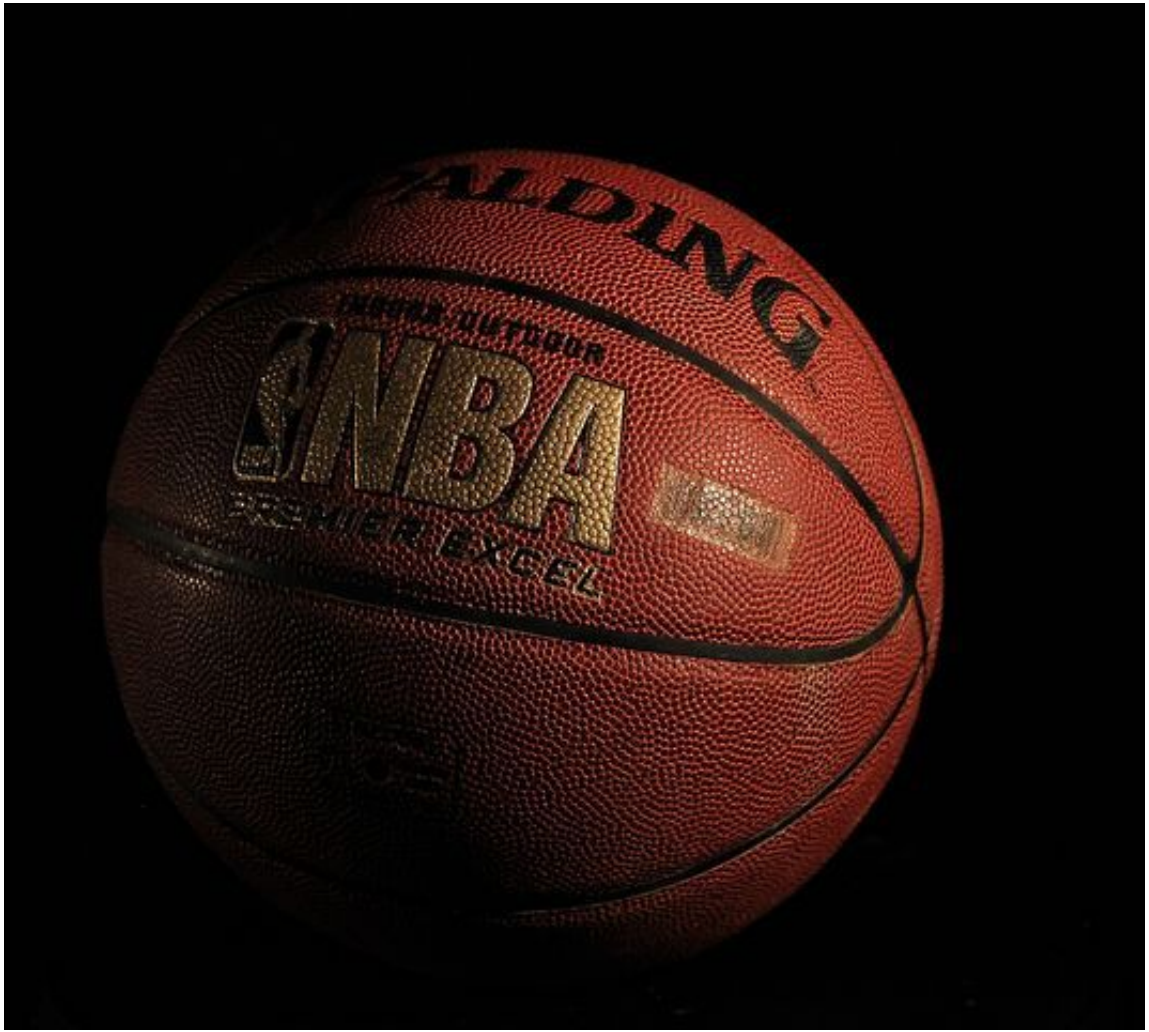
