

TO BET OR NOT TO BET:

ANALYZING THE CREDIBILITY OF FIXED-ODDS BETTING ON MATCH OUTCOMES

By: Maria Tsakalacos, Tseegi
Nyamdorj, and Fungai Jani



INTRODUCTION

- Soccer is the most watched sport in the world
 - 3.5 billion fans
- Sports betting is profitable
 - 70% of sports bets is from soccer
- What is the most popular league in the world?
 - Premier League
 - ▶ \$6.2 billion in revenue in 2021-2022 season
 - ▶ €68 billion of wagers world wide

PURPOSE

- The goal
 - To build a comprehensive predictive framework for forecasting match results in the Premier League
- How?
 - Premier League 2018-2019 season
 - Fixed betting odds from the league/season
- Added metrics
 - Player evaluation metrics
 - Streak of the team from last five matches

THE DATA

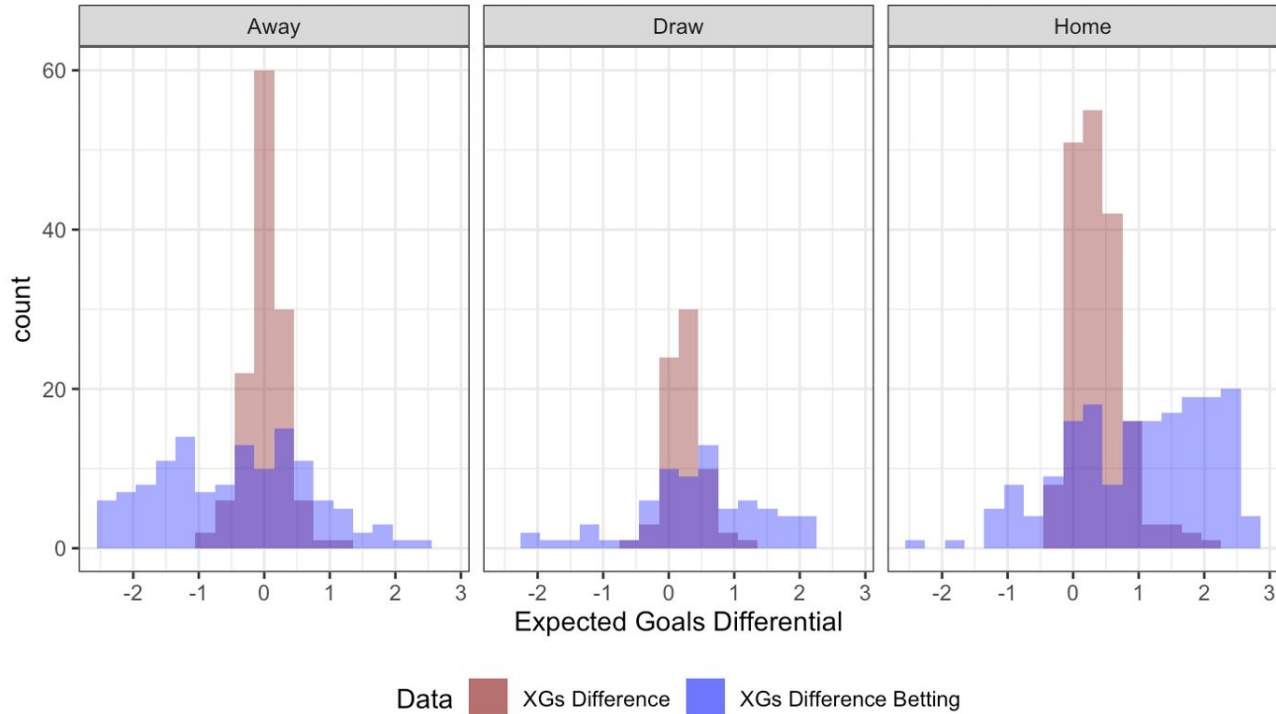
- “Basic” Premier League data
 - Game level data
 - Full time result, expected goals, home/away team, etc.
- Betting data
 - Game level data
 - Fixed betting odds for home win, away win, draw
 - ▶ Bet 365
- Merged two data sets together
 - Dropped unnecessary columns
- FIFA player ranking data

DATA ENGINEERING

- Merged data set:
 - *Expect_win*: based on betting data odds
 - *Expect_point_h*: home points based on betting data
 - *Full_time_result_point*: Creating dummy variables
 - *Expect_point_diff_bet*: Home - Away expected points
 - *Streak*: Streak of last five games per team
 - ▶ Weighted by historical data
- FIFA Ranking
 - Overall team ranking based on position

