



SENG 474

Predicting NBA Hall of Fame Inductees

By Mohammad Mohammad & Peter Wilson

Introduction

- Objective: Predicting Hall of Fame inductees in the NBA based on player statistics.
- Importance: Identifying future Hall of Fame players allows for recognition of exceptional talent and provides insights into the impact of player performance on team success.
- Key Questions:
 - Can we accurately predict Hall of Fame inductees based on player statistics, or do we consider mvp and all-star selection?
 - Which machine learning algorithms perform best in predicting Hall of Fame candidates?
 - What are the most influential statistics in determining Hall of Fame eligibility?



Reference: <https://www.freeiconspng.com/downloading/26237>



Related Work

- There are numerous professional analyzers that predict possible Hall of Fame Inductees through various methods and data considerations
- Top 10 active players that [Basketball-reference.com](https://www.basketball-reference.com) predicts will be inducted into the Hall of Fame:
 - LeBron James
 - Kevin Durant
 - Chris Paul
 - Stephen Curry
 - James Harden
 - Russel Westbrook
 - Dwight Howard
 - Anthony Davis
 - Carmela Anthony
 - Giannis Antetokounmpo
- We will discuss our results at the end and determine if our model gave the same predictions



Data Preparation/Preprocessing

- Aggregation of Player Stats: Consolidating player career statistics into a single row for comprehensive analysis.
- Calculation of Ratios: Deriving important ratios, such as field goal percentage, by dividing relevant player stats.
- Integration of Multiple CSV Files: Combining diverse data sources, including player stats, Hall of Fame records, MVP/Finals MVP data, and All-Star selections.
- Parsing Natural Text into CSV: Converting All-Star selections data from natural text format to structured CSV for seamless integration.

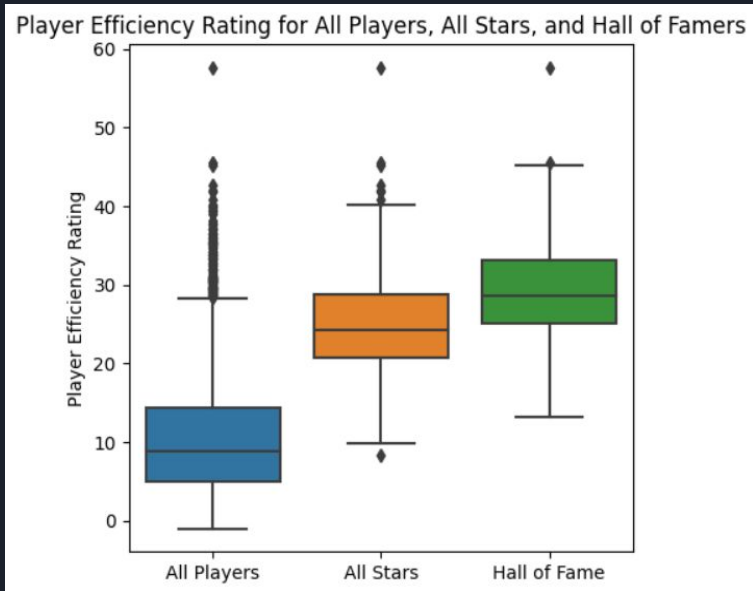
Data Mining Process

- Data Transformation: Converting aggregated player stats into a structured Data Frame with calculated ratios and per-game statistics.
- Data Enrichment: Retrieving additional player information such as seasons played, first and last season, and retired status to enhance the dataset.
- Further Data Preprocessing: Handling missing values by filling them with 0 to ensure complete data for analysis.
- Model Building and Evaluation: Creating a pipeline that combines data preprocessing and various classification models for predicting Hall of Fame inductees. Evaluating model performance through accuracy scores averaged over multiple iterations.
- Results Presentation: Storing average accuracy scores of different models in a DataFrame for further analysis and model comparison.