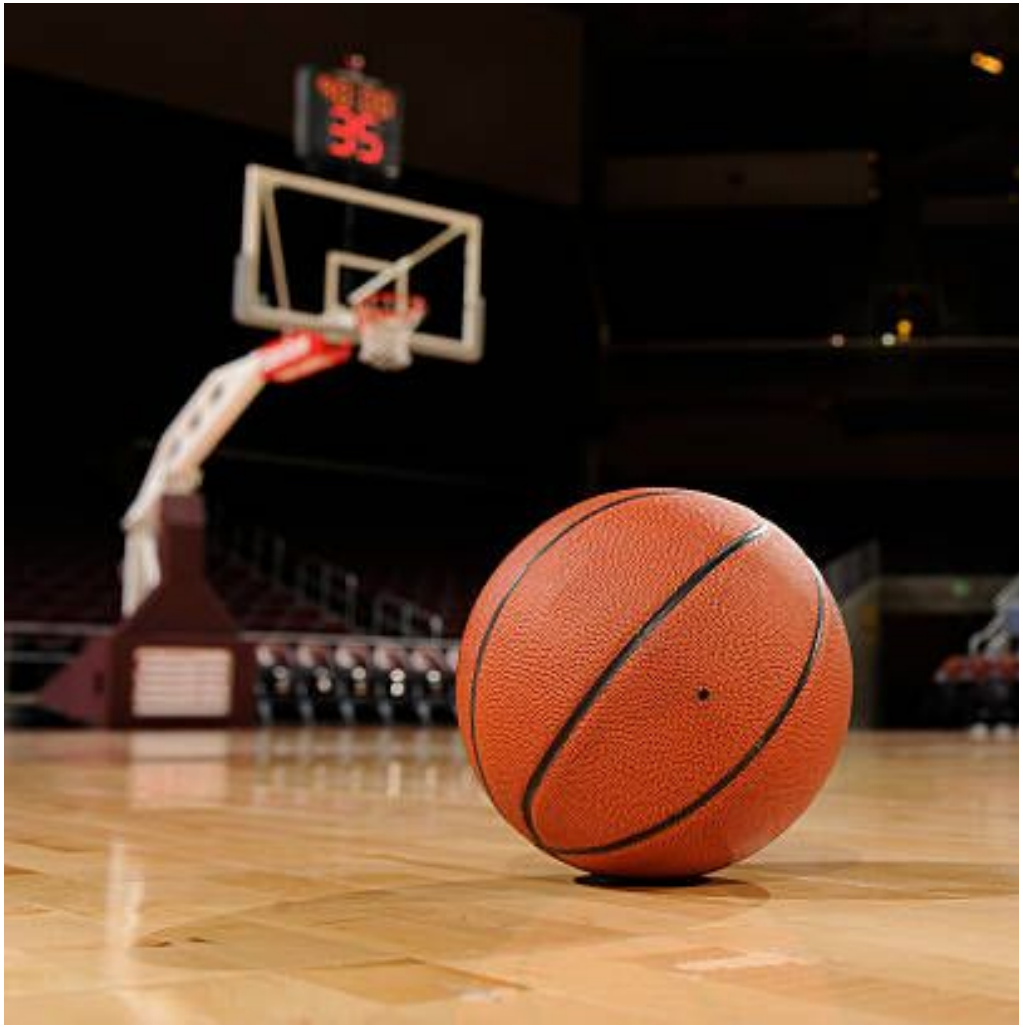