EXPLORATORY DATA ANALYSIS

Comparing Expected Goals Differential Between Basic and Betting Data



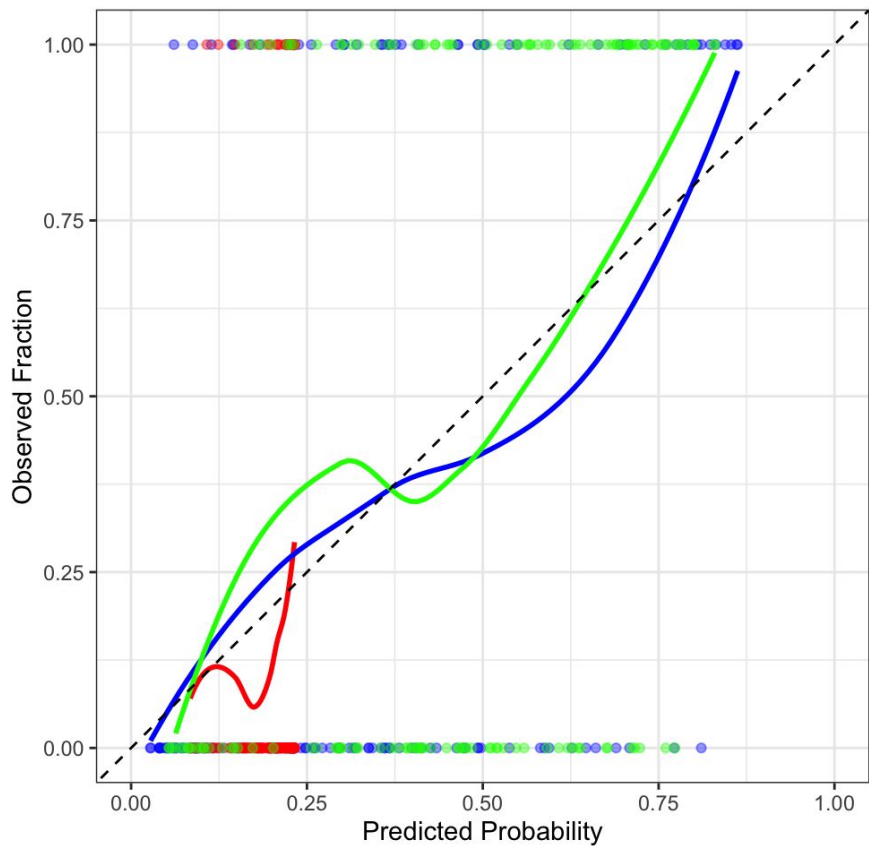
METHODS

- Dependent variable is categorical
 - Match Outcome
- Explanatory Variables
 - Expected goals
 - ▶ From basic, betting, positional rating
 - Streak
- Model types
 - Multinomial regression, Generalized Additive Models (GAMs), Random forest algorithm
- Trained/tested data
 - Train data - first $\frac{2}{3}$ of season
 - Test data - last $\frac{1}{3}$ of season

