**PREDICTING THE NEXT HALL OF FAME PLAYER USING LINEAR REGRESSION AND WEB SCRAPING**

**Oluwabori Ige**

**Abstract**

I made the decision to concentrate my linear regression assignment on the NBA because I love basketball so much. The performance rating of a player is mostly represented by their statistics, which invariably influence their selection for the Naismith Award at the end of the season. This project's objective is to use linear regression to forecast NBA players' odds of being inducted into the Hall of Fame.

**Design**

The links between different box score statistics should be investigated in order to forecast NBA players' prospects. It is important to take into account the multicollinearity between the stats of the inductee and the stats of the Naismith Hall of Famer. Additionally, feature engineering will be used to take into consideration unavailable dummy indicators and/or interaction factors.

I split the data into training and test sets after data cleaning and manipulation, and I tested three distinct models (OLS, Ridge Regression, and Lasso Regression). I created models using the training data and evaluated the models' performance using the test data. Along with r squared, mean absolute error, and intuitive fit, I also evaluated the models.

**Data**

I'll be collecting web content from the following websites:

1. Based on player statistics from Basketball-Reference for the 2022 season.
2. Naismith Hall of Famer stats inductees.

**Algorithm**

*Data cleaning*

Players whose rows are empty are eliminated. Usually, players that are technically on a roster but don't get much playing time are responsible for these.

*Data Manipulation*

The following variables were collected and considered: player, position, age, games/played, minutes played per game, made and attempted field goals per game, shooting and advances

**PER-GAMES**

G -games

GS – Games started

MP – Minutes played per Game

FG – Field goals per game

FGA – Field goals attempted per game

FG% - Field goal percentage

3P – 3-points field goals per game

3PA – 3-Point Field Goal Attempts

3P% - 3-Point Field Goal percentage

2P – 2-Point field goals per game

2PA- 2-point field goal attempts per game

2P% - 2-point field goal percentage

eFG% - Effective Field Goal Percentage

FT – free Throws per Game

FTA – Free Throw Attempts Per Game

FT% Free Throw Percentage

ORB – Offensive Rebounds per game

DRB – Defensive Rebounds Per Game

TRB – Total Rebound Per Game

AST – Assists Per Game

STL – Steals Per Game

BLK – Blocks per Game

TOV – Turnovers Per Game

PF – Personal Fouls Per Game

PTS – points per Game

**ADVANCED**

MP- Minutes played

PER – Player efficiency rating

TS% - True shooting percentage

3Par – 3-Point Attempt Rate

Ftr – Free Throw Attempt Rate

ORB% - Offensive rebound Percentage

DRB% - Defensive Rebound Percentage

TRB% - Total Rebound Percentage

AST% - Assist Percentage

STL% - Steal Percentage

BLK% - Block Percentage

TOV% - Turnover Percentage

USG% - Usage Percentage

OWS – Offensive Win Share

DWS – Defensive Win Share

WS – Win Share

WS/48 – Win Shares Per 48 minutes

OBPM – Offensive Box Plus/Minus

DBPM – Defensive Box Plus/Minus

BPM – Box Plus/Minus

VORP – Value over Replacement Player

**SHOOTING**

FG% - Field Goal Percentage

Dist. – Average distance(ft.) of FGA

% of FGA by Distance (2P, 0-3, 3-10, 10-16, 16-3, 3P)

FG% by Distance – (2P, 0-3, 3-10, 10-16, 16-3P, 3P)

% of FG Ast’d – (2P, 3P)

Dunks – (%FGA, #)

Corner 3s – (%3PA, 3P%)

Heaves – (Att, #)

2P -

**JAW4J9T8CXGGNK4R**