PREDICTING THE OUTCOME OF SOCCER MATCHES

LUCAS KIMBALL

BUSINESS PROBLEM

I have been hired by the People's Market, a betting website and blog, to predict the outcome of English Premier League soccer matches based on data, and via these predictions develop a profitable system of betting.

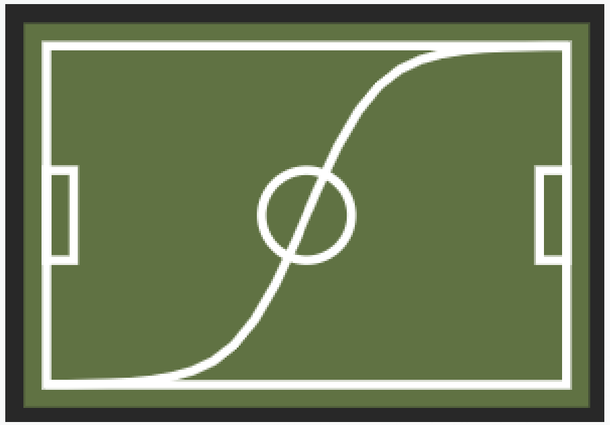
THE DATA

FBREF

Data Collected from:

- Football Reference (FBRef)
- Football-Data.uk
- Club Elo Rating

- Includes 2280 matches from the 17/18 season through the 22/23 season
- Closing odds from Pinnacle for each match
- Elo rating for home and away team for each match



FEATURE ENGINEERING

- Our initial data for each match includes the score and xG for each side
- I took all previous cumulative xG and goals scored and conceded for that season, to create a host of new predictive features
- Including goals to date, conceded to date, xG to date, xG conceded to date

MODELING

Multi-classification problem (Win, Loss, Draw)

01

BASELINE MODEL

Basic Logistic Regression using only numeric features

02

ENSEMBLE MODELING

An variety of classification models and hyperparameters including including Log Reg, RandomForest, XGBoost, CatBoost

03

CATBOOST

The CatBoost classifier performed the best out of the previously tested models, now it is trained again using additional categorical data, while iterating through additional hyperparameters.

PERFORMANCE

Our CatBoost Model with optimized hyper parameters has a .469 F1 Score, but that is massively weighed down by the accuracy of draws.

