

GA Capstone: Predicting NBA Accolades

By Eli Winton





Problem Statement

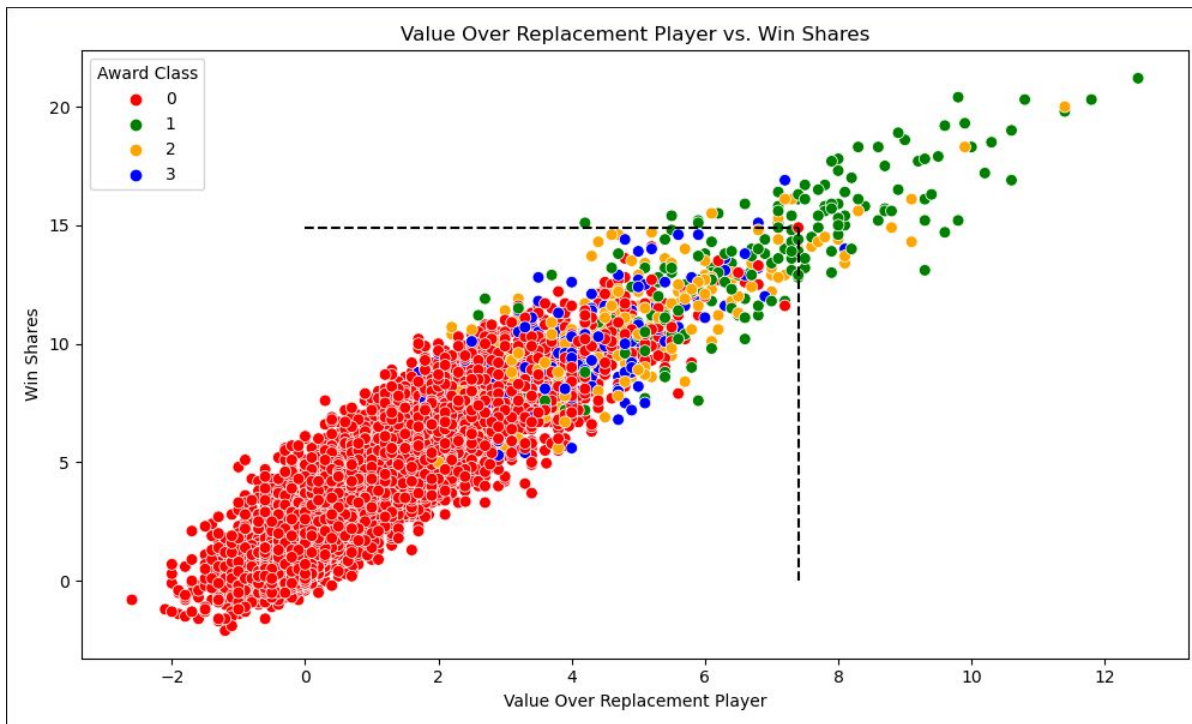
- All-NBA teams are the most prestigious awards in the NBA consisting of three teams of five players recognizing the top players every season.
- Making the teams has an enormous impact on a player's career, influencing their legacy, contract negotiations, and marketability.
- Predicting All-NBA team using various performance metrics, player attributes, and team dynamics.
- Three main steps: Scraping, combining all of the data, and Modeling.



Scraping/Concat

- All data was scraped from Basketball-Reference.com.
- Player data was scraped back to the 1979-1980 season as that was when the three point line was introduced.
- Cleaning and preprocessing: handling missing values, converting data types as needed, and dealing with duplicate entries for players that were traded mid-season.
- Standardizing player names, team names, and other categorical variables to maintain consistency.

Most Correlated Metrics





Modeling

Model	F1 Score	F1 Score Change
Baseline: SS, Random Forest	<ul style="list-style-type: none">• No Award: 0.99• 1st Team All-NBA: 0.68• 2nd Team All-NBA: 0.29• 3rd Team All-NBA: 0.08	
SS, Random Forest (class_weight = 'balanced')	<ul style="list-style-type: none">• No Award: 0.99• 1st Team All-NBA: 0.67• 2nd Team All-NBA: 0.19• 3rd Team All-NBA: 0.04	<ul style="list-style-type: none">• No Award: 0.0• 1st Team All-NBA: -0.01• 2nd Team All-NBA: -0.10• 3rd Team All-NBA: -0.04
SS, SMOTE(18,000:9,000), Random Forest	<ul style="list-style-type: none">• No Award: 0.99• 1st Team All-NBA: 0.63• 2nd Team All-NBA: 0.40• 3rd Team All-NBA: 0.32	<ul style="list-style-type: none">• No Award: 0.0• 1st Team All-NBA: -0.05• 2nd Team All-NBA: +0.11• 3rd Team All-NBA: +0.24



