Model Training Report: Track 1

Configuration Details

Target Variable

`target`

Selected Features

`goal_diff_80, home_shots_pre80, away_shots_pre80, home_fouls_pre80, away_fouls_pre80, home_corners_pre80, away_corners_pre80, home_yellow_cards_pre80, away_yellow_cards_pre80, home_sending_off_pre80, away_sending_off_pre80, home_corners_70_75, away_corners_70_75, home_shots_70_75, away_shots_70_75, home_corners_75_80, away_corners_75_80, home_shots_75_80, away_shots_75_80, odd_h, odd_d, odd_a`
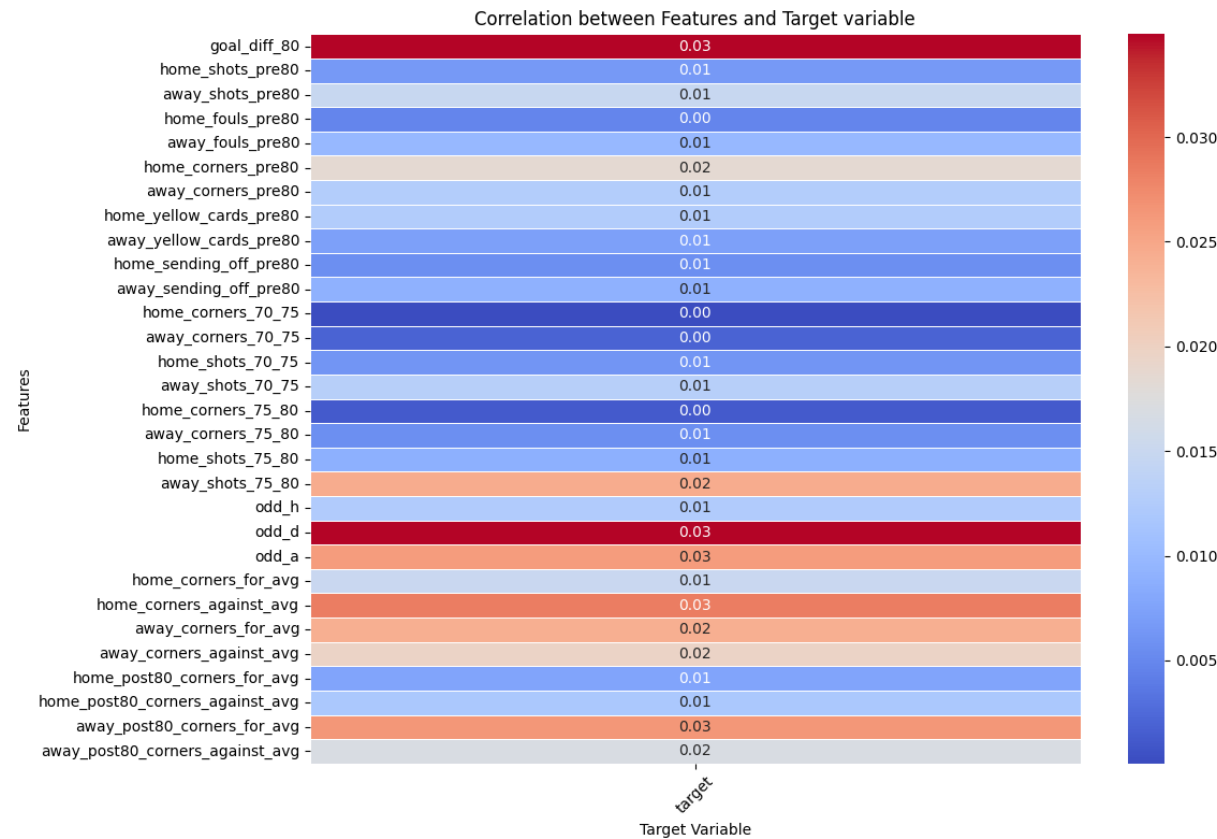
Constructed Features

`home_corners_for_avg, home_corners_against_avg, away_corners_for_avg, away_corners_against_avg, home_post80_corners_for_avg, home_post80_corners_against_avg, away_post80_corners_for_avg, away_post80_corners_against_avg`

Models Trained (& Hyperparameters)

- **random_forest**: {'bootstrap': True, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}
- **logistic_regression**: {'solver': 'liblinear', 'penalty': 'l2', 'C': 1.0}
- **svc**: {'C': 1.0, 'kernel': 'rbf', 'gamma': 'scale'}
- **xgboost**: {'n_estimators': 50, 'max_depth': 3, 'learning_rate': 0.1, 'subsample': 0.8}
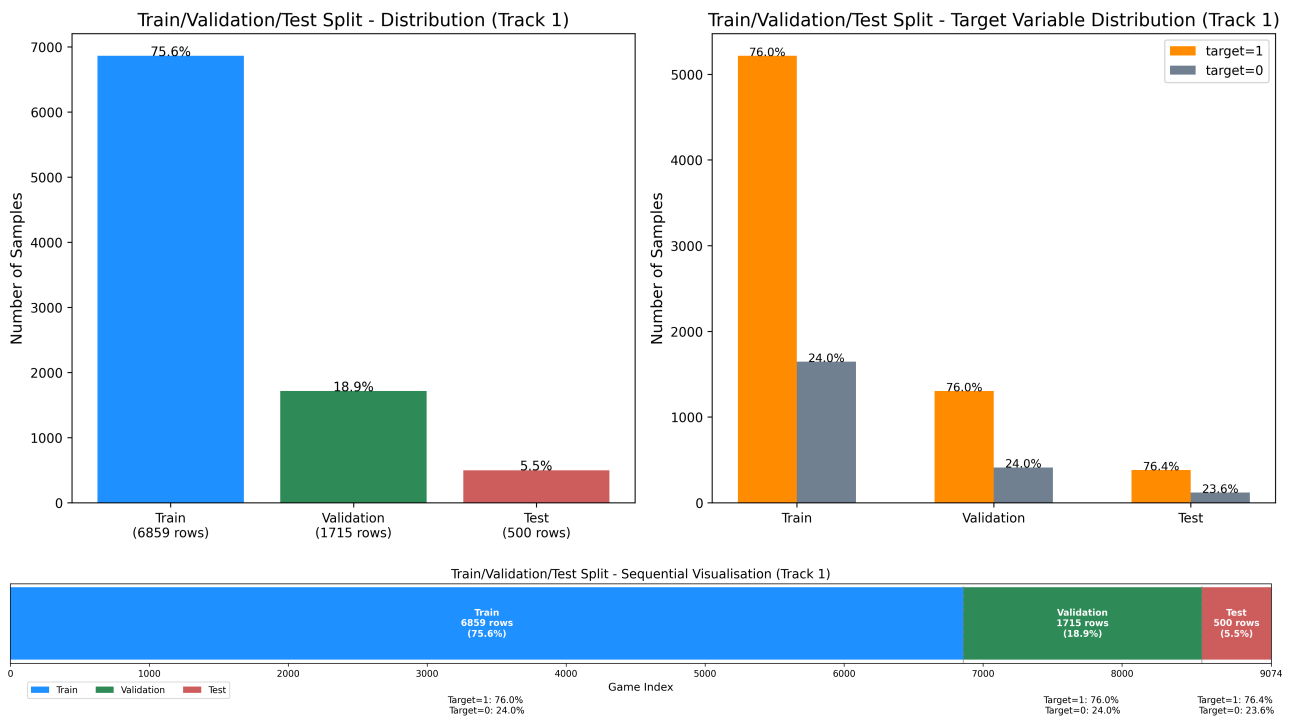
Feature Correlation