How do NBA All-Star Players Shine?

Predicting All-Star Players

Med C. Billian



Goals of this Project:

- Which data features are the best to classify NBA All-Star players?
- Model - predict whether an NBA player should be part of the All-Star roster using data statistics.

Impacts:

- Help players understand how to be in the All-Star roster.
- Help determine if a player is an All-Star “mistake” or an All-Star “snub”?

Methodology:

- BeautifulSoup to scrape data from:
  - https://www.basketball-reference.com
- Join (Google Sheets) and Clean Data
  - Player Statistics
  - Team Statistics
- EDA Using Tableau - Remove Outliers
- Features Selection using Logit and VIF

Methodology (Data Set):

I combined the Players Statistics data, Team Ranking data, and All-Star Games Data. I matched the player, the team ranking and the All-Star data within the same season.

1. Training Data
    - Data from 2000- 2019
    - 9,283 rows
2. Validation Data
    - Data from 2020
    - 528 row
3. Test Data
    - Data from 2021 (early this year)
    - 539 rows

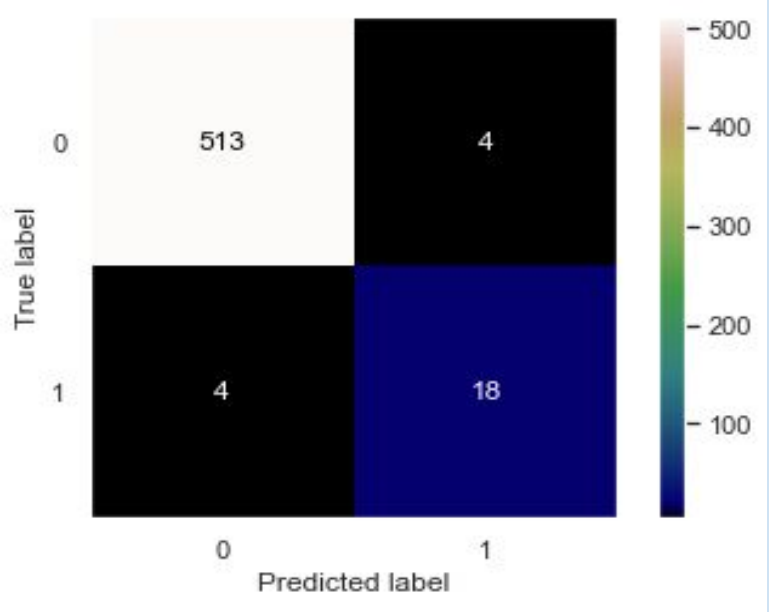
Methodology:

- 4 models using Training Data:
  - Logistic Regression with Scaled Data (Baseline)
  - Random Forest
  - GBM
  - XGB
- Evaluate 4 models using Validation Data.
  - Random Forest Final Model
- Fine Tune Final Model.
- Test final model.

Model Evaluation

Confusion Matrix and F1 Score:

- Random Forest model displayed the highest F1 using the validation data. **F1=0.698**
- Random Forest Model fitted with Training and Validation data scored better using Test Data **F1=0.818**

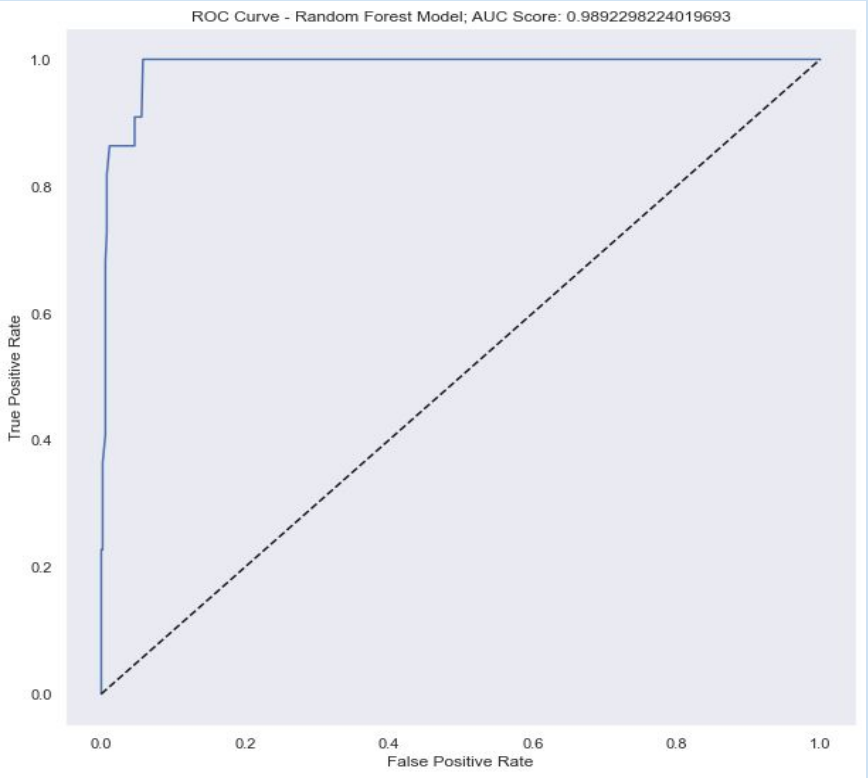


Model Evaluation

ROC Curve

- Random Forest Model using Test Data

  AUC Score=0.989



Model Interpretation

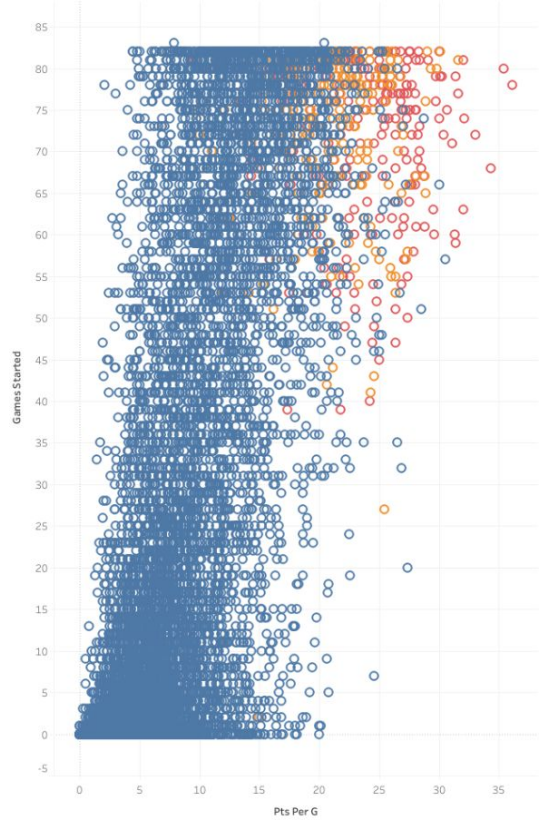
- Using features_importances property - features with best predictive power:
  - Total Points per Game
  - Games Started
  - Assists per Game
  - Team Standing

