Draft Analysis - What was the problem?

Can we predict early success in NBA with players' college stats?

The NBA draft happens every year in June. It is where teams in the National Basketball Association (NBA) choose players who have never played in the NBA before.

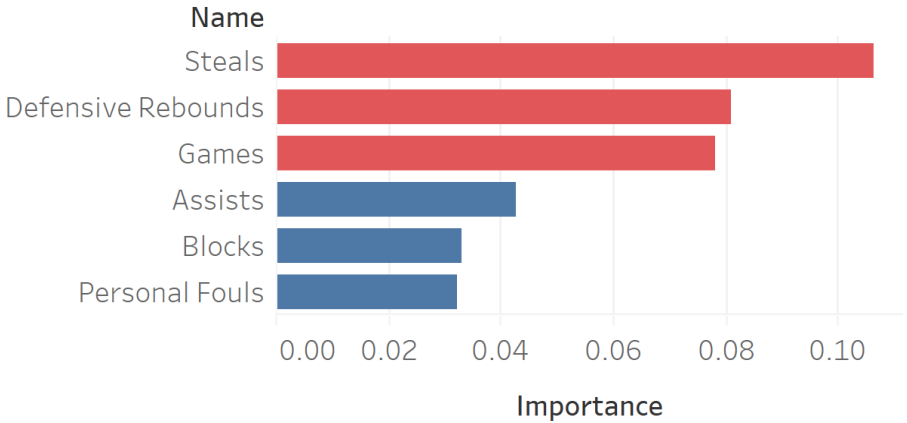
- Wikipedia

Gaining knowledge of which stats help predict success well can help teams better scout players. This analysis is going to provide insights of how college stats are indicative of future NBA success.

This could be used to mitigate the problem of underrated prospects in the draft as well as 'draft busts'. One of the deliverables is expected to be an NBA draft prospect, which can be compared to the real draft order and their actual performance in NBA.

Draft Analysis - What did I find?

Most and least predictive variables



The graph above only shows top 3 and bottom 3 variables. All but 'Games' are per game measures.
The most predictive variables for early NBA success (in terms of VORP) are steals and defensive rebounds per game and games in college.
The least predictive variables in the model are a player's assists, blocks and personal fouls per game.

2018 Projection Highlight

Top 5

Player	Position	VORP S18-19	VORP Projected	Draft Rank
Mohamed Bamba	C	0.4	3.7	6
Trae Young	PG	0.2	2.7	5
Alize Johnson	PF	-0.1	2.3	50
Jaren Jackson	PF/C	0.8	0.9	4
Khyri Thomas	SG	-0.1	0.9	38

Projected Ranking

Bottom 5

Jalen Brunson	PG	-0.1	-0.2	33
Landry Shamet	PG	0.5	-0.3	26
Aaron Holiday	PG	-0.1	-0.4	23
Lonnie Walker	SG	-0.1	-0.4	18
Grayson Allen	SG	-0.4	-0.5	21

Draft Analysis - How did I do the analysis?

1. Web Scraping

Data source: Basketball Reference, Sports Reference, Draft Express and NBA.com
Build scrapers for each website and extract data before compiling

2. Modeling

Data and variables: 2011 – 2017 NBA drafts
- International players were not included in the model since the level of competition and data availability varies in different leagues.
- Target variables: VORP (NBA prospects' value above replacement over their first three seasons in the league) as a proxy of early success
- Predictors: 18 attributes
  - College stats (per game level except for games)
  - Physical data (weight, height, wingspan)

Methods:
-Techniques: Train test split, standardization, log transformation, cross validation
-Algorithms: Regularized linear model, KNN and **Random Forest**

3. Prediction

Project VORP for 2018 draft and rank the players according to the estimated VORP.

Reference: https://www.cmusportsanalytics.com/2017-nba-draft-forecast/