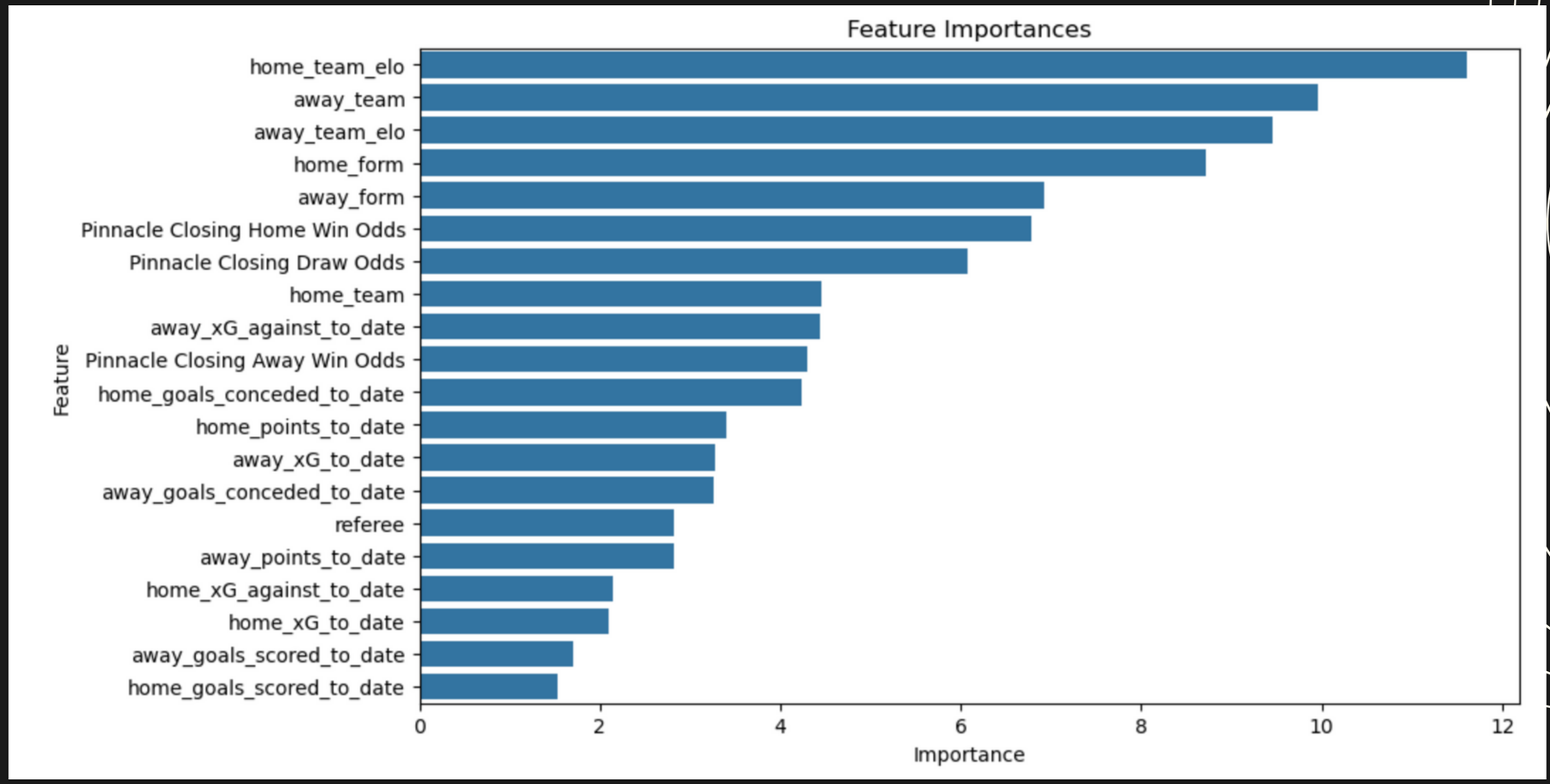
.469 F1 SCORE

.759 ACCURACY FOR HOME WINS

.603 ACCURACY FOR AWAY WINS

.018 ACCURACY FOR DRAWS

FEATURE IMPORTANCE



CONVERTING RESULTS TO BETS

01 EXTRACT OUTCOME PROBABILITIES FOR TEST SET

02 CALCULATE EXPECTED VALUE FOR ALL OUTCOMES

03 DEVELOP BETTING STRATEGIES

04 SIMULATE STRATEGIES ACROSS THE TEST SET

05 CALCULATE PROFIT AND ROI

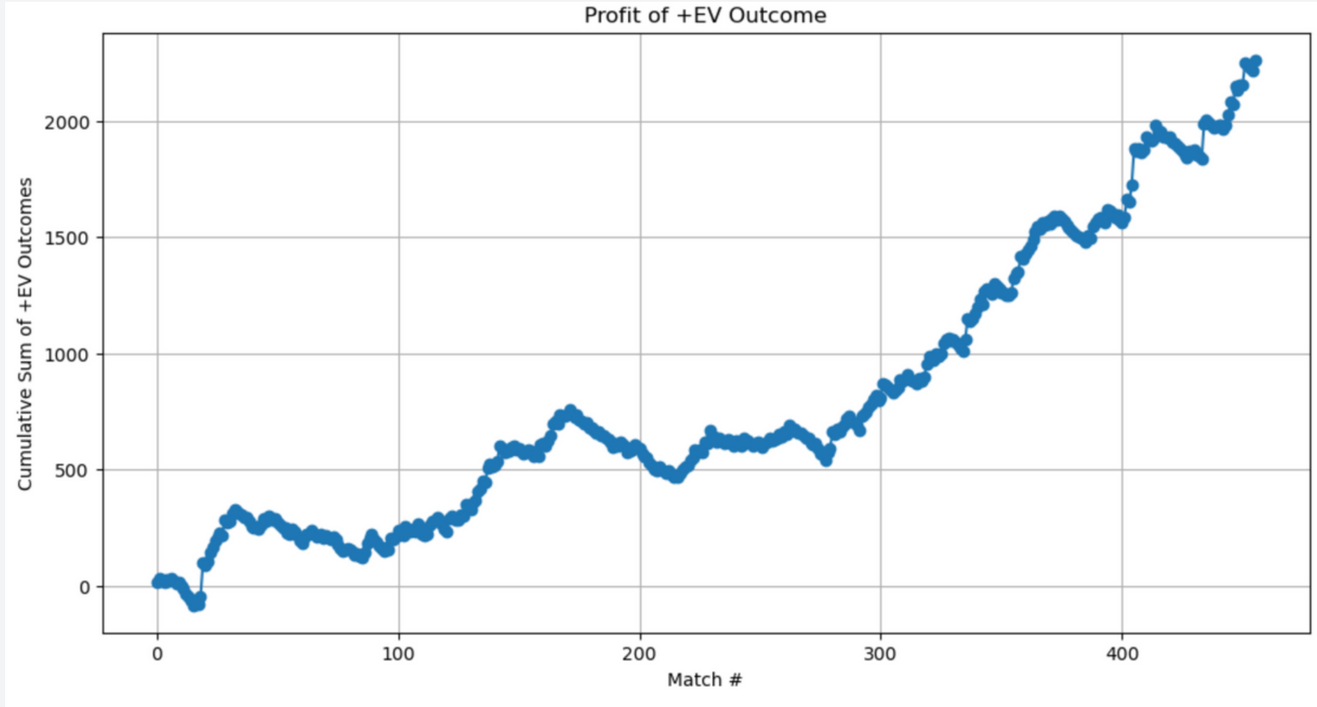
RESULTS

Strategy 1

- Bet $10 on every +EV outcome
- 611 Bets over 456 matches
- $2263.80 in profit
- 37.05% ROI