Correlation between Features and Target variable

Features	target
goal_diff_80	0.03
home_shots_pre80	0.01
away_shots_pre80	0.01
home_fouls_pre80	0.00
away_fouls_pre80	0.01
home_corners_pre80	0.02
away_corners_pre80	0.01
home_yellow_cards_pre80	0.01
away_yellow_cards_pre80	0.01
home_sending_off_pre80	0.01
away_sending_off_pre80	0.01
home_corners_70_75	0.00
away_corners_70_75	0.00
home_shots_70_75	0.01
away_shots_70_75	0.01
home_corners_75_80	0.00
away_corners_75_80	0.01
home_shots_75_80	0.01
away_shots_75_80	0.02
odd_h	0.01
odd_d	0.03
odd_a	0.03
home_corners_for_avg	0.01
home_corners_against_avg	0.03
away_corners_for_avg	0.02
away_corners_against_avg	0.02
home_post80_corners_for_avg	0.01
home_post80_corners_against_avg	0.01
away_post80_corners_for_avg	0.03
away_post80_corners_against_avg	0.02

Target Variable

Point-Biserial Correlation
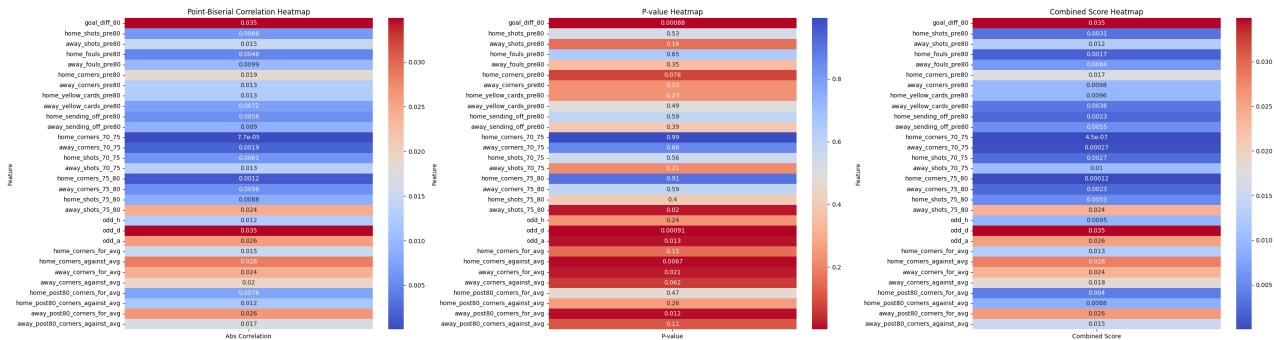
Dataset Train/Validation/Test Split Visualisation

Train/Validation/Test Split - Distribution (Track 1)

Train/Validation/Test Split - Target Variable Distribution (Track 1)

Train/Validation/Test Split - Sequential Visualisation (Track 1)

random_forest Model

Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
home_corners_against_avg	0.05962	home_sending_off_pre80	0.00403
away_post80_corners_for_avg	0.05762	away_sending_off_pre80	0.00557
away_corners_for_avg	0.05713	away_corners_70_75	0.00751
away_corners_against_avg	0.05665	away_corners_75_80	0.00789
home_corners_for_avg	0.05635	home_corners_70_75	0.00839
home_post80_corners_for_avg	0.05587		
home_post80_corners_against_avg	0.05500		
away_post80_corners_against_avg	0.05479		

Optimal Threshold

- **Optimal threshold:** 0.83
- **Expected Precision:** 0.823
- **Expected Recall:** 0.104

