Team 103

Predicting NBA Player Performance

Does this statistic mean anything?

Tim Duncan:
5 x NBA champion
3 x NBA Finals Most Valuable Player
2 x NBA Most Valuable Player

Problem Statement:

Perform a comprehensive analysis to predict the success of players drafted from NCAA, based on their relationship between college experience and their performance in NBA, using offensive and defensive statistics like point scores, assists, steals, blocks, field goals, Win Shares, and VORP.

Win Shares

Win Shares is a player statistic which attempts to divvy up credit for team success to the individuals on the team. This metric is designed to estimate the player's contribution in terms of wins.

VORP

Value over Replacement Player

This metric converts the Box Plus/Minus (BPM) rate into an estimate of each player's overall contribution to the team, measured vs. what a theoretical "replacement player" would provide, where a "replacement player" is defined as a player on minimum salary or not a normal member of a team's rotation.

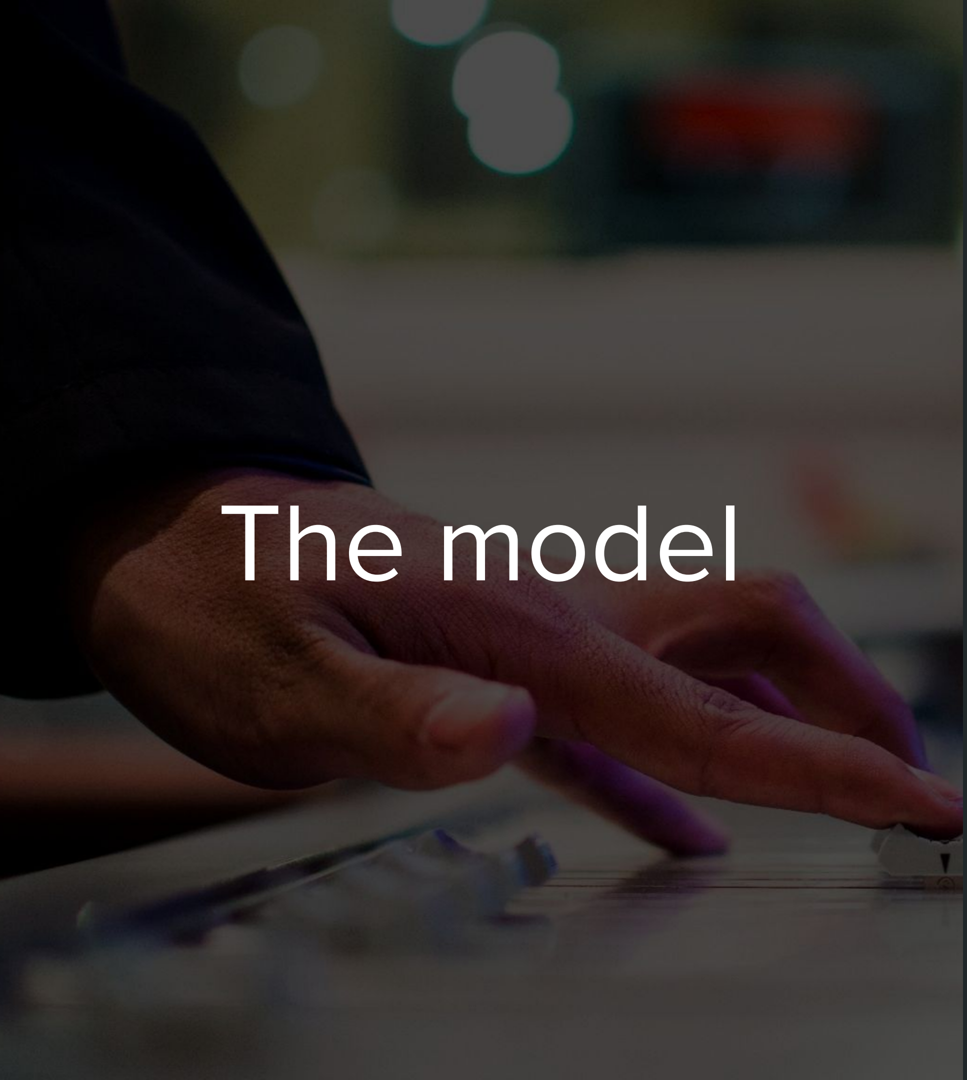
The NBA Data

Rk	Player	Pos	Age	Tm	G	MP	PER	TS%	3PAr	FTr	ORB%	DRB%	TRB%	AST%	STL%	BLK%	TOV%	USG%	OWS	DWS	WS	WS/48	OBPM	DBPM	BPM	VORP
1	Precious Achiuwa	PF	21	MIA	61	737	14.2	.550	.004	.482	11.5	20.6	16.1	6.1	1.3	4.0	13.5	19.5	0.3	1.0	1.3	.085	-3.6	-0.5	-4.1	-0.4
2	Jaylen Adams	PG	24	MIL	7	18	-6.5	.125	.250	.000	0.0	16.9	8.8	12.7	0.0	0.0	0.0	18.6	-0.1	0.0	-0.1	-0.252	-15.1	-4.6	-19.8	-0.1
3	Steven Adams	C	27	NOP	58	1605	15.1	.596	.010	.438	14.4	20.4	17.4	9.1	1.6	2.2	17.5	11.7	2.3	1.7	4.0	.119	-0.4	0.1	-0.3	0.7
