

Trust the pRocess: NBA Award Predictions

STAT 385 SP2019 - Team Hurst

Ajay Dugar (dugar3)

Eric Qian (ewqian2)

Joshua Immanuel (joshuai2)

Bhanuchandra Kappala (kappala2)

May 9th, 2019

Abstract

This project will involve looking at 30 years of individual NBA player statistics. We will run an analysis using this historical data to develop a regression function for each individual regular season award in the NBA (Most Valuable Player, Defensive Player of the Year, Rookie of the Year, and Sixth Man of the Year). We will use this to determine every player's likelihood of winning each award. We will visualize these results using radar charts in a Shiny app.

Contents

1	Introduction	2
1.1	Problem Statement	2
1.2	Relevance	2
1.3	Description of the data	2
1.4	Connection to the Course	2
2	Related Work	2
2.1	Previous Approaches	2
2.2	Originality	3
3	Methods	3
4	Results	3
5	Discussion	3
6	Conclusion	3
7	Appendix	4
7.1	Variables	4
7.2	Results	5
	References	6

•

1 Introduction

1.1 Problem Statement

While fans and media pundits try to guess the winners of the major NBA awards (Most Valuable Player, Defensive Player of the Year, Rookie of the Year, and Sixth Man of the Year), the issue is that there's no real criteria or concrete way of predicting the winners of these awards. In this project, we seek to use historical data to predict these winners.

1.2 Relevance

Valuing player assets are important. Being able to accurately determine player value is one of the most fundamental skills in basketball front offices. Being able to do so accurately can be the difference between winning and losing. Many different subjects (statistics, differential equations, analytics) can be applied to these problems. With the success of Daryl Morey and Sam Hinkie, we see that these approaches have been successful in the NBA. Many interesting and unsolved issues in the NBA require statistical analysis, which is why we chose this project.

Our shared passion of basketball and the NBA led us to this topic. This is a topic that some of us have talked about and previous academic curiosity have led us to this topic. We wanted to know if we could use a variety of variables to accurately predict NBA Award winners.

1.3 Description of the data

The data is a series of dataframes of regular and advanced statistics spanning from 1989 and 2019. The statistics are listed and explained in the Appendix. Our data will come from the NBA Reference website. It will be 30 years of individual player data. The long time period and amount of data for our data points will allow us for a greater degree of accuracy and precision in the evaluation of our model.

1.4 Connection to the Course

We will be dealing with large datasets, validating data, and building statistical models. All of these skills form the foundations of statistical programming and we will apply what we have learned in our project.

2 Related Work

2.1 Previous Approaches

While not specifically related to this project, the application of sabermetrics has become a point of interest for many baseball teams. Likewise, we have begun to see this same transition in basketball analytics. The long term view of teams that have embraced analytics has been very good: the Philadelphia 76ers and Houston Rockets are both top seeded teams that have a significant chance at being the Championship team. While we haven't seen rigorous academic analysis of predicting NBA awards, there have been some implementations done by various individuals. There are also award trackers that just use raw data - specifically VORP (Value over Replacement Player) and PER (Player Efficiency Rating) and the player with the highest value is predicted to win. While these ways of predicting awards works somewhat well, they tend to have problems undervaluing certain players (the inventor of PER notes that it is not an end all metric - it rewards inefficient shooting and doesn't sufficiently reward effective defense) so our idea will be unique.

2.2 Originality

A lot of the statistics in use with modern publications are based on the research of one individual: Daryl Morey. We will derive our own variables and determine how impactful they are in determining wins using a regression model. By doing so, we will establish other variables and metrics that may be better suited for different situations and will allow for better comparison of players across time periods and eras.

3 Methods

First we use the ballr package (Elmore 2018) to pull the statistical data from Basketball Reference website (“Basketball Statistics and History,” n.d.). Then we clean and combine the data using the concepts of “tidy data” (Wickham and Henry 2019). We introduce dummy variables with respect to position (C, SG, SF, PF, PG). For the interactive interface, we have created a Shiny App. The user can choose both the year (between 1989 and 2019) and the award to be predicted. In order to visualize the data, we give radar plots of different attributes important for each of the awards. Next to this is the probability that that person would win the award.

4 Results

We have effectively created predictions for every year from 1989 using our model. The model boasts at least above a 70% success rate - the model is also exceptionally good at picking the top player for each award every year. Our top five predictions tend to wane off in accuracy after the first prediction. Seeing how NBA awards are coach/media votes, it’s hard to numerically gauge whether those players were really in the order of most effective. Take for example, Steph Curry’s unanimous MVP year of 2016: He only has around a 70% of winning, though you could say it was probably estimated to be impossible for him to lose that year.

5 Discussion

It is not particularly difficult to predict the overall winner of most NBA awards. Typically, the winner is pretty clear, but even if the winner is not obvious, our model does a good job of predicting who the top candidates are with suprising accuracy. Our results suggest that our goal has been accomplished - with systematic data from every year, we can predict the winner with a surprisingly high level of accuracy.

6 Conclusion

Overall, the project was generally successful. We took historical data and found accurate results for older years and probably for this year. For the year where MVP has not been decided yet, our choice is Giannis Antetokounmpo who is most analysts pick for the MVP despite James Harden’s incredible scoring season. # Conclusion If you can accurately predict NBA awards, there is significant money to be made betting on winners of these awards.

