Hall of Fame Prediction

Ryan Lewis & Chaz Frazer

TABLE OF CONTENTS

01 Business Question

Business Question:

Can you predict if NBA players will be elected into the Hall of Fame based on career statistics and accolades?

02

Overview

What Does It Take to Make the HOF?

Overview

Data Set:

- Kaggle Data (via basketball-reference.com) of seasonal player stats from 1950 - 2017
- aggregated data set down to career totals & averages
- added MVP awards, DPOY awards, All-star selections & Finals appearances
- filtered on HOF eligible players only (retired in 2014)

HOF FAQ:

- Out of the **4,374 NBA players** to appear in at least one game, only **177** have been elected to the Hall of Fame

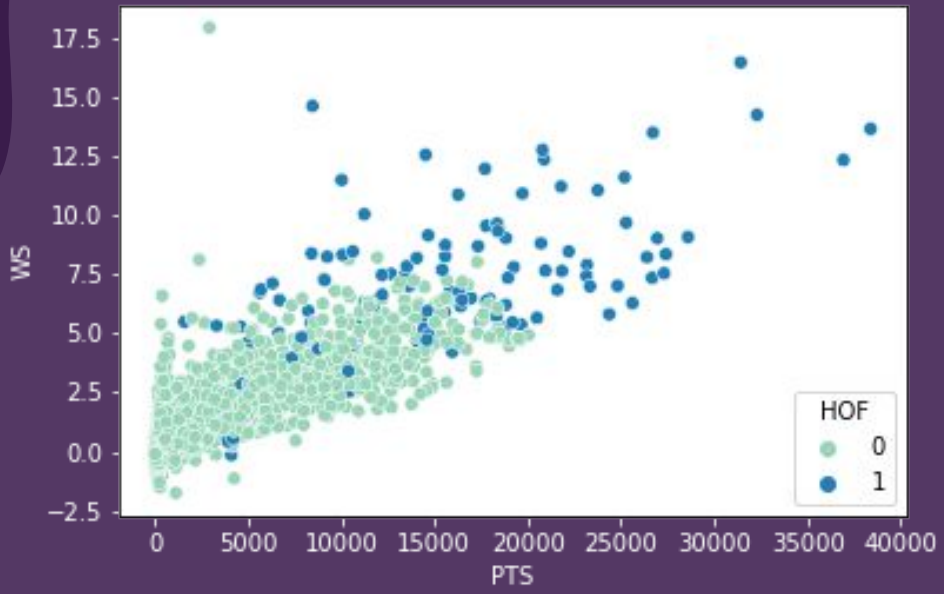
Classification Metric Focus: Precision

- Goal is to ensure have a **high positive predictive rate**, meaning when our model predicts a player is an HOFer, we want that to be as accurate as possible so we have few false positives.

03 EDA

Visualizations of NBA Statistics

Distribution of Points vs OWS



All-time leaders in Points are especially poised to be HOF's

All-time leaders in Points are especially poised to be HOF's

Extreme outliers of HOF players are almost godlike compared to the average NBA player

There is an extreme outlier in Offensive Win-Share regarding a non-HOF player named Alex Groza.

- Groza was a member of the NCAA Title winning Kentucky Wildcats and played only 2 seasons in the NBA.
- A Rookie of the Year winner, Groza was implicated in the 1949 NCAA Point Shaving Scandal, and was subsequently banned from the NBA for life.
- Groza Averaged 22.5 ppg in his rookie year.

All-Stars, Champions, and Point Totals

Clear and distinct separation of HOF players based on their career stats

Lots of players are on championship teams, but All-Star Appearances and Leaders in Points set HOF legends apart from the pack

Players with top stats but without an NBA Championship (ie: Karl Malone), make up for that metric with their career accomplishments in other areas

These statistics are only a small subset of the key indicators in what we used to predict Hall of Famers

- Feature Engineering was used to create new statistics, as well as interactions of existing and new features
- Advanced Statistics were calculated and imputed for players when merging multiple data sources into one dataframe
- NBA All-Time Leaders in Points, Rebounds, Assists, and Free Throws were manually created

04 Model & Results

Feature Selection

Original Features

57 columns worth of player data

Feature Selection

- Correlation
- Recursive Feature Elimination
- Random Forest Feature Importance

Final Features

- All-star count
- MVP Count
- Free Throws
- Win Shares
- Finals Appearances
- DPOY Count
- Finals MVP Count
- Assists
- Rebounds
- Category Leaders

Model Selection

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.939163	0.418919	0.861111	0.563636
Weighted K-NN	**0.9784854**	**0.952381**	**0.555556**	**0.701754**
Decision Tree	0.969582	0.700000	0.583333	0.636364
XGBoost	0.978454	0.827586	0.666667	0.738462
Best GS DT	0.915082	0.340206	0.916667	0.496241

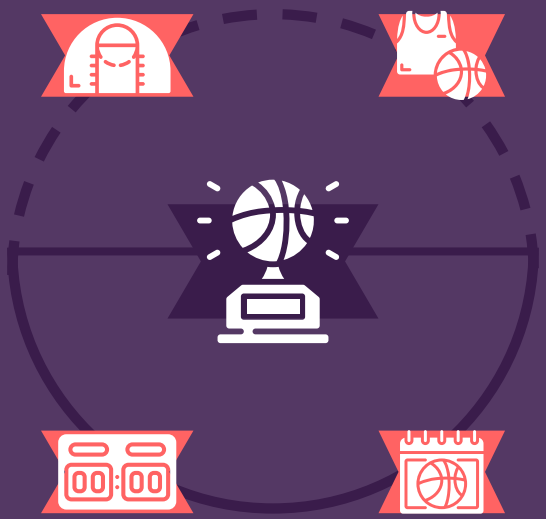
Weighted K-NN Results

Accuracy = 97.8%

Our model does a good job predicting TP & TN, but with such heavy class imbalance we can't rely solely on accuracy

Recall = 55.6%

Recall is low, as we favored a high precision score, resulting in more FN's



Precision = 95.2%

Our model has a high positive predictive rate, meaning that when it predict a player to be an HOF, it's correct 95.2% of the time

F1 Score = 70.1%

A high F1 score shows that overall our model does well predicting NBA HOFers

Final Model

A KNN Model using nearest neighbors parameter of 5 was used to get our best model, focusing on Precision metric

- In the 3D graph plot above, diamond shaped data points represent our HOF players, while color denotes MVP winners
- Precision focuses on the value of our accurate positive predictions for players inducted into the HOF
- Free Throws had high multicollinearity with Points, so we used the FT metric instead as a unique approach to these models

Next Steps

- Update data set to include data from 2018-2020 seasons
- Incorporate additional advanced analytics stats, foreign born status
- Use our final KNN model to predict on the current & not yet eligible NBA players

"Being an All-Star is EVERYTHING"

–Allen Iverson

THANK YOU

Chaz Frazer - https://github.com/Mynusjanai

Ryan Lewis - https://github.com/rylewww


CARTER
15
4,374
Players to Have Ever Played in the NBA
Vince Carter has played against 40% of them...