Classification Reports
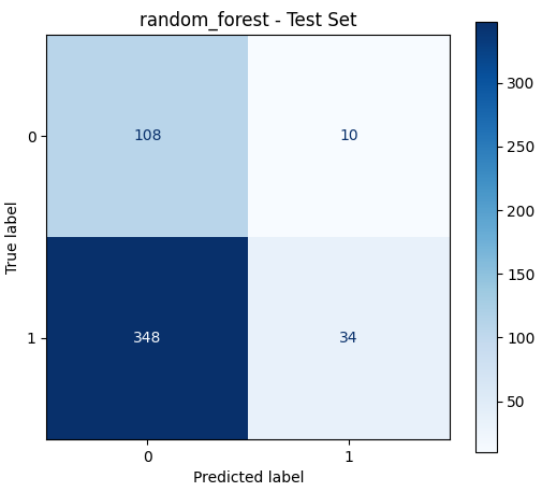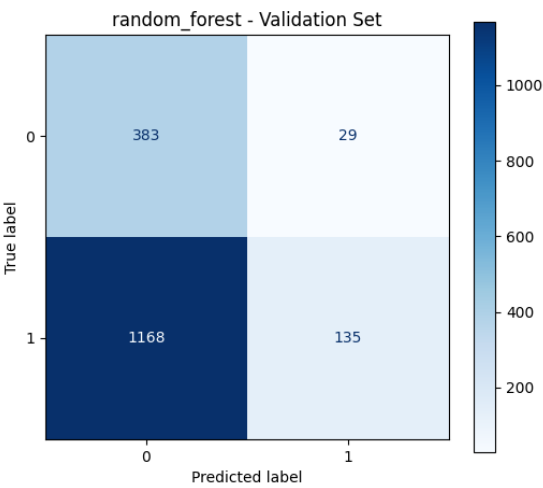
Classification Report (*Validation Set*)

```
          precision    recall  f1-score   support

     0         0.25      0.93      0.39       412
     1         0.82      0.10      0.18      1303

accuracy                          0.30      1715
```
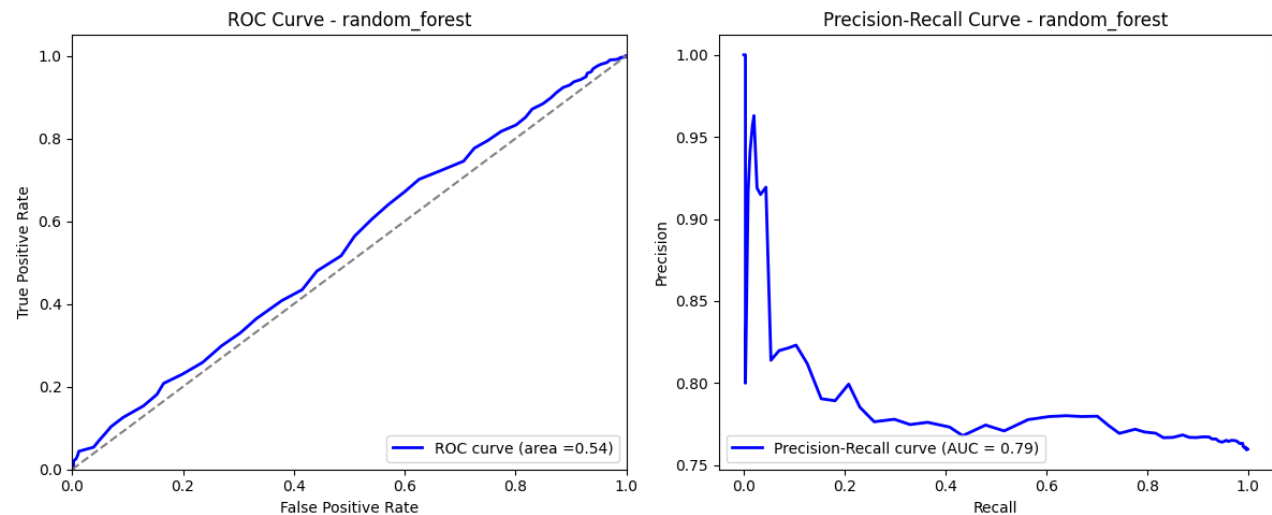
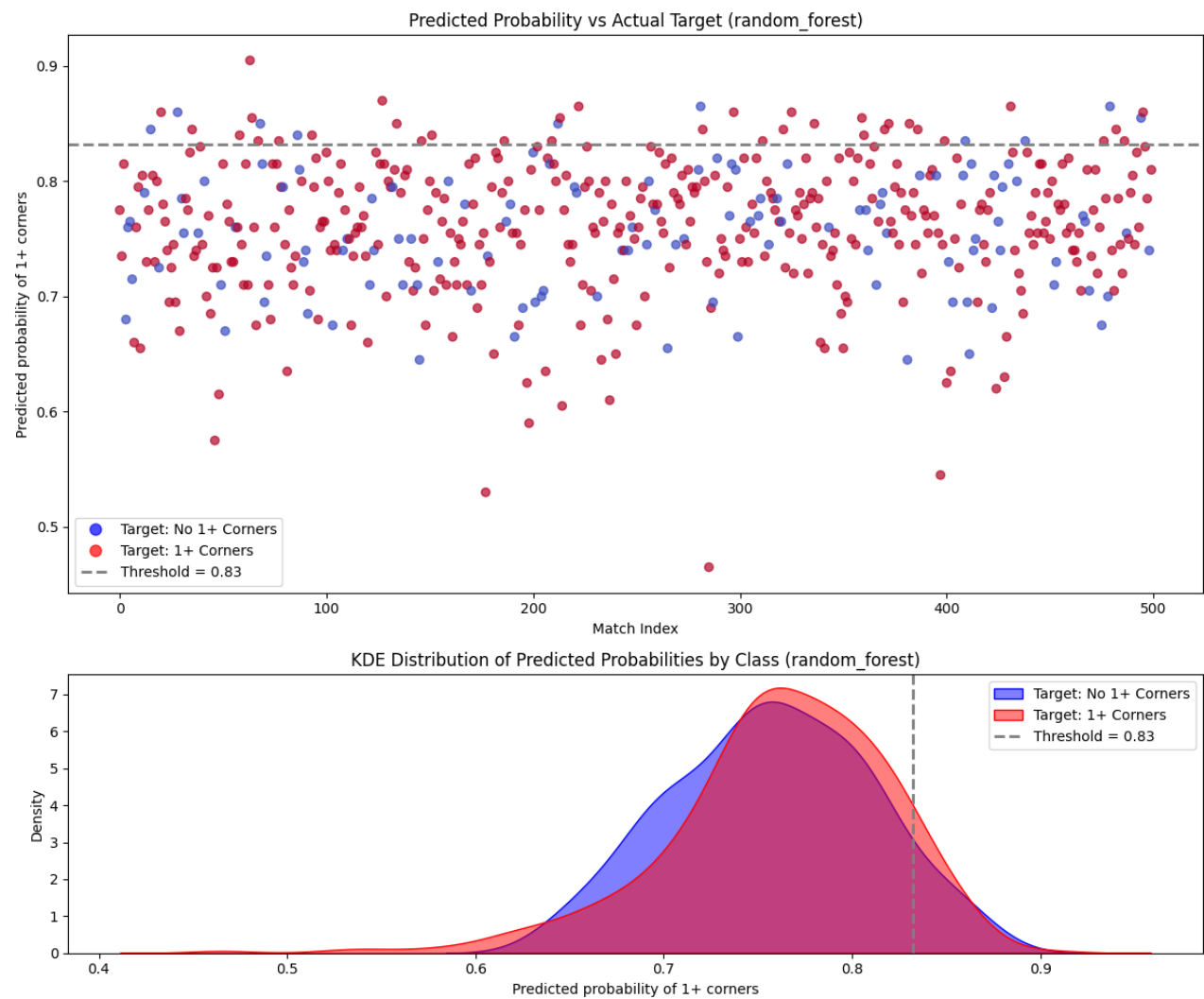macro avg 0.54 0.52 0.29 1715 weighted avg 0.68 0.30 0.23 1715

Classification Report (*Test Set*)

```
          precision    recall  f1-score   support

     0         0.24      0.92      0.38       118
     1         0.77      0.09      0.16       382

accuracy                          0.28       500
```

macro avg 0.50 0.50 0.27 500 weighted avg 0.65 0.28 0.21 500

ROC and Precision-Recall Curves

Prediction Distribution Graphs

Predicted Probability vs Actual Target (random_forest)

KDE Distribution of Predicted Probabilities by Class (random_forest)

Backtesting Results

--- Overall Summary ---

🏦 **Initial Bankroll:** £1000.00

💰 **Final Bankroll:** £866.98

💸 **Total Staked:** £4093.03

📈 **Total Profit:** £-133.02

📊 **ROI:** -13.30%

✅ **Win rate:** 78.57% over 42 bets

📉 **Sharpe Ratio:** -0.04

🎯 **Edge Over Bookies:** -1.20%

Backtesting Bankroll Growth (random_forest)
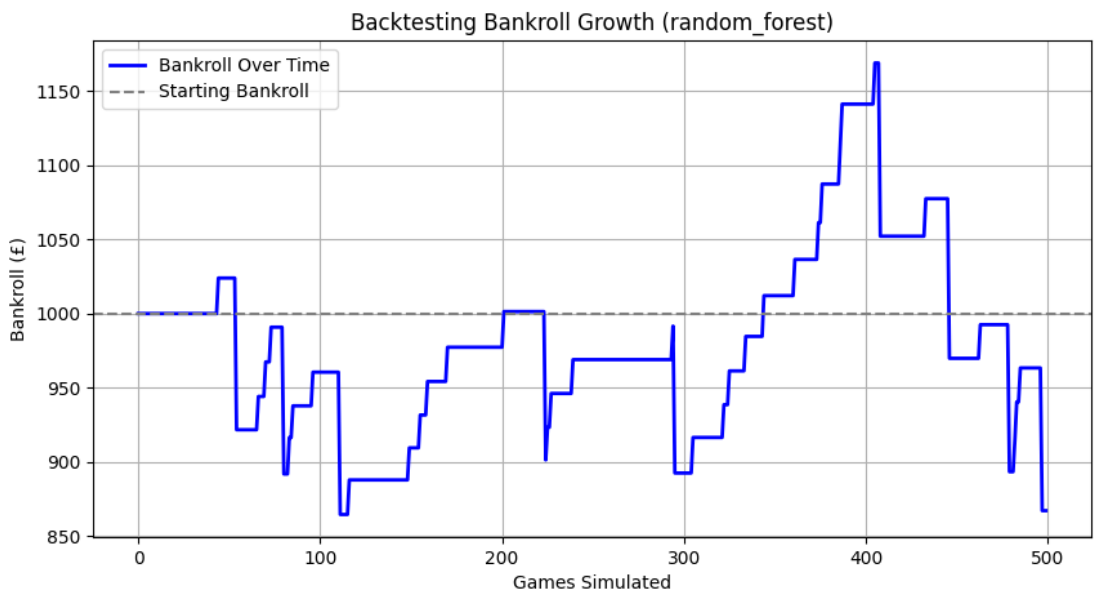
logistic_regression Model

Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
away_corners_against_avg	0.54319	goal_diff_80	-0.78387
away_post80_corners_for_avg	0.51372	odd_h	-0.64380
away_shots_75_80	0.42683	odd_d	-0.53019
home_corners_for_avg	0.40431	odd_a	-0.34382
home_corners_against_avg	0.27775	away_shots_70_75	-0.30718
away_corners_for_avg	0.24780		
home_corners_pre80	0.24122		
home_shots_75_80	0.20315		

Optimal Threshold

- **Optimal threshold:** 0.55
- **Expected Precision:** 0.821
- **Expected Recall:** 0.162

Classification Reports
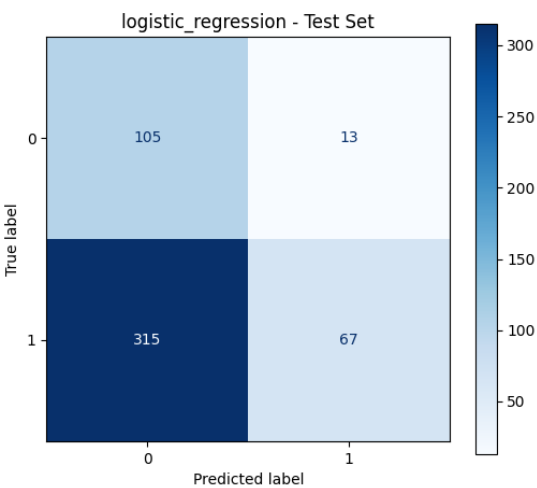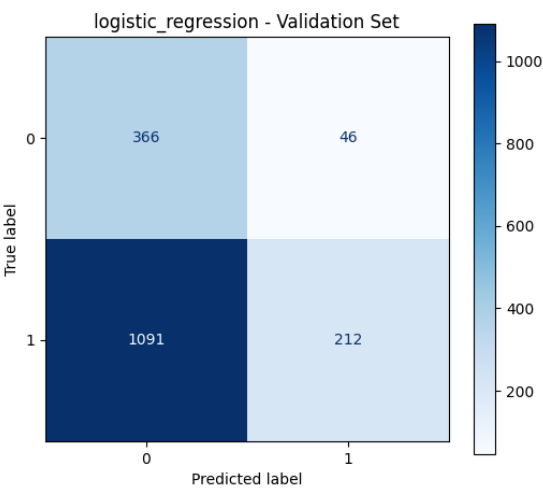
Classification Report (*Validation Set*)

```
          precision    recall  f1-score   support

      0        0.25      0.89      0.39       412
      1        0.82      0.16      0.27      1303

accuracy                          0.34      1715
```
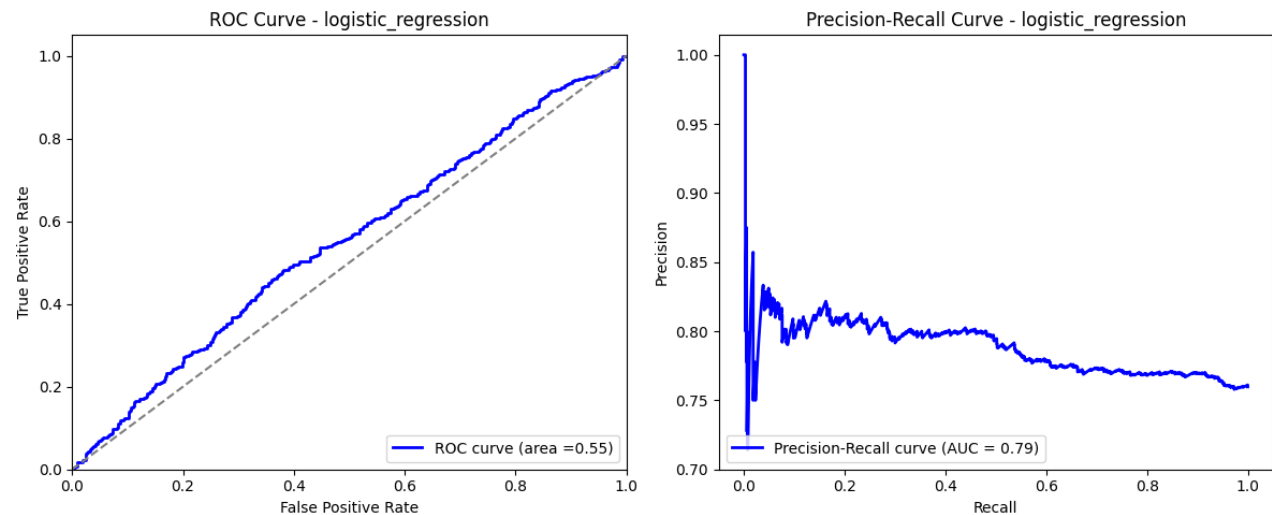
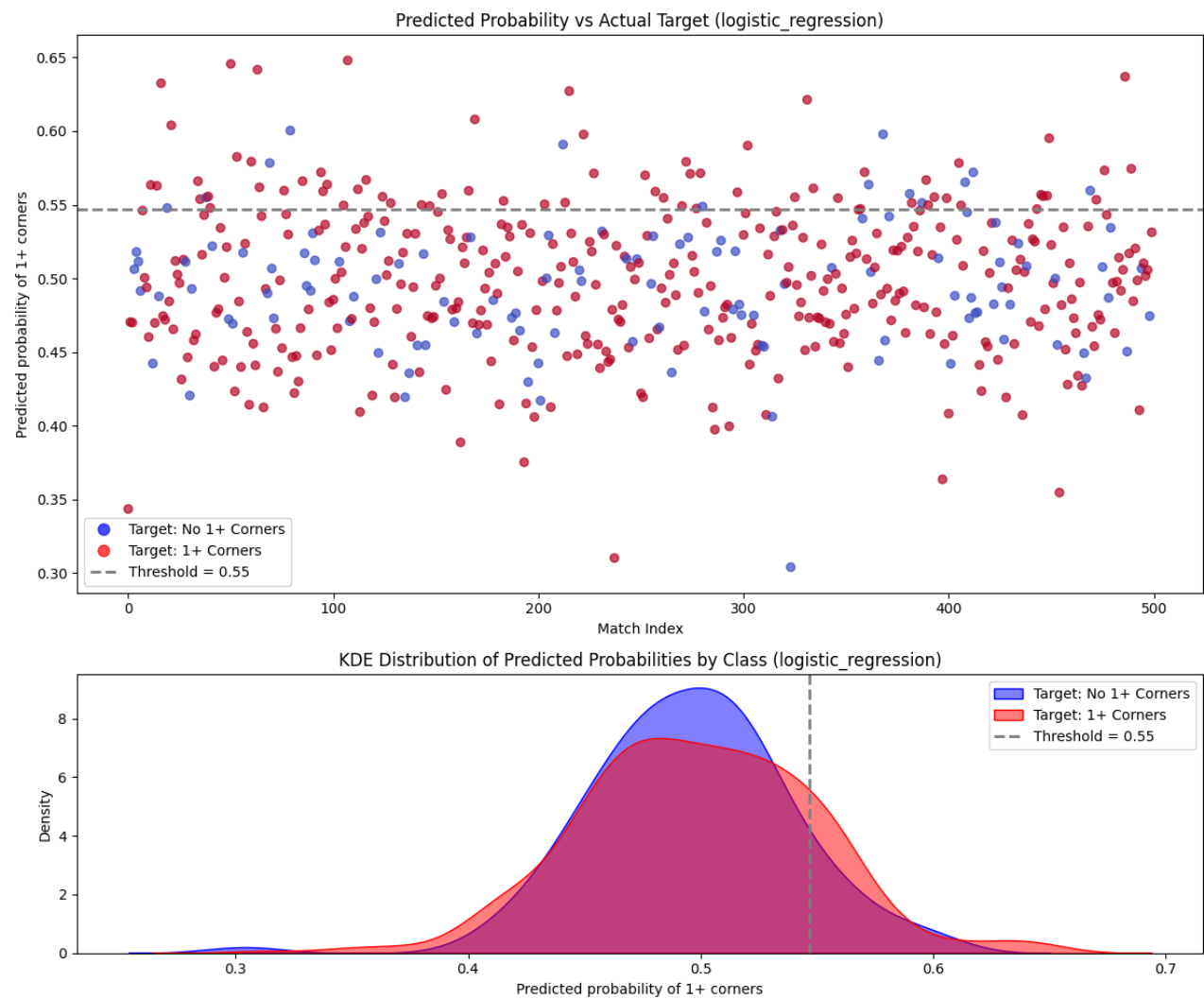macro avg 0.54 0.53 0.33 1715 weighted avg 0.68 0.34 0.30 1715

