Predicting NBA Game Outcomes

Analysis for prospective/new NBA Analysts

Noah-John Hizon
5/10/22

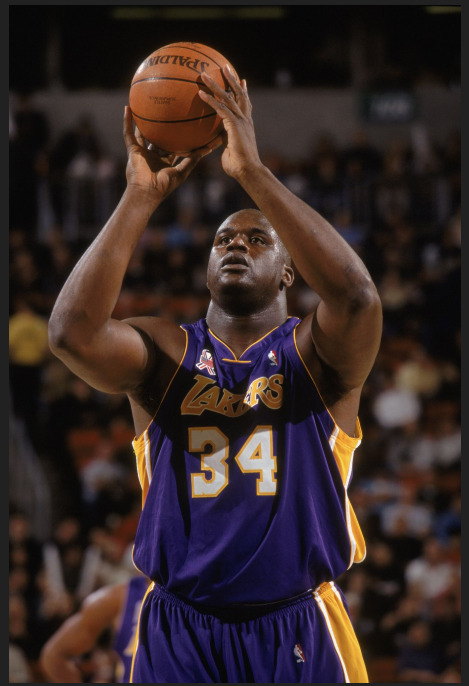
Outline

- Business Problem
- Data Understanding
- Classifier Model
- Classifier Results
- Recommendations
- Next Steps

Introduction

What factors affect the outcome of an NBA game?

- Game Statistics
  - Shooting percentages
  - Rebound rates
  - Assist rates
- External factors
  - Rivalries
  - Media scrutiny



Business Problem

In an ever-fluctuating game state, analysts may be flustered when events occur that discount their game predictions.

- Mid-game injuries, coaching changes, and mid-season trades
- Misinterpretation of game stats
- Discounted credibility

Current prominent NBA analysts: Shaquille O'neal, Ernie Johnson, Kenny Smith, Charles Barkley, Stephen A. Smith, and Max Kellerman

Data Understanding

SOURCE: NBA Game Data 2019-2022

- Each row in the table represents a unique game, with performance summarized for both Home/Away teams
- Variables selected based on relevance to game result, without direct involvement of points scored

- **Columns of interest: Top 50 Contract Players, Double-doubles, FT% 3PT% 2PT%, Assists, Rebounds, Blocks, Turnovers, Steals**
- **Excluded columns: AVG Plus/Minus, Offensive Rating, Defensive Rating, Point Difference**

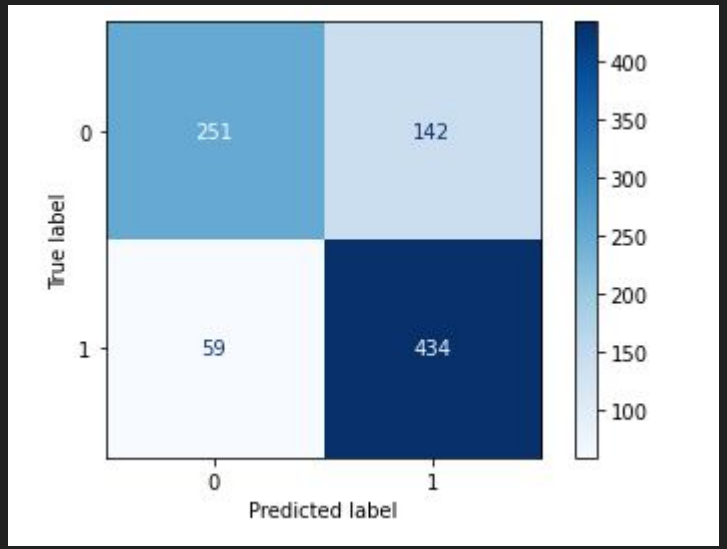
Model Record Count: **3,544.** No data was excluded, as all NBA regular season games should be considered as valid data with a distinct win/loss result.

Regression Modeling

An iterative process of classifier model comparison was used to collectively assess the effect of relevant game metrics. The classification model will also predict game results through a binary outcome variable, **GAME RESULT,** represented by 0 or 1.

