NBA Draft Prediction

Leveraging college basketball statistics to predict the NBA draft pick

Authors: Kajal Tiwary, Austin Eaton, Jackson Piccione, Abby Fremaux

GEORGETOWN UNIVERSITY

Introduction

Problem to Address

How to predict NBA draft selections using college basketball player statistics

Importance

Teams can develop more comprehensive strategies to acquire the best talent

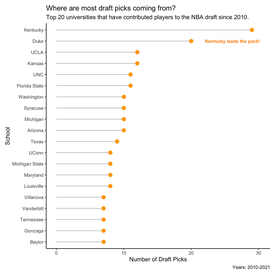
Background

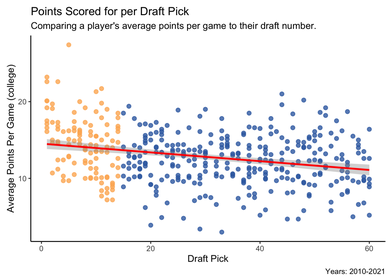
What: Selection of players to join the NBA
When: Annually every summer (June)
Who: Individuals 19 or older
How: Teams with worse records likely select first
Where: Changes year to year

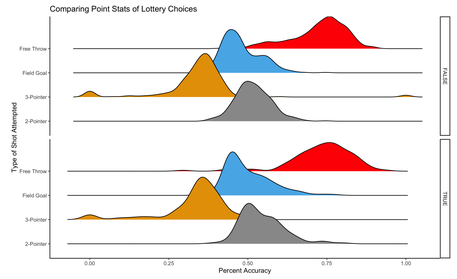
Dataset

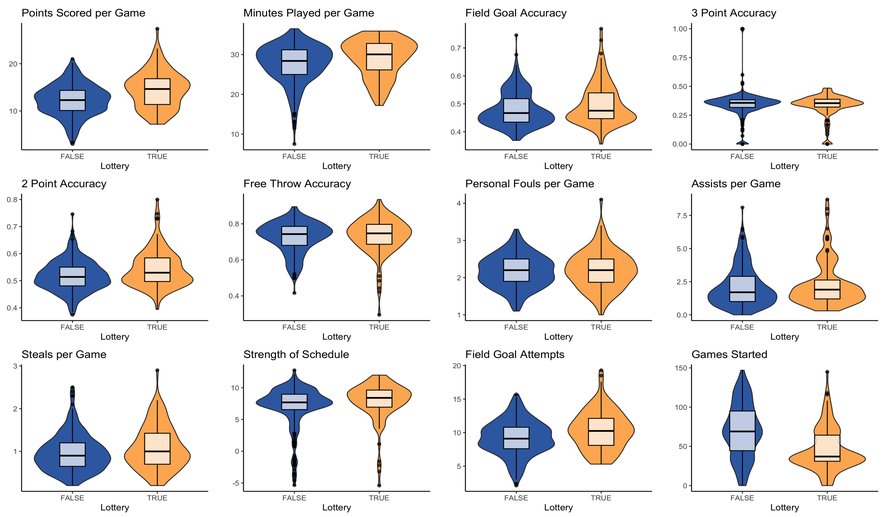
Pk	School	STL	BLK	TOV	PF	SOS	name	lottery
1	Kentucky	1.8	0.5	4	1.9	6.82	john wall	TRUE
2	Ohio State	1.6	0.7	3.5	2.7	7.86	evan turner	TRUE
3	Georgia Tech	0.9	2.1	2.5	2.6	9.02	derrick favors	TRUE
5	Kentucky	1	1.8	2.1	3.2	6.82	demarcus cousins	TRUE
7	Georgetown	1.5	1.5	2.9	2.5	9.26	greg monroe	TRUE
9	Butler	1.3	0.9	2	2.1	3.57	gordon hayward	TRUE
10	Fresno State	2	0.9	2.7	2.8	1.13	paul george	TRUE
11	Kansas	0.5	2.3	1.2	2.1	8.12	cole aldrich	TRUE
12	Kansas	1.5	0.5	1.9	1.8	8.25	xavier henry	TRUE
13	UNC	0.4	2.1	1.4	1.9	9.09	ed davis	TRUE
14	Kentucky	0.7	1.6	1.6	2.1	6.24	patrick patterson	TRUE
15	VCU	0.6	2.7	1.5	3.1	-0.69	larry sanders	FALSE
16	Nevada	0.9	0.8	2.1	2.2	1.98	luke babbitt	FALSE
18	Kentucky	1.4	0.3	3	2.2	6.82	eric bledsoe	FALSE
19	Texas	1.3	0.5	1.5	2.4	7.38	avery bradley	FALSE
21	Iowa State	0.5	1	2	2	7.34	craig brackins	FALSE

- College basketball statistics scraped from Basketball Reference
- Years: 2000-2014
- Dimensions: XX Rows, XX Columns



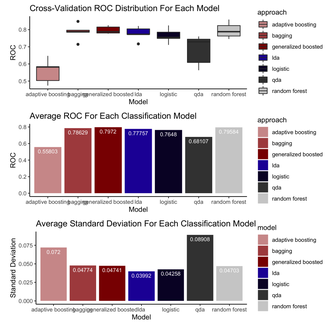




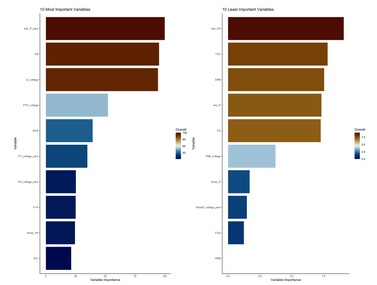


Results

Classification Model Comparison:



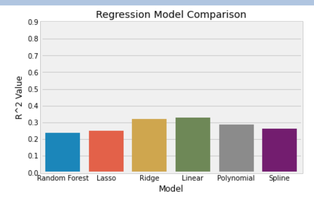
Classification Interpretation:

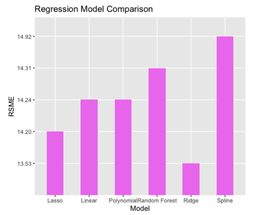


Model Approach:

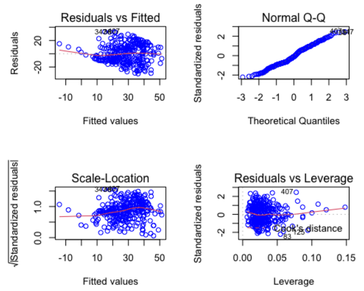
- Rank and pick variables removed to avoid multicollinearity
- Missing values removed
- Lottery converted to factor
- 5-fold cross validation because of data size
- Random grid search
- Hyperparameter tuning

Regression Model Comparison:





Regression Interpretation:



Model Approach:

- Best subset selection for OLS models only
- Six regression models run
- Most optimal model determined
- Residuals analyzed and model was re-assessed (e.g., outliers removed)

Conclusions

1. Basketball statistics can accurately predict whether a player will be in the lottery
2. Games started, two point percentage, number of games are most important in lottery prediction
3. Basketball statistics cannot accurately predict draft pick order

Future Considerations

Future Considerations

1. Assess NBA performance post-draft
2. Collect additional demographic information
3. Evaluate prospect predictions to actuals
4. Segment analysis by position

Sources: Basketball statistics and history. Basketball. (n.d.). Retrieved April 22, 2022, from https://www.basketball-reference.com/