NBA HOME TEAM GAME OUTCOME CLASSIFICATION

Jay Kwon

Metis Data Science and Engineering (flex program)

Module 4 – Classification

March 23, 2022

Goal:

- **Use classification models to predict whether the home team wins upcoming NBA games.** **home team wins = positive outcome = 1**

  1. Assess the models and optimize best model for precision

  2. Interpret model – most important features; can this model be used to turn a profit by wagering on the games?

Background and Motivation:

- **Sports Betting:** can potentially turn a profit by wagering on the outcomes of NBA games
- must consistently beat the odds that the oddsmakers lay
- insights gained would be beneficial to anyone interested in NBA, sports betting, making money

EVALUATION METRICS

1st -- Precision:

- only concerned when we actually place wagers
- False Positives = lose money
- True Positives = make money
- maximize precision = minimize FP and maximize TP

2nd -- Recall:

- must make enough wagers in a reasonable timeframe

  - weather variance of binomial distribution to reduce risk-of-ruin (losing everything)

  - bet sizes must be small enough relative to capital to make enough wagers

Soft Predictions:

- probabilities may be used as proxy for expected value (EV) calculations
- given the probability and payouts is the EV positive?
- if positive bet on home team winning; else do not bet on the game

DATASET

- Raw data with historical game outcomes and statistics: https://www.kaggle.com/nathanlauga/nba-games?select=games.csv
- rolling previous 10 game averages were obtained for each game
- 24526 rows and 26 features (all quantitative)
- time period: Oct, 2003 – Nov, 2021 (18.5 NBA seasons)

- **Home team advantage**

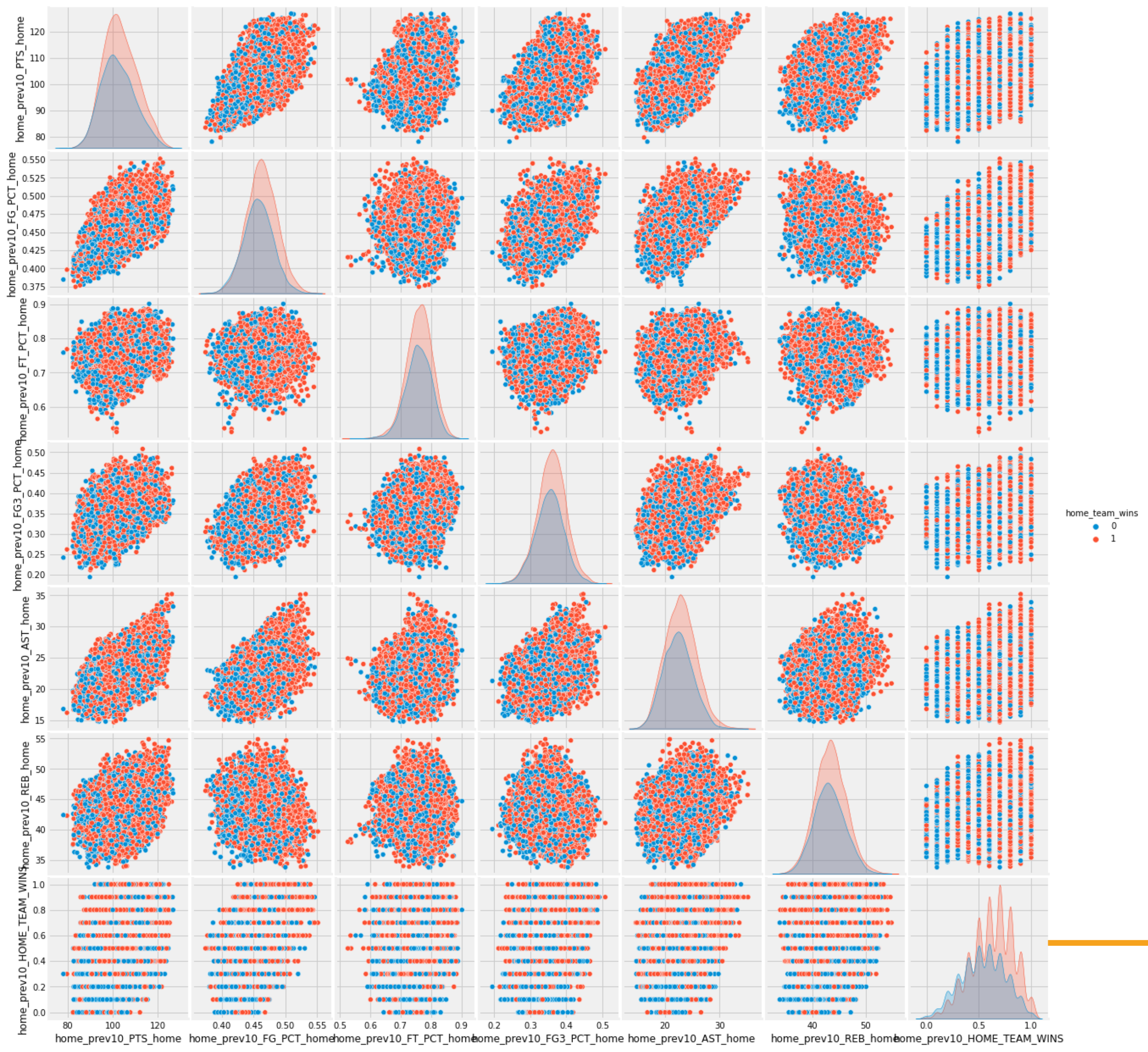
  - empirical probability of home team winning 59% of the time