4	Bam Adebayo	C	23	MIA	64	2143	22.7	.626	.010	.443	7.7	22.6	15.3	26.9	1.7	3.2	15.0	23.7	5.6	3.2	8.8	.197	2.9	2.0	4.9	3.7
5	LaMarcus Aldridge	C	35	TOT	26	674	15.7	.556	.270	.159	3.0	15.8	9.4	11.0	0.8	3.7	7.9	22.2	0.5	0.6	1.1	.080	-0.2	-0.2	-0.3	0.3
5	LaMarcus Aldridge	C	35	SAS	21	544	15.1	.545	.302	.149	3.3	15.4	9.2	10.2	0.7	2.8	7.0	22.7	0.3	0.5	0.8	.067	-0.2	-0.7	-0.9	0.2
5	LaMarcus Aldridge	C	35	BRK	5	130	18.2	.611	.104	.208	1.8	17.8	10.2	14.3	1.1	7.4	11.8	19.9	0.2	0.2	0.4	.135	0.1	2.1	2.2	0.1
6	Ty-Shon Alexander	SG	22	PHO	15	47	4.2	.349	.750	.167	4.9	19.0	12.1	15.4	0.0	1.9	18.9	15.0	-0.1	0.0	0.0	-0.048	-4.8	-1.7	-6.5	-0.1
7	Nickeil Alexander-Walker	SG	22	NOP	46	1007	12.5	.522	.478	.144	1.4	14.1	7.8	14.7	2.2	2.1	12.4	23.2	-0.3	1.0	0.7	.035	-1.4	0.1	-1.3	0.2
8	Grayson Allen	SG	25	MEM	50	1259	12.8	.586	.662	.220	1.6	12.0	6.7	11.5	1.7	0.6	9.6	16.8	1.5	1.2	2.7	.101	-0.2	0.1	-0.2	0.6
9	Jarrett Allen	C	22	TOT	63	1864	20.1	.661	.039	.602	11.6	26.3	18.9	8.7	0.8	4.3	14.1	16.6	4.3	2.1	6.4	.166	1.3	-0.2	1.1	1.4
9	Jarrett Allen	C	22	BRK	12	320	21.5	.730	.000	.938	13.8	28.6	21.6	8.3	1.1	5.2	19.3	15.5	1.0	0.5	1.4	.216	1.7	0.9	2.6	0.4
9	Jarrett Allen	C	22	CLE	51	1544	19.8	.649	.046	.549	11.2	25.9	18.3	8.8	0.8	4.1	13.1	16.8	3.4	1.6	5.0	.155	1.2	-0.4	0.8	1.1
10	Al-Farouq Aminu	PF	30	TOT	23	434	8.9	.469	.374	.222	5.1	21.9	13.3	10.1	2.1	1.9	20.5	13.6	-0.4	0.5	0.1	.010	-4.9	1.1	-3.8	-0.2
10	Al-Farouq Aminu	PF	30	ORL	17	367	9.9	.482	.348	.191	5.5	20.8	12.9	11.3	2.3	2.2	20.6	14.3	-0.3	0.4	0.1	.014	-4.4	1.3	-3.1	-0.1
10	Al-Farouq Aminu	PF	30	CHI	6	67	3.0	.369	.600	.500	3.3	27.8	15.7	3.5	1.4	0.0	19.7	9.8	-0.1	0.1	0.0	-0.013	-7.5	0.2	-7.3	-0.1

