Model Training Report: Track 2

Configuration Details

Target Variable

target

Selected Features

is_home_trailing, leading_team_corners_70_75, trailing_team_corners_70_75, leading_team_shots_70_75, trailing_team_shots_70_75, leading_team_fouls_70_75, trailing_team_fouls_70_75, leading_team_corners_75_80, trailing_team_corners_75_80, leading_team_shots_75_80, trailing_team_shots_75_80, leading_team_fouls_75_80, trailing_team_fouls_75_80
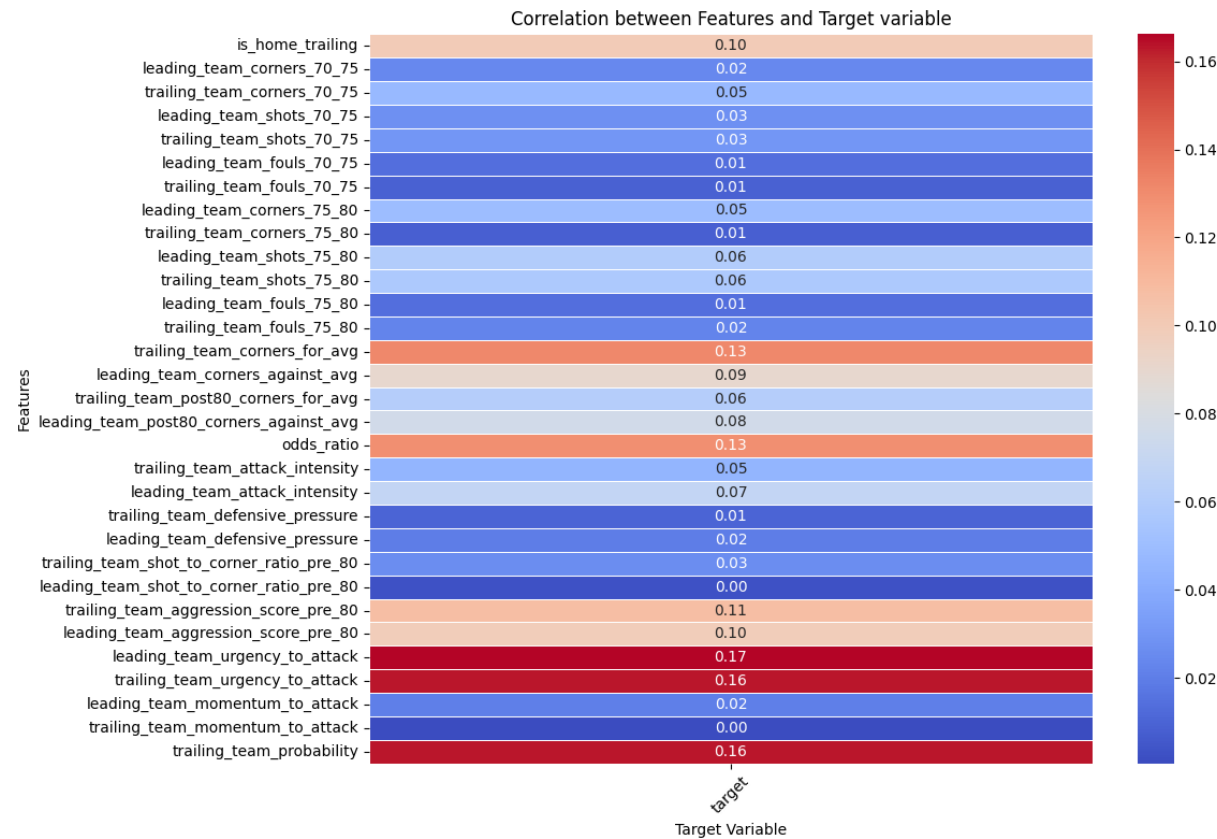
Constructed Features

trailing_team_corners_for_avg, leading_team_corners_against_avg, trailing_team_post80_corners_for_avg, leading_team_post80_corners_against_avg, odds_ratio, trailing_team_attack_intensity, leading_team_attack_intensity, trailing_team_defensive_pressure, leading_team_defensive_pressure, trailing_team_shot_to_corner_ratio_pre_80, leading_team_shot_to_corner_ratio_pre_80, trailing_team_aggression_score_pre_80, leading_team_aggression_score_pre_80, leading_team_urgency_to_attack, trailing_team_urgency_to_attack, leading_team_momentum_to_attack, trailing_team_momentum_to_attack, trailing_team_probability

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}
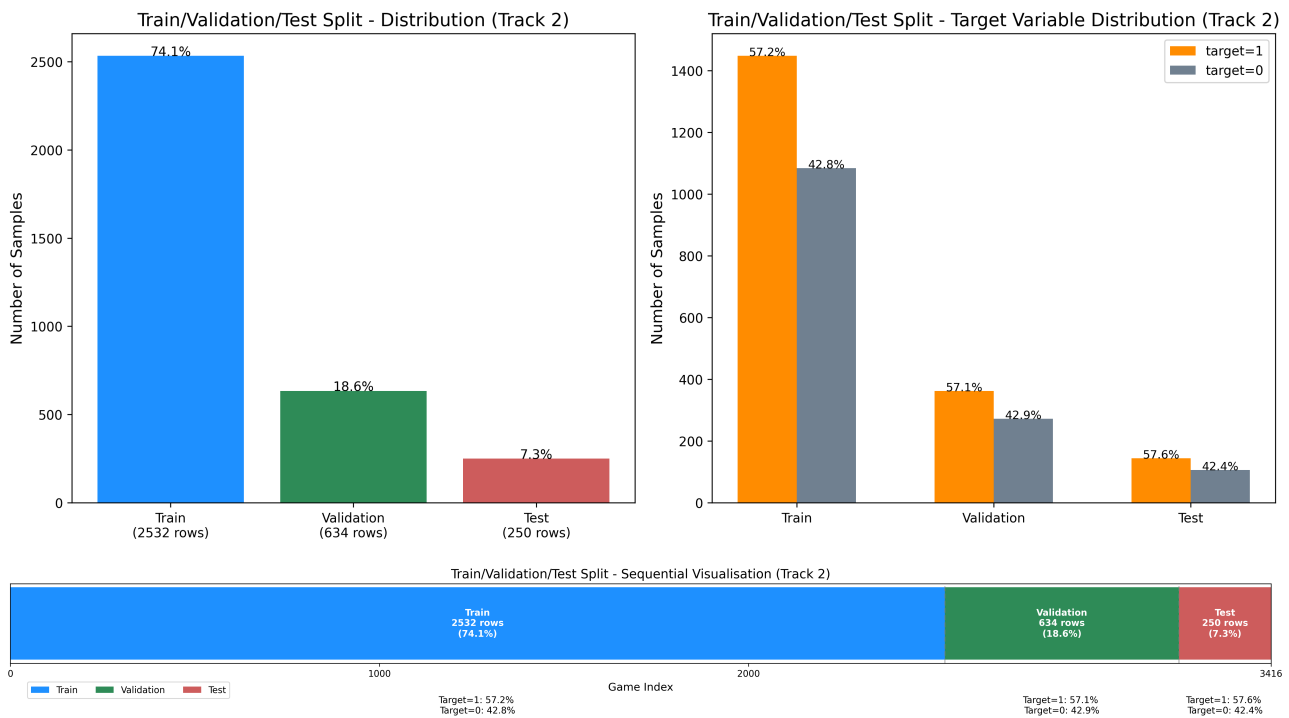
Feature Correlation

