-1.155154	-0.392095
6	peteral01	Alec Peters	-0.179805	0.840693	1.077937	0.325611	0.633266	0.951162	0.162884	0.574082	...	-0.780301	-0.558074	-0.833550
7	johnsal02	Alize Johnson	-0.034765	0.731815	0.056028	1.165665	-0.348857	-0.324623	1.367865	2.760846	...	-0.719496	-0.158729	-1.306538
8	harrian01	Andrew Harrison	1.415632	0.351556	-0.389791	-0.312830	-1.793415	-1.012060	-1.005766	-1.468815	...	-0.823733	1.020999	-0.360562
9	wiggian01	Andrew Wiggins	0.545394	1.040572	0.212354	0.574267	-0.604297	-0.089284	0.459588	-0.663166	...	-0.024585	-0.735899	0.080893

To rank college players effectively, the ideal metric must capture offensive and defensive impact when the player is on the court relative to when they are not, while also adjusting for minutes played, so as to isolate individual performance as accurately as possible. As such, we rank college players by "Points Over Replacement Per Adjusted Game" (PORPAG, and DPORPAG for the defensive equivalent). We define PORPAG as:

(Offensive rating – 88) * %Possession * Minutes% *65

Where the 88 is the offensive rating (points produced per 100 possessions) of the hypothetical replacement player and the 65 is the constant for possessions in an average game. However, as Bart Torvik points out[2] on his analytics blog, PORPAG does not account for the fact that it is far easier to be more efficient at a low usage, where usage is defined as:

100*(Field Goals Attempted +1/3 of Assists +

[1] Note that all figures displayed in Figure 1 have been scaled. Please see appendix for glossary of stat abbreviations.

[2] https://barttorvik.com/porpagatu.html

(0.44*Free Throws Attempted) + Turnovers) / Possessions.

and efficiency is defined as:

(Points + Rebounds + Assists + Steals + Blocks − Missed Field Goals − Missed Free Throws - Turnovers) / Games Played

Subsequently, we adjust for usage before scaling to arrive at our final metric. The top 90th percentile of college players drafted after 2013 in both adjusted PORPAG and adjusted DPORPAG, specified by dummy variables 'top_o' and 'top_d' (Figure 2), are as follows:

	player_id	player_name	adj_DPORPAG	adj_PORPAG	top_d	top_o
0	gordoaa01	Aaron Gordon	1.667584	-0.259165	1	0
1	holidaa01	Aaron Holiday	-0.319128	-0.053946	0	0
2	nesmiaa01	Aaron Nesmith	-0.698642	0.174340	0	0
3	naderab01	Abdel Nader	-0.592256	-1.312317	0	0
4	schofad01	Admiral Schofield	-0.639545	-0.522169	0	0
...	...	...	...	...	...	...
291	colliza01	Zach Collins	-0.078246	-0.339010	0	0
292	lavinza01	Zach LaVine	-1.327851	-1.182476	0	0
293	nnajize01	Zeke Nnaji	1.044681	0.848894	0	0
294	smithzh01	Zhaire Smith	-0.197045	-0.094319	0	0
295	willizi01	Zion Williamson	2.818566	3.268828	1	1

Figure 2

Top 10% of defensive college players drafted after 2013:

- Aaron Gordon
- Ben Simmons
- Bol Bol
- Caleb Swanigan
- Cole Anthony
- Collin Sexton
- D'Angelo Russell
- Daniel Hamilton
- De'Aaron Fox
- DeAndre' Bembry
- Deandre Ayton
- Delon Wright
- Ignas Brazdeikis
- Isaiah Whitehead
- Ja Morant
- Jabari Parker
- Jahlil Okafor
- Jarrett Culver
- Jaylen Brown
- Jayson Tatum
- Josh Okogie
- Julius Randle
- Marcus Smart
- Marvin Bagley III
- Miles Bridges
- Precious Achiuwa
- Stanley Johnson
- Trae Young
- Vernon Carey Jr.
- Zion Williamson

Top 10% of offensive college players drafted after 2013:

- Ben Simmons
- Bol Bol
- Cameron Payne
- Carsen Edwards
- Cassius Winston
- Collin Sexton
- D'Angelo Russell
- Deandre Ayton
- Delon Wright
- Derrick White
- Doug McDermott
- Grant Riller
- Isaiah Stewart
- Ja Morant
- Jabari Parker

