NBA Accolades Analysis

By: Mitchell Ren

Overview

- Business Context
- Data Exploration
- Modeling
- Results Evaluation
- Key Takeaways & Next Steps

Business Context

*Each year, members of the NBA media vote on superlative awards like **Most Valuable Player and Defensive Player of the Year.***

Opportunity

The NBA has yet to clearly define its voting criteria.

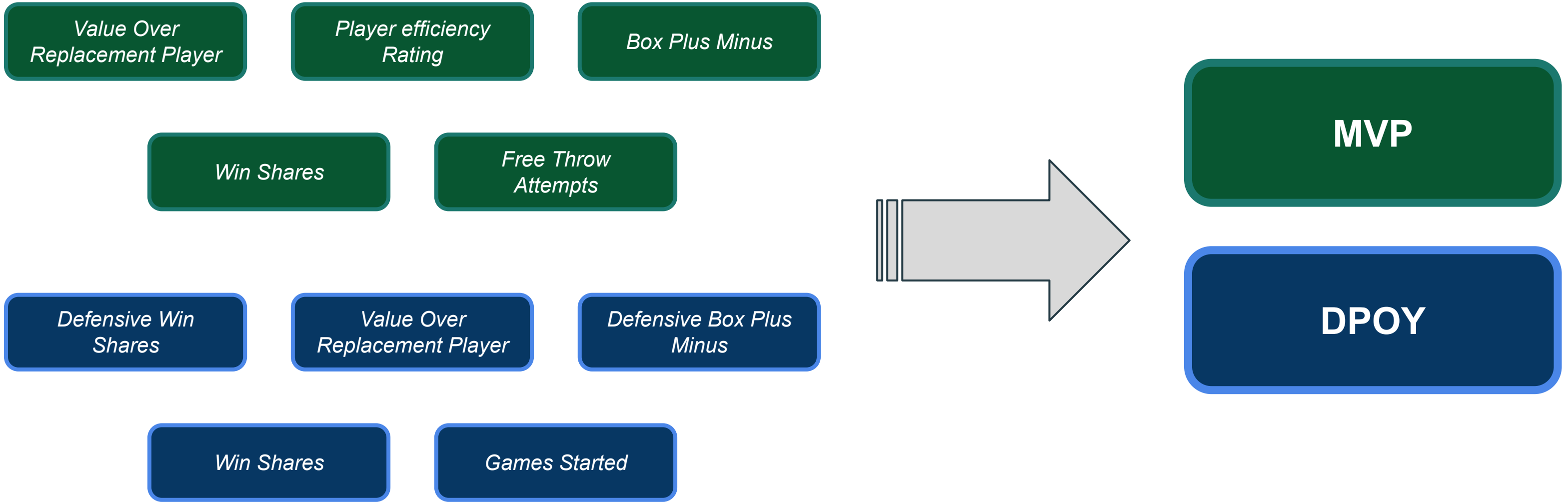
Approach

Conduct **supervised learning** in order to understand:

- Which **statistical categories are most important** to predicting MVP and DPOY?
- How successfully can we **make future predictions** based on historical selections?