Correlation between Features and Target variable

Features	target
is_home_trailing	0.10
leading_team_corners_70_75	0.02
trailing_team_corners_70_75	0.05
leading_team_shots_70_75	0.03
trailing_team_shots_70_75	0.03
leading_team_fouls_70_75	0.01
trailing_team_fouls_70_75	0.01
leading_team_corners_75_80	0.05
trailing_team_corners_75_80	0.01
leading_team_shots_75_80	0.06
trailing_team_shots_75_80	0.06
leading_team_fouls_75_80	0.01
trailing_team_fouls_75_80	0.02
trailing_team_corners_for_avg	0.13
leading_team_corners_against_avg	0.09
trailing_team_post80_corners_for_avg	0.06
leading_team_post80_corners_against_avg	0.08
odds_ratio	0.13
trailing_team_attack_intensity	0.05
leading_team_attack_intensity	0.07
trailing_team_defensive_pressure	0.01
leading_team_defensive_pressure	0.02
trailing_team_shot_to_corner_ratio_pre_80	0.03
leading_team_shot_to_corner_ratio_pre_80	0.00
trailing_team_aggression_score_pre_80	0.11
leading_team_aggression_score_pre_80	0.10
leading_team_urgency_to_attack	0.17
trailing_team_urgency_to_attack	0.16
leading_team_momentum_to_attack	0.02
trailing_team_momentum_to_attack	0.00
trailing_team_probability	0.16

Target Variable

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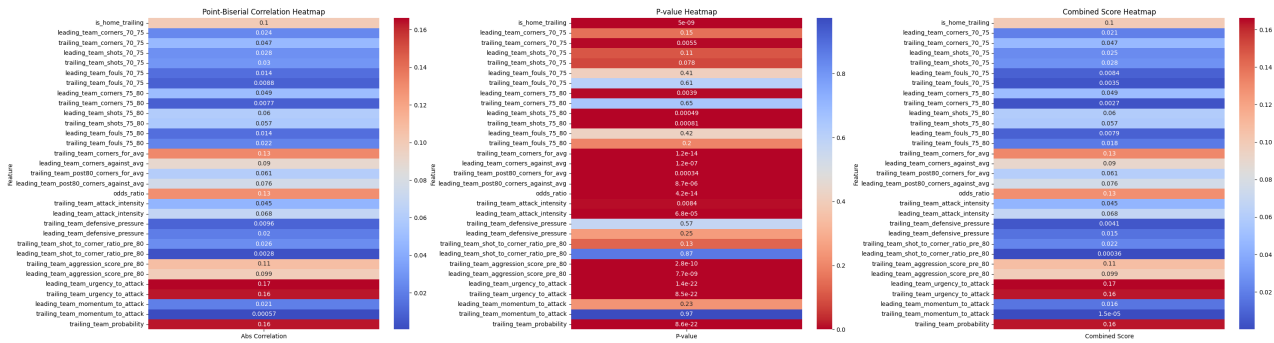
Dataset Train/Validation/Test Split Visualisation

Train/Validation/Test Split - Distribution (Track 2)

Train/Validation/Test Split - Target Variable Distribution (Track 2)

Train/Validation/Test Split - Sequential Visualisation (Track 2)

random_forest Model

Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
trailing_team_corners_for_avg	0.06732	leading_team_corners_75_80	0.00710
leading_team_post80_corners_against_avg	0.06265	leading_team_corners_70_75	0.00718
trailing_team_post80_corners_for_avg	0.06215	trailing_team_corners_70_75	0.00830
trailing_team_aggression_score_pre_80	0.06085	trailing_team_corners_75_80	0.00854
leading_team_corners_against_avg	0.06027	is_home_trailing	0.00854
leading_team_aggression_score_pre_80	0.05839		
odds_ratio	0.05619		
trailing_team_shot_to_corner_ratio_pre_80	0.05465		

Optimal Threshold

- **Optimal threshold:** 0.64
- **Expected Precision:** 0.650
- **Expected Recall:** 0.323

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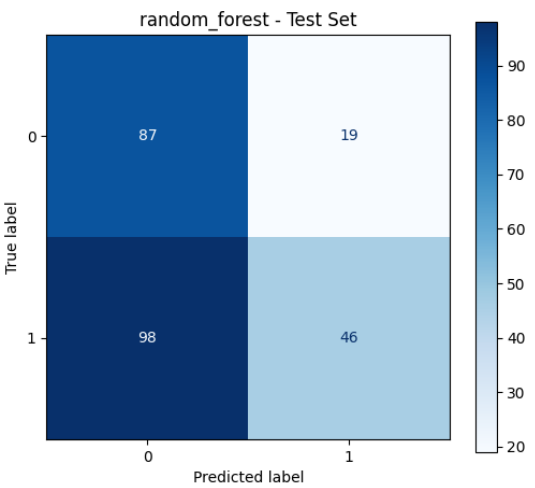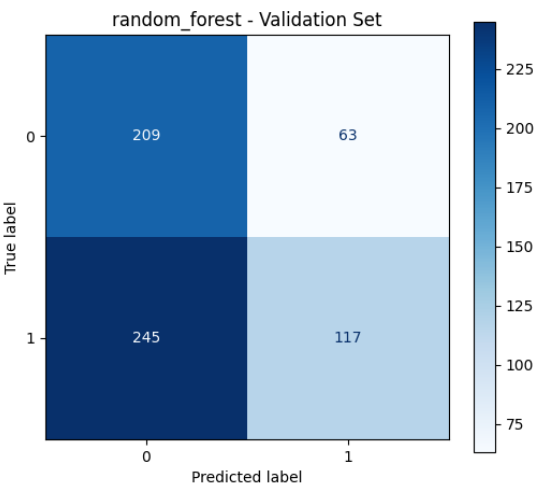
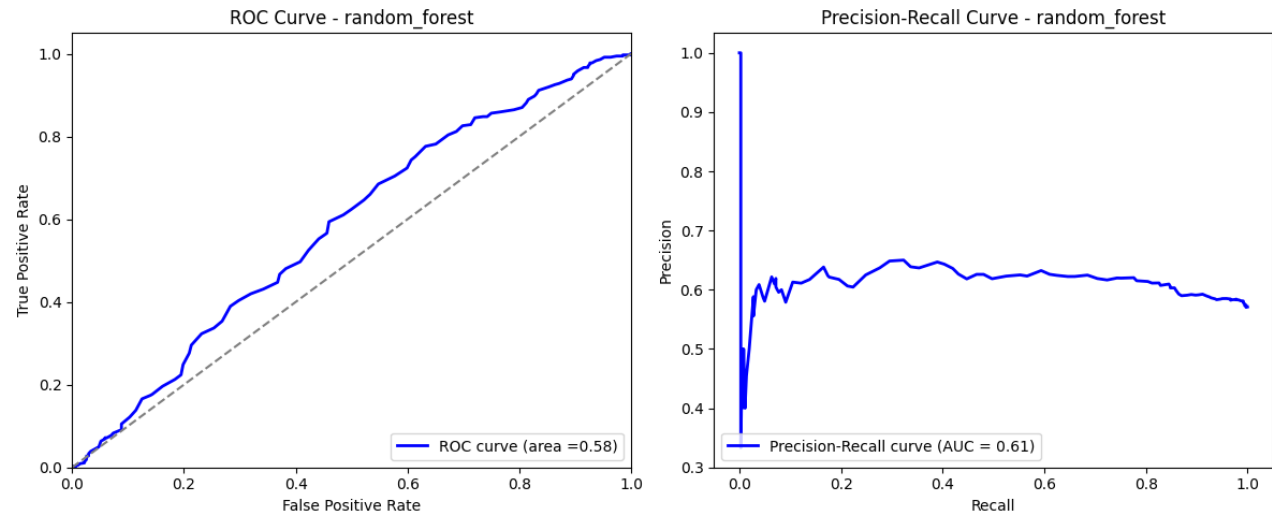## Classification Report (*Validation Set*)

```
          precision    recall  f1-score   support

      0        0.46      0.77      0.58       272
      1        0.65      0.32      0.43       362

accuracy                          0.51       634
```