The NCAA Data

Season | Conference

Rk	Player	G	GS	MP	FG	FGA	FG%	2P	2PA	2P%	3P	3PA	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
1	Loren Jackson	31	31	34.2	6.3	13.5	.465	3.3	6.5	.505	3.0	6.9	.428	4.4	5.0	.877	0.6	2.1	2.7	4.5	1.0	0.0	2.4	2.3	19.8
2	Tyler Cheese	30	19	30.8	5.3	12.0	.438	3.8	7.8	.487	1.5	4.2	.346	3.7	4.5	.822	1.6	3.3	4.8	3.4	1.1	0.3	2.6	2.6	15.7
3	Xeyrius Williams	31	31	31.6	4.8	12.6	.378	2.6	5.5	.480	2.1	7.1	.299	2.2	2.5	.908	1.7	7.7	9.5	1.3	0.8	0.9	1.7	2.4	13.9
4	Channel Banks	31	31	31.4	3.5	8.3	.415	1.0	2.4	.425	2.5	6.0	.411	1.5	2.1	.682	0.4	2.7	3.1	1.5	1.1	0.2	1.0	2.2	10.8
5	Camron Reece	30	12	16.1	2.3	3.8	.605	2.3	3.8	.605	0.0	0.0		1.3	2.1	.619	1.5	2.9	4.4	0.2	0.2	0.4	1.5	2.6	5.9
6	Deng Riak	31	20	21.3	1.4	2.8	.489	1.3	2.4	.548	0.1	0.5	.200	0.9	1.2	.737	1.4	3.5	4.9	0.4	0.3	0.7	1.1	2.5	3.8
7	Greg Tribble	31	0	16.4	1.1	2.4	.459	1.0	2.0	.492	0.1	0.4	.273	0.7	1.5	.489	0.2	1.5	1.7	1.1	0.3	0.1	1.1	1.9	3.0
8	Jaden Sayles	26	0	5.4	0.7	1.4	.500	0.7	1.4	.500	0.0	0.0		0.8	1.1	.786	0.4	1.1	1.5	0.0	0.0	0.3	0.2	0.3	2.2
9	Jeremy Roscoe	1	0	9.0	1.0	4.0	.250	1.0	3.0	.333	0.0	1.0	.000	0.0	0.0		0.0	1.0	1.0	0.0	0.0	0.0	0.0	1.0	2.0
10	Ali Ali	31	11	11.6	0.4	1.0	.387	0.3	0.6	.500	0.1	0.4	.231	0.3	0.4	.615	0.5	1.3	1.8	0.7	0.4	0.0	0.8	0.8	1.1
11	Enrique Freeman	7	0	1.9	0.1	0.1	1.000	0.1	0.1	1.000	0.0	0.0		0.4	0.6	.750	0.1	0.4	0.6	0.0	0.0	0.1	0.0	0.1	0.7
12	Mikal Dawson	17	0	4.4	0.1	0.7	.167	0.1	0.3	.400	0.0	0.4	.000	0.1	0.2	.500	0.4	0.5	0.9	0.1	0.1	0.1	0.4	0.5	0.4
13	Marquelle McIntyre	7	0	1.6	0.1	0.1	1.000	0.1	0.1	1.000	0.0	0.0		0.0	0.0		0.0	0.3	0.3	0.1	0.0	0.0	0.1	0.1	0.3



