

# Model Training Report: Track 2

---

## Configuration Details

---

### Target Variable

target

### Selected Features

is\_home\_trailing, trailing\_team\_shots\_pre80, trailing\_team\_fouls\_pre80, trailing\_team\_corners\_pre80, trailing\_team\_yellow\_cards\_pre80, trailing\_team\_sending\_off\_pre80, leading\_team\_corners\_70\_75, trailing\_team\_corners\_70\_75, leading\_team\_shots\_70\_75, trailing\_team\_shots\_70\_75, leading\_team\_corners\_75\_80, trailing\_team\_corners\_75\_80, leading\_team\_shots\_75\_80, trailing\_team\_shots\_75\_80

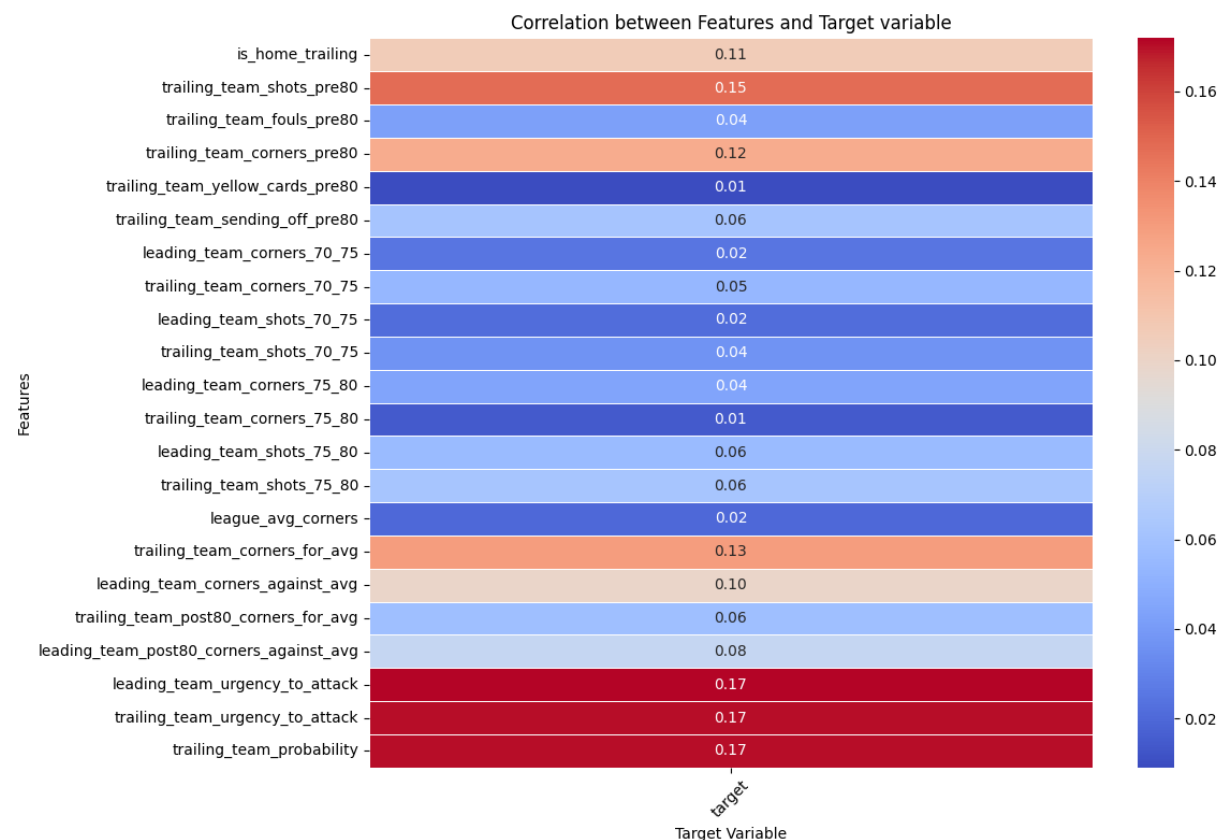
### Constructed Features

league\_avg\_corners, trailing\_team\_corners\_for\_avg, leading\_team\_corners\_against\_avg, trailing\_team\_post80\_corners\_for\_avg, leading\_team\_post80\_corners\_against\_avg, leading\_team\_urgency\_to\_attack, trailing\_team\_urgency\_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}