- Jahlil Okafor
- Jamal Murray
- James Wiseman
- Jawun Evans
- Jerian Grant
- Lauri Markkanen
- Malik Monk
- Marcus Smart
- Markelle Fultz
- Marvin Bagley III
- Monte Morris
- T.J. Warren
- Trae Young
- Vernon Carey Jr.
- Zion Williamson

Professional players:

Figure 3

	player_id	player_name	raptor_offense	raptor_defense	war_total	top_d	top_o	top_war
0	achiupr01	Precious Achiuwa	-0.868631	0.539676	-0.631396	0	0	0
1	adamsjo01	Jordan Adams	-0.313494	0.543459	-0.261756	0	0	0
2	alexani01	Nickeil Alexander-Walker	0.192744	0.115793	-0.181715	0	0	0
3	allengr01	Grayson Allen	0.120354	-0.413780	-0.344195	0	0	0
4	allenja01	Jarrett Allen	0.250224	0.256513	0.425470	0	0	0
5	allenka01	Kadeem Allen	0.295476	0.256634	-0.293697	0	0	0
6	anderju01	Justin Anderson	0.121669	0.260815	-0.275771	0	0	0
7	anderky01	Kyle Anderson	0.185858	0.661893	1.169810	0	0	0
8	anigbik01	Ike Anigbogu	-3.413702	-5.980354	-0.576506	0	0	0
9	antetko01	Kostas Antetokounmpo	-1.994156	2.812279	-0.469236	1	0	0
10	anthoco01	Cole Anthony	-0.057979	-0.612216	-1.444651	0	0	0
11	aytonde01	Deandre Ayton	0.478578	0.555042	1.594747	0	0	1
12	azubuud01	Udoka Azubuike	-2.898437	-0.733005	-0.673141	0	0	0
13	bacondw01	Dwayne Bacon	-0.338755	-0.198671	-1.034403	0	0	0
14	baglema01	Marvin Bagley III	-0.131646	0.023697	-0.519401	0	0	0
15	bairsca01	Cameron Bairstow	-1.183704	-0.093904	-0.569723	0	0	0
16	baldwwa01	Wade Baldwin IV	-1.658365	-0.224912	-0.890975	0	0	0
17	balllo01	Lonzo Ball	0.583711	0.470372	1.272357	0	0	0
18	banede01	Desmond Bane	0.788095	-0.178779	0.267243	0	0	0
19	bateske01	Keita Bates-Diop	-0.194690	0.075185	-0.612928	0	0	0

At the professional level, the aim is to again capture player impact in offense and defense with a non-convoluted metric that adjusts for time on the court. FiveThirtyEight's RAPTOR (Robust Algorithm (using) Player Tracking (and) On/Off Ratings)[3] is the ideal statistic for capturing this information given that it reflects the extent to which a player's presence on the court boots his team's performance, while also using player tracking data, making it a modern improvement to the standard box scores (e.g. easily recorded statistics like three point shooting percentage) that neglect a significant amount of play-by-play information containing a important signal regarding a given player's skill. The

[3] https://fivethirtyeight.com/features/how-our-raptor-metric-works/

metric was created in 2019 and while it might appear opaque at first glance, it does in fact exclusively use publicly available information[4] and provides a key insight into how NBA teams evaluate the impact of their own players. In using RAPTOR for this analysis, we are essentially assessing which college players were able to excel in adjusted PORPAG and reaffirm their standard when assessed with a more rigorous tracking metric that evaluates their movement even when they do not have the ball. While the nuances of how RAPTOR is calculated are relevant to most modern basketball analysis, this paper does not include a deep-dive into the metric. Rather, we take RAPTOR to be the most accurate reflection of player impact at the professional level and move on to the analysis section. As with our definition of 'elite' for college players, we define the elite category for professionals as the players in the top 90th percentile of offensive and defensive RAPTOR rankings.