The model

Logistic Regression

Final NCAA Dataset

	Player	Position	School	G	GS	MP	FG	FGA	FG_percent	Twos	TwosA	Twos_perc	Threes	ThreesA	Threes	FT	FTA	FT_pe	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	Drafted
3	A.C. Reid	Guard	Liberty	52	6	1162	103	301	0.342	25	57	0.439	78	244	0.32	32	45	0.71	6	107	113	79	33	5	62	99	316	0
4	A.J. Astroth	Guard	Overall	37	9	484	45	97	0.464	40	75	0.533	5	22	0.23	29	46	0.63	39	92	131	17	15	5	27	50	124	0
2	A.J. Atkinson	Guard	Davidson	16	1	36	3	5	0.6	2	3	0.667	1	2	0.5	2	2	1	3	8	11	0	0	0	2	6	9	0
1	A.J. Bowers	Forward	Chattanooga	6	0	30	0	1	0	0	0		0	1	0	2	2	1	2	4	6	0	1	1	1	5	2	0
4	A.J. Caldwell	Guard	Chattanooga	69	15	1731	120	294	0.408	43	92	0.467	77	202	0.38	23	38	0.61	37	203	240	129	42	15	56	107	340	0
3	A.J. Cheeseman	Forward	Overall	33	12	559	85	185	0.459	83	173	0.48	2	12	0.17	70	109	0.64	43	76	119	15	14	16	44	68	242	0
4	A.J. Davis	Forward	Overall	61	52	1814	176	461	0.382	126	303	0.416	50	158	0.32	190	271	0.7	87	283	370	113	51	18	130	146	592	0

Player name, position, school, box score statistics, and draft status (0 for no, and 1 for yes)

Model 1

```
Call:
glm(formula = Drafted ~ G + GS + MP + FG + FGA + FG_percent +
    Twos + TwosA + Twos_percent + Threes + ThreesA + Threes_percent +
    FT + FTA + FT_percent + ORB + DRB + TRB + AST + STL + BLK +
    TOV + PF + PTS, family = binomial(link = "logit"), data = trainSet)
```

Statistically significant variables are: G (Games), GS (Games Started), MP (Minutes Played), FG (Field Goals), FG_percent (Field Goal Percentage), FT_percent (Free Throw Percentage), DRB (Defensive Rebounds), AST (Assists), BLK (Blocks), and TOV (Turnovers)

Coefficients: (4 not defined because of singularities): Threes, ThreesA, TRB, and PTS

“Threes” (Three-point shots made) is almost perfectly correlated with “ThreesA” (0.99011326) which makes sense, the number of three points made has a strong linear relationship with how many three point shots are attempted.

“TRB” (Total Rebounds) is the sum of “ORB” (Offensive Rebounds) and “DRB” (Defensive Rebounds) which is why “TRB” is almost perfectly correlated with those predictor variables.

Model 1 - Confusion Matrix

```
          Reference
Prediction    0    1
         0 2903  111
         1    7    4

           Accuracy : 0.961
```

Model 1 with a 0.5 threshold has an accuracy of 96.1%

Maybe a lower threshold will increase accuracy?

```
          Reference
Prediction    0    1
         0 2891  111
         1   19    4

           Accuracy : 0.957
```

Model 1 with a 0.4 threshold has an accuracy of 95.7%.

