MODELING ALL-NBA TEAM VOTING

Eric J. Wissner (Group 7)

gtID: ******824

ejwissner@gatech.edu

PROBLEM STATEMENT

Despite advances in how analytics are used to understand and measure National Basketball Association player performance, "All-NBA Team" honors are determined by voting conducted by a panel of 100 sports reporters and broadcasters.

To what extent can an analytic model be trained to effectively represent historical All-NBA Team voting results?

Can this model be used to assess the results of subsequent All-NBA Team voting?

DATA OVERVIEW

20 years of NBA player data and All-NBA Team voting results were obtained via web-scraping from basketball-reference.com.

After removing data for players with less than 10 games played in a year, the total data set contained 8,717 records – 8,223 for the first 19 years (training set) and 494 records for the 2021-2022 NBA season (prediction set).

In addition to traditional basketball statistics (e.g., field goals, free throws, and rebounds), advanced metrics, including composite scores intended to capture a player's total performance, were used as predictor variables.

A binary indicator of whether each player earned All-NBA Team honors in that year served as the response variable.

Source: *https://basketball-reference.com*

PREDICTOR VARIABLES

The following predictor variables were included for each player per season.

Season Totals

Variable	Description	Variable	Description
Age	Age	EffFGPct	Effective field goal percentage
Gms	Games played	FT	Free throws
GmsStarted	Games started	FTA	Free throws attempted
MP	Minutes played	FTPct	Free throw percentage
FG	Field goals made	ORB	Offensive rebounds
FGA	Field goals attempted	DRB	Defensive rebounds
FGPct	Field goal percentage	TRB	Total rebounds
X3P	Three-point shots made	AST	Assists
X3PA	Three-point shots attempted	STL	Steals
X3PPct	Three-point shot percentage	BLK	Blocks
X2P	Two-point shots made	TOV	Turnovers
X2PA	Two-point shots attempted	PF	Personal fouls
X2PPct	Two-point shot percentage	PTS	Points

Advanced Statistics

Variable	Description	Variable	Description
PER	Player efficiency rating	TOVPct	Turnover percentage
TSPct	True shooting percentage	USGPct	Usage percentage
X3PAr	Three-point attempt rate	OWS	Offensive win shares
FTr	Free throw attempt rate	DWS	Defensive win shares
ORBPct	Offensive rebound percentage	WS	Total win shares
DRBPct	Defensive rebound percentage	WSper48	Win shares per 48 minutes
TRBPct	Total rebound percentage	OBPM	Offensive box plus/minus
ASTPct	Assist percentage	DBPM	Defensive box plus/minus
STLPct	Steal percentage	BPM	Total box plus/minus
BLKPct	Block percentage	VORP	Value over replacement player

Source: *https://basketball-reference.com*

METHODOLOGY

The following classification models were trained (with parameters tuned as noted):

- Logistic Regression
- LASSO (lambda tuned so cross-validation error is within 1 standard error of the minimum error)
- KNN (number of neighbors tuned to reduce cross-validated training error)
- PCA with KNN (number of principal components tuned to explain at least 80% of variance in data; number of neighbors tuned to reduce cross-validated training error)
- Classification Tree
- Random Forest (number of trees used and number of variables included to minimize cross-validated training error)
- Boosting (number of trees/iterations included in the leaning process tuned to optimal quantity)

The best model was determined by averaging the training error rates produced during ten-fold cross-validation.

RESULTS

The Boosting model returned the best cross-validated training error rate (0.50%) and was used to assess the actual results of 2021-2022 All-NBA Team voting.

Analytic Model	Cross-Validated Training Error Rate
Logistic Regression	0.0151
LASSO	0.0170
KNN	0.0163
PCA with KNN	0.0203
Classification Tree	0.0215
Random Forest	0.0152
Boosting	0.0050

ASSESSING 2021-2022 ALL-NBA TEAM VOTING

KIA PERFORMANCE AWARDS

2021-22 KIA ALL-NBA TEAM

Position	Player (Team)	1st Team Votes (5 Points)	2nd Team Votes (3 Points)	3rd Team Votes (1 Point)	Total Points
2021-22 KIA ALL-NBA FIRST TEAM					
Forward	Giannis Antetokounmpo (Milwaukee) ✓	100	0	0	500
Guard	Luka Dončić (Dallas) ✓	88	12	0	476
Center	Nikola Jokić (Denver) ✓	88	12	0	476
Guard	Devin Booker (Phoenix) ✓	82	16	2	460
Forward	Jayson Tatum (Boston) ✓	49	47	4	390
2021-22 KIA ALL-NBA SECOND TEAM					
Center	Joel Embiid (Philadelphia) ✓	57	43	0	414
Guard	Ja Morant (Memphis) ✓	13	76	8	301
Forward	Kevin Durant (Brooklyn) ✓	10	68	22	276
Guard	Stephen Curry (Golden State) ✓	9	69	22	274
Forward	DeMar DeRozan (Chicago) ✓	2	39	57	184
2021-22 KIA ALL-NBA THIRD TEAM					
Center	Karl-Anthony Towns (Minnesota) ✓	0	38	60	174
Forward	LeBron James (L.A. Lakers) ✓	2	35	54	169
Guard	Chris Paul (Phoenix) ✗	0	16	66	114
Guard	Trae Young (Atlanta) ✓	0	11	77	110
Forward	Pascal Siakam (Toronto) ✗	0	7	42	63