RESULTS

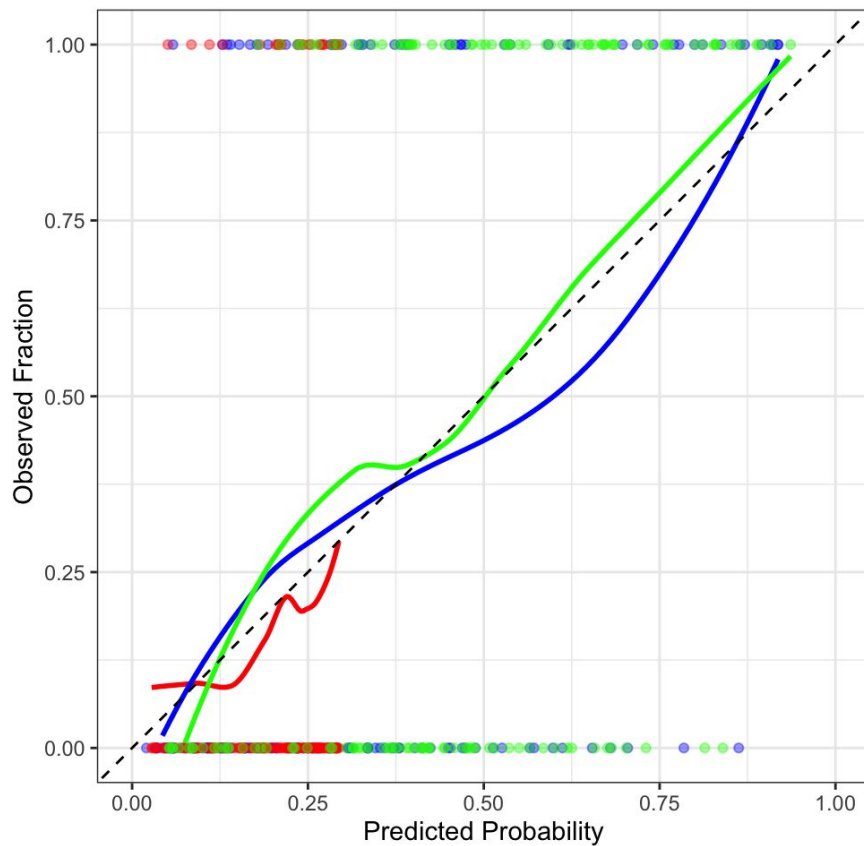
	Brier_Score	Type_of_Model
	<dbl>	<chr>
Pos GAMs	0.2552390	GAMs
multinomial betting pos	0.2565047	multinomial
base GAMs	0.2574400	GAMs
multinomial full	0.2583760	multinomial
multinomial betting pos streak	0.2588285	multinomial
multinomial base	0.2591502	multinomial
Full GAMs	0.2620212	GAMs
streak GAMs	0.2634418	GAMs
streak xg GAMs	0.2648771	GAMs
multinomial pos	0.2718021	multinomial
rand forest pos	0.4409670	random forest
rand forest streak Xg	0.4444513	random forest
rand forest base	0.4980714	random forest

Calibration Curves of Multinomial Base Model for Different Outcomes



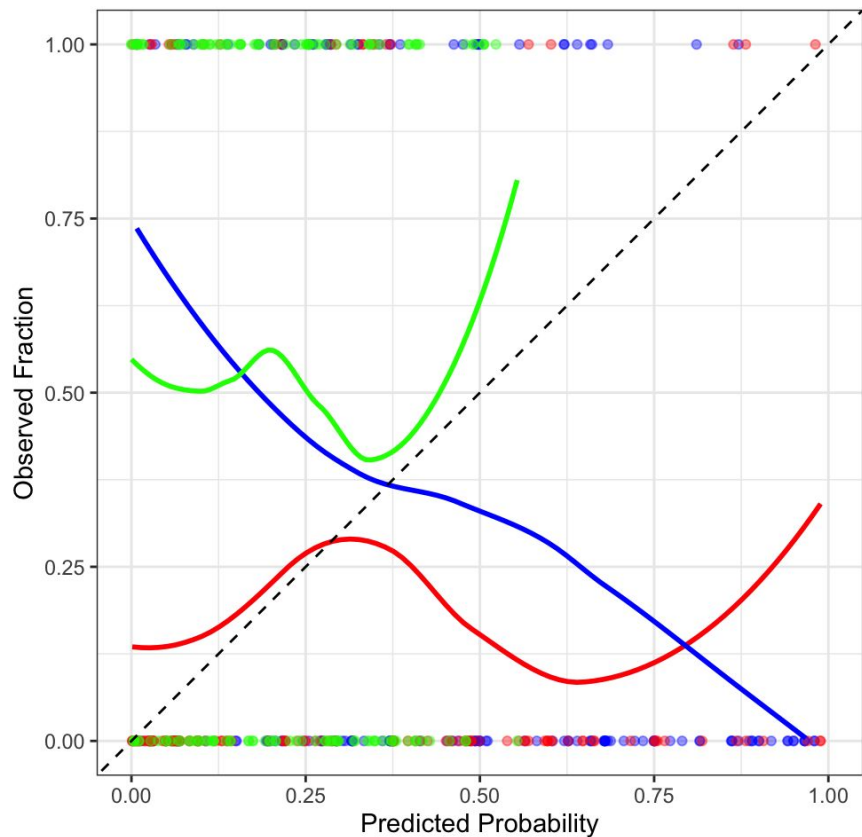
Expected Outcome — Draws — Losses — Wins

