# SOCCER BETTING SIMULATION: OPTIMIZING LONG-TERM PROFIT

JOE OMATOI

# OVERVIEW

A BETTING SIMULATION ON OVER 8000 GAMES WAS IMPLEMENTED.

THREE STRATEGIES WERE ANALYZED AND COMPARED:

1.  ARBITRAGE BETTING

2.  POSITIVE EV BETTING

3.  ML-BASED BETTING

THE GOAL IS TO ASSESS GENERAL PROFITABILITY AND THE BEST METHOD FOR LONG-TERM PROFIT

# RAW DATA

DATA FOR 8264 LEAGUE MATCHES IN THE TOP 5 EUROPEAN FOOTBALL LEAGUES IN THE 2019-2020, 2020-2021, AND 2021-2022 SEASONS:

1. FIXTURE DATA (SPORTS REFERENCE)

   • DATE, RESULT, HOME TEAM, AWAY TEAM, ATTENDANCE, VENUE, LEAGUE, REFEREE, GOALS HOME, GOALS AWAY

2. ODDS DATA (FOOTBALL-DATA.CO.UK)

   • MARKET MAXIMUM, MARKET AVERAGE, AND 6 BOOKMAKER ODDS (HOME DRAW, AWAY)

# PROCESSED DATA

| date | result | home | away | b365h | b365d | b365a | max_h | max_d | max_a | avg_h | avg_d | avg_a |
|------|--------|------|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 2019-08-09 | A | Monaco | Lyon | 3.00 | 3.3 | 2.37 | 3.0 | 3.60 | 2.52 | 2.89 | 3.40 | 2.43 |
| 2019-08-10 | A | West Ham | Manchester City | 12.00 | 6.5 | 1.22 | 13.0 | 6.75 | 1.29 | 11.84 | 6.28 | 1.25 |
| 2019-08-10 | H | Burnley | Southampton | 2.62 | 3.2 | 2.75 | 2.8 | 3.33 | 2.85 | 2.68 | 3.22 | 2.78 |

# 1) ARBITRAGE BETTING: FORMULAS

SPOT ARBITRAGE OPPORTUNITY:

$$1/odds_h + 1/odds_d + 1/odds_a < 1$$

REQUIRED WAGER AMOUNT TO ENSURE RISK FREE BET:

$$wager_h = wager_t/\left(1 + (odds_h/odds_d) + (odds_h/odds_a)\right)$$

$$wager_d = wager_t/\left(1 + (odds_d/odds_h) + (odds_d/odds_a)\right)$$

$$wager_a = wager_t/\left(1 + (odds_a/odds_h) + (odds_a/odds_d)\right)$$

# 1) ARBITRAGE BETTING: CALCULATION

| date | wager_h | wager_d | wager_a | wager_total | profit | cum_profit |
|------|---------|---------|---------|-------------|--------|------------|
| 2019-08-10 | NA | NA | NA | NA | NA | NA |
| 2019-08-16 | 329.44 | 49.62 | 23.53 | 402.59 | 9.21 | 11.81 |
| 2019-08-16 | 267.59 | 89.20 | 57.61 | 414.40 | 0.37 | 12.18 |
| 2019-08-17 | 341.71 | 59.74 | 25.13 | 426.57 | 0.56 | 12.74 |
| 2019-08-17 | 83.39 | 100.45 | 255.48 | 439.31 | 2.66 | 15.40 |