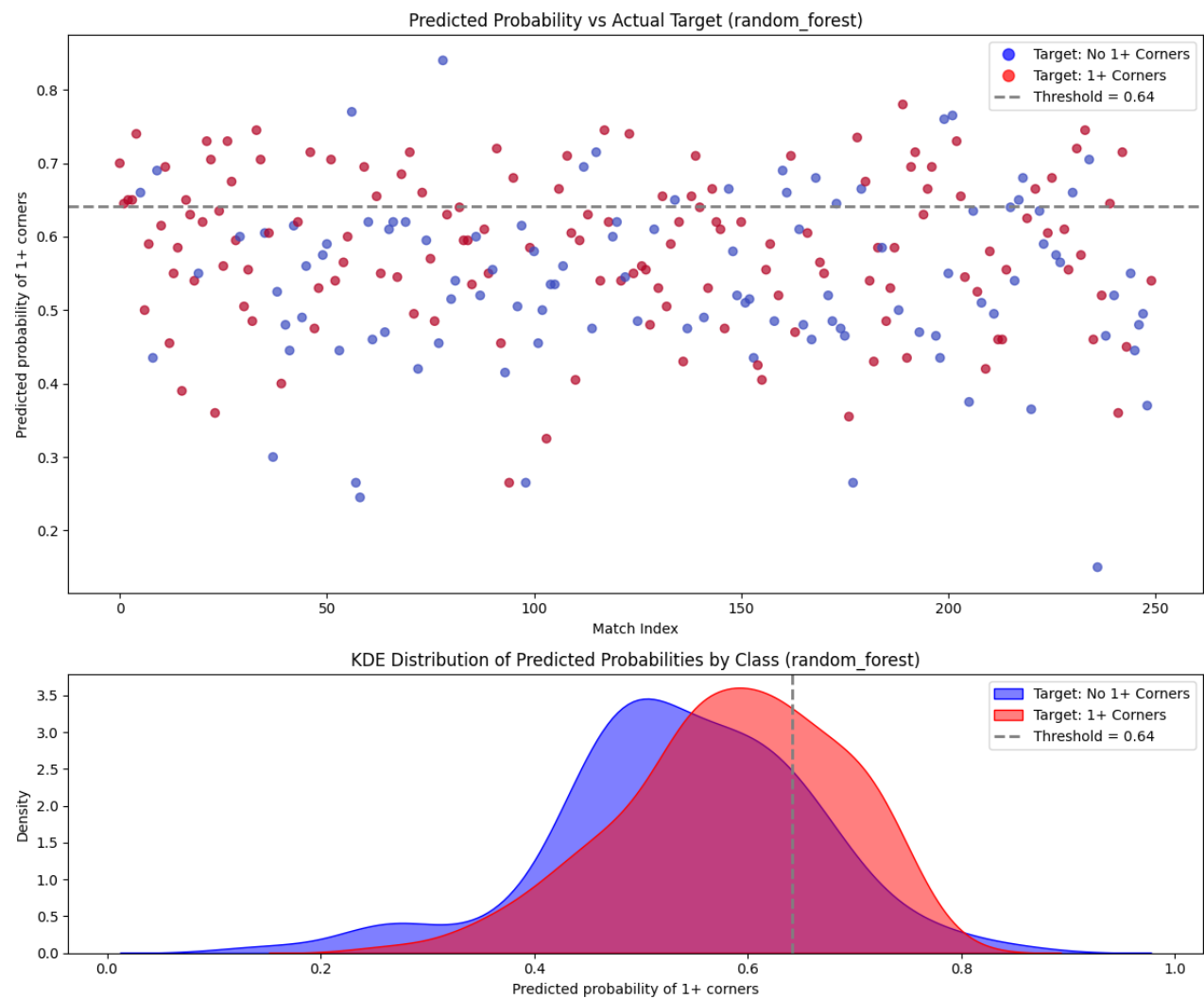macro avg 0.56 0.55 0.50 634 weighted avg 0.57 0.51 0.49 634

Classification Report (*Test Set*)

```
          precision    recall  f1-score   support

      0        0.47      0.82      0.60       106
      1        0.71      0.32      0.44       144

accuracy                          0.53       250
```

macro avg 0.59 0.57 0.52 250 weighted avg 0.61 0.53 0.51 250

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

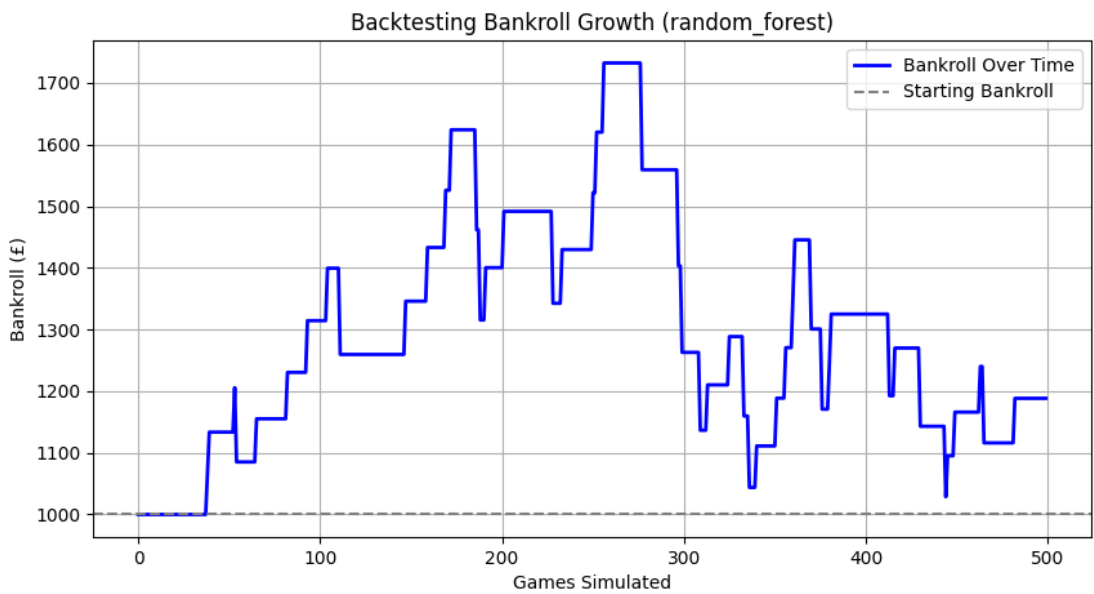💰 **Final Bankroll:** £1188.41

💸 **Total Staked:** £6182.46

📈 **Total Profit:** £188.41

📊 **ROI:** 18.84%

✅ **Win rate:** 64.58% over 48 bets

📉 **Sharpe Ratio:** 0.09

🎯 **Edge Over Bookies:** 4.39%

Backtesting Bankroll Growth (random_forest)