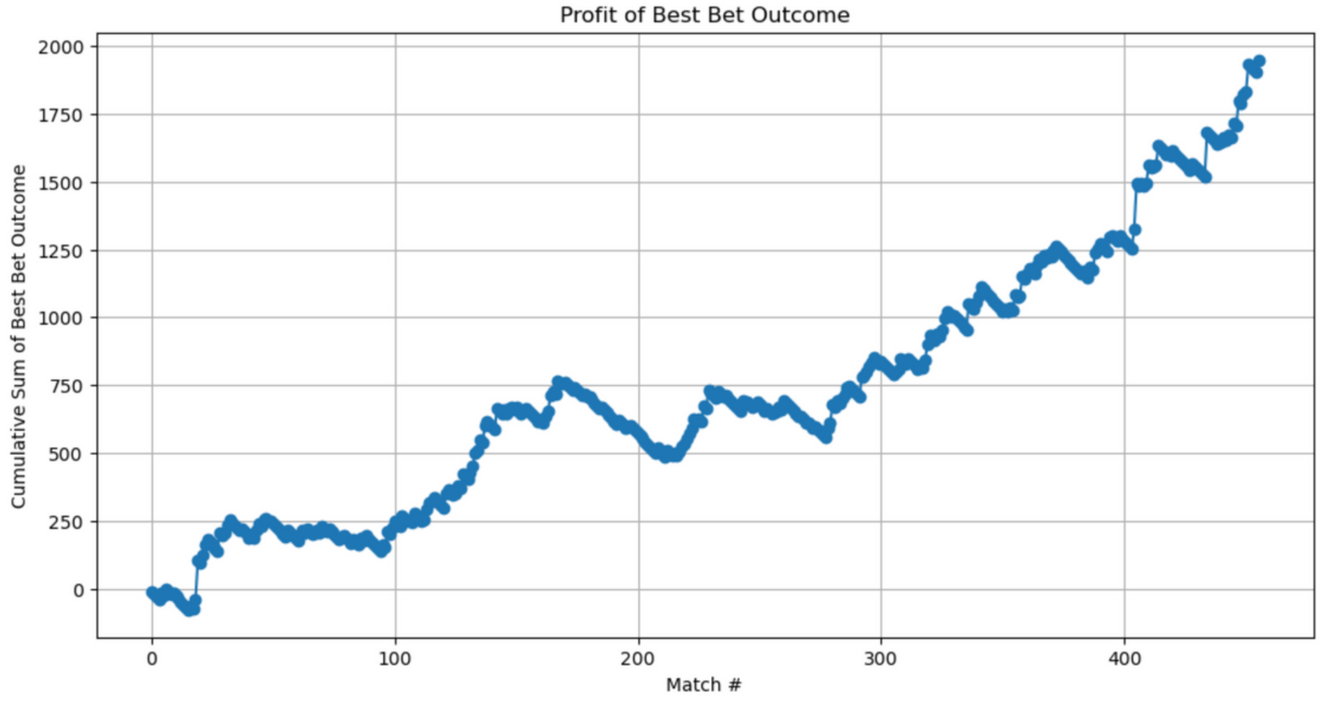
Strategy 2

- Bet $10 on the best EV outcome for each match
- 456 bets over 456 matches
- $1948.30 in profit
- 43.10% ROI



46.7% Win rate per match

38.8% Win rate per match



NEXT STEPS

Test out the betting strategies on upcoming games

Add additional predictors such as the value of a starting lineup

Create a real-time win probability model, incorporating game state, xThreat, Field Tilt

Questions?

lucaskimball98@gmail.com

https://github.com/LMK08

https://www.linkedin.com/in/lucas-kimball-/