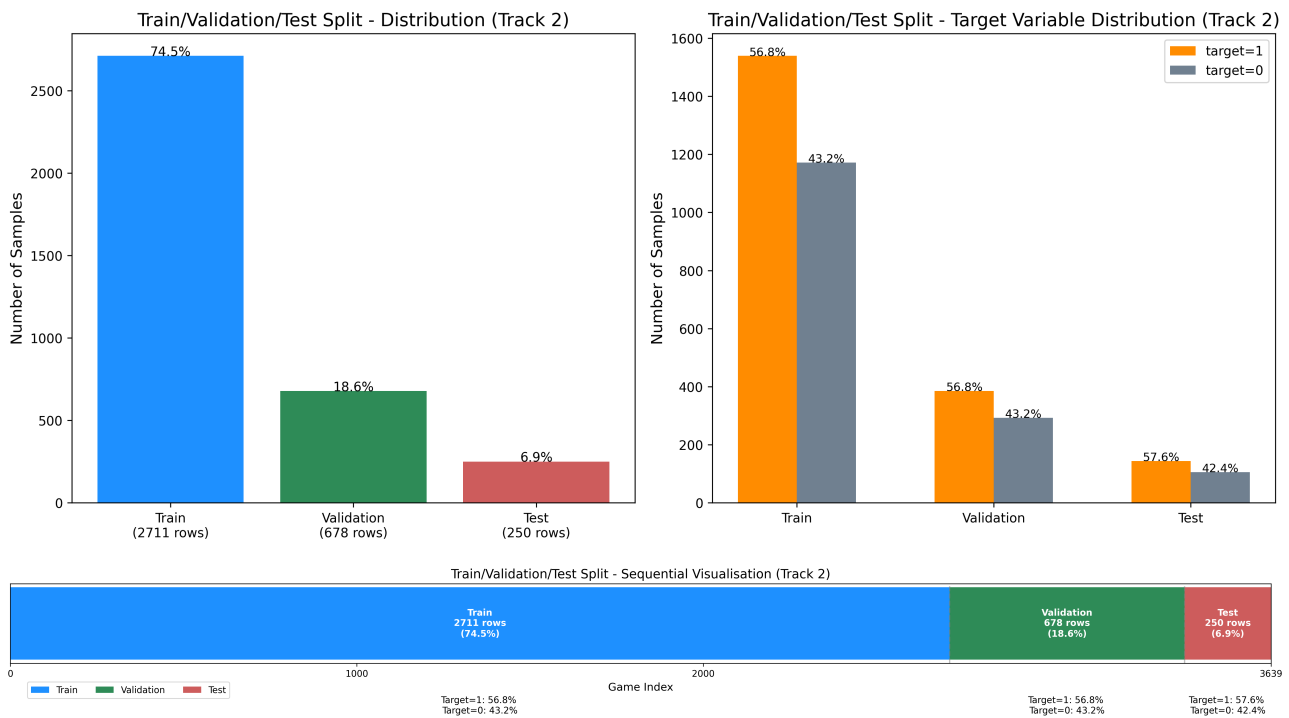
# Feature Correlation



# Point-Biserial Correlation



# Dataset Train/Validation/Test Split Visualisation



# random\_forest Model

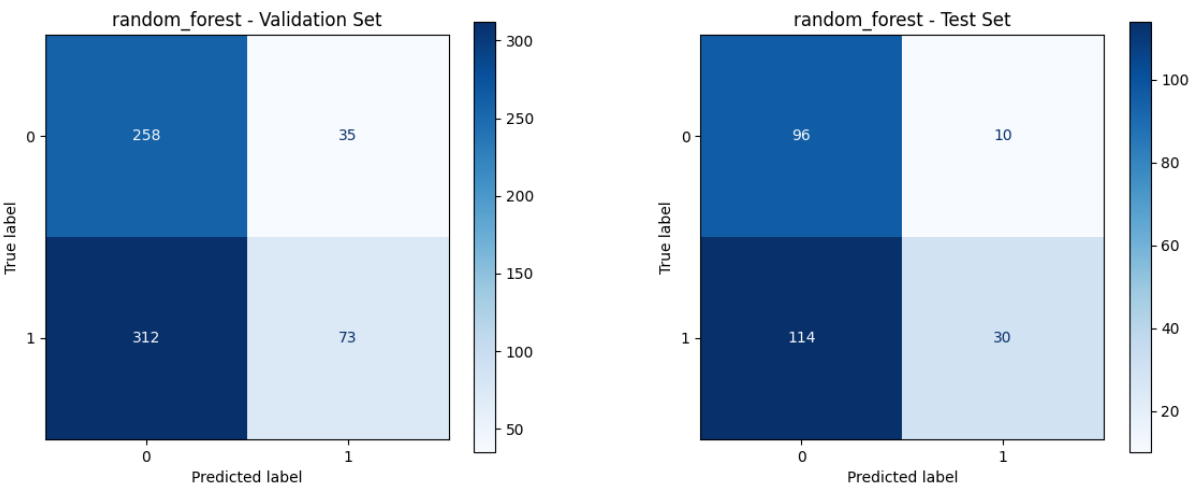
## Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
trailing_team_corners_for_avg	0.09211	trailing_team_sending_off_pre80	0.01062
leading_team_post80_corners_against_avg	0.08799	is_home_trailing	0.01088
leading_team_corners_against_avg	0.08687	leading_team_corners_70_75	0.01264
trailing_team_post80_corners_for_avg	0.08428	leading_team_corners_75_80	0.01284
leading_team_urgency_to_attack	0.08108	trailing_team_corners_70_75	0.01294
trailing_team_probability	0.07809		
trailing_team_urgency_to_attack	0.07716		
trailing_team_shots_pre80	0.06514		

## Optimal Threshold

- **Optimal threshold:** 0.69
- **Expected Precision:** 0.676
- **Expected Recall:** 0.190

## Classification Reports



---

### Classification Report (*Validation Set*)

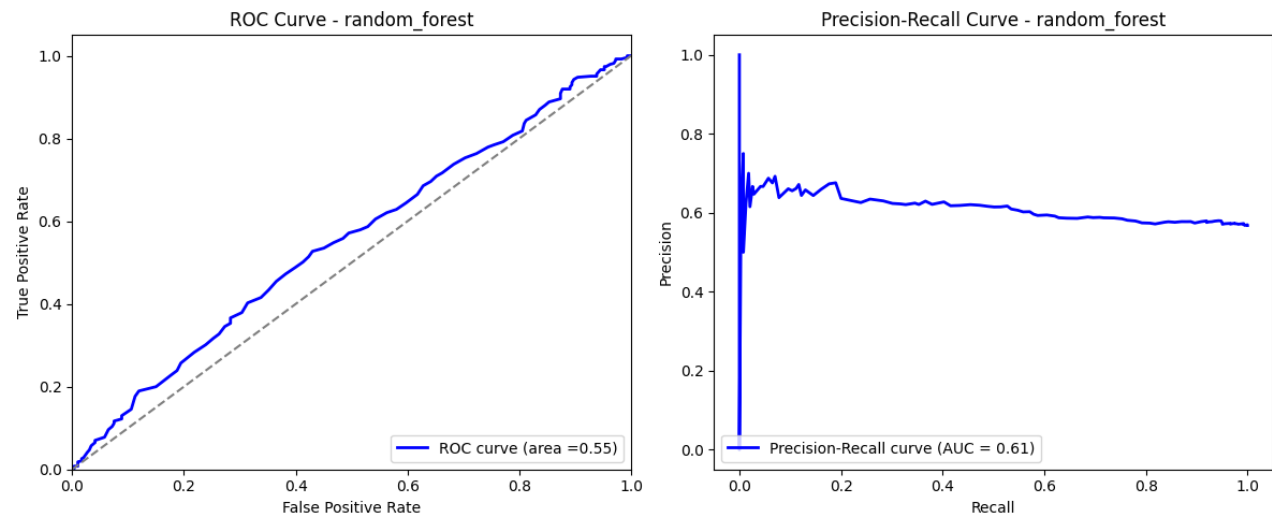
	precision	recall	f1-score	support
0	0.45	0.88	0.60	293
1	0.68	0.19	0.30	385
accuracy			0.49	678
macro avg	0.56	0.54	0.45	678
weighted avg	0.58	0.49	0.43	678