logistic_regression Model

Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
leading_team_aggression_score_pre_80	1.09959	trailing_team_aggression_score_pre_80	-1.09408
leading_team_post80_corners_against_avg	1.08427	leading_team_urgency_to_attack	-0.65559
trailing_team_corners_for_avg	0.82142	leading_team_corners_against_avg	-0.64684
trailing_team_fouls_75_80	0.45156	odds_ratio	-0.61816
trailing_team_fouls_70_75	0.43620	trailing_team_corners_75_80	-0.52236
trailing_team_corners_70_75	0.28272		
trailing_team_shots_75_80	0.27866		
trailing_team_momentum_to_attack	0.20384		

Optimal Threshold

- **Optimal threshold:** 0.64
- **Expected Precision:** 0.741
- **Expected Recall:** 0.119

Classification Reports
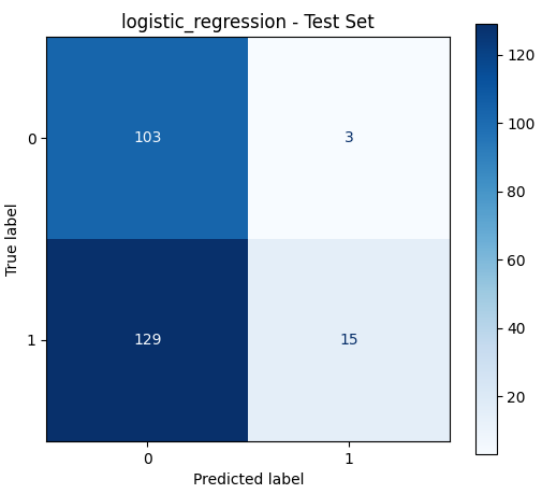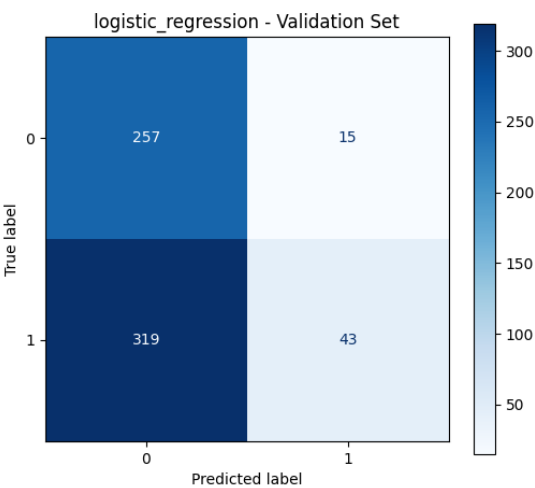
Classification Report (*Validation Set*)

```
            precision    recall  f1-score   support

        0        0.45      0.94      0.61       272
        1        0.74      0.12      0.20       362

 accuracy                           0.47       634
```
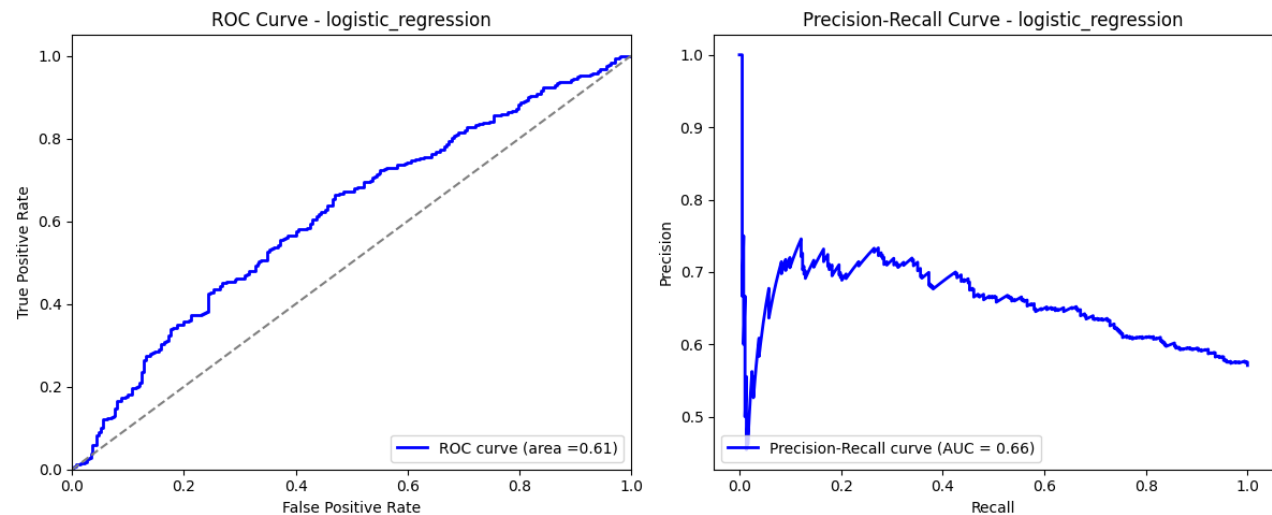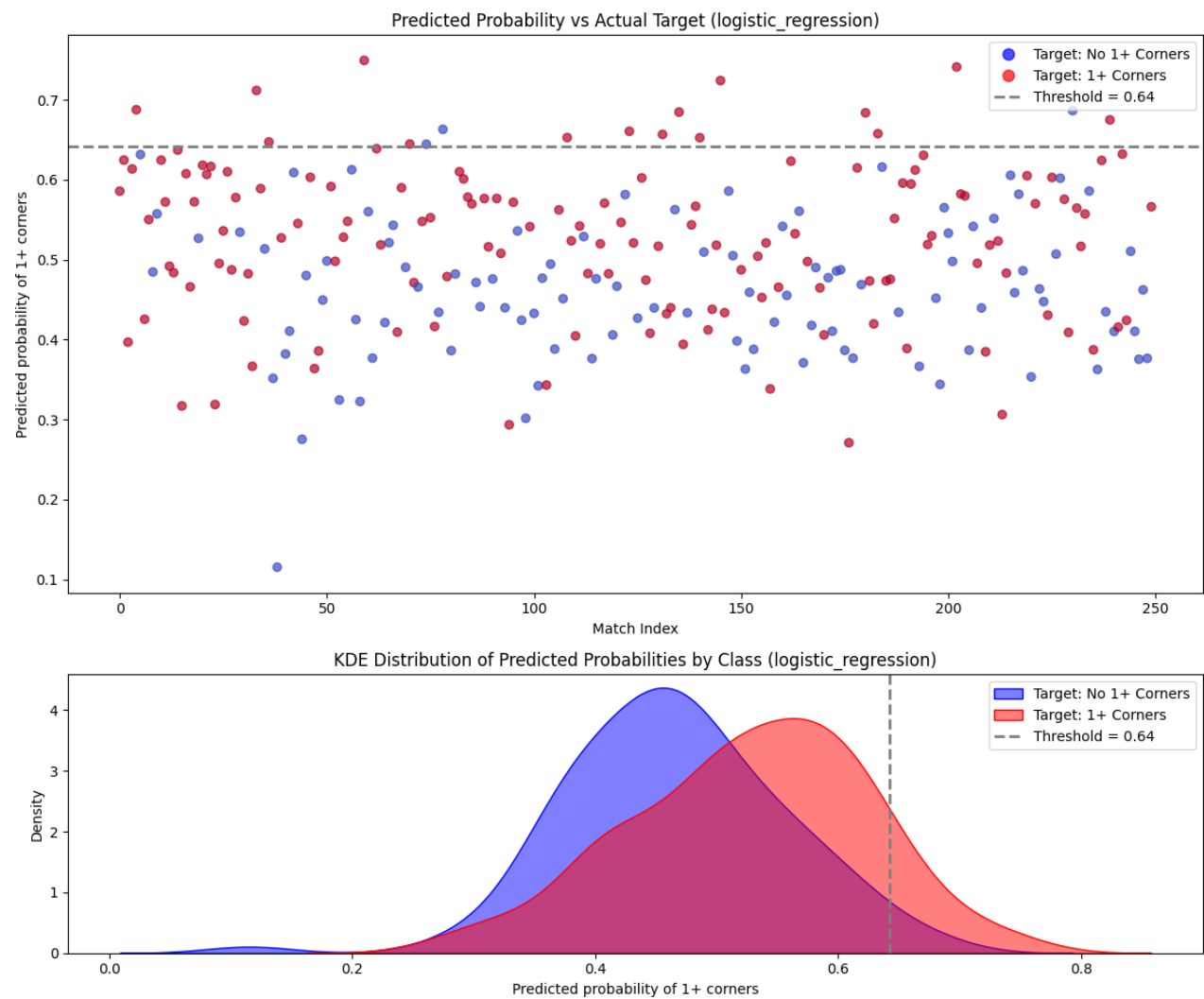
macro avg 0.59 0.53 0.41 634 weighted avg 0.61 0.47 0.38 634