- **Why ML?**
- **Final Model: Random Forest Classifier**

Confusion matrix depicting final model performance on test data



LEGEND:

0 in True label, 0 in Predicted label - True Negative

0 in True label, 1 in Predicted label - False Positive

1 in True label, 1 in Predicted label - True Positive

1 in True label, 0 in Predicted label - False Negative

Regression Results

Final classifier shows that the variables most impactful towards NBA Home team game wins are (in order): 3PT%, 2PT%, Rebounds, and Assists

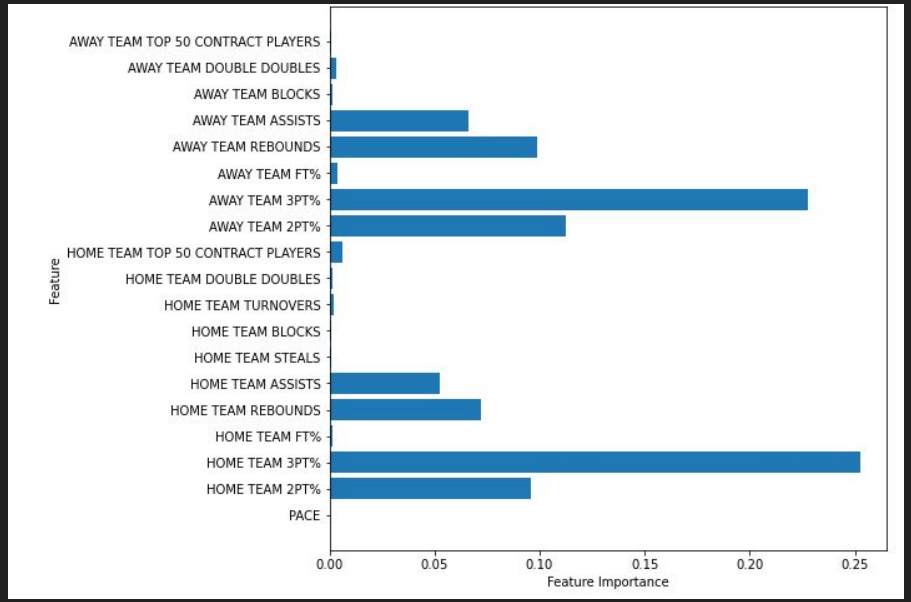
This makes sense, given the nature of the metrics

Created model was able to predict game outcomes with an accuracy of 77%

Model is also relatively accurate in the face of new data

Random Forest Classifier Results

- As seen, 3PT% is seen to have the largest scale of importance.
- Rebounds and 2PT% are similar in scope
- Top 50 Contract Players and Double-doubles have slight importance.



Random Forest Classifier Results

- Model has similar performance on the data used to train it, and in the face of new data
- Accuracy of 77% on test data (new), and 80% on training data
- Recall is relatively low on predicted losses

Test Data Classification Report

	precision	recall	f1-score	support
0	0.81	0.64	0.71	393
1	0.75	0.88	0.81	493
accuracy			0.77	886
macro avg	0.78	0.76	0.76	886
weighted avg	0.78	0.77	0.77	886

Training Data Classification Report

	precision	recall	f1-score	support
0	0.82	0.72	0.77	1217
1	0.79	0.87	0.83	1441
accuracy			0.80	2658
macro avg	0.81	0.80	0.80	2658
weighted avg	0.80	0.80	0.80	2658

Recommendations

When predicting nightly game outcomes, analysts should consider (in order of priority):

- Team 3PT%
- Team 2PT%
- Average Team Rebounds
- Average Team Assists

In addition, young NBA players that were exceptional in one/many of these metrics prior to being drafted should be scrutinized for potential future star talent

Next Steps

- Analyzing EFG%
- Offensive Rebounds
- Quantify intangible factors that involve NBA players
  - Player/Team Rivalries
  - Media Scrutiny

Thank You!

Email: noahiz917@gmail.com

Github: noahiz917

LinkedIn: linkedin.com/in/noah-john-hizon/