### Classification Report (*Test Set*)

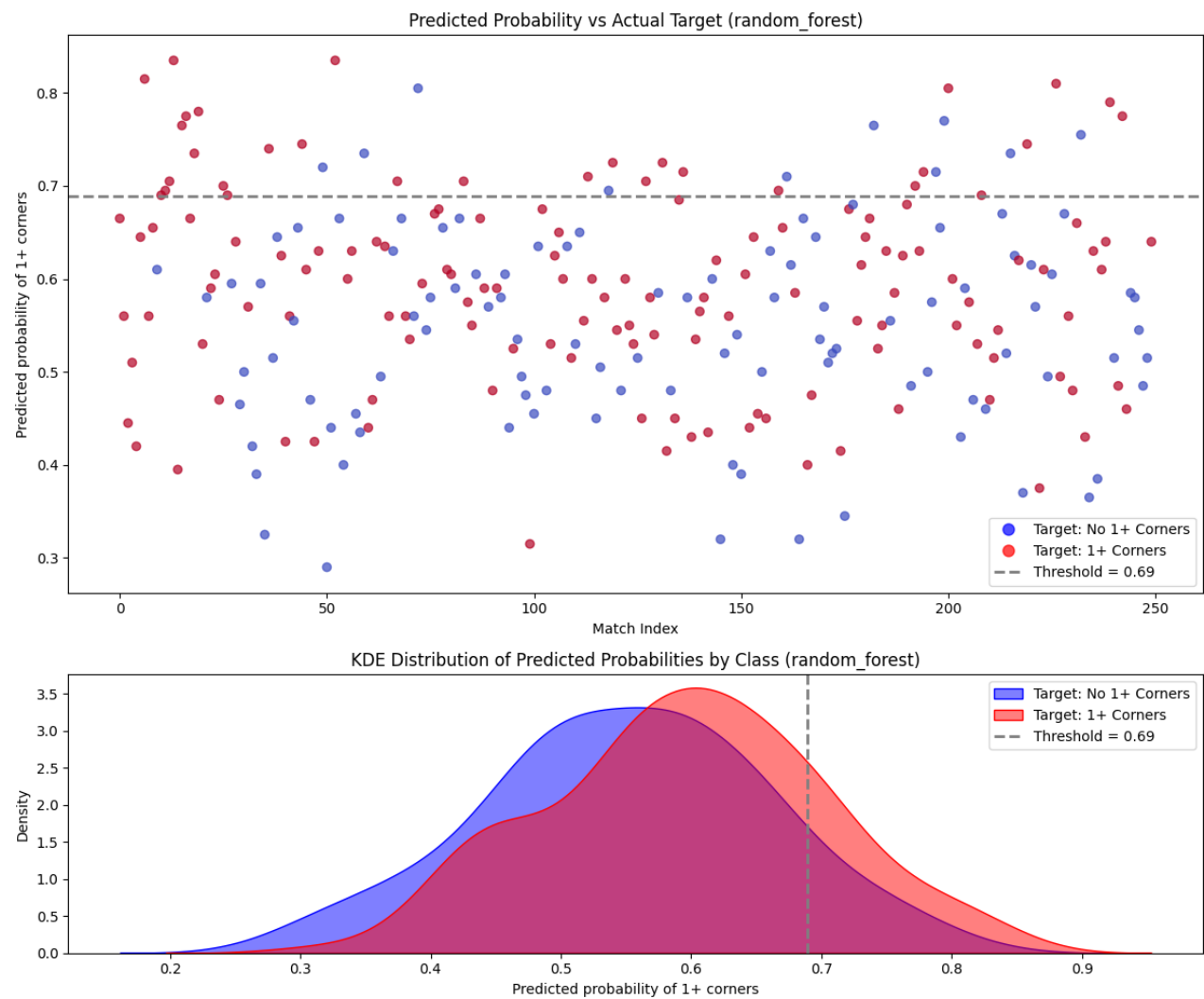
	precision	recall	f1-score	support
0	0.46	0.91	0.61	106
1	0.75	0.21	0.33	144
accuracy			0.50	250
macro avg	0.60	0.56	0.47	250
weighted avg	0.63	0.50	0.45	250