Classification Report (*Test Set*)

```
            precision    recall  f1-score   support

        0        0.44      0.97      0.61       106
        1        0.83      0.10      0.19       144

 accuracy                           0.47       250
```

macro avg 0.64 0.54 0.40 250 weighted avg 0.67 0.47 0.37 250

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

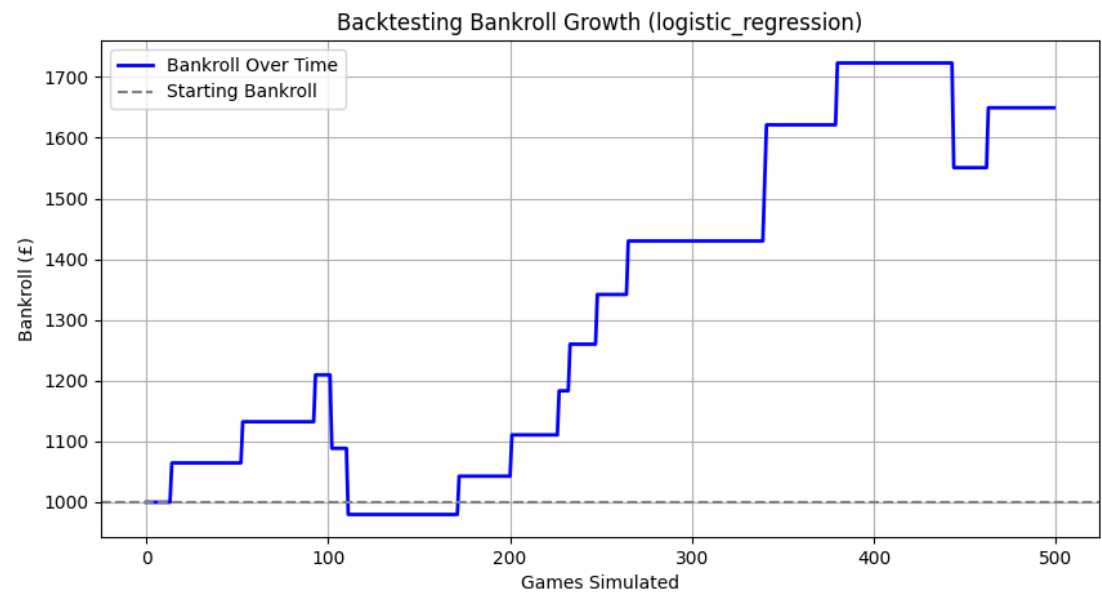💰 **Final Bankroll:** £1649.30

💸 **Total Staked:** £2026.13

📈 **Total Profit:** £649.30

📊 **ROI:** 64.93%

✅ **Win rate:** 81.25% over 16 bets

📉 **Sharpe Ratio:** 0.51

🎯 **Edge Over Bookies:** 21.06%

Backtesting Bankroll Growth (logistic_regression)

svc Model

Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
is_home_trailing	0.00000	trailing_team_probability	0.00000
leading_team_post80_corners_against_avg	0.00000	trailing_team_corners_70_75	0.00000
trailing_team_momentum_to_attack	0.00000	leading_team_shots_70_75	0.00000
leading_team_momentum_to_attack	0.00000	trailing_team_shots_70_75	0.00000
trailing_team_urgency_to_attack	0.00000	leading_team_fouls_70_75	0.00000
leading_team_urgency_to_attack	0.00000		
leading_team_aggression_score_pre_80	0.00000		
trailing_team_aggression_score_pre_80	0.00000		

Optimal Threshold

- **Optimal threshold:** 0.67
- **Expected Precision:** 0.688
- **Expected Recall:** 0.182