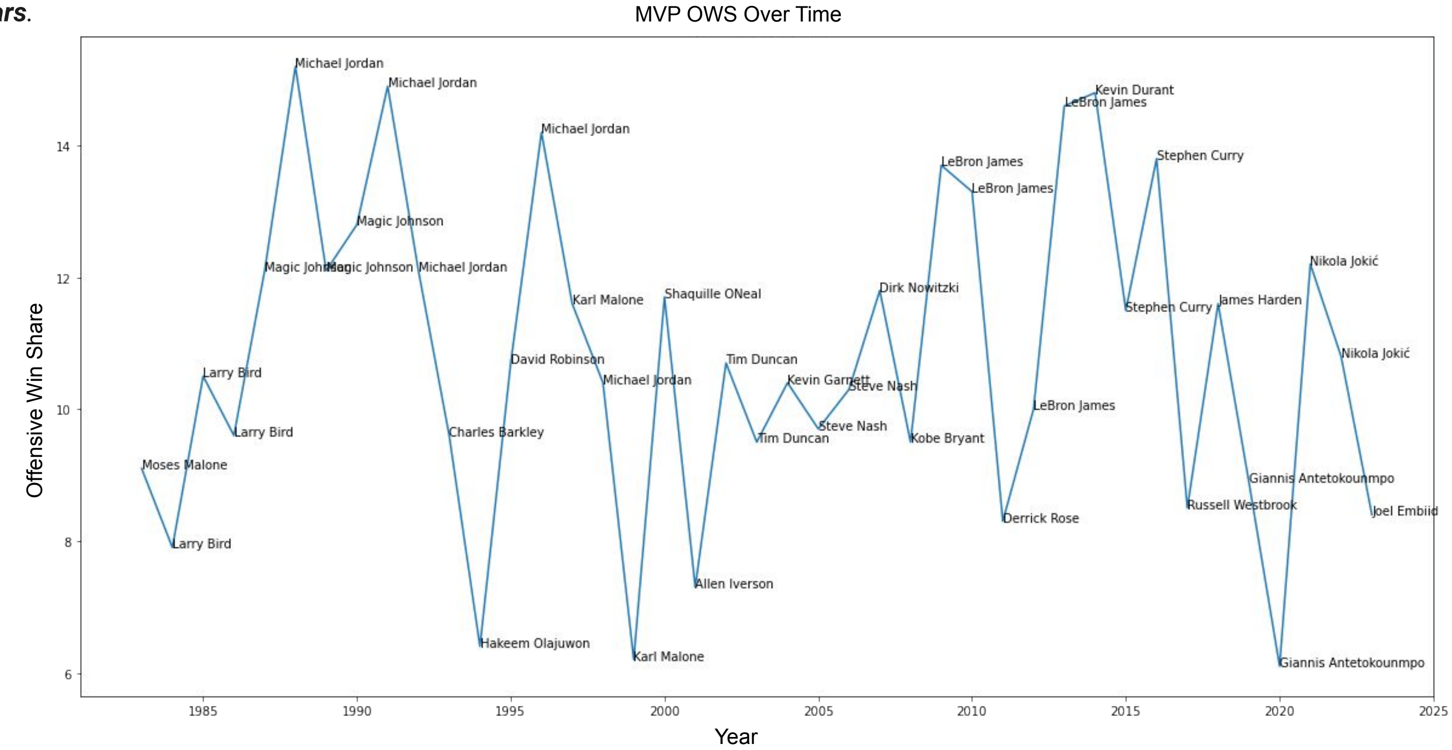
Data Exploration

*I scraped seasonal player data from **Basketball-Reference.com** for the last 40 seasons - the final, consolidated dataset included over **20,000 rows and 140 features**.*