Top 10% of defensive players drafted after 2013:

- Kostas Antetokounmpo
- Bruce Brown
- Vernon Carey Jr.
- Donte DiVincenzo
- Henry Ellenson
- Joel Embiid
- Rondae Hollis-Jefferson
- Josh Huestis
- Jonathan Isaac
- Cory Jefferson
- Isaiah Joe
- Brice Johnson
- Dakari Johnson
- Romeo Langford
- Cody Martin
- De'Anthony Melton
- Larry Nance Jr.
- Chuma Okeke
- Jakob Poeltl
- Marcus Smart
- Isaiah Stewart
- Jarnell Stokes
- Tyrell Terry
- Matisse Thybulle
- Myles Turner
- Tyler Ulis
- Jarred Vanderbilt
- Thomas Welsh
- Derrick White
- Delon Wright

Top 10% of offensive players drafted after 2013:

- Devin Booker
- Malcolm Brogdon
- Anthony Brown
- Jalen Brunson
- Rakeem Christmas
- Jordan Clarkson
- John Collins
- Joel Embiid
- De'Aaron Fox
- Shai Gilgeous-Alexander
- Devonte' Graham
- Tyrese Haliburton
- Jalen Harris
- Buddy Hield
- Luke Kennard
- Skylar Mays
- Donovan Mitchell
- Ja Morant
- Monte Morris
- Jamal Murray
- Michael Porter Jr.
- Immanuel Quickley
- Grant Riller

[4] https://www.kaggle.com/datasets/schmadam97/nba-playbyplay-data-20182019

- Jarnell Stokes
- Jayson Tatum
- Karl-Anthony Towns
- Derrick White
- Zion Williamson
- Robert Woodard II
- Trae Young

Analysis

The target category (ŷ) that we identify consists only of the players who were elite in college and stayed elite at the professional level. These players are:

Elite defensive players in college and the NBA:

- Delon Wright
- Marcus Smart
- Vernon Carey Jr.

Elite offensive players in college and the NBA:

- Derrick White
- Grant Riller
- Ja Morant
- Jamal Murray
- Monte Morris
- Trae Young
- Zion Williamson

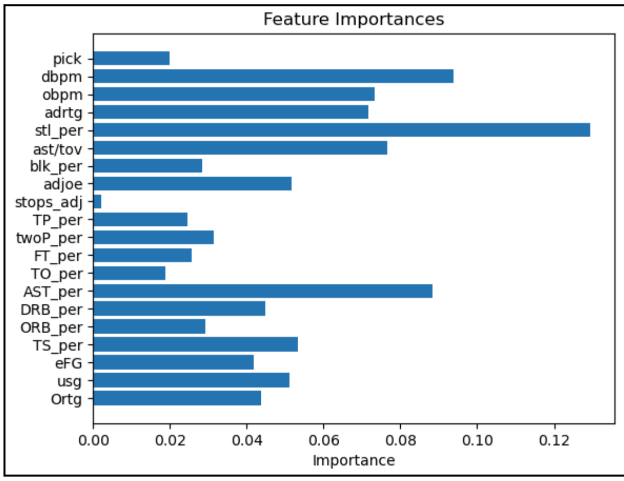
What makes these star players so consistent? Given we have too few observations in our data to perform any predictive analysis, we train a random forest (RF) classifier on our box college stats to observe the relative importance of different metrics. We remove all RAPTOR and PORPAG stats before training, given these metrics were used to define the categorical dependent variable. Note, grid search provides the following optimal hyperparameters for our RF:

{'max_depth': 2, 'minimum_samples_leaf': 1, 'minimum_samples_split': 2, 'n_estimators': 50}