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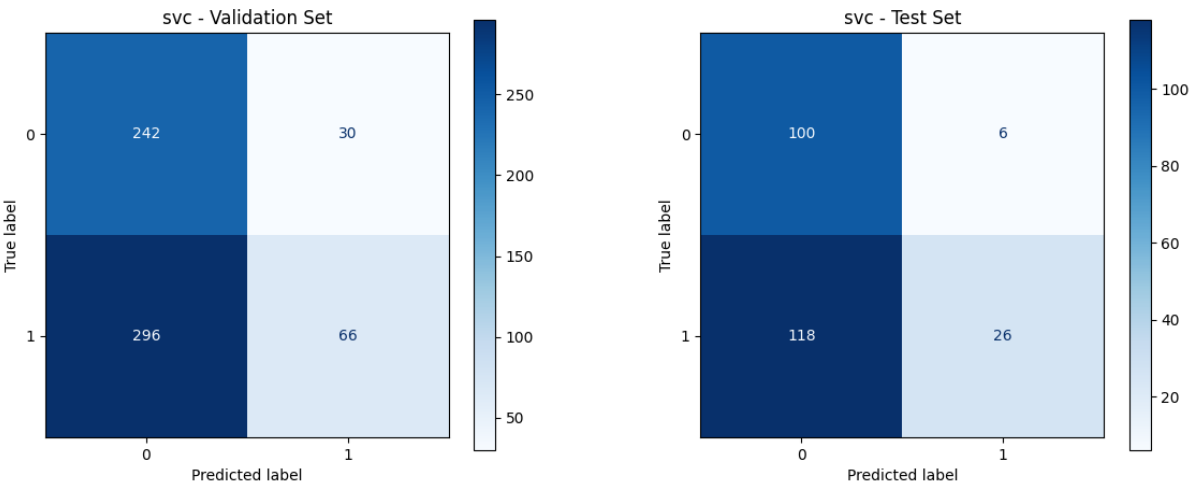
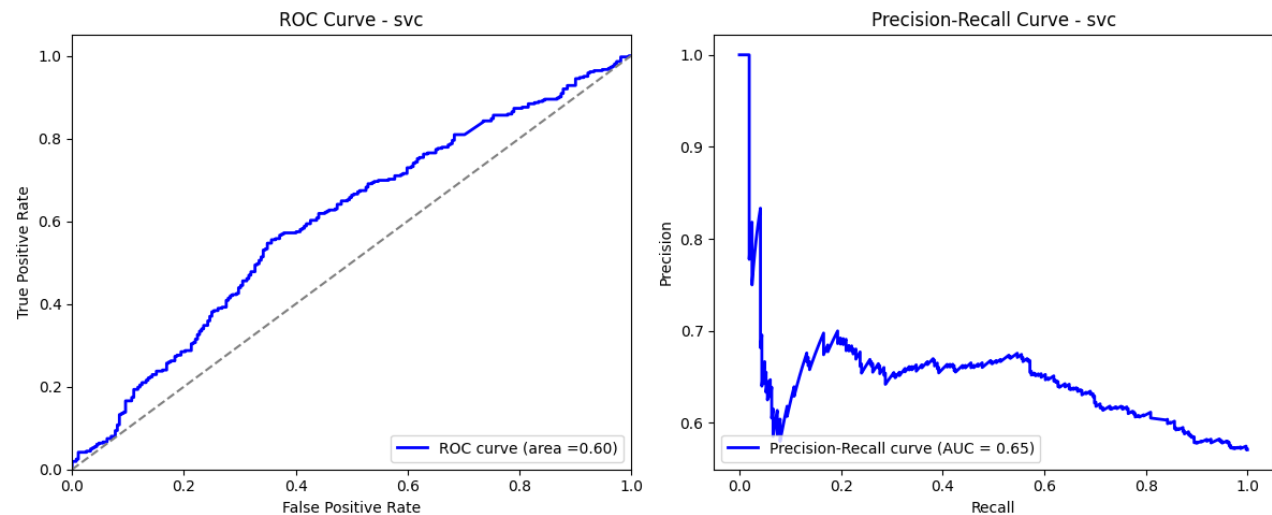## Classification Report (*Validation Set*)

```
          precision    recall  f1-score   support

    0        0.45        0.89      0.60        272
    1        0.69        0.18      0.29        362

accuracy                          0.49        634
```

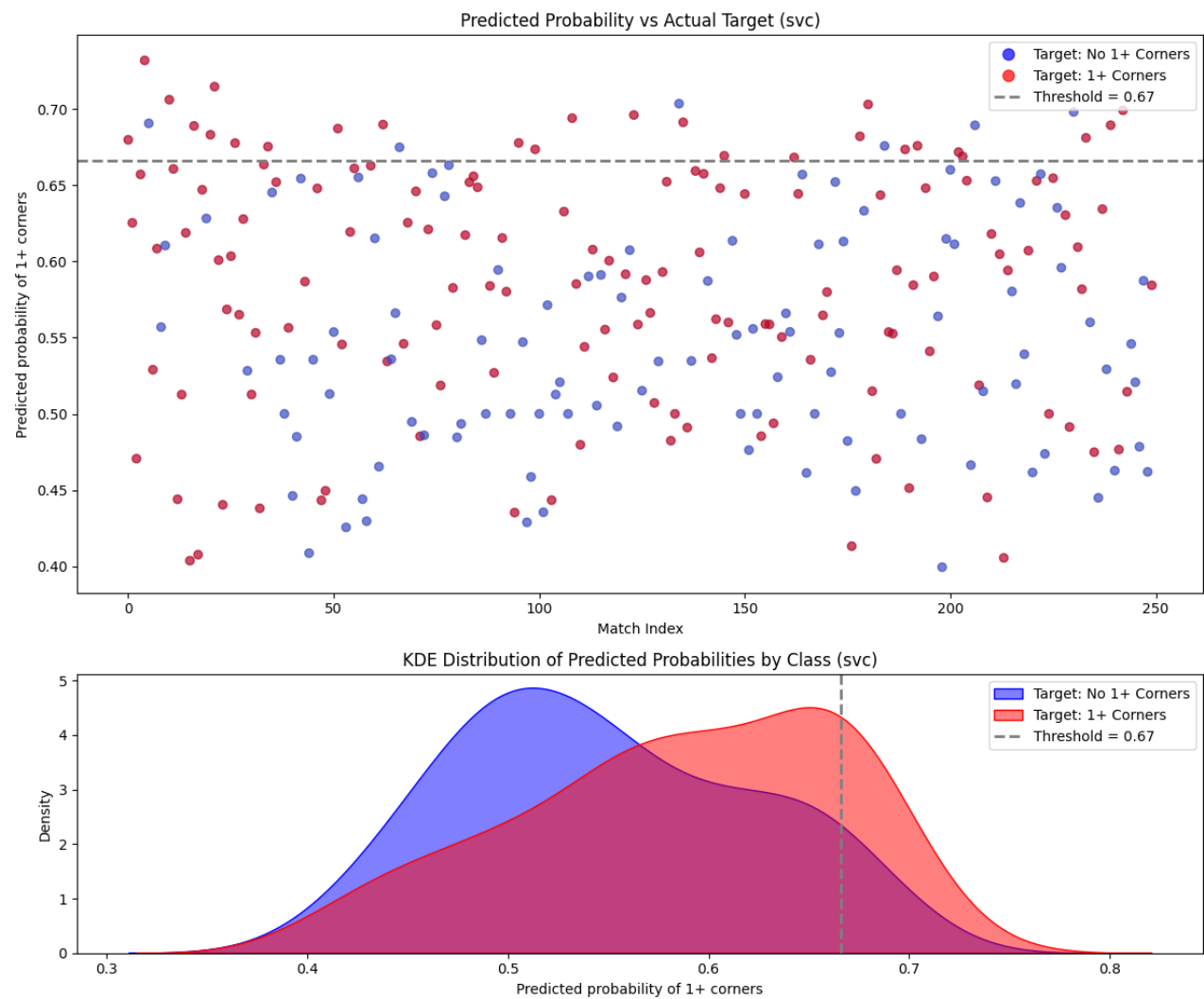macro avg 0.57 0.54 0.44 634 weighted avg 0.59 0.49 0.42 634