# 1) ARBITRAGE BETTING: SIMULATION



Unbiased arbitrage (compounded): cumulative profit from $400 bets
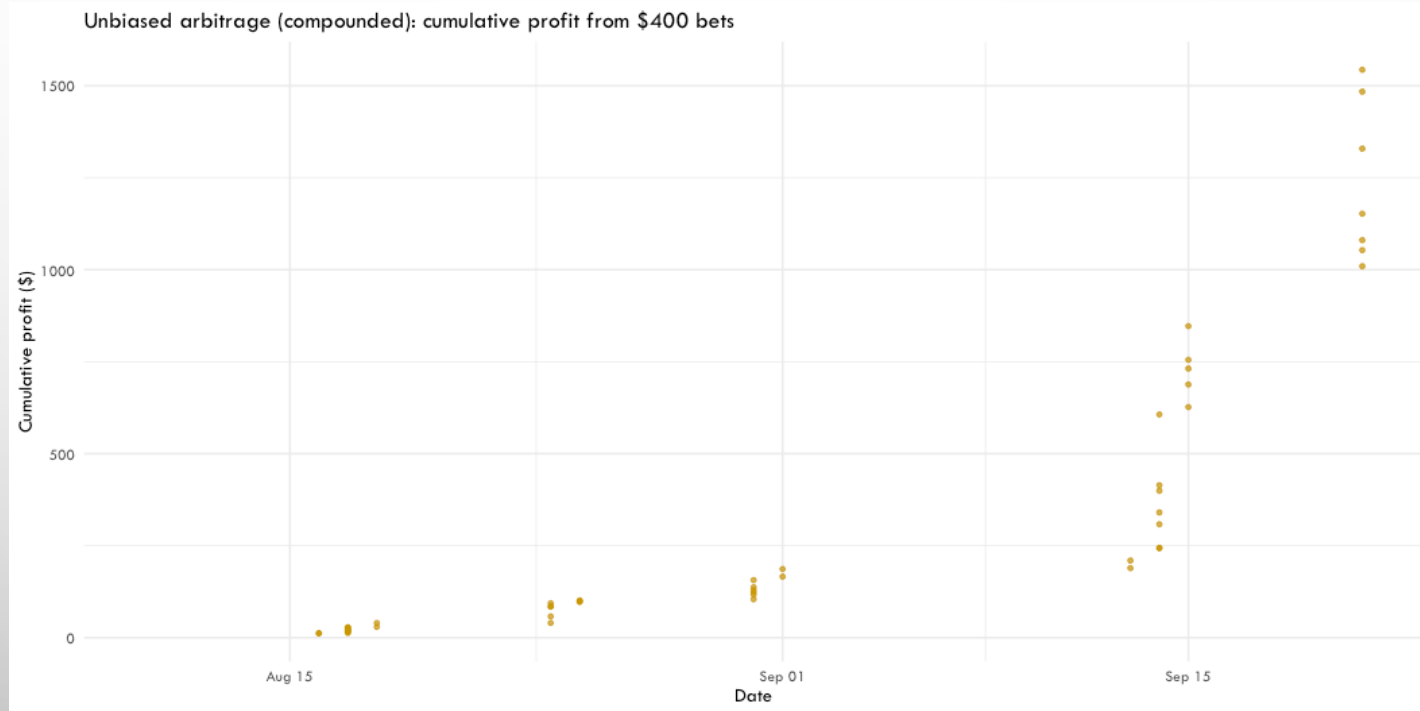
# 1) ARBITRAGE BETTING: EVALUATION

- ADVANTAGE: GUARANTEED AND COMPOUNDED PROFIT

- DISADVANTAGE: NOT ALLOWED FOR LONG (USUALLY GET BANNED)

# 2) POSITIVE EV BETTING: FORMULAS

$$EV_h = 10(odds_h - 1)p_h + (-10)(1 - p_h)$$
$$EV_d = 10(odds_d - 1)p_d + (-10)(1 - p_d)$$
$$EV_a = 10(odds_a - 1)p_a + (-10)(1 - p_a)$$

($p$ WAS CALCULATED USING ITS CORRESPONDING $odds_{(avg)}$, DETAILED CALCULATION IN SOURCE CODE)

# 2) POSITIVE EV BETTING: CALCULATION

| date | ev_h | ev_d | ev_a | bet | wager | result | profit | cum_profit |
|------|------|------|------|-----|-------|--------|--------|------------|
| 2019-08-09 | -0.13 | 0.07 | -0.14 | D | 10 | A | -10 | -10 |
| 2019-08-10 | 0.52 | 0.30 | -0.11 | H | 10 | A | -10 | -20 |
| 2019-08-10 | 0.01 | -0.09 | -0.17 | H | 10 | H | 18 | -2 |
| 2019-08-10 | -0.12 | -0.16 | 0.35 | A | 10 | A | 36 | 34 |
| 2019-08-10 | -0.16 | -0.02 | 0.08 | A | 10 | D | -10 | 24 |