Classification Report (*Test Set*)

```
          precision    recall  f1-score   support

      0        0.25      0.89      0.39       118
      1        0.84      0.18      0.29       382

accuracy                          0.34       500
```

macro avg 0.54 0.53 0.34 500 weighted avg 0.70 0.34 0.31 500

ROC and Precision-Recall Curves

Prediction Distribution Graphs

Predicted Probability vs Actual Target (logistic_regression)

KDE Distribution of Predicted Probabilities by Class (logistic_regression)

Backtesting Results

--- Overall Summary ---

🏦 **Initial Bankroll:** £1000.00

💰 **Final Bankroll:** £1247.32

💸 **Total Staked:** £9531.86

📈 **Total Profit:** £247.32

📊 **ROI:** 24.73%

✅ **Win rate:** 83.33% over 78 bets

📉 **Sharpe Ratio:** 0.08

🎯 **Edge Over Bookies:** 3.56%
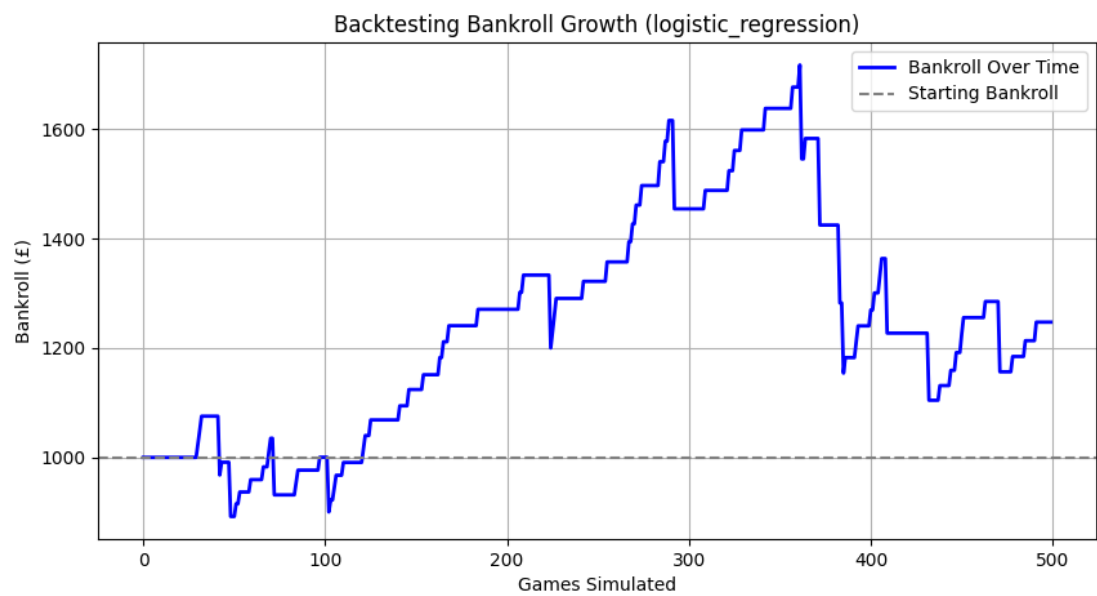
svc Model

Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
goal_diff_80	0.00000	away_post80_corners_against_avg	0.00000
home_shots_pre80	0.00000	away_shots_pre80	0.00000
away_post80_corners_for_avg	0.00000	home_fouls_pre80	0.00000
home_post80_corners_against_avg	0.00000	away_fouls_pre80	0.00000
home_post80_corners_for_avg	0.00000	home_corners_pre80	0.00000
away_corners_against_avg	0.00000		
away_corners_for_avg	0.00000		
home_corners_against_avg	0.00000		

Optimal Threshold

- **Optimal threshold:** 0.76
- **Expected Precision:** 0.775
- **Expected Recall:** 0.515