Classification Report (*Test Set*)

```
          precision    recall  f1-score   support

    0        0.46        0.94      0.62        106
    1        0.81        0.18      0.30        144

accuracy                          0.50        250
```

macro avg 0.64 0.56 0.46 250 weighted avg 0.66 0.50 0.43 250

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:** £1919.15

💸 **Total Staked:** £2938.72

📈 **Total Profit:** £919.15

📊 **ROI:** 91.92%

✅ **Win rate:** 80.95% over 21 bets

📉 **Sharpe Ratio:** 0.51

🎯 **Edge Over Bookies:** 20.76%

Backtesting Bankroll Growth (svc)
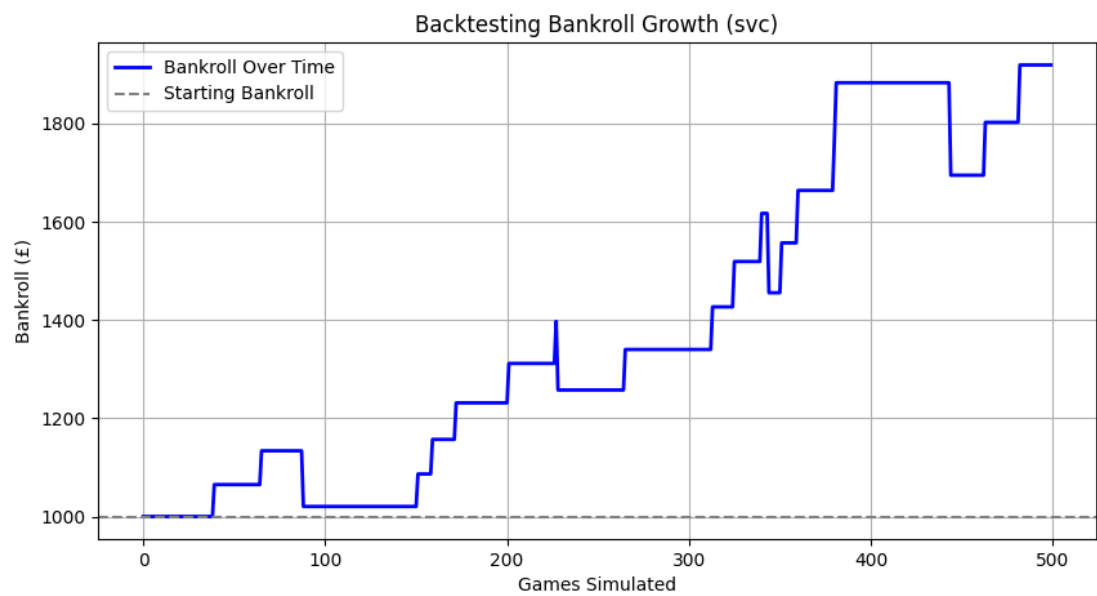
xgboost Model

Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
trailing_team_probability	0.19660	leading_team_shots_75_80	0.00000
leading_team_urgency_to_attack	0.04253	leading_team_corners_70_75	0.00000
trailing_team_corners_75_80	0.04011	trailing_team_corners_70_75	0.01749
trailing_team_urgency_to_attack	0.03927	leading_team_corners_against_avg	0.02030
odds_ratio	0.03723	trailing_team_shots_75_80	0.02088
is_home_trailing	0.03482		
trailing_team_aggression_score_pre_80	0.03443		
leading_team_momentum_to_attack	0.03407		

Optimal Threshold

- **Optimal threshold:** 0.69
- **Expected Precision:** 0.756
- **Expected Recall:** 0.163

Classification Reports
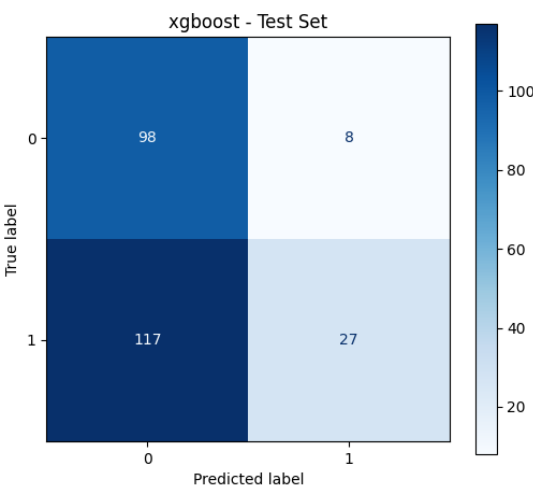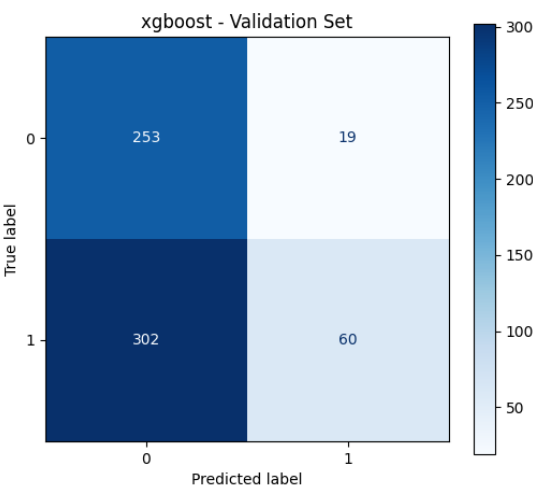
Classification Report (*Validation Set*)

```
              precision    recall  f1-score   support

           0       0.46      0.93      0.61       272
           1       0.76      0.17      0.27       362

    accuracy                           0.49       634
```
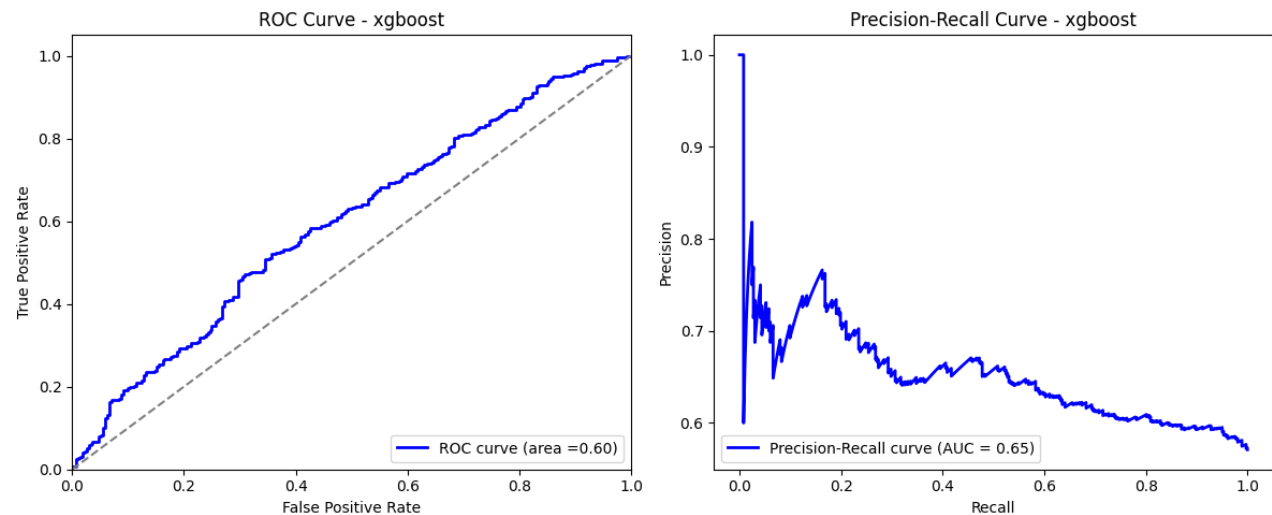
macro avg 0.61 0.55 0.44 634 weighted avg 0.63 0.49 0.42 634