111 incorrectly undrafted players for both thresholds

Model 2

```
model2 <- glm(Drafted ~ G + GS + MP + FG + FG_percent + Twos + FT_percent +
            ORB + DRB + AST + STL + BLK + TOV,
           family=binomial(link='logit'), data=trainSet)
```

```
         Reference
Prediction    0    1
         0 3396  119
         1   11    3

              Accuracy : 0.9632
```

According to the Confusion Matrix output our model's accuracy did improve to 96.32%.

However, it still does not correctly draft 119 players.

Predicting the 2022 NBA Draft

Using the NCAA 2022 dataset Model 1 predicts 9 players to be drafted to the NBA.

No Armando Bacot, but he is not looking to enter the NBA yet either.

Name	Drafted
Chet Holmgren	1
Trayce Jackson-Davis	1
Keegan Murray	1
JT Shumate	1
Norchad Omier	1
Jayveous McKinnis	1
Johni Broome	1
Max Abmas	1
Jalen Pickett	1

'I'm back': Armando Bacot stays at UNC for senior season to chase NCAA championship glory

Notable Draftees Prior to 2011

- Stephen Curry
- Blake Griffin
- Kevin Durant
- Carmelo Anthony

Our model did not predict clearly elite college players with *guaranteed* NBA potential.

This implies there could be more to NBA drafts then just box score statistics.

- Draymond Green
- James Harden
- Derrick Rose
- Chris Paul
- Brook Lopez
- Russell Westbrook

Notable Draftees Prior to 2011

- Stephen Curry
- Blake Griffin
- Kevin Durant
- Carmelo Anthony

Our model did not predict clearly elite college players with *guaranteed* NBA potential.

This implies there could be more to NBA drafts then just box score statistics.

- Draymond Green
- Kevin Durant
- Chris Paul
- Brook Lopez
- Carmelo Anthony
- Russell Westbrook

Improving model’s accuracy

We could use other predictor variables that are not box score statistics like:

- Player’s position
- Order of the NBA teams for draft picks
- The age and health of the current NBA teams’ rosters

Our dataset focused only on NCAA college players when it is possible for foreign players, G league athletes, and even high school players to be drafted to the NBA.

Let’s take a look at a decision tree regressor model for predicting player performance.

