

NBA Hall of Fame Prediction App: Mid-Project Review

Max Holiber

5/15/2019

MS in Analytics
Northwestern University

Highlights

- Data Transformation and Filtering
 - Filtered out irrelevant records (minimum games played criteria, etc.)
 - Transformed raw dataset into data aggregated at the player level
 - Split data into historical players (used for model training) and current players (will be used in the app for predictions)
- Feature Generation:
 - Created features to be used in the modeling related to player career averages for a given statistic, e.g. points, rebounds, etc.
- Predictions: Model Selection
 - Tuned Boosted Tree and Random Forest for optimal parameters via cross validation
 - Compared Boosted Tree, Random Forest, and Logistic Regression in cross validation to obtain optimal model based on F-score: Logistic Regression
- Player Similarity List
 - Wrote code to output list of most similar players to a given player based on cosine similarity, correlation, euclidean distance --- euclidean distance outputted the most logical results

Review Progress

- Completed stories in the first sprint:
 - Data cleaning and dataset creation
 - Modeling: tuned hyperparameters via 10 fold CV and compared models for final model selection
 - Player similarity list: function to output list of most similar players based on stats to a given player by computing euclidean distance

Demo/Analysis

Logistic Regression to Predict Current or recently retired players that will be in HOF

```
## print HOF players
for pl in df_non_model['Player']:
    if log_predict(pl)[0] == 1:
        print(pl + " " + str(log_predict(pl)))
```

```
Amar'e Stoudemire (array([1]), array([[0.38186754, 0.61813246]]))
Blake Griffin (array([1]), array([[0.17835399, 0.82164601]]))
Chris Paul (array([1]), array([[0.23981588, 0.76018412]]))
Deron Williams (array([1]), array([[0.46998113, 0.53001887]]))
Derrick Rose (array([1]), array([[0.18553422, 0.81446578]]))
Dirk Nowitzki (array([1]), array([[0.24445321, 0.75554679]]))
Dwight Howard (array([1]), array([[0.30193746, 0.69806254]]))
Dwyane Wade (array([1]), array([[0.29873309, 0.70126691]]))
Elton Brand (array([1]), array([[0.47525768, 0.52474232]]))
James Harden (array([1]), array([[0.2098263, 0.7901737]]))
John Wall (array([1]), array([[0.48161149, 0.51838851]]))
Kevin Durant (array([1]), array([[0.04065112, 0.95934888]]))
Kevin Garnett (array([1]), array([[0.27870852, 0.72129148]]))
Kevin Love (array([1]), array([[0.2949604, 0.7050396]]))
Kobe Bryant (array([1]), array([[0.43863303, 0.56136697]]))
LeBron James (array([1]), array([[0.02829758, 0.97170242]]))
Pau Gasol (array([1]), array([[0.16191907, 0.83808093]]))
Russell Westbrook (array([1]), array([[0.1866955, 0.8133045]]))
Stephen Curry (array([1]), array([[0.22975427, 0.77024573]]))
Steve Nash (array([1]), array([[0.47705286, 0.52294714]]))
Tim Duncan (array([1]), array([[0.08569227, 0.91430773]]))
Tracy McGrady (array([1]), array([[0.34952062, 0.65047938]]))
```

62% HOF probability

Prediction: HOF (1)

Player similarity list (* indicates HOF)

```
euc_dist(10, 'Tim Duncan')
```

	Player	Euc
653	Moses Malone*	3.704046
357	Hakeem Olajuwon*	3.837047
792	Shaquille O'Neal*	4.757553
132	Charles Barkley*	4.805328
681	Patrick Ewing*	5.366574
256	Dolph Schayes*	5.544051
225	David Robinson*	5.953430
88	Bob McAdoo*	6.107167
86	Bob Lanier*	6.263106
503	Karl Malone*	6.339288

Lessons Learned

- Python
 - Sklearn packages have very efficient functions for Cross-Validation and hyperparameter tuning
- Predictions
 - Solely using stats as predictors of HOF probability is not a very strong predictor of HOF -- championships, awards, etc. would increase accuracy
- Development
 - Functions are essential to producing a production ready model -- need to convert code to functions

Recommendations

- Stories to be completed in the **next** sprint:
 - Modularize code into functions
 - Unit testing and logging
 - Configure code to read raw data from S3 bucket
 - Store model output results in RDS
 - Develop user input to predict for new player and output player similarity list
 - Configure web app and get it up and running