Classification Report (*Test Set*)

```
              precision    recall  f1-score   support

           0       0.46      0.92      0.61       106
           1       0.77      0.19      0.30       144

    accuracy                           0.50       250
```
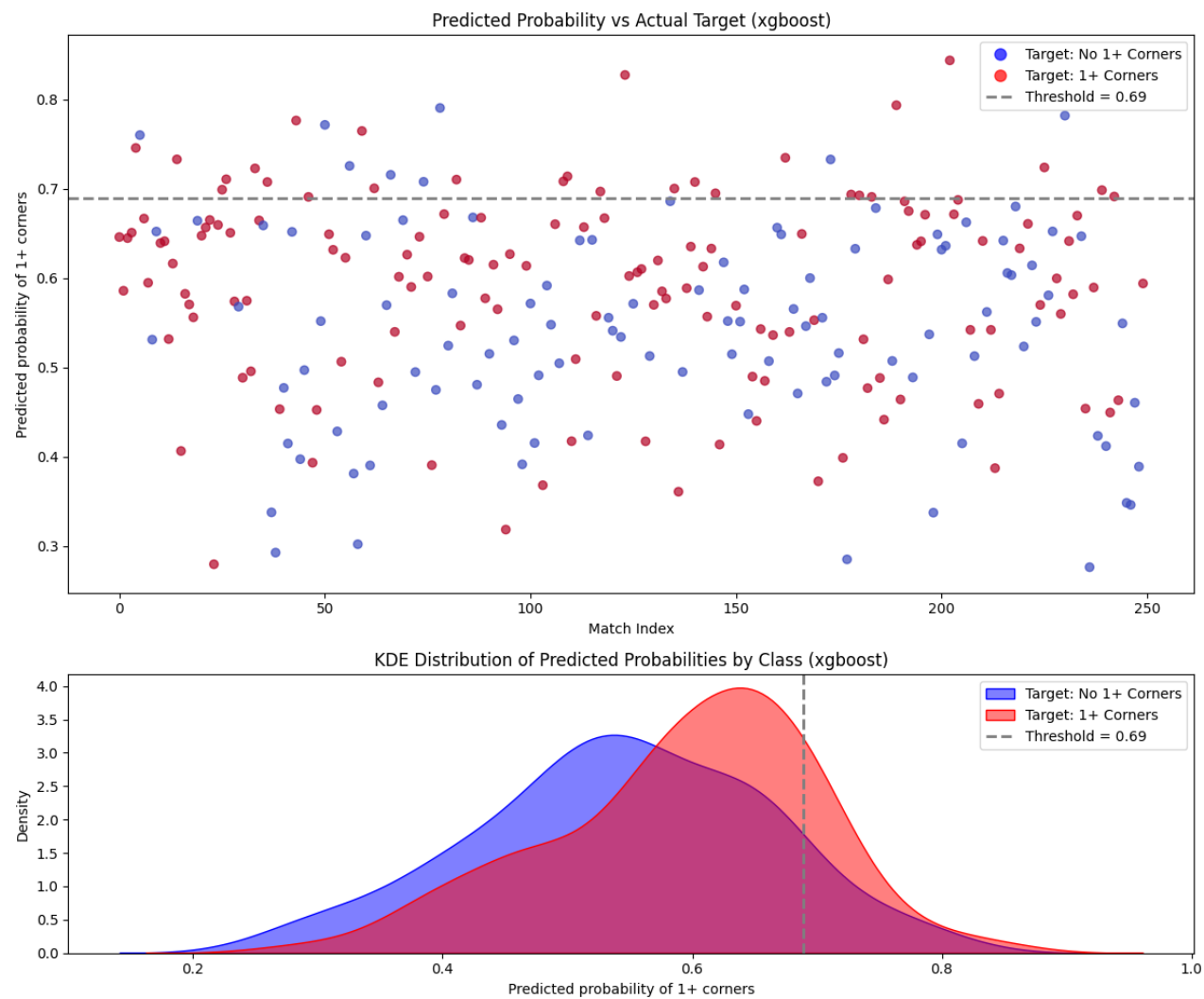
macro avg 0.61 0.56 0.46 250 weighted avg 0.64 0.50 0.43 250

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:** £1909.61

💸 **Total Staked:** £3582.14
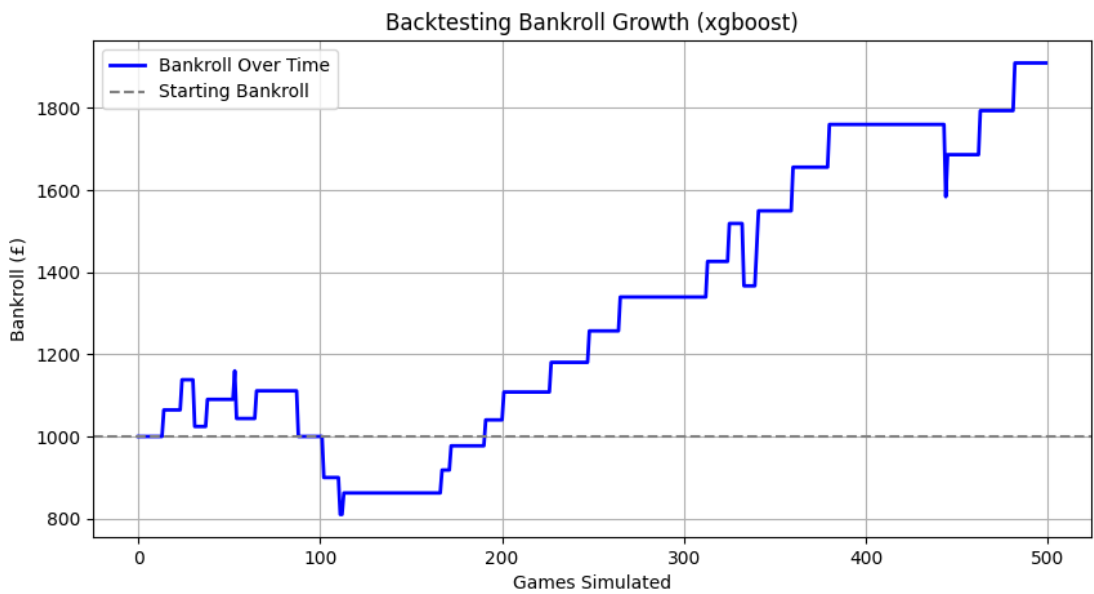
📈 **Total Profit:** £909.61

📊 **ROI:** 90.96%

✅ **Win rate:** 75.86% over 29 bets

📉 **Sharpe Ratio:** 0.35

🎯 **Edge Over Bookies:** 15.67%

Backtesting Results Summary

Backtesting Bankroll Growth – Track Comparison