Defense

Offense

Figure 4



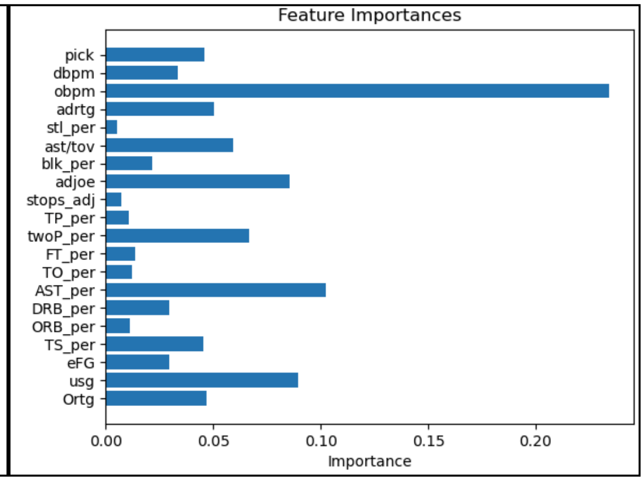


Figure 5

It is interesting to note, firstly, that *steal percentage* is the most significant predictor of consistent defensive impact between the college and NBA level, and secondly that draft pick position is a relatively insignificant predictor of future skill for defensive players while for offensive players it is an average predictor. Intuitively, the former observation makes sense. A steal is one of the most impactful defensive moves a player can pull off, given that it leads to a turnover with no stoppage, allowing for the stealing team to rush their opposition and make a score attempt in transition against a scattered defense, hence why we see a significant gap between the effect of block percentage and steal percentage on our dependent variable. It is also a move that requires a significant degree of both athleticism[5], and "basketball IQ", a term first defined by former coach of the Atlanta Hawks as "seeing stuff better, understanding what people are trying to do, […] making the right passes, the right decisions."[6] While basketball IQ is mostly used as a catch-all categorization of the intangible qualities that make a player good, we can go further and ascribe certain discrepancies in variance captured to the on-court smarts of a defensive player. For instance, the results above show that steal percentage is not only the strongest predictor of defensive skill, but also crucially more than 25% more significant than simple defensive box plus/minus (which measures the difference, per 100 possessions, in points allowed with a player on the court versus off

[5] A steal results in the defensive team instantly gaining control of the ball, which is only true some of the time when a scoring attempt is blocked. Often, the difference between a block and a steal is that the defending player has the explosively and reach to get two hands on the ball rather than one.

[6] '*Basketball IQ' and the racial coding of the word',* David Leonard, Andscape, 2016. https://andscape.com/features/basketball-iq-and-the-racial-coding-of-the-word/.

the court[7]). While of course this difference is not entirely explained by the added game intelligence of the stealing individual, given that good guarding as a team will make a steal more likely, it nonetheless provides a good proxy for the extra edge contributed by an individual provides on account of the added skill required in a steal compared to a block or rebound.

Moreover, the fact that draft pick position is more important in determining the future skill of elite offensive players also elucidates a known fact about basketball that is difficult to quantify: the draft favors offensive specialists over defensive specialists. In essence, the objective of this paper could be defined as an attempt to identify what creates star players — a challenge that scouts and NBA analyst work tirelessly to tackle. Were the draft a perfect system, the pick position would far outweigh other box score metrics and this analysis would be redundant. Our results do not suggest that the draft is necessarily a bad system, but rather that defensive talent is harder to observe and more a product of combinatory factors. The simplest way to observe this is to compare the standard deviation of the importance of our offensive and defensive features to predicting $\hat{y}$:

Offensive: 0.05 **Defensive: 0.03**

So if the ability to stay elite beyond the college level is decided primarily by steal percentage, assist percentage, defensive box plus/minus, and offensive box plus/minus, how might a scout adapt their strategy to find *rising* stars? We can define this new category as players who begin outside of the top 90th percentile in both adjusted PORPAG and DPORPAG but end up in the top 90th percentile of wins above replacement (WAR) as professionals, defined as the wins that a player individually adds to a team over a season.

The top 10% of players in terms of WAR, drafted after 2013, are:

- Aaron Gordon
- Ben Simmons
- Buddy Hield
- Deandre Ayton
- Delon Wright
- Derrick White
- Devin Booker
- Devonte' Graham
- Donovan Mitchell
- Donte DiVincenzo
- Isaiah Stewart
- Jakob Poeltl
- Jamal Murray
- Jaylen Brown
- Jayson Tatum

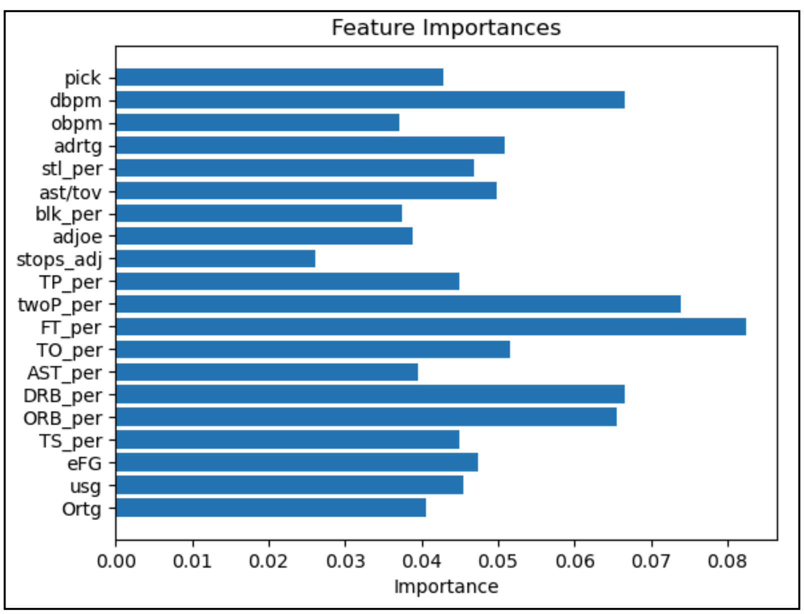
[7] As defined by NBAstuffer, https://www.nbastuffer.com/.