---

### ROC and Precision-Recall Curves



# Prediction Distribution Graphs



## Backtesting Results

--- Overall Summary ---

🏠 Initial Bankroll: £1000.00

💰 Final Bankroll: £1080.75

📊 Total Staked: £2978.22

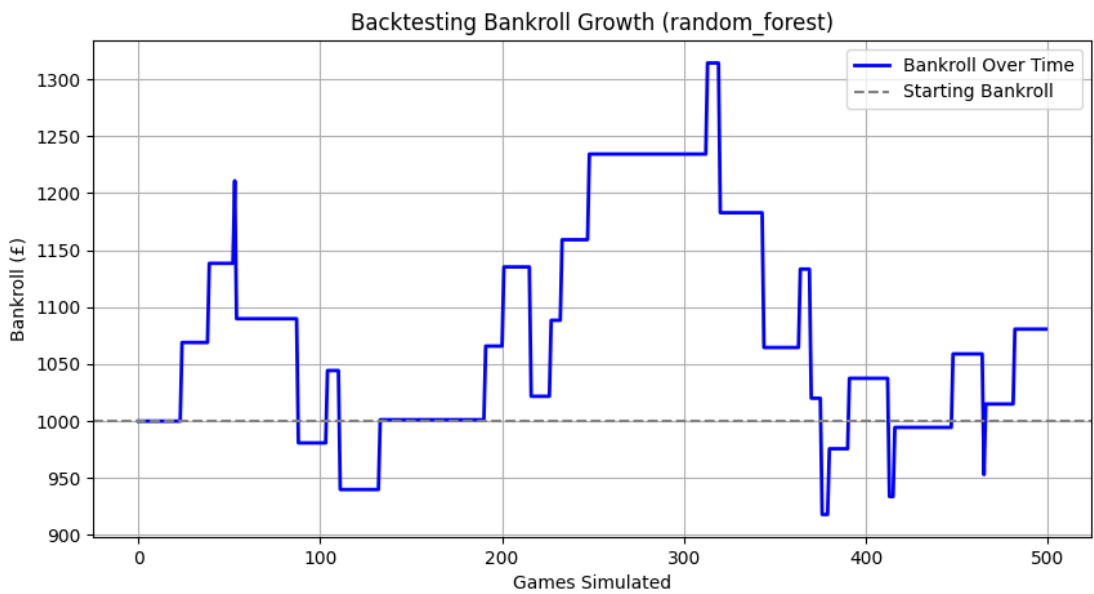
📈 Total Profit: £80.75

🇮🇹 ROI: 8.07%

✅ Win rate: 64.29% over 28 bets

📉 Sharpe Ratio: 0.07

🎯 Edge Over Bookies: 4.10%



# logistic\_regression Model

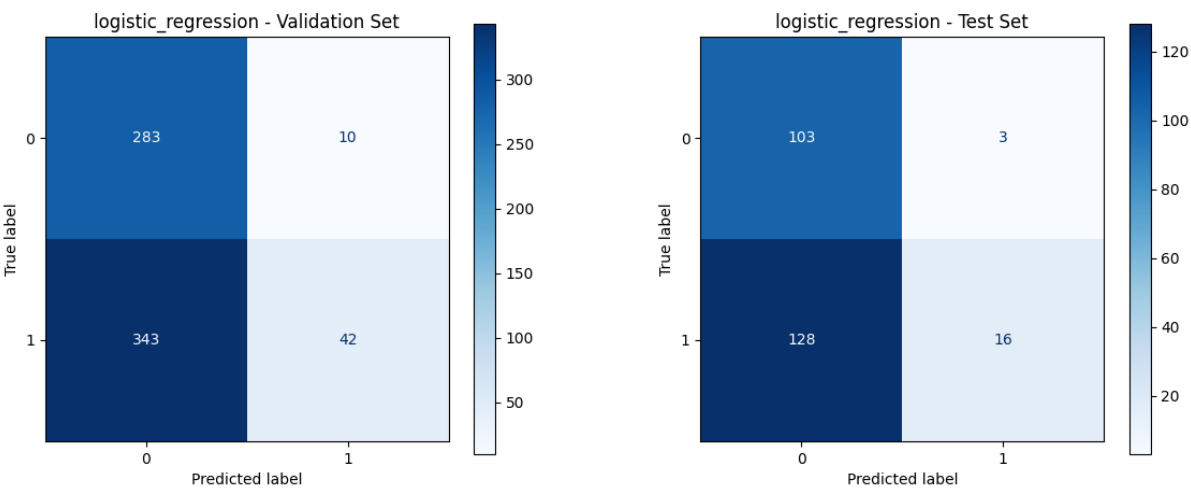
## Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
leading_team_post80_corners_against_avg	0.99292	leading_team_urgency_to_attack	-0.93857
trailing_team_corners_for_avg	0.92576	trailing_team_sending_off_pre80	-0.89286
trailing_team_shots_pre80	0.92378	leading_team_shots_75_80	-0.54443
trailing_team_corners_pre80	0.38503	trailing_team_fouls_pre80	-0.43672
trailing_team_yellow_cards_pre80	0.33106	trailing_team_corners_75_80	-0.31921
trailing_team_corners_70_75	0.21730		
trailing_team_shots_75_80	0.19720		
leading_team_shots_70_75	0.05570		

## Optimal Threshold

- **Optimal threshold:** 0.67
- **Expected Precision:** 0.811
- **Expected Recall:** 0.112

## Classification Reports





---

### Classification Report (*Validation Set*)

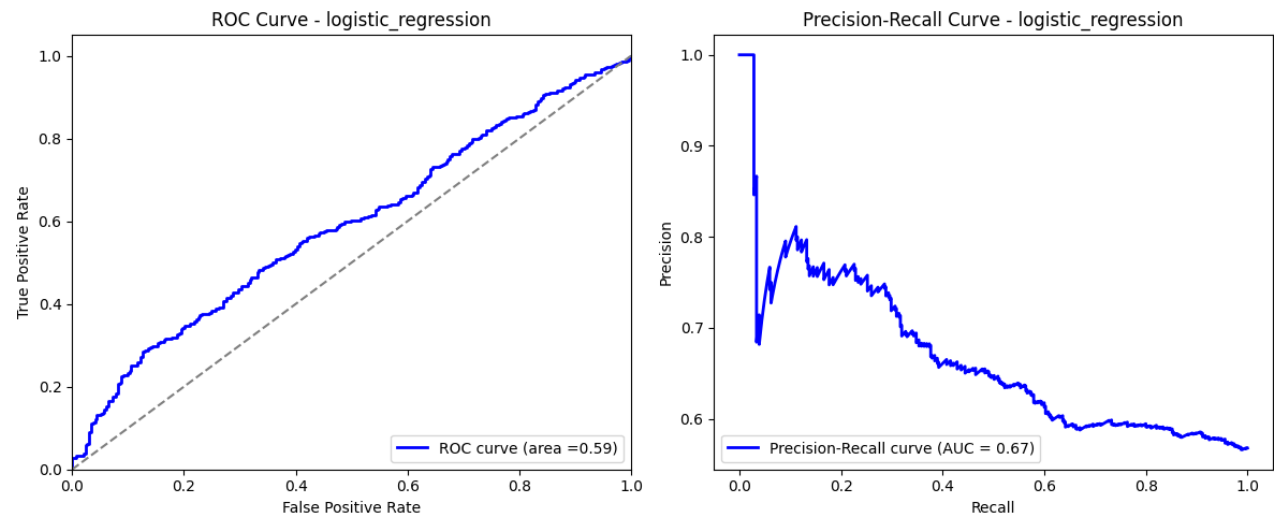
	precision	recall	f1-score	support
0	0.45	0.97	0.62	293
1	0.81	0.11	0.19	385
accuracy			0.48	678
macro avg	0.63	0.54	0.40	678
weighted avg	0.65	0.48	0.38	678

