NBA Season Wins

Capstone: Jose Harper March 10th 2022

Agenda

1) Business Problem and Stakeholder
2) Data Understanding
3) Model evaluation
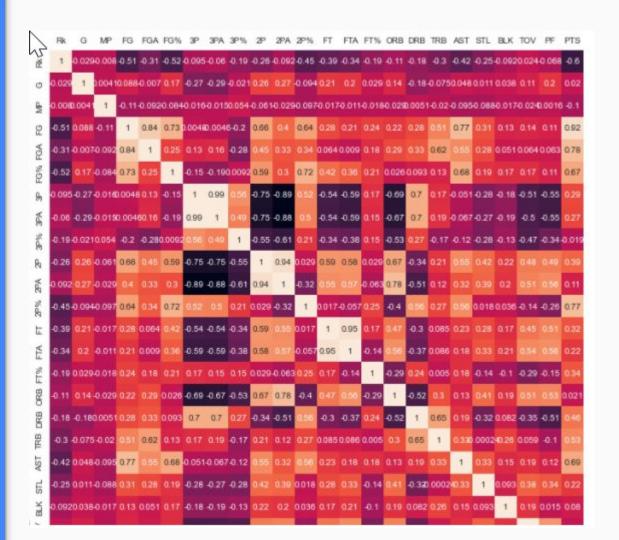4) Conclusion and Strategy
5) Future Analysis

Business Problem

1) Determine best metrics for winning the most games per NBA season
2) Current Research focuses on winning nightly games
3) Use mainly classification algorithms for game by game wins, loss or draw.

Stakeholder

- NBA basketball franchise Houston Rockets
- Hired to explore ways of predicting upcoming season wins (rank)
- Using the data from the 35 prior seasons

Data Understanding

1) High Multicolinearity for many features
2) 1057 rows of 35 NBA seasons
3) Data omitted : (a) Playoffs made (b) Second half of 2022 season metrics (c) Draft Picks (d) Teams average height and weight (e) Players salary

Modeling Process

1) Data Preperation
2) Train and Tune models
3) Retrain and Validate
4) Final Model evaluations

Model Process continued:

Baseline Regression Model

Linear Regression Model 4

R-squared: 38%

R-squared: 63%

RMSE: 46.8

RSME: 18

Final Results and Key features

1) Best Model: 4th LR regression

2) Highly correlated features for winning

3) 5 features to focus on: (1) 3P%, (2) FG%, (3) STL, (4) ORB, (5) DRB

1. 3 Point percentage

2. Field Goal Percentage

3. Steals

4. Offensive Rebounds

5. Defensive Rebounds

Conclusion/Recomendations

	G	MP	FG	FGA	FG%	3P	3PA	3P%	2P	2PA	2P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
Team	63	15170	2441	5427	.450	844	2454	.344	1597	2973	.537	1103	1555	.709	612	2045	2657	1456	461	284	1051	1284	6829

1) Areas that need improvement at 3P%, ORB, DFR, FT%, AST, STL, BLK, FG%
2) Using the best models feature importances: Look for athletic big men that have high 3P%, and/or Athletic guards with long wingspan

Future Analysis

1) Use website api
2) Use only 10 most recent NBA seasons for data
3) Add features such as : Age, weight, height, wingspan, experience, teams salary and minutes per game played

Thanks!

Contact:

Jose Harper
Harper.Jose@gmail.com