- Joel Embiid
- John Collins
- Josh Richardson
- Karl-Anthony Towns
- Larry Nance Jr.
- Malcolm Brogdon
- Marcus Smart
- Matisse Thybulle
- Michael Porter Jr.
- Mikal Bridges
- Montrezl Harrell
- Pascal Siakam
- Shai Gilgeous-Alexander
- Trae Young
- Zion Williamson

… meaning our ŷ subset (I.e. players in this list who are not in our predefined elite adjusted PORPAG and DPORPAG cohort) is:

- Buddy Hield
- Devin Booker
- Devonte' Graham
- Donovan Mitchell
- Donte DiVincenzo
- Jakob Poeltl
- Joel Embiid[8]
- John Collins
- Josh Richardson
- Karl-Anthony Towns
- Larry Nance Jr.
- Malcolm Brogdon
- Matisse Thybulle
- Michael Porter Jr.
- Mikal Bridges
- Montrezl Harrell
- Pascal Siakam
- Shai Gilgeous-Alexander

… giving the following significance of features:



[8] Declared MVP just 8 days ago. Naturally, this cohort will include some of the most impressive career trajectories in the league, given the skill development that is being identified by our constraints.

Free throw percentage, defensive box plus/minus, and rebound percentage all provide interesting talking points in light of these results. The free throw was a virtually insignificant metric when determining the likelihood that an exceptional player stays exceptional, but when we confine our dependent variables to rising starts, this scoring opportunity becomes especially impactful, perhaps suggesting that a baseline level of shooting accuracy is required to continue developing at the rate that Devin Booker and Joel Embiid did between college and their peaks in the NBA. Another crucial takeaway from these results is that the defensive strength of a player is almost twice as significant, when determining their future trajectory, as offensive ability, building on a previous observation which suggested that defensive talent is less comprehensible from a quantitative standpoint. While the WAR analysis does not contest this, we observe that defensive impact at the college level remains highly significant when identifying developing talent. This is only reaffirmed by the importance of rebound percentage, in so far as the ability to regain possession off of the board indicates a deeper understanding of the game (referring back to "Basketball IQ"), more so than say three point or free throw percentage.

Conclusion

The approach taken in this paper is predicated on the partitioning of data. By spending extra time extracting the precise subsets of players that reflect talent profiles and trajectories, we can begin to discern the insights necessary for making high quality scouting decisions. The marginal value this kind of analysis provides when compared to other more traditional methods, is more often intended to find what *is not* relevant to the differentiation of players, than to find what is. This is most aptly exemplified by our identification of steal percentage as the most significant driver of constant elite defensive performance which, while only being a useful statistic over a multi-season sample size given the rarity of steals compared to blocks and stops, is still frequently overlooked. Ultimately this analysis gives a good basis for further investigation into the interconnected quality of some metrics that we have covered and also provides a good direction for further categorization of players when using supervised learning to predict career trajectories. Next, one might apply the same approach to additional wholistic career analyses, such as which factors impact a player's susceptibility to injury, or even compare the findings of this paper to players who did not join the NBA through the draft, thereby initiating a deeper dive into the strengths and weaknesses of that system.

Appendix

1. Glossary for appendix of abbreviations:
    - **GM**, **GP**; **GS**: games played; games started
    - **PTS**: points
    - **TP_per**: two point percentage
    - **FGM**, **FGA**, **FG%**: field goals made, attempted and percentage
    - **FTM**, **FTA**, **FT%**: free throws made, attempted and percentage
    - **3FGM**, **3FGA**, **3FG%**: three-point field goals made, attempted and percentage
    - **REB**, **OREB**, **DREB**: rebounds, offensive rebounds, defensive rebounds
    - **AST**: assists
    - **ADJOE**: adjusted offensive efficiency
    - **ORTG**: offensive rating
    - **STL**: steals
    - **BLK**: blocks
    - **TO**: turnovers
    - **EFF**: efficiency
    - **GP**: games played
    - **MP:** Minutes played
    - **AST/TOV**: assist to turnover ratio