### Classification Report (*Test Set*)

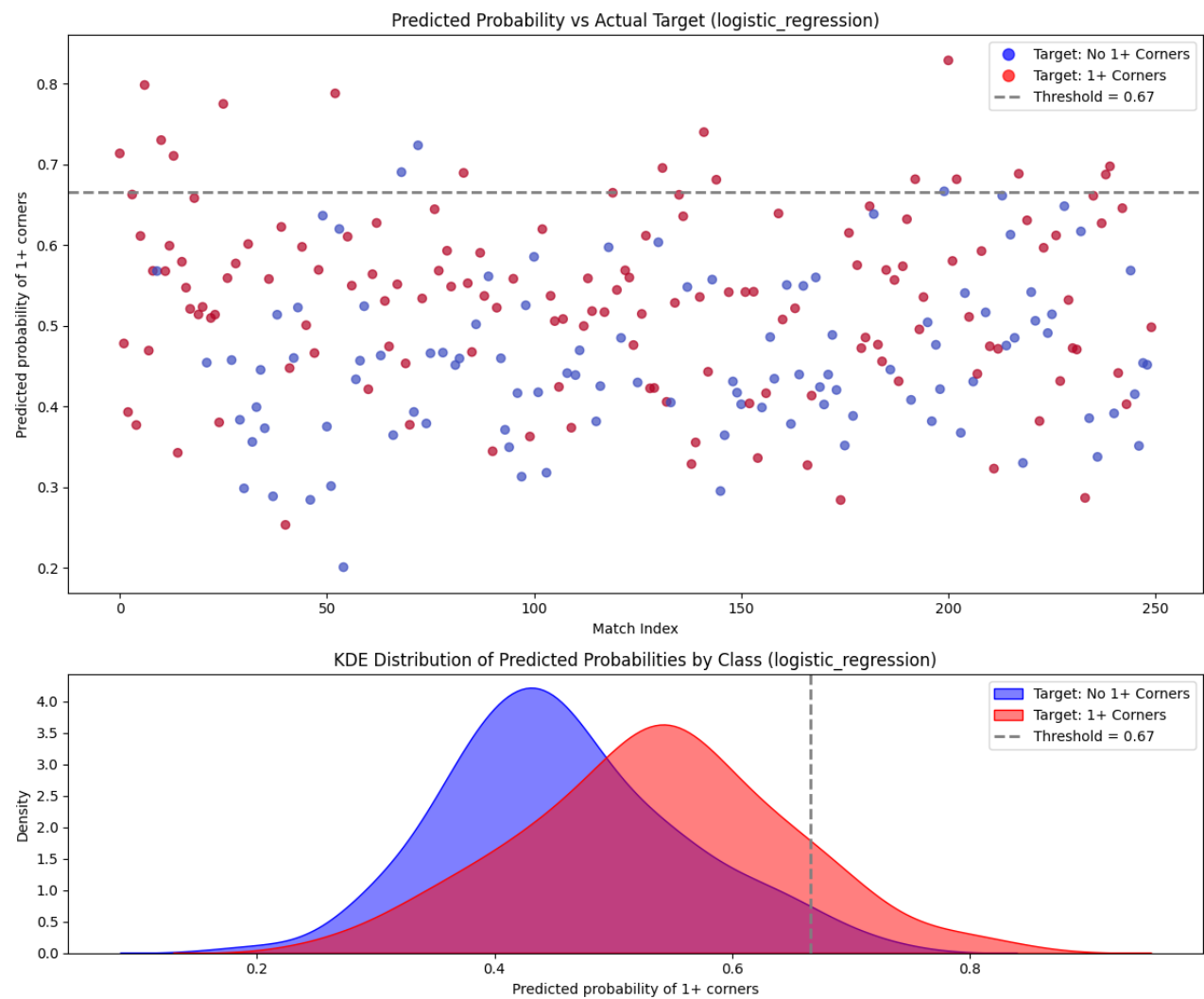
	precision	recall	f1-score	support
0	0.45	0.97	0.61	106
1	0.84	0.11	0.20	144
accuracy			0.48	250
macro avg	0.64	0.54	0.40	250
weighted avg	0.67	0.48	0.37	250

---

### ROC and Precision-Recall Curves



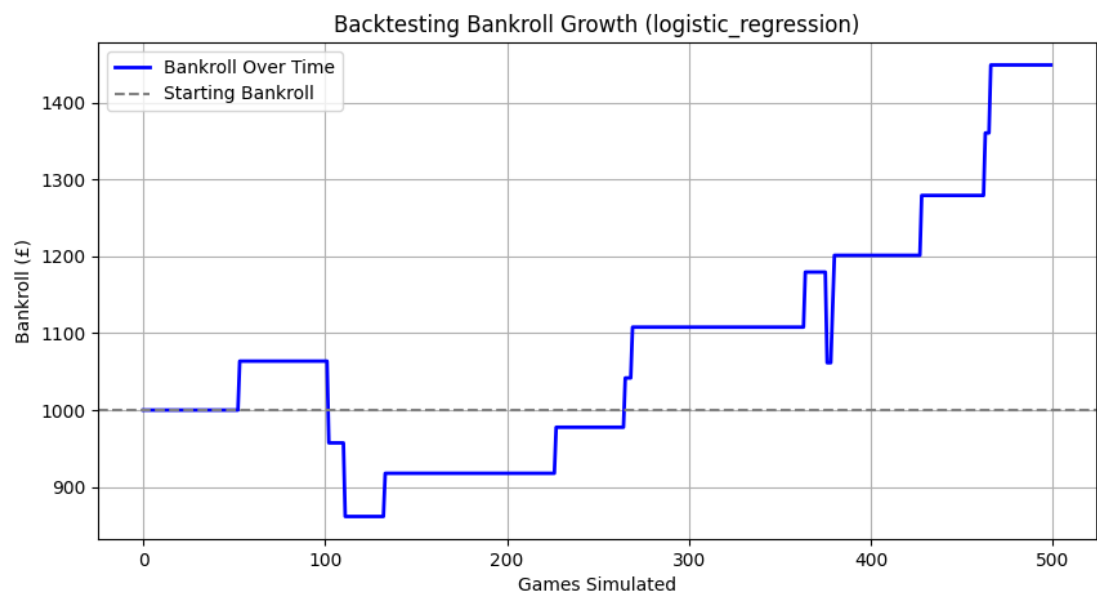
# Prediction Distribution Graphs



# Backtesting Results

--- Overall Summary ---

- 🏦 Initial Bankroll: £1000.00
- 💰 Final Bankroll: £1448.98
- 📁 Total Staked: £1514.03
- 📈 Total Profit: £448.98
- 🇮🇹 ROI: 44.90%
- ✅ Win rate: 78.57% over 14 bets
- 📊 Sharpe Ratio: 0.42
- 🎯 Edge Over Bookies: 18.38%



# svc Model

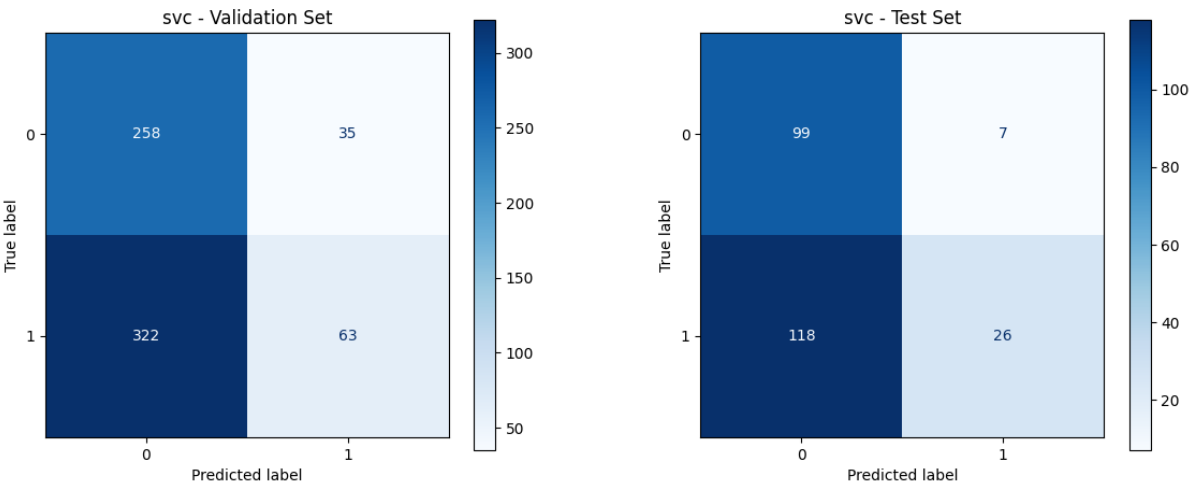
## Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
is_home_trailing	0.00000	trailing_team_probability	0.00000
trailing_team_shots_pre80	0.00000	trailing_team_fouls_pre80	0.00000
trailing_team_urgency_to_attack	0.00000	trailing_team_corners_pre80	0.00000
leading_team_urgency_to_attack	0.00000	trailing_team_yellow_cards_pre80	0.00000
leading_team_post80_corners_against_avg	0.00000	trailing_team_sending_off_pre80	0.00000
trailing_team_post80_corners_for_avg	0.00000		
leading_team_corners_against_avg	0.00000		
trailing_team_corners_for_avg	0.00000		