  - not a big class imbalance

Tools used:

- Sklearn, Pandas, Matplotlib, Seaborn

PAIRPLOT OF SELECT FEATURES



- Home team’s **previous 10 home game average statistics** not too promising as classification features

- previous 10 game win-rate looks promising

separation of 0 vs 1 distributions

Classification Model Evaluation

Test Accuracies:

- Logistic Regression: **0.65**
- Random Forest: **0.63**
- Grad Boosted Trees: **0.64**

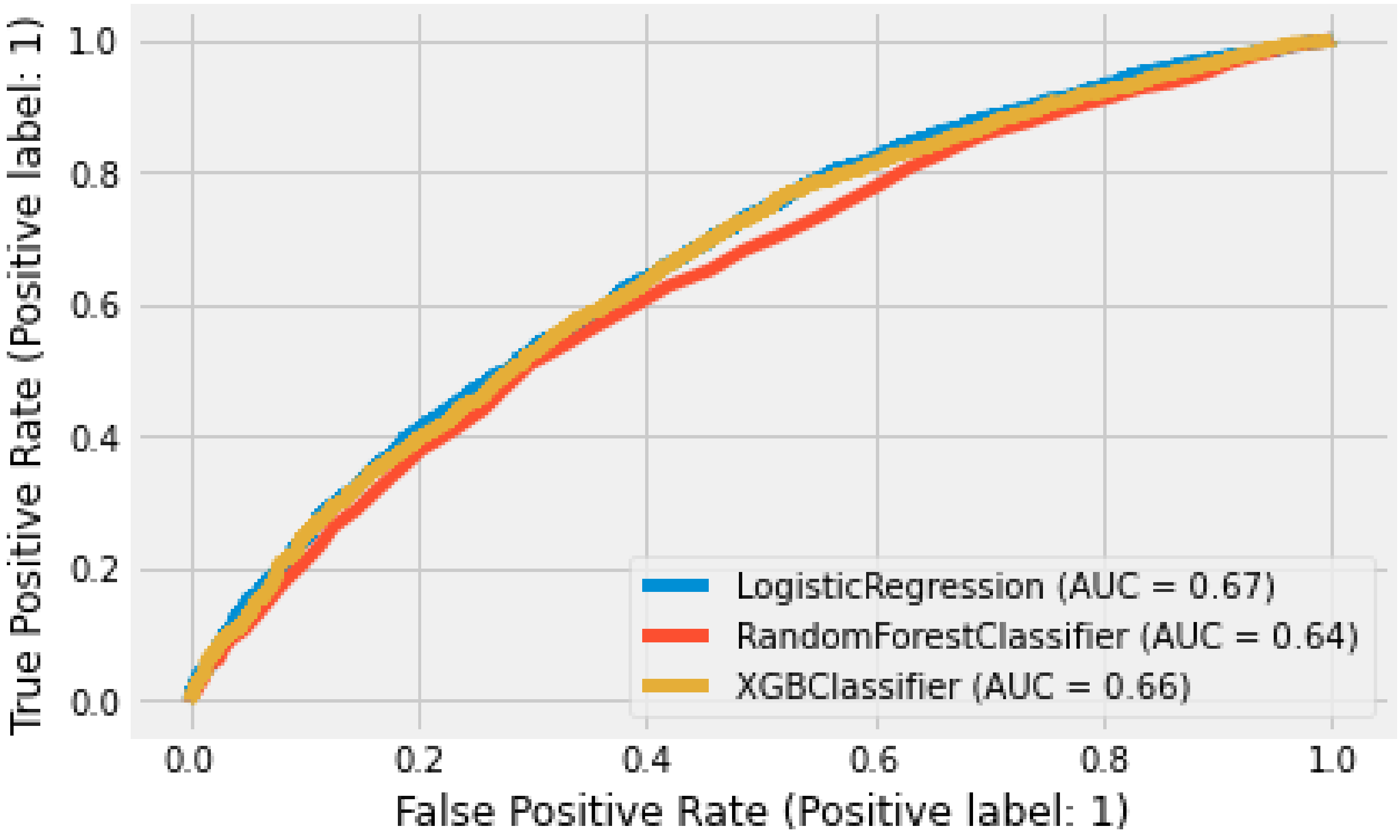
* empirical probability: 0.59

Precision:

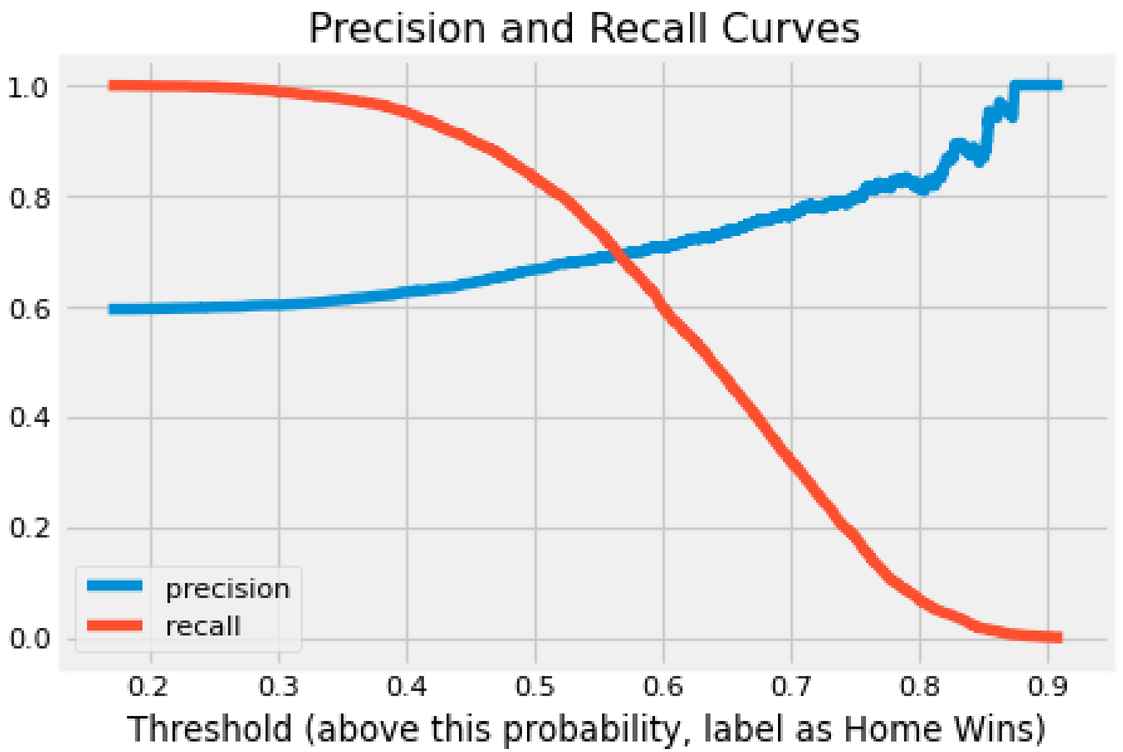
- Logistic Regression: **0.67**
- Random Forest: **0.66**
- Grad Boosted Trees: **0.66**