Modeling

Model	F1 Score	F1 Score Change
Baseline: SS, Random Forest	<ul style="list-style-type: none">• No Award: 0.99• 1st Team All-NBA: 0.68• 2nd Team All-NBA: 0.29• 3rd Team All-NBA: 0.08	
SS, SMOTE(18,000:14,000), Random Forest	<ul style="list-style-type: none">• No Award: 0.99• 1st Team All-NBA: 0.64• 2nd Team All-NBA: 0.38• 3rd Team All-NBA: 0.33	<ul style="list-style-type: none">• No Award: 0.0• 1st Team All-NBA: -0.04• 2nd Team All-NBA: +0.09• 3rd Team All-NBA: +0.25
SS, ADASYN(18,000:9,000), Random Forest	<ul style="list-style-type: none">• No Award: 0.99• 1st Team All-NBA: 0.68• 2nd Team All-NBA: 0.35• 3rd Team All-NBA: 0.34	<ul style="list-style-type: none">• No Award: 0.0• 1st Team All-NBA: 0.0• 2nd Team All-NBA: +0.06• 3rd Team All-NBA: +0.26

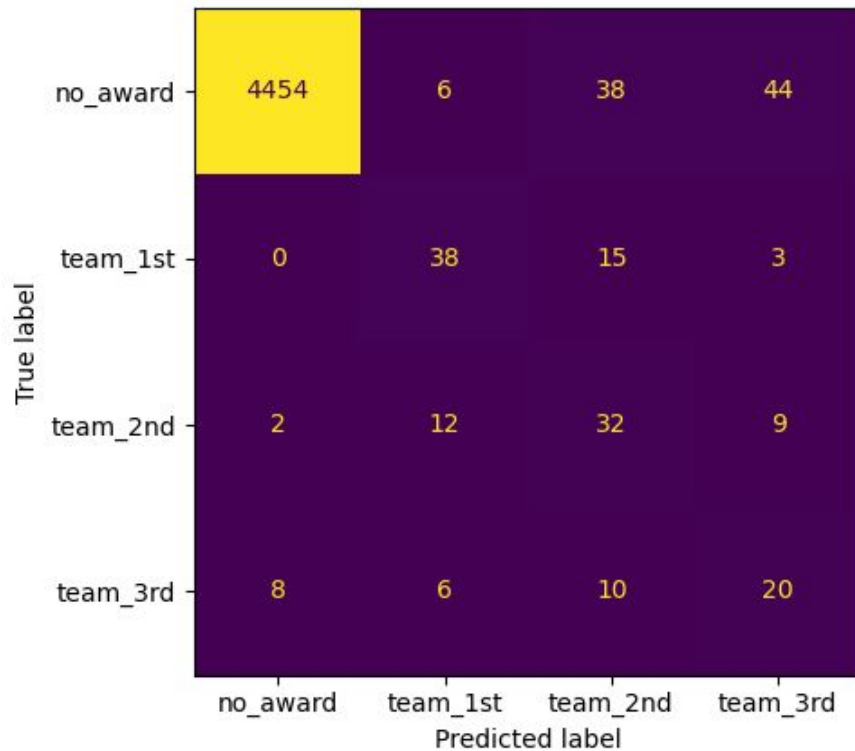


Modeling

Model	F1 Score	F1 Score Change
Baseline: SS, Random Forest	<ul style="list-style-type: none">• No Award: 0.99• 1st Team All-NBA: 0.68• 2nd Team All-NBA: 0.29• 3rd Team All-NBA: 0.08	
SS, SMOTE(18,000:9,000), XGB	<ul style="list-style-type: none">• No Award: 0.99• 1st Team All-NBA: 0.65• 2nd Team All-NBA: 0.40• 3rd Team All-NBA: 0.20	<ul style="list-style-type: none">• No Award: 0.0• 1st Team All-NBA: -0.03• 2nd Team All-NBA: +0.11• 3rd Team All-NBA: +0.12
SS, SMOTEENN, Random Forest	<ul style="list-style-type: none">• No Award: 0.99• 1st Team All-NBA: 0.64• 2nd Team All-NBA: 0.43• 3rd Team All-NBA: 0.33	<ul style="list-style-type: none">• No Award: 0.0• 1st Team All-NBA: -0.04• 2nd Team All-NBA: +0.14• 3rd Team All-NBA: +0.25



Best Model Confusion Matrix





Recommendations

- Additional feature engineering: injury data, salary information, and team performance metrics
- Wider range of machine learning algorithms and continuing to fine-tune hyperparameters
- Dynamic time-series modeling techniques to incorporate real-time updates of player statistics
- More data!!