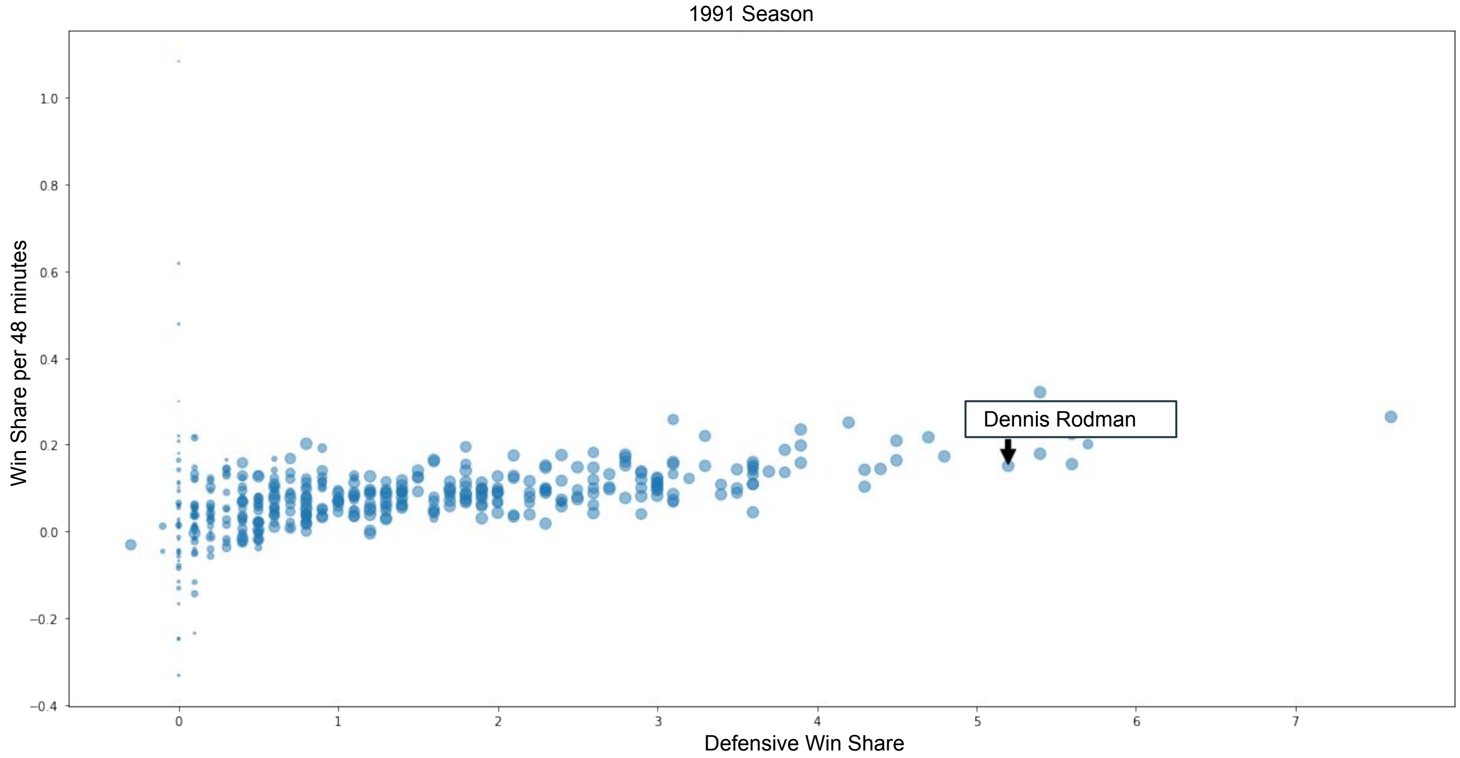
Data Exploration (Cont.)

The below graph shows that ***offensive win share (OWS) was a key consideration for certain years and not very much so for other years***.



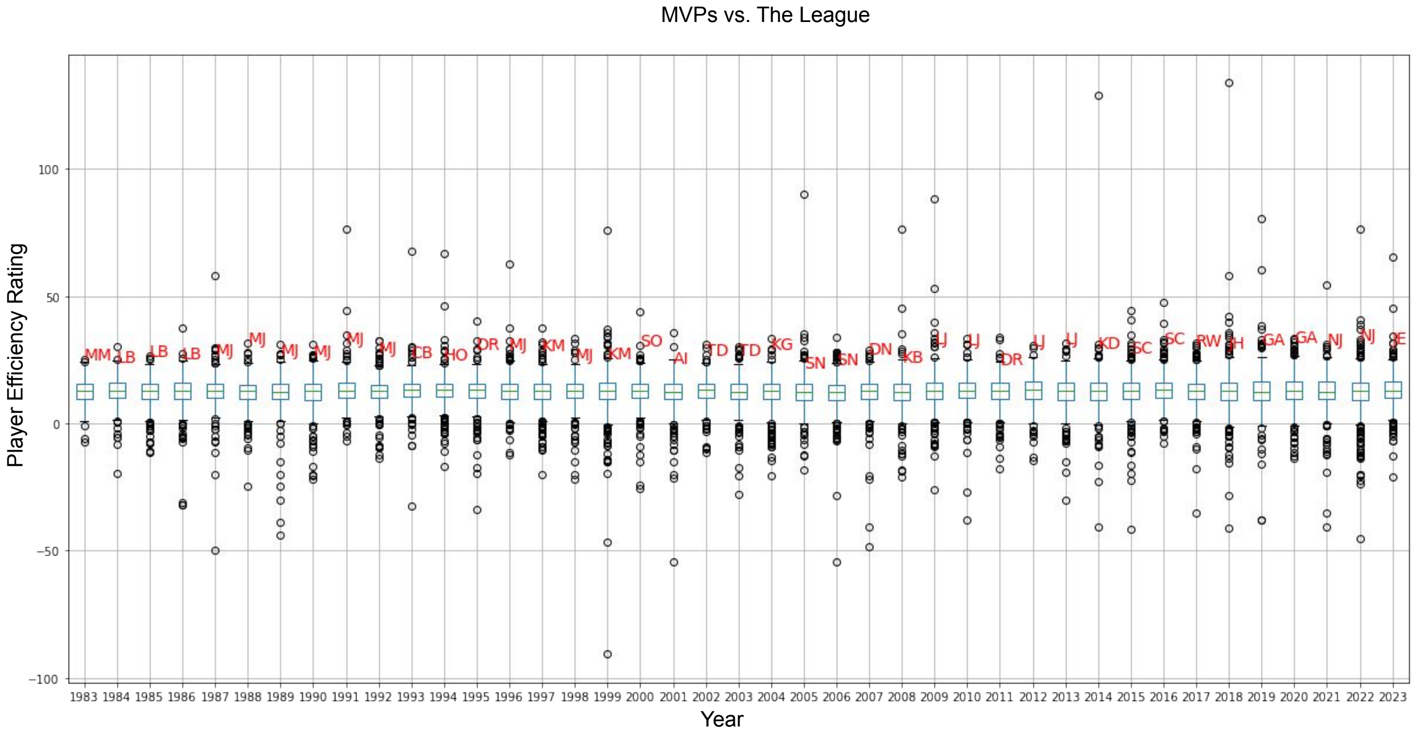
Data Exploration (Cont.)