Predicting NBA Success

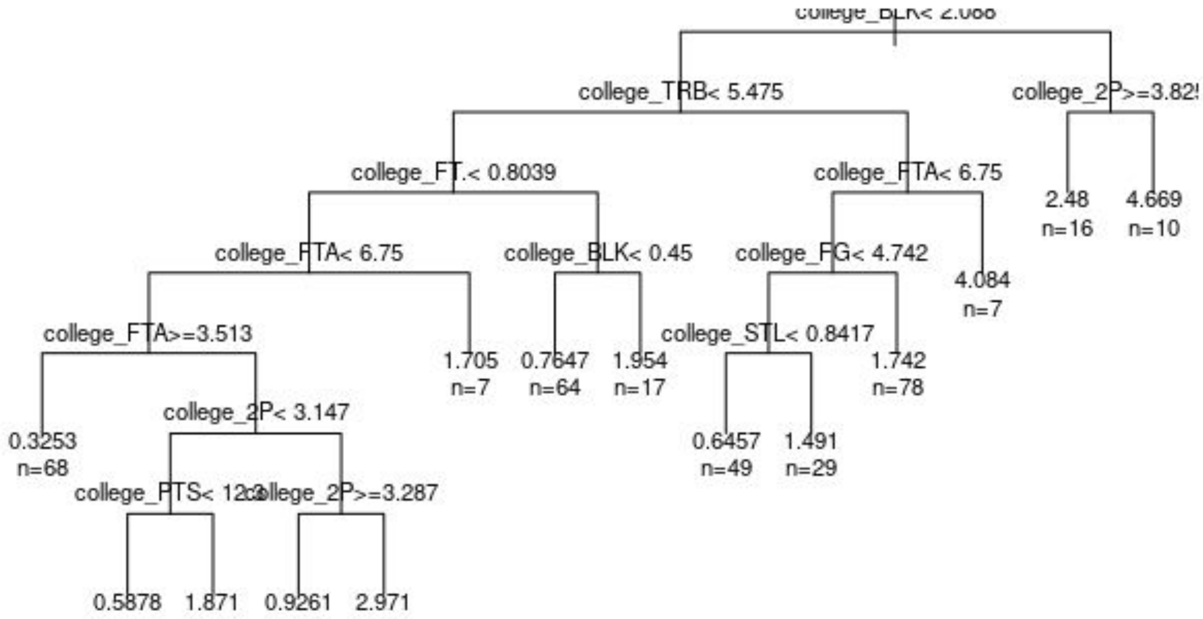
- Decision Tree
- Linear Regression

Decision Tree Regressor: WS

Significant Variables: FG (Field Goals), 2pt FG, 3pt FG, FTA (Free Throw Attempts), FT% (Free Throw Percentage), TRB (Total Rebounds), AST (Assists), STL (Steals), BLK (Blocks), and PTS (Points Scored)

Model	MAE	SSE
All College Variables	1.15	481
Selected College Variables	1.11	465

NBA Win Share Decision Tree Regression

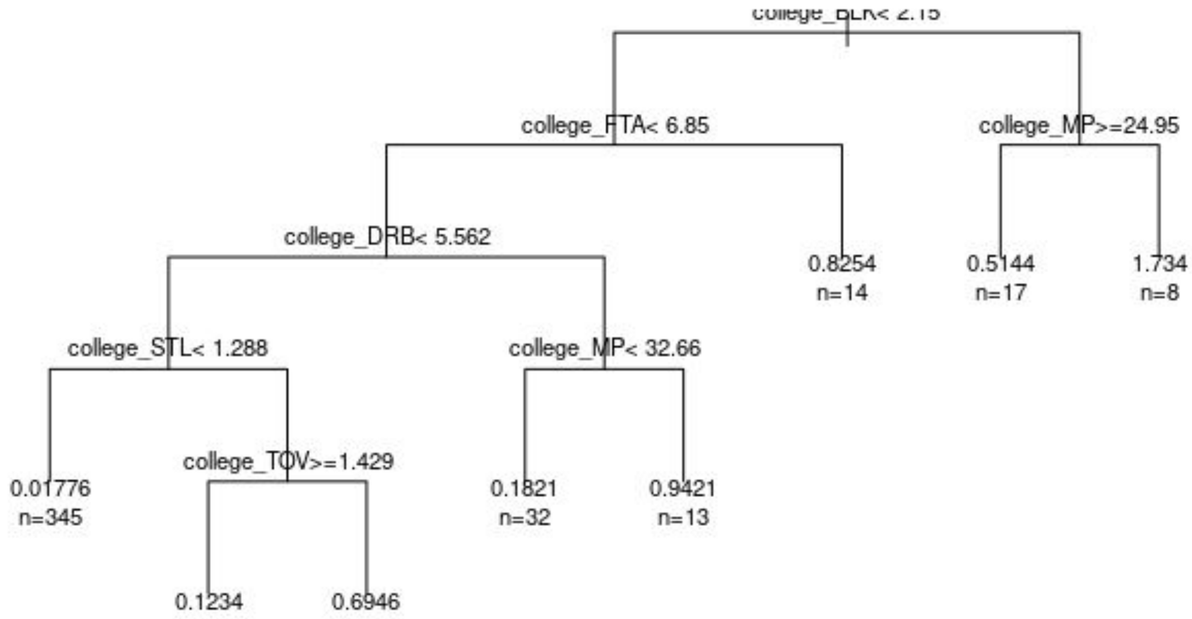


Decision Tree Regressor: VORP

Significant Variables: MP (minutes played), FG (Field Goals), 2pt FG, 3pt FG, 3pt FGA (3pt Field Goal Attempts), 3pt FG% (3pt Percentage), FT (Free Throws Made), FTA (Free Throw Attempts), DRB (Defensive Rebounds), TRB (Total Rebounds), AST (Assists), STL (Steals), BLK (Blocks), TOV (Turnovers)