Classification Reports
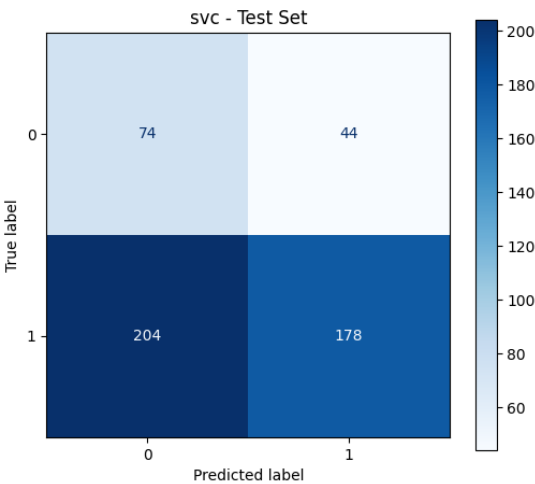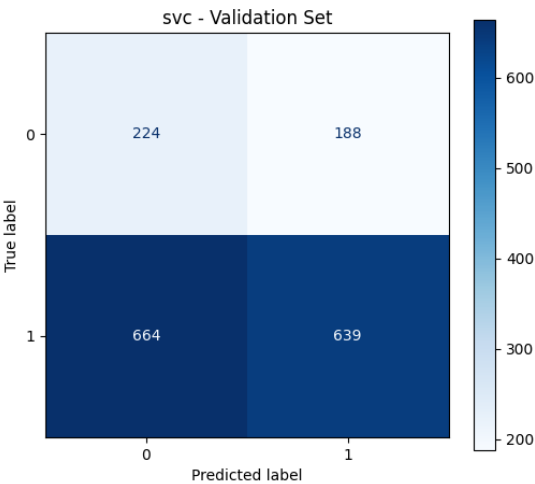
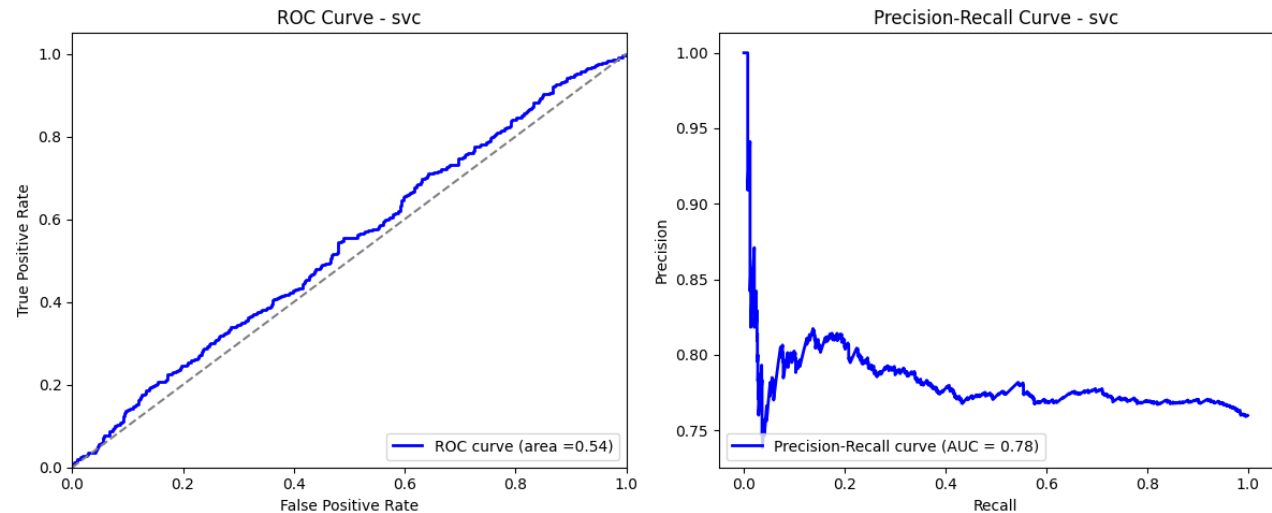## Classification Report (*Validation Set*)

```
           precision    recall  f1-score   support

        0       0.25      0.54      0.34       412
        1       0.77      0.49      0.60      1303

 accuracy                           0.50      1715
```

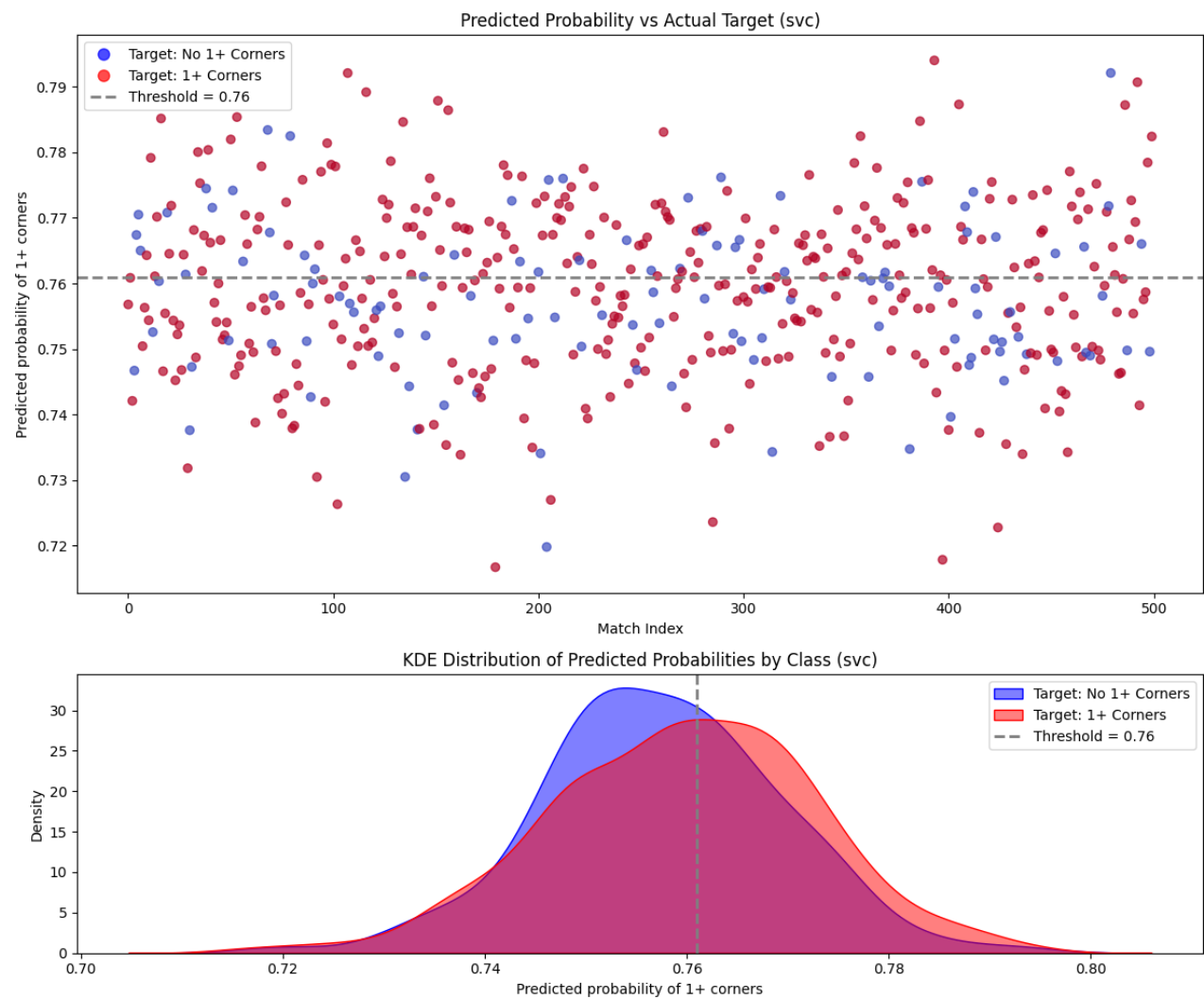macro avg 0.51 0.52 0.47 1715 weighted avg 0.65 0.50 0.54 1715

Classification Report (*Test Set*)

```
           precision    recall  f1-score   support

        0       0.27      0.63      0.37       118
        1       0.80      0.47      0.59       382

 accuracy                           0.50       500
```

macro avg 0.53 0.55 0.48 500 weighted avg 0.68 0.50 0.54 500

ROC and Precision-Recall Curves

Prediction Distribution Graphs

Predicted Probability vs Actual Target (svc)

KDE Distribution of Predicted Probabilities by Class (svc)

