



Who's In First



A Major League Baseball Award
Prediction App



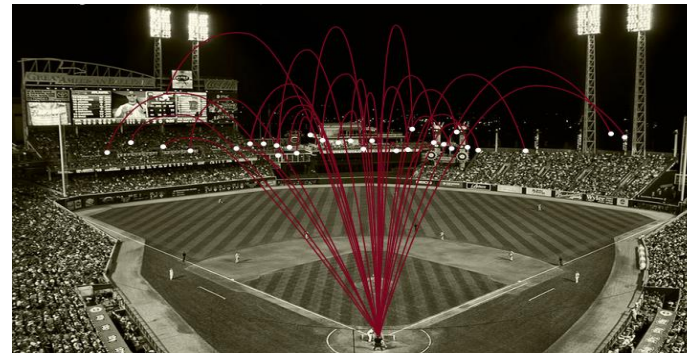
Motivation

1. Increase Major League Baseball (MLB) fan engagement by allowing fans to track their favorite players when it comes to who will win the Most Valuable Player and Cy Young Awards
2. Greater engagement means greater MLB revenues:
 - Increased MLB game attendance
 - Improved sales of MLB player and team merchandise
 - Increased TV viewership ad revenues
3. The success of this app will be judged by an increase in fan engagement leading to increased team revenues
4. To further increase fan engagement, future developments could allow users to personalize their experience by creating lists of their favorite players and teams



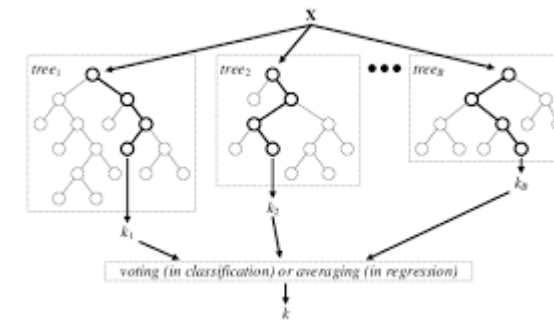
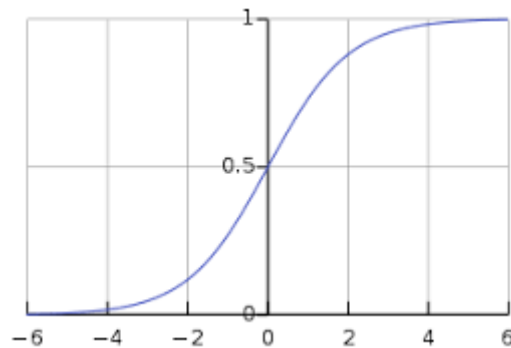
Data Used

- The data for the app was obtained by calling the API at <https://appac.github.io/mlb-data-api-docs/>
- The data from this API is official Major League Baseball statistics for all players, teams, and seasons
- To train the models, pitching and batting statistics dating back to 1960 were used
- PECOTA projections were used to predict the order of the current season's players
 - PECOTA projections are projections for end-of-season statistics for each player originally developed by Nate Silver
- Demographic data displayed in the app for teams and players also comes from this API



Models Used/Success Criteria

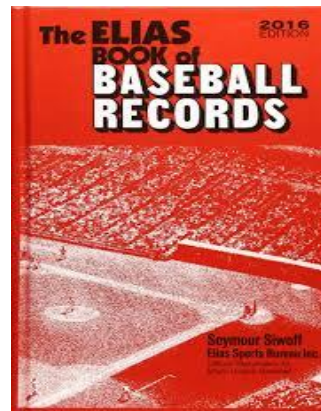
- Both a Logistic Regression model and a Random Forest model were used to order players based on their likelihood of winning an award
- The results of each model were averaged to determine the order of the players
- Because so few players actually win an award (less than .1%), I was trying find a model that maximizes F1-Score
- The Logistic Regression model gave me a ranking, but the F1-Score was very low (14%)
- The Random Forest model greatly improved F1-score (31%), but the binary results did not allow me to order players
- Combining the results from both models, by averaging, gave me a model with an F1-Score of 24%, while still allowing me to rank players



Interesting Insight

Some Major League Baseball records look as though they may never be broken

- Barry Bonds hit 73 home runs in the 2001 season
 - This season no player is projected to hit more than 43 home runs
- It was common for pitchers to get 30 or more wins in a season in the early 1900's, with the record being 46 wins in one season
 - This season no pitcher is projected to get more than 18 wins
- Nolan Ryan struck out 383 batters in 1972
 - This season no pitcher is projected to get more than 292 strikeouts



Thank You!



Contact Information:

Alex Burzinski
aburzinski2019@u.northwestern.edu