# 2) POSITIVE EV BETTING: SIMULATION



Positive EV: cumulative profit from $10 bets
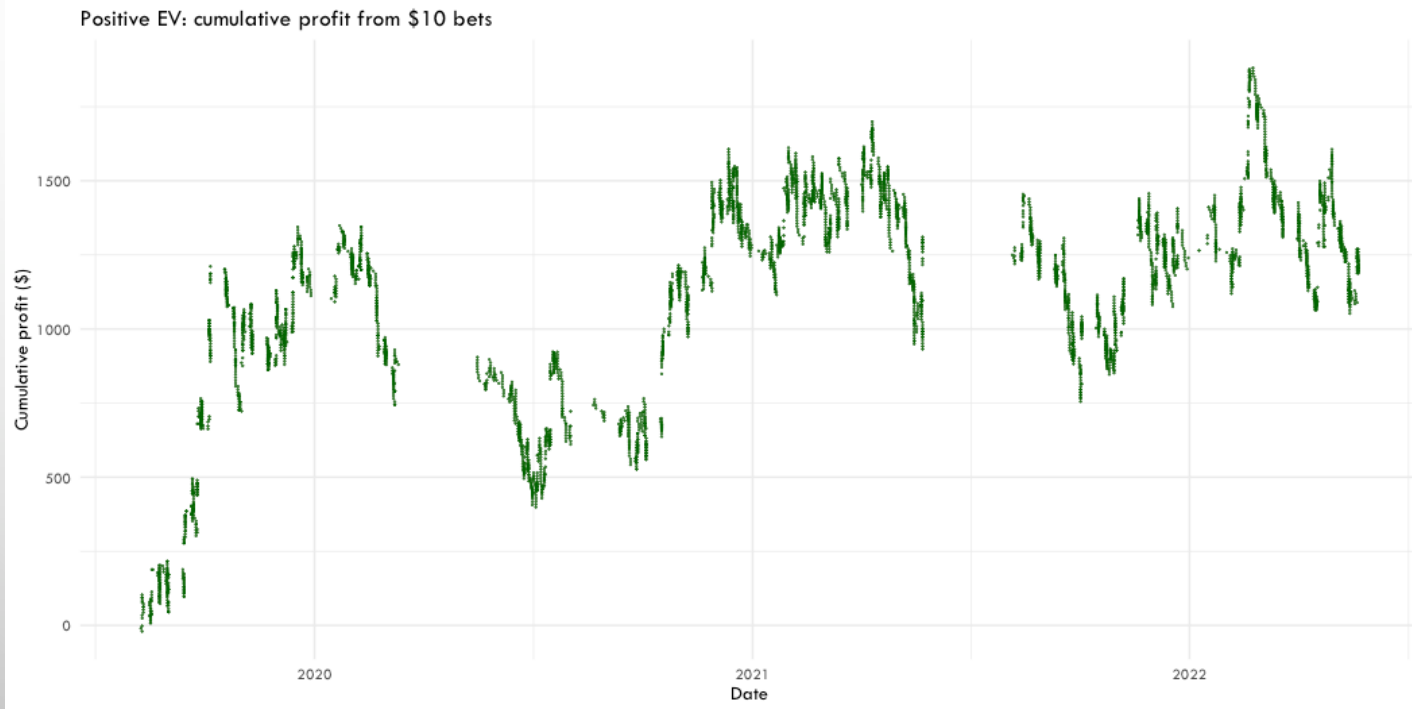
# 2) POSITIVE EV BETTING: EVALUATION

- ADVANTAGE: LONG-RUN PROFIT

- DISADVANTAGE: SHORT-RUN FLUCTUATIONS REQUIRE MANY BETS AND MAKE BIG BETS RISKY

# 3) MACHINE LEARNING BETTING: TRAINING

**RESAMPLING**

- 6612 GAMES FOR TRAINING (80% OF TOTAL)