Variable	Importance
pts_per_g	0.421812
gs	0.147859
ast_per_g	0.141913
tm_srs	0.132245
blk_per_g	0.093305
age	0.062866

PPGvsTmSrs by Med Billian

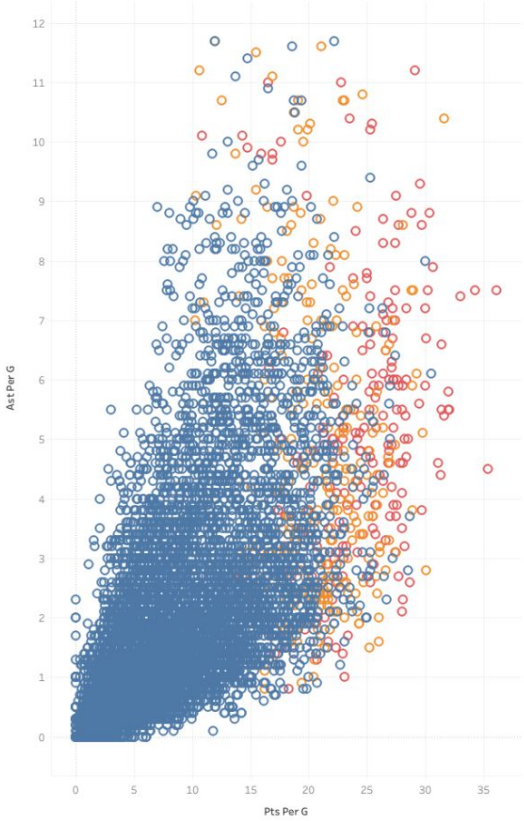
Points Per Game VS Games Started

NBA All-Stars Classification



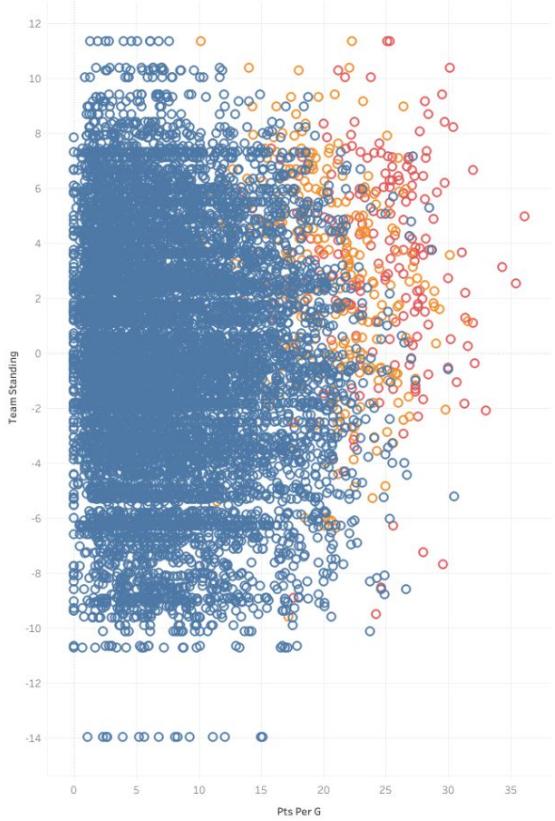
Points Per Game VS Assists Per Game

NBA All-Stars Classification



Points Per Game VS Team Standing

NBA All-Stars Classification



All-Star Position
none
Reserves
Starters

Sample Use of Classification Model

- In 2007, Shaquille O'Neal was the starting center of the Eastern All-Star team
- Based on Shaquille O'Neal's 2007 statistics, he shouldn't be part of the All-Star team
- It was controversial that Shaquille O'Neal was voted over young star Dwight Howard.

MORE QUESTIONS

- How can I measure player's popularity?
- Should I use social media data? How?

Data Source:
https://www.basketball-reference.com

Image Source:
https://pixabay.com