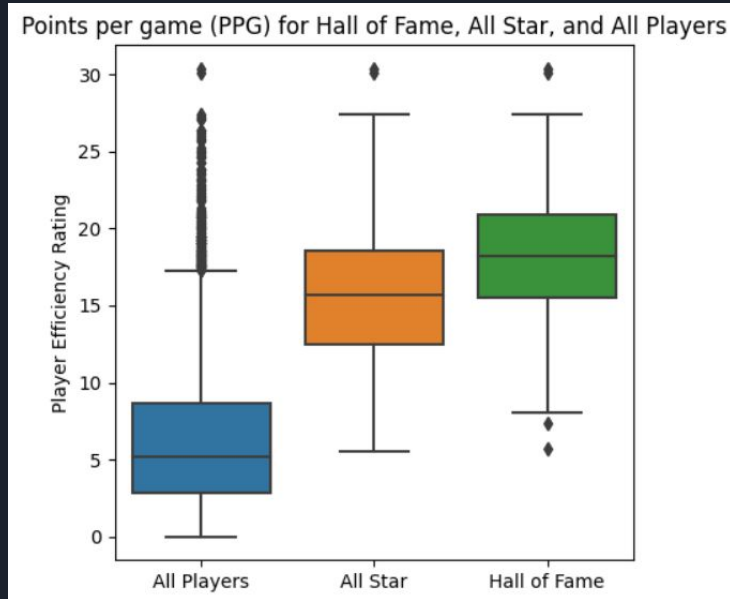
Reference: <https://justtotaltech.com/data-mining-techniques/>

Visual Representations



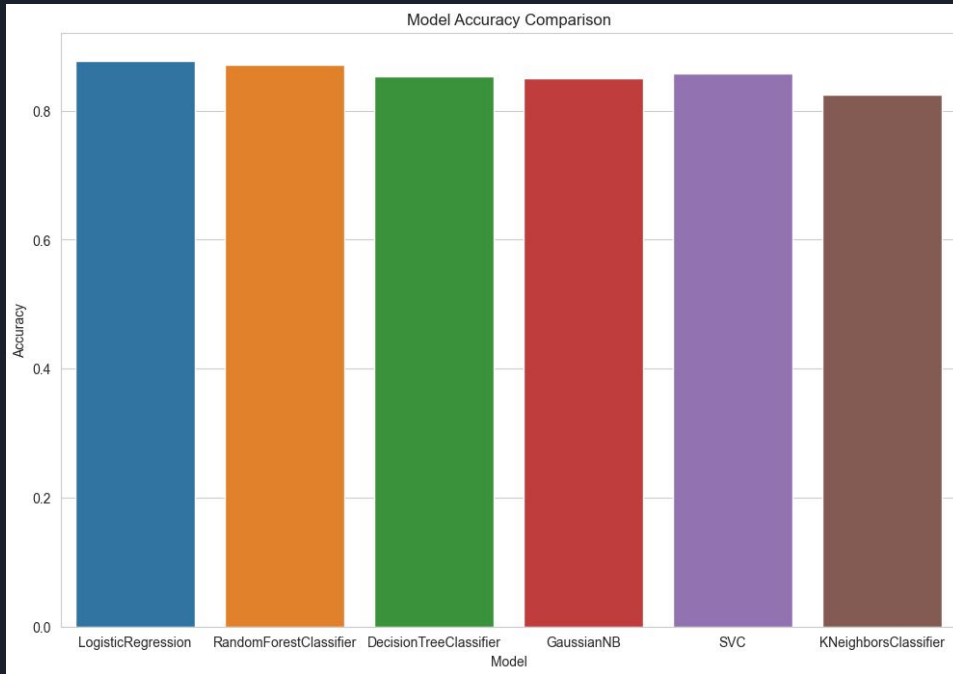
- The graph on the left depicts the Player Efficiency Rating (PER) between All Players, All-Star select players, and Hall of Famers.
- We can observe a clear correlation between All-Star selected players and Hall of Famers, as the average PER of both is drastically greater than that of All Players

Visual Representations - 2



- The graph on the left depicts the Points per Game (PPG) for All Players, All-star selected players, and Hall of Famers
- In this graph, we observe a significant discrepancy between All Players when compared to All-Stars and Hall of Famers. This is expected as the number of All-Stars and Hall of Famers is greatly less, and these categories have much higher scorers than the average NBA player

Model Selection



- In the figure on the left we portray which models we use and the respective accuracy of each one.
 - All models performed well, averaging at 85.57%
- Our best algorithm was Logistic Regression, obtaining 87.72%, closely followed by Random Forest Classifier with 87.11%
- We used these models because they are commonly used for binary classification



Conclusion

	PLAYER_NAME	prediction
2909	Carmelo Anthony	1
2947	LeBron James	1
3003	Dwight Howard	1
3100	Chris Paul	1
3224	Kevin Durant	1
3329	Russell Westbrook	1
3347	Stephen Curry	1
3361	James Harden	1
3546	Anthony Davis	1
3610	Giannis Antetokounmpo	1

- Overall, our model was successful and allowed for ~87% accuracy in predicting whether a player is inducted into the Hall of Fame
- On the left we have our result on the Top 10 active players with high probability to be in the Hall of Fame after retirement, showing that our model predicted these players successfully!