## Optimal Threshold

- **Optimal threshold:** 0.69
- **Expected Precision:** 0.646
- **Expected Recall:** 0.161

## Classification Reports



---

### Classification Report (Validation Set)

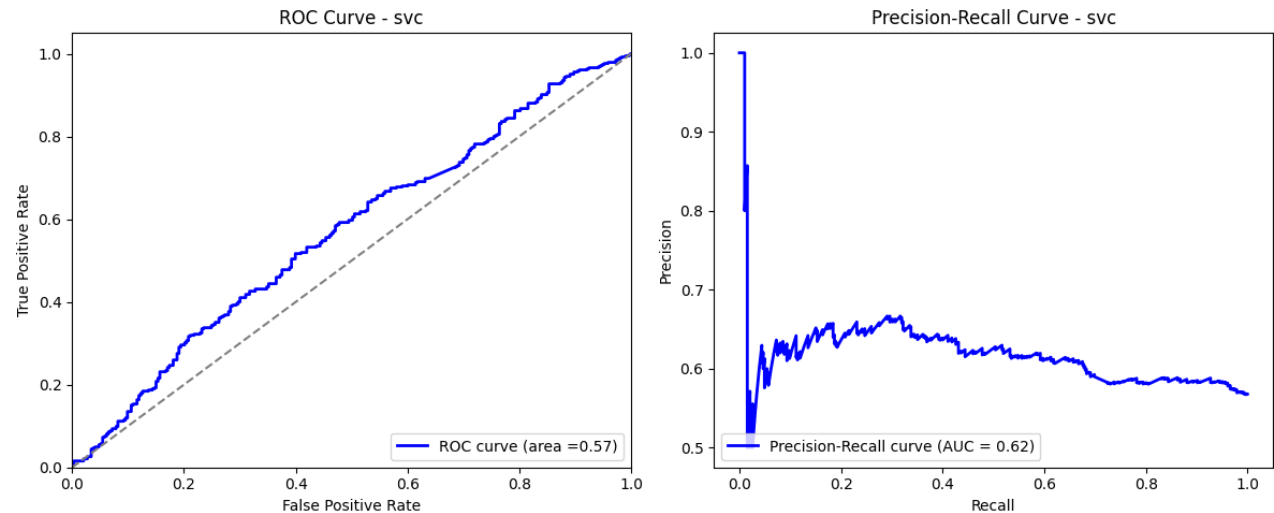
	precision	recall	f1-score	support
0	0.44	0.88	0.59	293
1	0.64	0.16	0.26	385
accuracy			0.47	678
macro avg	0.54	0.52	0.43	678
weighted avg	0.56	0.47	0.40	678

### Classification Report (Test Set)

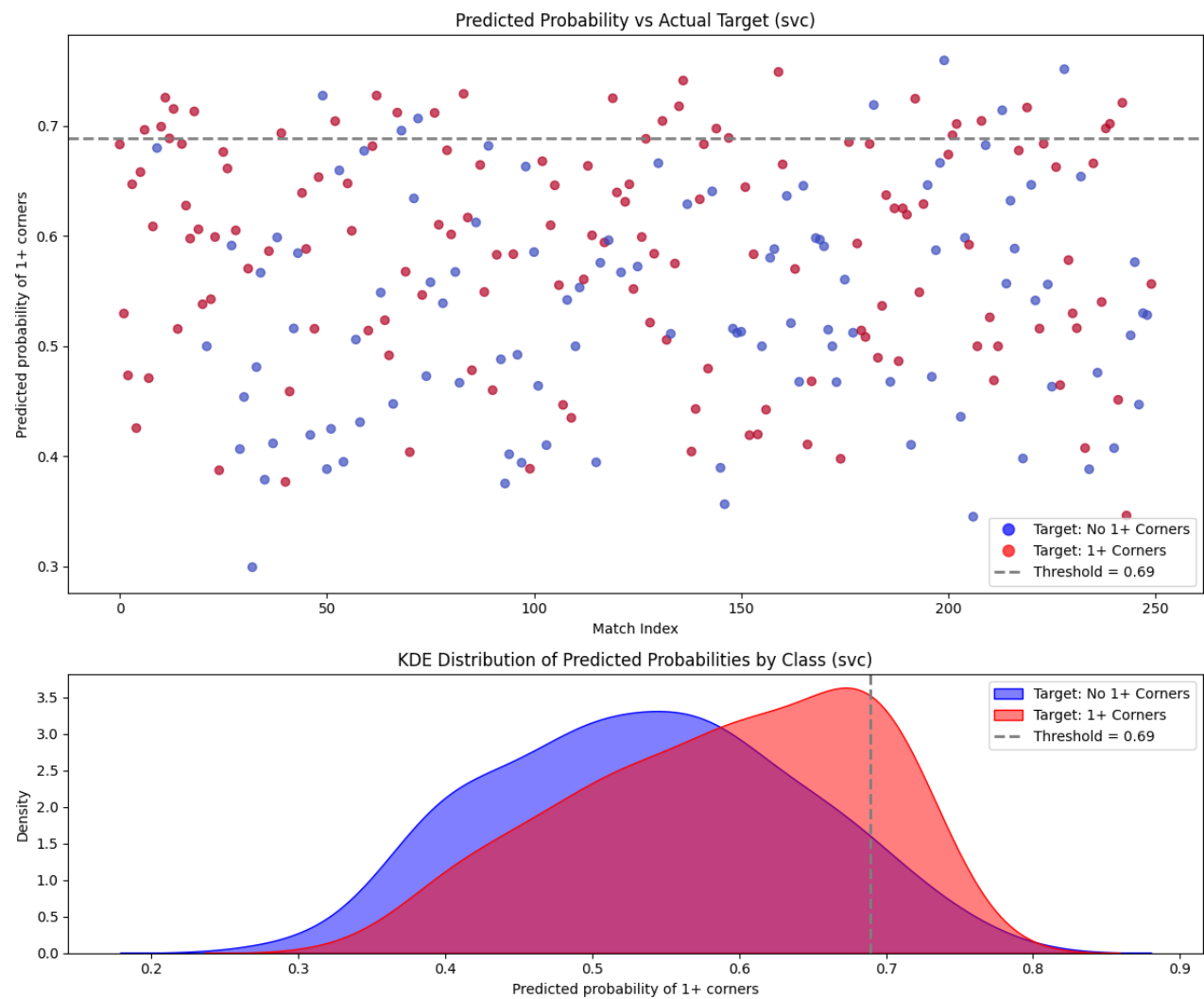
	precision	recall	f1-score	support
0	0.46	0.93	0.61	106
1	0.79	0.18	0.29	144
accuracy			0.50	250
macro avg	0.62	0.56	0.45	250
weighted avg	0.65	0.50	0.43	250