*The below graph shows that **although Dennis Rodman won DPOY in 1991, there were other players more outstanding**.*



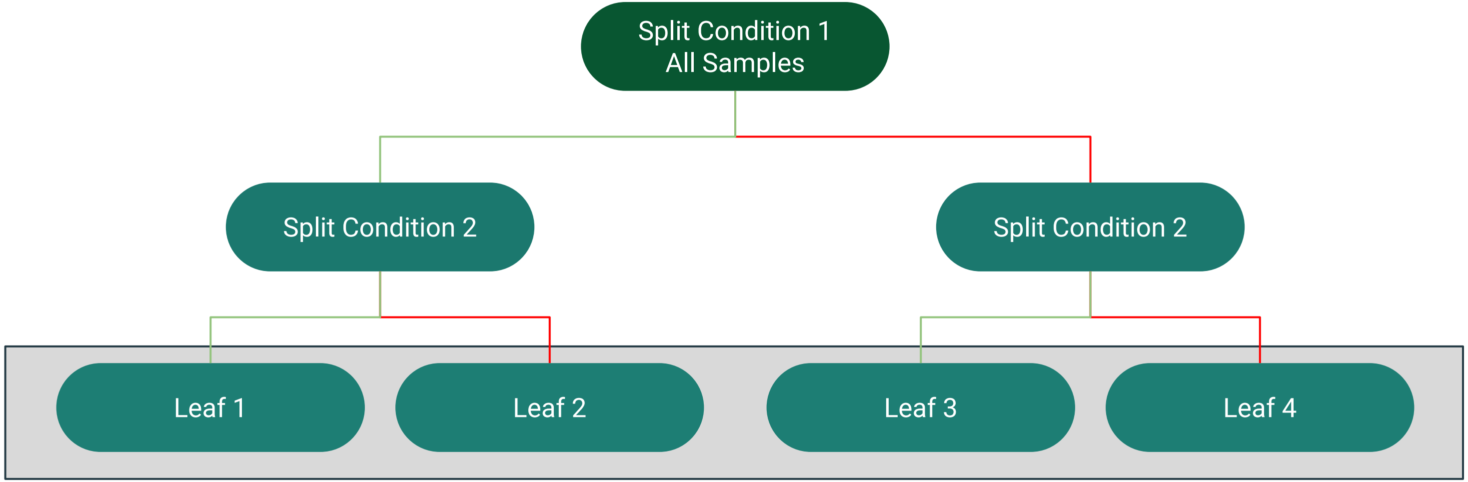
Data Exploration (Cont.)