Backtesting Results

--- Overall Summary ---

🏦 **Initial Bankroll:** £1000.00

💰 **Final Bankroll:** £817.57

💸 **Total Staked:** £16136.22

📈 **Total Profit:** £-182.43

📊 **ROI:** -18.24%

✅ **Win rate:** 80.28% over 218 bets

📉 **Sharpe Ratio:** 0.01

🎯 **Edge Over Bookies:** 0.50%
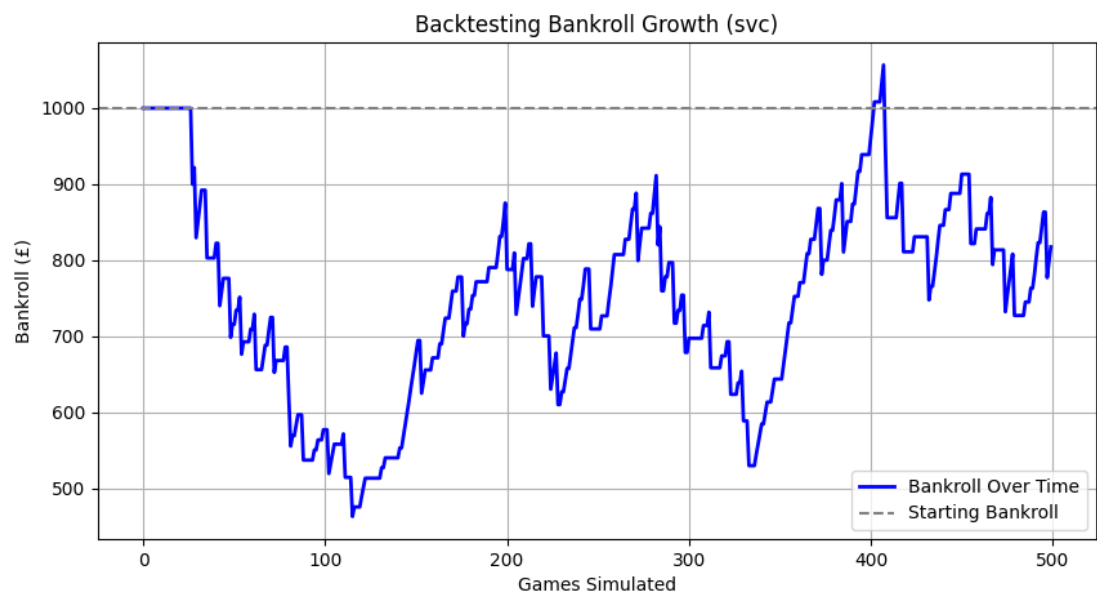
xgboost Model

Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
away_sending_off_pre80	0.05806	home_corners_75_80	0.00000
odd_h	0.04447	home_sending_off_pre80	0.00000
goal_diff_80	0.04267	away_corners_75_80	0.00000
away_corners_for_avg	0.04081	home_yellow_cards_pre80	0.02579
home_post80_corners_against_avg	0.04058	away_yellow_cards_pre80	0.02814
away_post80_corners_against_avg	0.04019		
odd_a	0.03989		
away_corners_against_avg	0.03984		

Optimal Threshold

- **Optimal threshold:** 0.81
- **Expected Precision:** 0.784
- **Expected Recall:** 0.117