Model	MAE	SSE
All College Variables	1.12	390
Selected College Variables	0.34	62

NBA VORP Decision Tree Regression



Linear Regression

WinShare Linear Regression Model – Predictors like Field Goal Attempt(FGA), 2-Point Field Goal Attempt(2PA), 3-Point Field Goal Attempt(3PA), Total Rebounds(TRB), Assist(AST), Steals(STL), Turnovers(TOV) and Points scored(PTS) significantly impact the model’s output.

VORP Linear Regression Model – Predictors like Minutes Played (MP), 2-Point Field Goal Attempt(2PA), 3-Point Field Goal Attempt(3PA), Field Goal Attempt(FGA), Free Throws(FT), Defensive Rebounds(DRB), Assist(AST), Steals(STL), Blocks(BLK), Turnovers(TOV) and 3-Point percentage(3P%) significantly impact the model’s output.

Output from WinShare and VORP linear regression models show that better Offensive and Defensive scores in college, of players drafted from NCAA, does positively influence their success rate in the NBA while turnovers and fouls have a negative effect.

R-Squared, Adj. R-Squared and p-value of the 2 linear models

Models	Using all college predictors for each model			Using different selected college predictors for each model		
	R^2	Adj. R^2	p-value	R^2	Adj. R^2	p-value
WinShare	18.91%	14.93%	4.55e-12	13.52%	11.45%	7..108e-11
VORP	18.53%	14.53%	1.15e-11	13.42%	11.17%	2.413e-10

Do our results make sense?
Let's use the Four Factors of Basketball to find out!

Efficient Shooting	Rebounding
Turnovers	Free throw attempts per field goal attempts (FTA/FGA)

College Four Factors of Basketball Factors in If NBA Drafts the Player Logistic Model

Efficient Shooting

- High block numbers (3.26e-11)
- High assists (p=8.86e-5)
- High field goal % (p=0.0002)
- High field goals made (p=0.0118)
- High free throw % (p=0.0188)

Rebounding

- Defensive rebounds (p=0.0259)

Turnovers

- Low turnovers (p=0.0001)

FTA/ FGA

- **NONE!**

*Games played (p<2e-16), minutes played (p=0.0069), and games started (p=0.0391)

College Four Factors of Basketball Factors in Two Advanced NBA Stats Decision Tree Models

Efficient Scoring

- Blocks (x2)
- 2 pointers made
- Free throw %
- Points
- Free throws made

Rebounding

- Defensive rebounds
- Total rebounds

Turnovers

- Steals (x2)
- Turnovers

FTA/ FGA

- Free throw attempts (x2)

*Minutes played also appeared once

College Four Factors of Basketball Factors in Two Advanced NBA Stats Linear Regression Models

Efficient Scoring

- High assists (x2)
- High blocks
- High points
- Low 3 point %

Rebounding

- High defensive rebounds
- High total rebounds

Turnovers battle

- Low turnovers (x2)
- High steals (x2)

FTA/ FGA

- High 2 pointers attempted (x2)
- High 3 pointers attempted (x2)
- Low field goals attempted (x2)

*Low minutes played also appeared once

Conclusions

NBA draft vs future NBA performance:

- NBA teams overrating playing
- NBA teams underrating FTA/ FGA
- NBA teams underrating steals
- NBA teams undervalue points

Future work

- Add physical traits
- Position based models
- Add on off stats

Does our results make sense?

Yes, we think so!


Thank you!