The below graph visualizes how much ***MVPs were outliers compared to the rest of the field*** *within a given season.*



Modeling

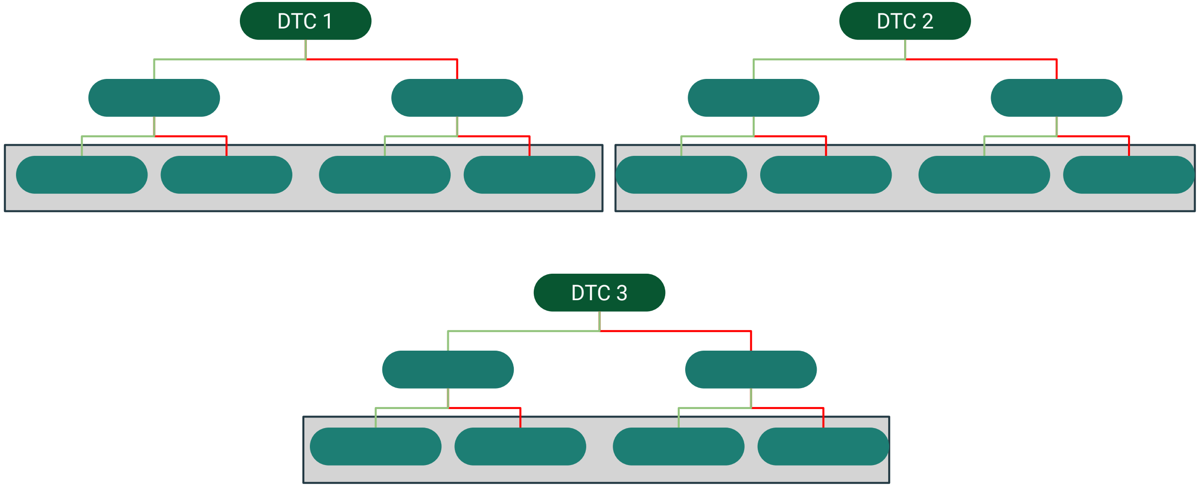
*A Decision Tree Classifier model learns patterns in the data through a series of **condition-based splits**. It splits the data into sub-samples and makes **predictions** from what it learns.*



Decision Tree Classifier

Modeling (Cont.)

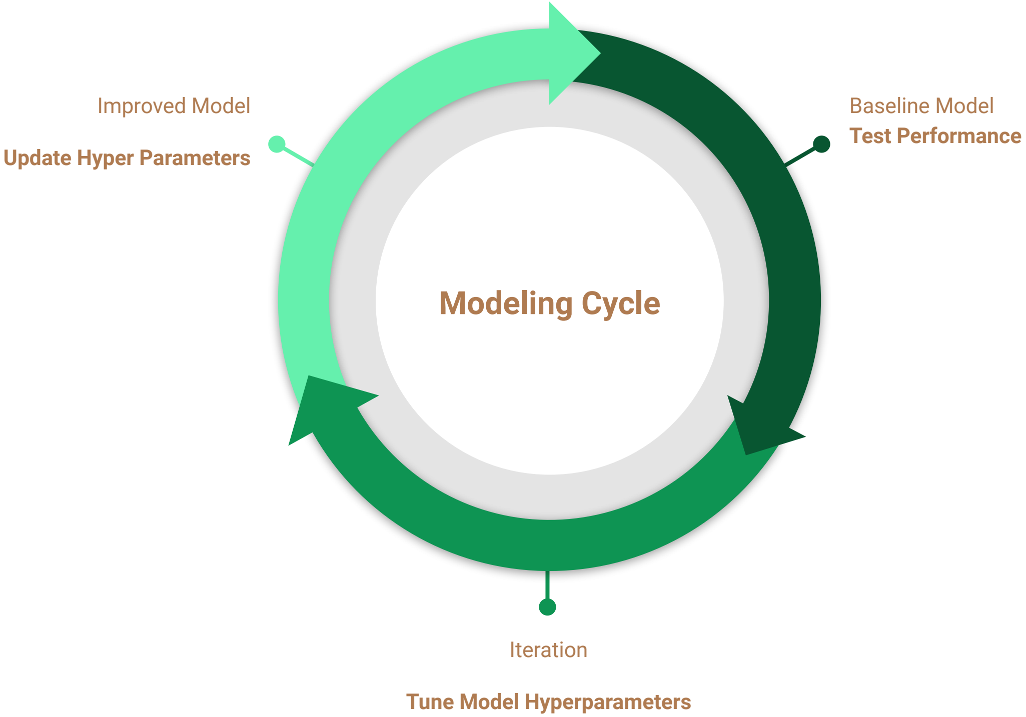
A Random Forest Classifier model combines learnings from ***multiple Decision Tree Classifiers*** *to make its predictions.*