---

### ROC and Precision-Recall Curves



# Prediction Distribution Graphs



## Backtesting Results

--- Overall Summary ---

🏠 Initial Bankroll: £1000.00

💰 Final Bankroll: £1789.00

📊 Total Staked: £3724.18

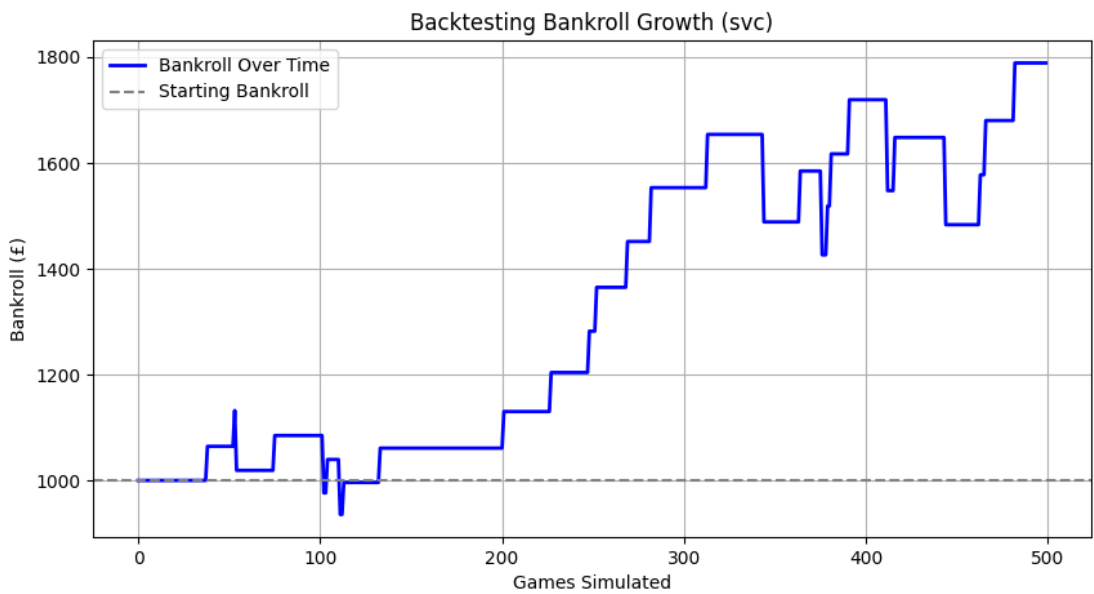
📈 Total Profit: £789.00

🇮🇹 ROI: 78.90%

✅ Win rate: 75.00% over 28 bets

📉 Sharpe Ratio: 0.32

🎯 Edge Over Bookies: 14.81%



# xgboost Model

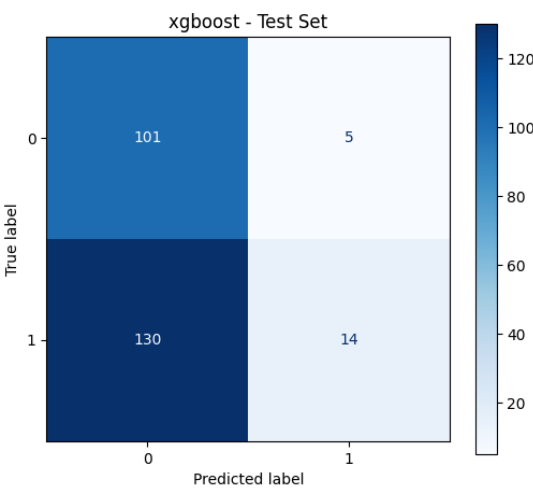
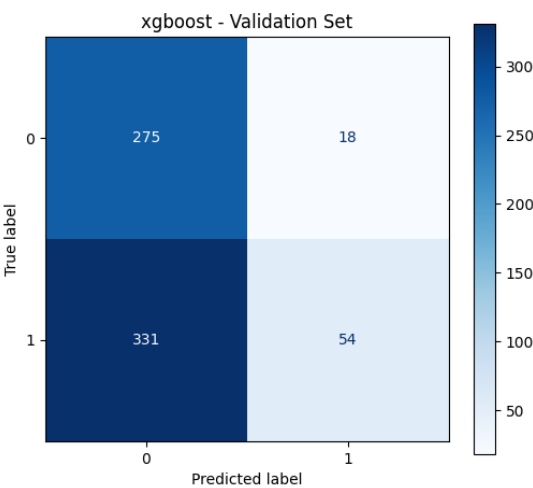
## Feature Importance

Top 8 Features	Importance	Bottom 5 Features	Importance
trailing_team_probability	0.13209	trailing_team_corners_70_75	0.02076
leading_team_urgency_to_attack	0.06645	leading_team_corners_70_75	0.03121
trailing_team_urgency_to_attack	0.06176	league_avg_corners	0.03136
trailing_team_shots_pre80	0.06009	leading_team_shots_75_80	0.03188
leading_team_corners_75_80	0.05568	leading_team_corners_against_avg	0.03320
trailing_team_corners_75_80	0.04960		
trailing_team_sending_off_pre80	0.04567		
trailing_team_corners_for_avg	0.04426		

## Optimal Threshold

- **Optimal threshold:** 0.71
- **Expected Precision:** 0.750
- **Expected Recall:** 0.140