* 0.5 threshold

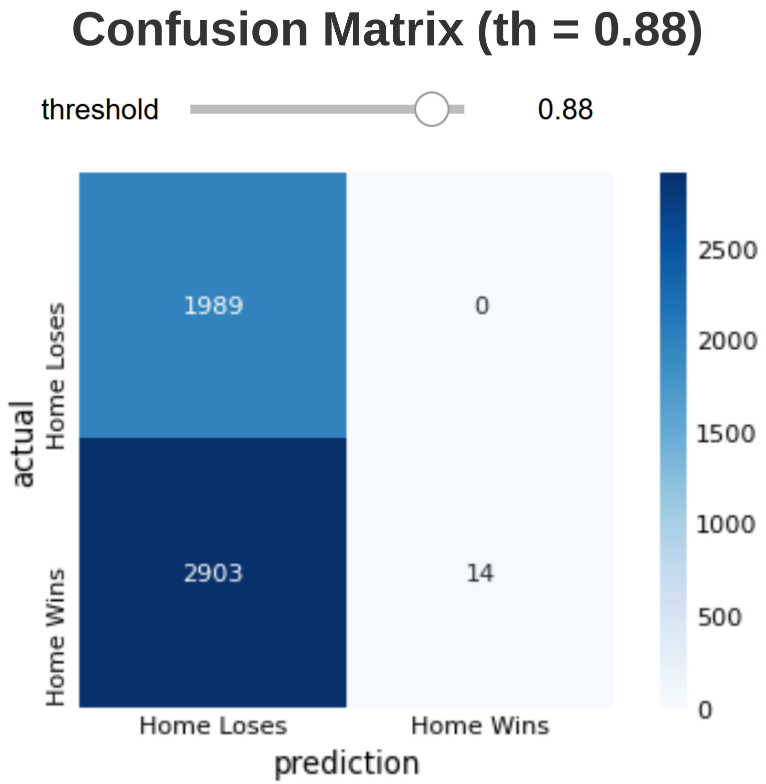
ROC AUC curves



MAXIMIZING PRECISION FOR LOG-REG



- higher threshold:
  higher precision but lower recall



- 100% precision when th = 0.88
- only 14 games out of 24,000+
- need higher recall

Practical Application: wagering on NBA games

Breakeven Odds at 88% precision:

Expected Value = (win prob x (payout – wager)) – (lose prob x wager)

- let's say we **wager $100**
- use soft probability threshold as proxy for real game outcome probabilities
- when EV = 0, we breakeven
- EV = 0 = 0.88 x (payout – wager) – 0.12 x $100
- (payout – wager) = **$13.63** ← breakeven point: must be laid **0.136 to 1 ; -733.68** in sports betting terms
- **For every $100 wagered, we must be offered a profit > $13.63 for our model to be profitable (positive EV, "beat the odds")**

Practical Problem of Time – Low Recall:

- over 18.5 years, we found 14 games that had 88%+ chance of the home team winning
- to decrease the probability of going broke (tail risk) we must make enough wagers to endure variance
- in order to make many smaller wagers, we need more games to bet on (increase recall)

Conclusions and Future Projects

Conclusions:

- profitable scenarios exist for deploying our model successfully
- trying different ML techniques and tinkering with the hyperparameters led to similar performances
- need more features with better predictive power to increase F1 score
- previous 10-game win-rate is a good predictive feature

Future Projects:

- find more features: injury data, offensive and defensive rankings, player specific data
- obtain historic data for sports wagering odds
- calculate EV for each prediction and adjust threshold and bet sizes to find profitable strategies