- V-FOLD CROSS VALIDATION WITH 5 FOLDS AND 5 REPEATS

- PRECISION (TRUE POSITIVES/TOTAL PREDICTED POSITIVES) AS PERFORMANCE MEASURE
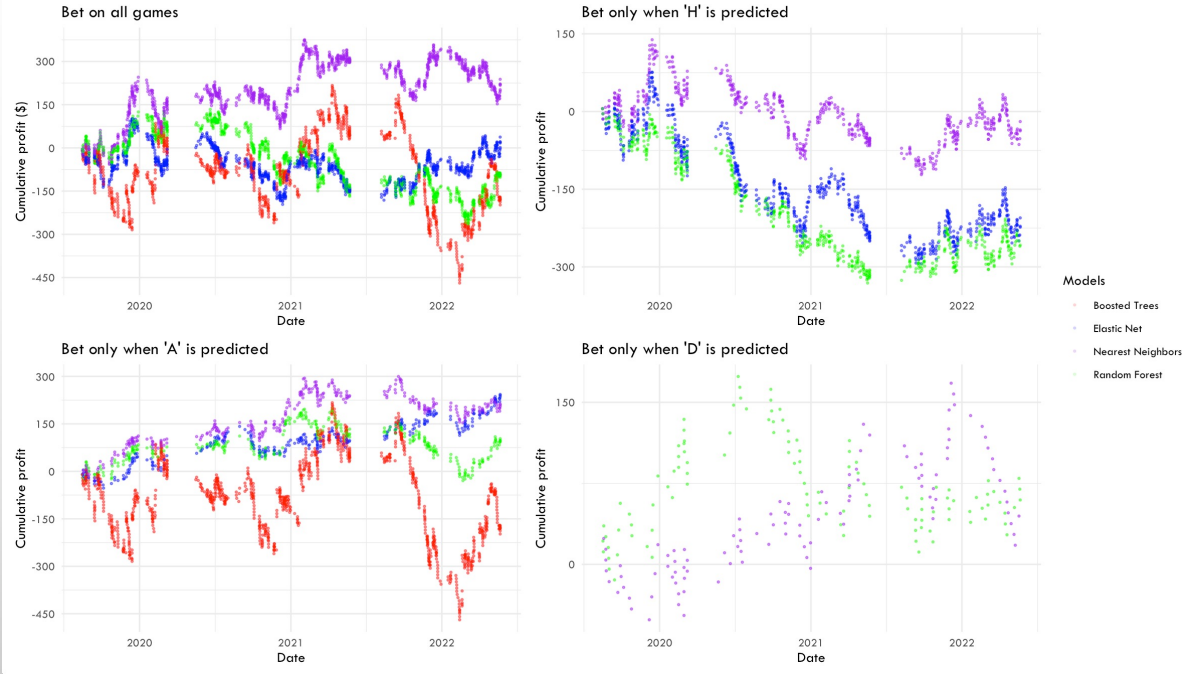
# 3) MACHINE LEARNING BETTING: TRAINING

**MODELS**

- RANDOM FOREST

- BOOSTED TREES

- ELASTIC NET

- NEAREST NEIGHBORS

**PREDICTORS**

PCA ON ALL 6 BOOKMAKERS' ODDS

# 3) MACHINE LEARNING BETTING: SIMULATION



ML: cumulative profits from $10 bets

# 3) MACHINE LEARNING BETTING: EVALUATION

- ADVANTAGE: HIGH PROFITABILITY ACHIEVED WITH ONLY BETTING WHEN DRAW OR AWAY IS PREDICTED

- DISADVANTAGE: LIMITED GAMES TO BET PER SEASON

## FINAL COMPARISON

*PROFITS OF MACHINE LEARNING METHODS ARE MULTIPLIED BY 8264/1010 SINCE ONLY 1010 GAMES WERE AVAILABLE (TESTING DATA WAS 20% OF THE TOTAL AND NOT ALL GAMES HAD MARKET MAX/AVG ODDS)

*PROFIT OF THE POSITIVE EV METHOD IS MULTIPLIED BY 8264/4963 SINCE ONLY 4963 GAMES HAD MARKET MAX/AVG ODDS.

| Method | Estimated profit per season ($)* | Estimated Profit ($)* | Profit ($) | Number of bets |
|---|---|---|---|---|
| Random Forest (A&D) | 791.7621 | 2375.2863 | 290.3 | 436 |
| Nearest Neighbors (A&D) | 702.3036 | 2106.9109 | 257.5 | 431 |
| Positive EV | 661.4419 | 1984.3258 | 1191.7 | 4565 |
| Nearest Neighbors | 649.3922 | 1948.1766 | 238.1 | 1010 |
| Elastic Net | 102.5500 | 307.6499 | 37.6 | 1010 |
| Random Forest | 59.7299 | 179.1897 | 21.9 | 1010 |
| Boosted Trees | -540.0238 | -1620.0713 | -198.0 | 1010 |

# THANK YOU!

- ACCESS SOURCE CODE ON MY GITHUB

- EMAIL ME WITH QUESTIONS AND COMMENTS