Additionally, by looking at this topic, NBA teams can get a better idea of what a winning player looks like, and can construct teams accordingly. This project will involve looking at 30 years of individual NBA player statistics. We will develop a number of new advanced analytics and statistics to determine player efficient and value. Using these numbers, we will rank players and determine who should win prominent NBA awards. We will visualize these results using bar graphs and sentiment analysis.

7 Appendix

7.1 Variables

- player - name of the player
- pos - position
- age - age
- g - games played
- mp - minutes played
- per - player efficiency rating

$$\begin{aligned} per = & (1/mp) * [3p + (2/3) * ast + (2 - factor * (ast_{team}/fg_{team})) * fg + (ft * 0.5 * (1 + (1 - (ast_{team}/fg_{team}))) \\ & + (2/3) * (ast_{team}/fg_{team}))) - vop * tov - vop * drb\% * (fga - fg) - vop * 0.44 * (0.44 + (0.56 * drb\%)) * (fta - ft) \\ & + vop * (1 - drb\%) * (drb) + vop * drb\% * orb + vop * stl + vop * drb\% * blk - pf * ((ft_{league}/pf_{league}) - 0.44 * (fta_{league}/pf_{league}) * vop)] \end{aligned}$$

- tspercent - true shooting percent

$$tspercent = \frac{pts}{2 * (fga + (0.44 * fta))} * 100$$

- x3par - 3 point attempt rate
- ows - offensive win shares
- dws - defensive win shares
- trbpercent - total rebound percentage
- astpercentage - assist percentage
- blkpercent - block percentage
- tovpercent - turnover percentage
- usgpercent - usage percentage
- ftr - free throw rate
- ws - win shares
- ws_48 - win shares per 48 minutes
- gs - games started
- fg - field goals made
- fga - field goals attempted
- fgpercent - field goal percentage
- orb - offensive rebounds
- drb - defensive rebounds
- trb - total rebounds
- x3p - 3 pointers made
- x3pa - 3 pointers attempted
- x3pppercent - 3 point percentage
- x2p - 2 pointers made
- x2pa - 2 pointers attempted
- x2pppercent - 2 point percentage
- ast - assists
- stl - steals
- ft - free throws
- fta - free throw attempts
- ftpercent - free throw percentage
- blk - blocks
- tov - turnovers
- pf - personal fouls

- pts - points per game
- efgpercent - effective field goal percentage
- obpm - offensive box plus/minus
- dbpm - defensive box plus/minus
- bpm - total box plus/minus
- vorp - value over replacement player

NBA Awards Predictor

Award Selection

What award do you want to predict?

MVP

Year (e.g. 1980)

2019

Load Preview Awards

Summary

Radar Plot of 1st Place

Radar Plot of 2nd Place

Radar Plot of 3rd Place

Radar Plot of 4th Place

Radar Plot of 5th Place

Radar Plot of 6th Place

Top Six Predictions

	Player	Position	Age	Probability
6	Giannis Antetokounmpo	PF	24	69.2025753
105	James Harden	PG	29	21.2244648
61	Anthony Davis	C	25	2.5880189
142	Nikola Jokic	C	23	2.1805201
73	Joel Embiid	C	24	0.8311908
247	Karl-Anthony Towns	C	23	0.7645522

7.2 Results

References

- Allaire, JJ. 2019. *Rsconnect: Deployment Interface for R Markdown Documents and Shiny Applications*.
- “Analytics 101.” n.d. <https://www.nbastuffer.com/analytics-101/>.
- Bache, Stefan Milton, and Hadley Wickham. 2014. *Magrittr: A Forward-Pipe Operator for R*.
- “Basketball Analytics.” n.d. *Basketball Analytics*. <http://www.basketballanalyticsbook.com/>.
- “Basketball Statistics and History.” n.d. *Basketball*. <https://www.basketball-reference.com/>.
- Dowle, Matt, and Arun Srinivasan. 2019. *Data.table: Extension of 'Data.frame'*.
- Elmore, Ryan. 2018. *Ballr: Access to Current and Historical Basketball Data*.
- Inc., Plotly Technologies. 2015. “Collaborative Data Science.” Montreal, QC: Plotly Technologies Inc. 2015. <https://plot.ly>.
- “NBA Stats.” n.d. National Basketball Association. <https://stats.nba.com/>.
- “Squared Statistics: Understanding Basketball Analytics.” n.d. *Squared Statistics: Understanding Basketball Analytics*. <https://squared2020.com/>.
- Wickham, Hadley. 2011. “The Split-Apply-Combine Strategy for Data Analysis.” *Journal of Statistical Software* 40 (1): 1–29. <http://www.jstatsoft.org/v40/i01/>.
- . 2018. *Stringr: Simple, Consistent Wrappers for Common String Operations*.
- . 2019. *Rvest: Easily Harvest (Scrape) Web Pages*.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2019. *Dplyr: A Grammar of Data Manipulation*.
- Wickham, Hadley, and Lionel Henry. 2019. *Tidyr: Easily Tidy Data with 'Spread()' and 'Gather()' Functions*.