Classification Reports
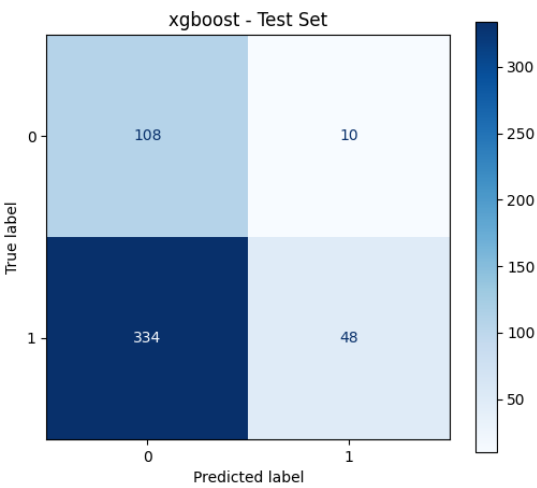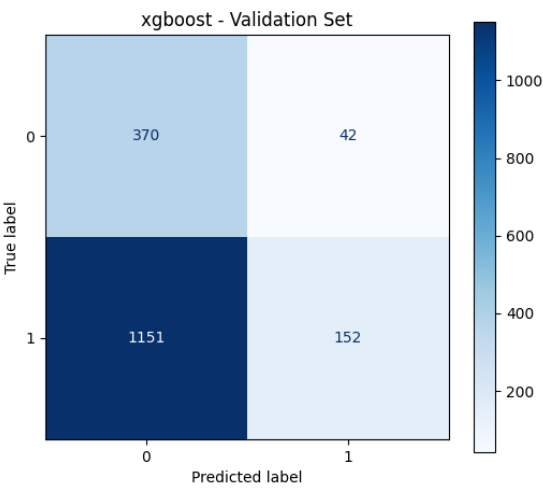
Classification Report (*Validation Set*)

```
         precision    recall  f1-score   support

      0       0.24      0.90      0.38       412
      1       0.78      0.12      0.20      1303

accuracy                          0.30      1715
```
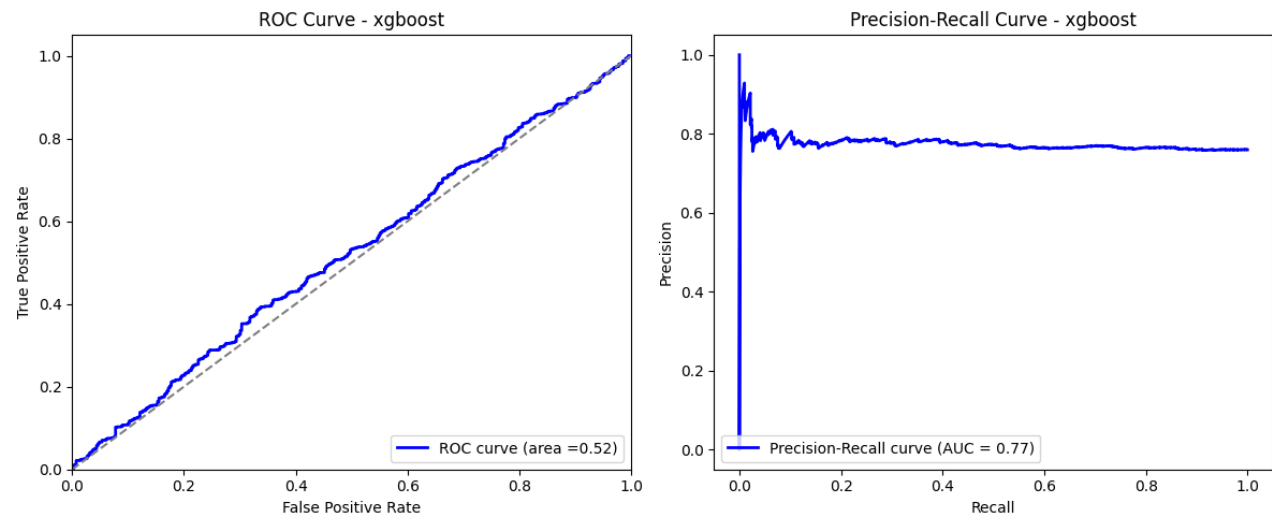
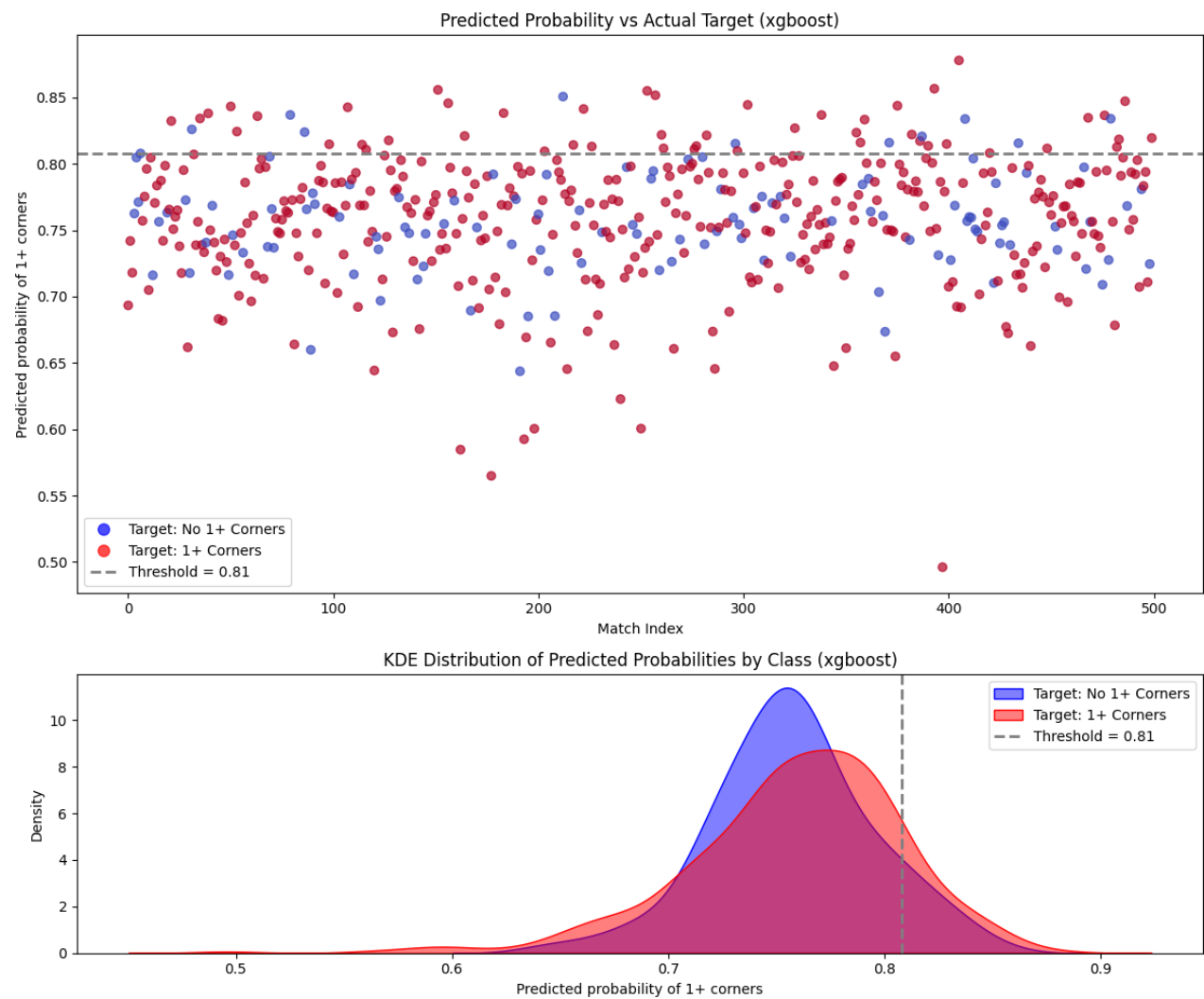macro avg 0.51 0.51 0.29 1715 weighted avg 0.65 0.30 0.25 1715

Classification Report (*Test Set*)

```
         precision    recall  f1-score   support

      0       0.24      0.92      0.39       118
      1       0.83      0.13      0.22       382

accuracy                          0.31       500
```

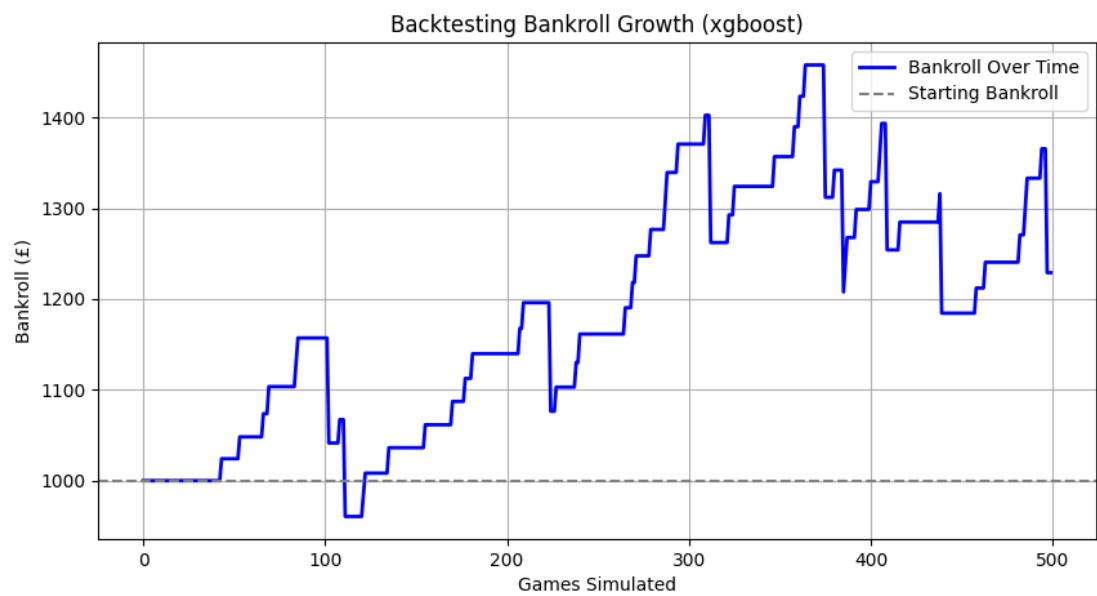macro avg 0.54 0.52 0.30 500 weighted avg 0.69 0.31 0.26 500

ROC and Precision-Recall Curves

Prediction Distribution Graphs

Predicted Probability vs Actual Target (xgboost)

KDE Distribution of Predicted Probabilities by Class (xgboost)

Backtesting Results

--- Overall Summary ---

🏦 **Initial Bankroll:** £1000.00

💰 **Final Bankroll:** £1229.07

💸 **Total Staked:** £6924.50

📈 **Total Profit:** £229.07

📊 **ROI:** 22.91%

✅ **Win rate:** 84.21% over 57 bets

📉 **Sharpe Ratio:** 0.10

🎯 **Edge Over Bookies:** 4.44%

Backtesting Results Summary

Backtesting Bankroll Growth – Track Comparison