Random Forest Classifier

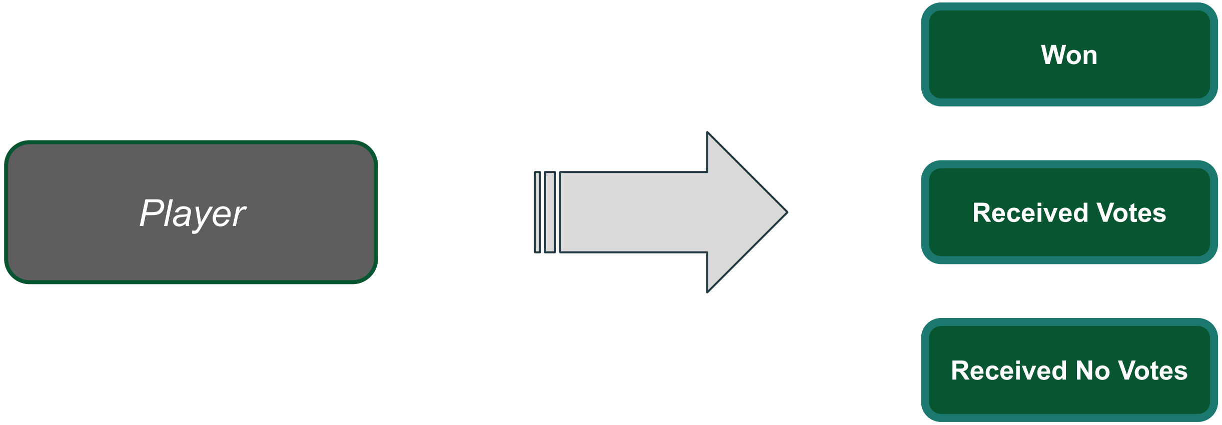
Modeling (Cont.)

*I took an **iterative** approach to modeling and applied it to each model version.*



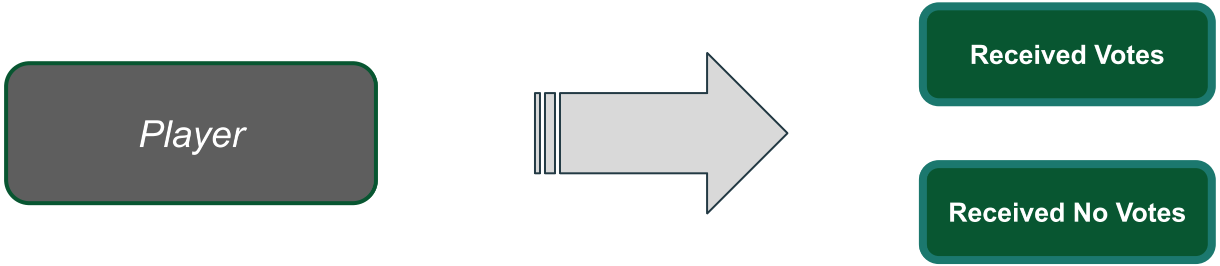
Modeling (Cont.)

I originally approached this problem as a ***multi-class*** *one. I built baseline models to categorize players into one of three categories:* ***Won (the award), Received Votes, Received No Votes****.*



Modeling (Cont.)

*I then tested changing the target variables to be **binary** as opposed to multi-class to see how that impacted model performance.*



Results & Evaluation (Multi-Class) - MVP

*The **Random Forest Classifier with SMOTE resampling and hyperparameter tuning** showed the best performance.*

Model	F1-Score (Macro Average)
Baseline Multi-class DTC	Train: 100% | Test: 64%
Multi-class DTC with SMOTE Resampling	Train: 100% | Test: 70%
Multi-class DTC with SMOTE Resampling and Hyperparameter Tuning	Train: 98% | Test: 71%
Multi-class RFC with SMOTE Resampling and Hyperparameter Tuning	Train: 100% | Test: 74%

Results & Evaluation (Binary) - MVP & DPOY

*The final RFC I built for **MVP predictions performed better** than did the one I built for DPOY predictions.*

Model	F1-Scores (Macro Average)	F1-Scores (Test)
Binary RFC with SMOTE Resampling and Hyperparameter tuning (MVP)	Train: 92% | Test: 83%	Received No Votes: 99% Received Votes: 68%
Binary RFC with SMOTE Resampling and Hyperparameter tuning (DPOY)	Train: 98% | Test: 71%	Received No Votes: 97% Received Votes: 45%