Calibration Curves of GAMs Model with XGs from Offense/Defense Rating for Different Outcomes



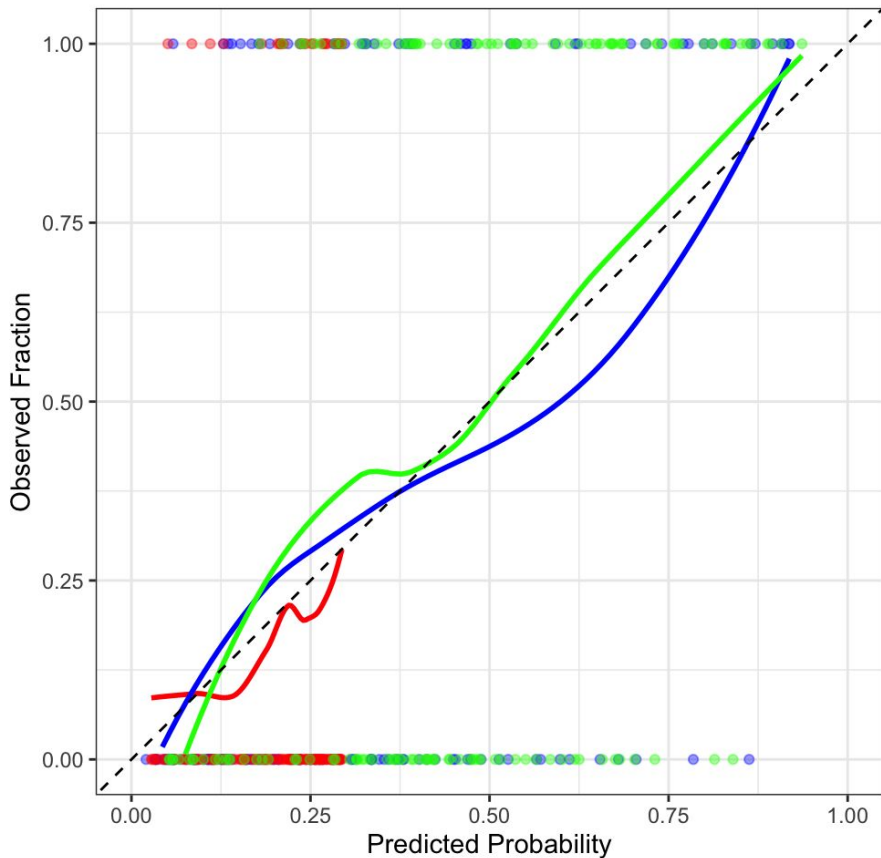
Expected Outcome — Draws — Loses — Wins

Calibration Curves of Base Random Forest Model for Different Outcomes



Expected Outcome — Draws — Loses — Wins

Calibration Curves of GAMs Model with XGs from Offense/Defense Rating for Different Outcomes



Expected Outcome — Draws — Loses — Wins

CONCLUSION

- Betting data creates good predictions
 - Expected goals from betting data
- Adding metrics create better prediction
 - player/team evaluation
- Team streak and expected goals from basic data
 - Over fits models
 - Not a great metric
- Next time you bet, look at the betting odds!
 - Or... make a new model based on it

LIMITATIONS

- Unpredictable factors in the game
 - In play injuries, lineup, and unexpected tactical adjustments
 - Could limit accuracy of predictions
- Predictions from previous season
 - Every season is different
- Every league is different
- Different betting websites could alter predictions

REFERENCES

Benjamin Holmes, Ian G. McHale, Forecasting football match results using a player rating based model, *International Journal of Forecasting*, 2023.

Ojeagbase, Solomon. "Premier League Generated over €68b of Wagers, More than Serie A and Bundesliga Combined." *Complete Sports*, 28 Sept. 2021, www.completesports.com/premier-league-generated-over-e68b-of-wagers-more-than-serie-a-and-bundesliga-combined/.

Ulmer, Ben, Matthew Fernandez, and Michael Peterson. "Predicting soccer match results in the english premier league." Doctoral dissertation, Doctoral dissertation, Ph. D. dissertation, Stanford (2013).

Razali, Nazim, et al. "Predicting football matches results using Bayesian networks for English Premier League (EPL)." *Iop conference series: Materials science and engineering*. Vol. 226. No. 1. IOP Publishing, 2017.

Title page picture - <https://trueidtv.trueid.net/en/premier-league>