Players Selected Using Boosted Model

Pos	Player	Probability	
F	Giannis Antetokounmpo	99.54%	
C	Nikola Jokic	99.12%	
G	Luka Doncic	96.75%	
C	Joel Embiid	96.71%	
F	DeMar DeRozan	92.25%	
F	Kevin Durant	91.85%	
G	Trae Young	90.94%	
G	Ja Morant	89.80%	
C	Rudy Gobert	89.74%	Not voted in
G	Stephen Curry	87.37%	
F	Jayson Tatum	83.84%	
C	Karl-Anthony Towns	72.25%	
C	LeBron James	66.64%	
F	Jimmy Butler	64.54%	Not voted in
G	Devin Booker	51.27%	

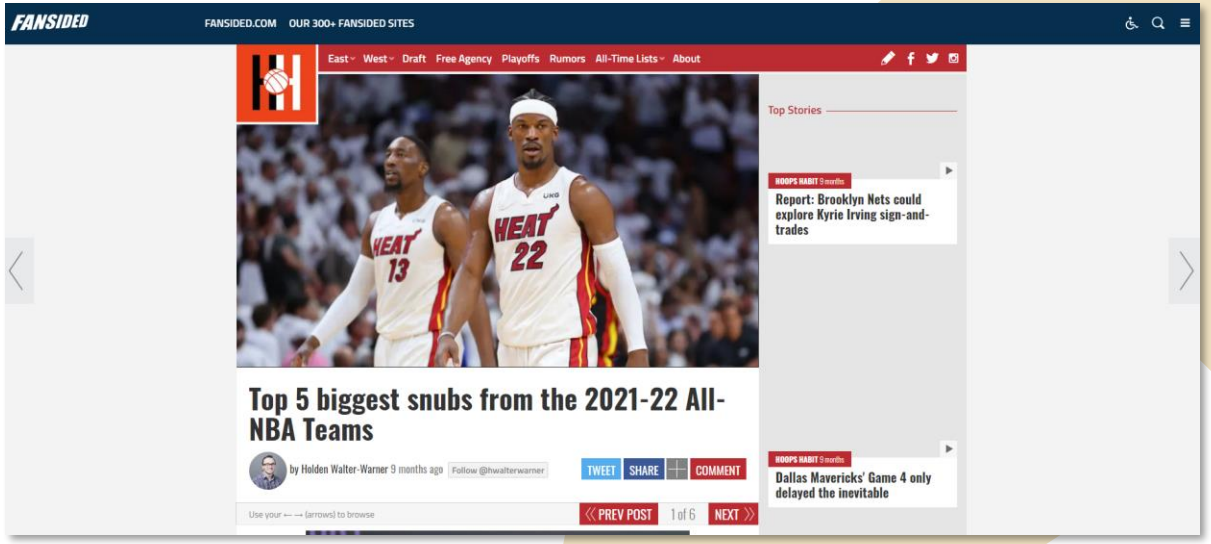
All-NBA Team Members Excluded from Boosted Model

Pos	Player	Probability
G	Chris Paul	47.14%
F	Pascal Siakam	5.79%

Source: *https://www.nba.com/news/antetokounmpo-jokic-doncic-2021-22-kia-all-nba-first-team*

INDEPENDENT CORROBORATION OF RESULTS

According to an article on Hoops Habit, a Fansided website, Rudy Gobert and Jimmy Butler were the two biggest "snubs" from the 2021-2022 All-NBA Team.



Source: *Walter-Warner, H. (2022, May 28). Top 5 biggest snubs from the 2021-22 all-NBA teams. Hoops Habit. Retrieved from https://hoopshabit.com/2022/05/28/snubs-all-nba-teams/*

CONCLUSION

Using individual player data for the 2021-2022 NBA season, the boosted analytic model was able to accurately predict 13 out of the 15 All-NBA Team selections for that season.

Given the wealth of data now available on player performance, it may be time to determine All-NBA Team members based on advanced analytics and not on potentially biased human voting.

The ensemble Boosting method would be a good option because, as a black box approach, it is not directly interpretable; meaning players could not game the system by padding specific statistics. Additionally, as a model trained by previous All-NBA Team voting patterns, it inherently would honor the traditions of the past by incorporating its complex reasoning in a modern way.