Key Takeaways

Feature importances *reveal a good starting point for defining* ***objective, statistical criteria*** *for MVP and DPOY voting.*

MVP

Top 5 feature importances:
- Value Over Replacement Player (VORP)
- Player Efficiency Rating (PER)
- Box Plus Minus (BPM)
- Win Shares (WS)
- Free Throw Attempts (FTA)

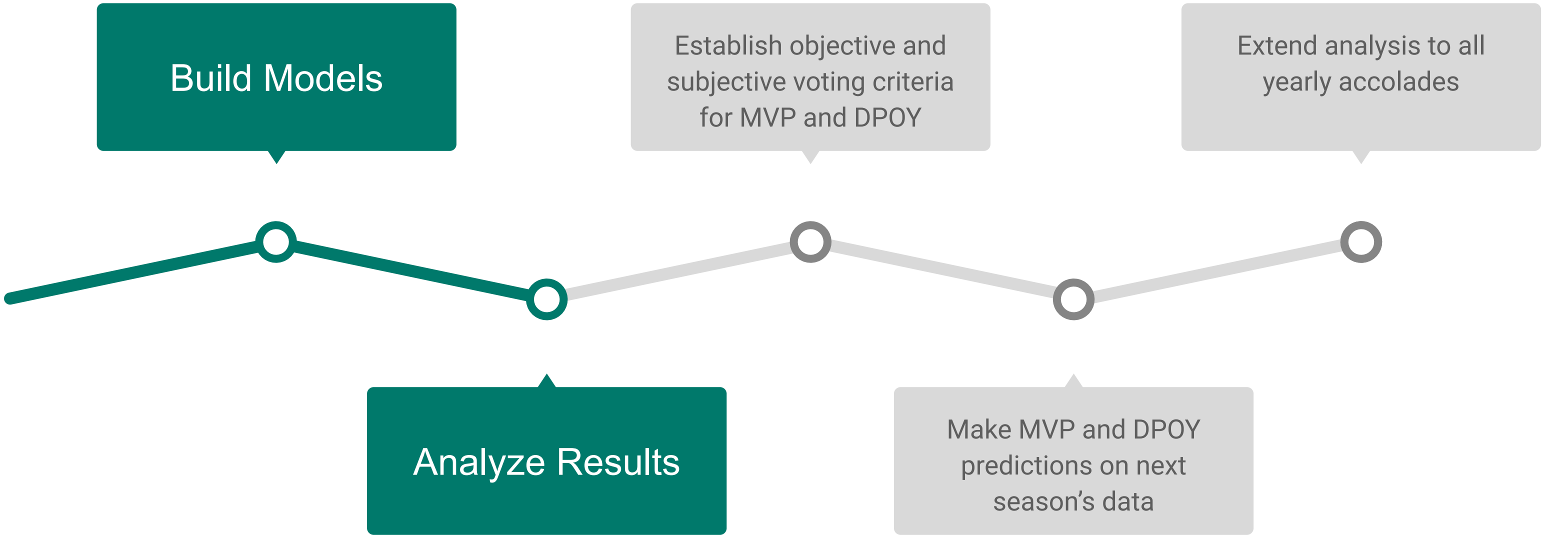
DPOY

Top 5 feature importances:
- Defensive Win Share (DWS)
- Value Over Replacement Player (VORP)
- Defensive Box Plus Minus (DBPM)
- Win Shares (WS)
- Games Started (GS)

Recommendations

01	Use feature importances from binary RFC models as basis for objective criteria	• Clearly define which statistical categories should be evaluated when selecting MVP and DPOY
02	Implement binary RFCs to make future predictions	• Use model predictions on future data as baseline for player selection pool
03	Establish subjective voting criteria	• Clearly define subjective criteria so that media can incorporate narrative components when voting
04	Share full analysis with fans, players to provide transparency	• Leading up to winner announcement - share statistical analysis as well as key narrative considerations on NBA countdown or ESPN

Next Steps


Build Models
Analyze Results
Establish objective and subjective voting criteria for MVP and DPOY
Make MVP and DPOY predictions on next season's data
Extend analysis to all yearly accolades

Thank You

Contact: mitchell.mathew.ren@gmail.com