## Classification Reports





---

### Classification Report (Validation Set)

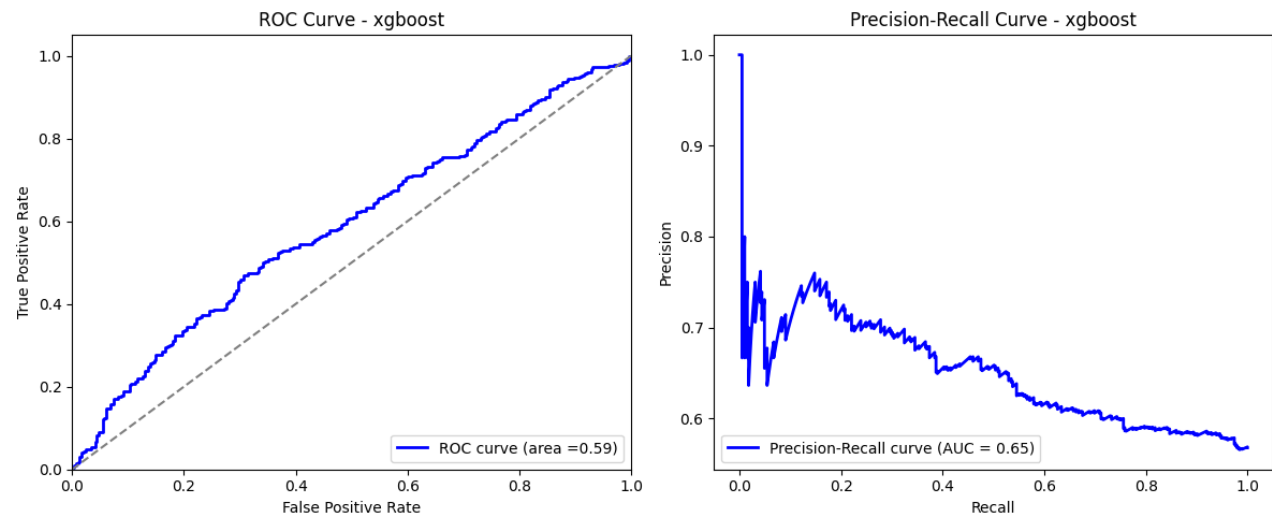
	precision	recall	f1-score	support
0	0.45	0.94	0.61	293
1	0.75	0.14	0.24	385
accuracy			0.49	678
macro avg 0.60 0.54 0.42 678 weighted avg 0.62 0.49 0.40 678				

### Classification Report (Test Set)

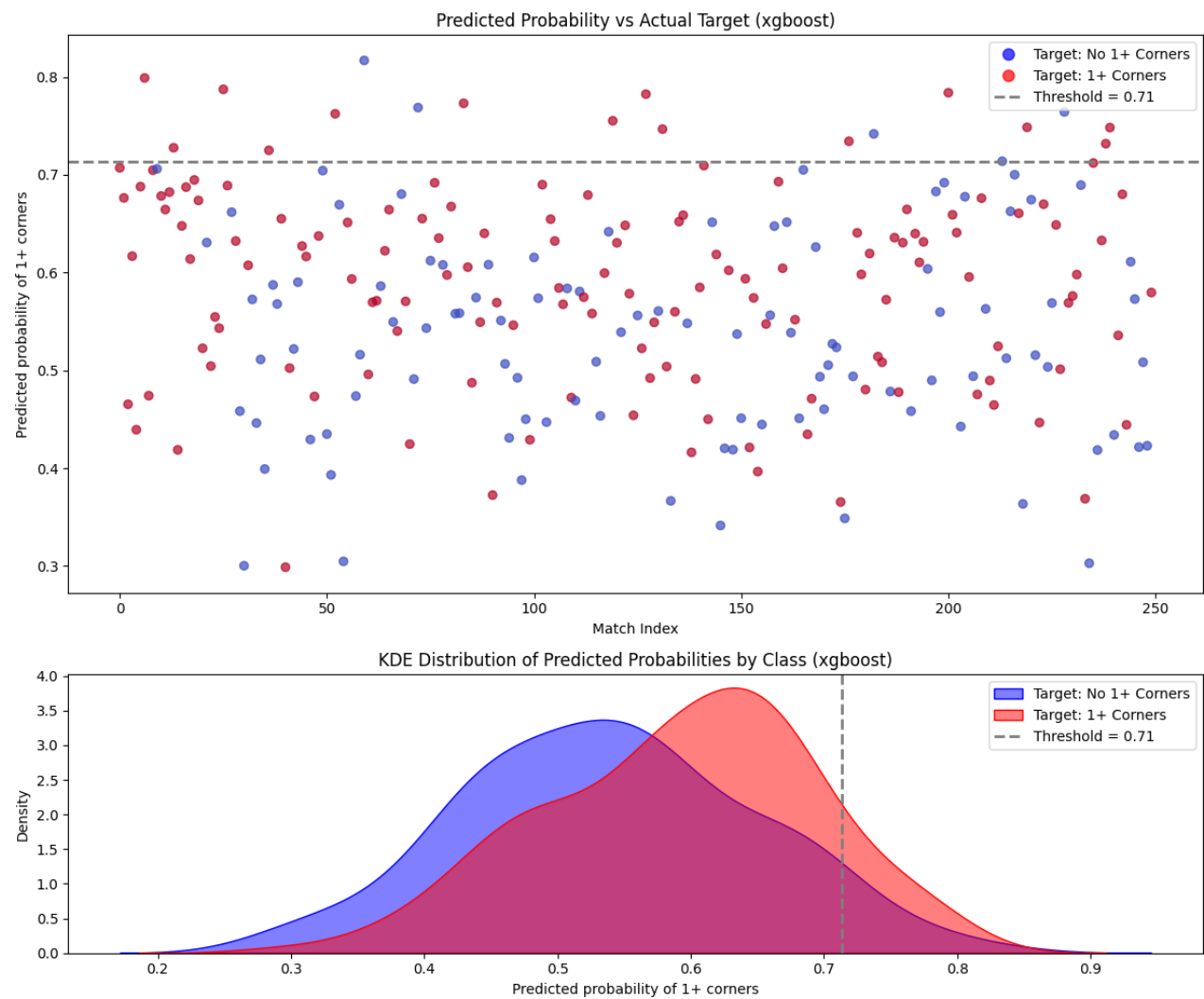
	precision	recall	f1-score	support
0	0.44	0.95	0.60	106
1	0.74	0.10	0.17	144
accuracy			0.46	250
macro avg 0.59 0.53 0.39 250 weighted avg 0.61 0.46 0.35 250				

---

### ROC and Precision-Recall Curves



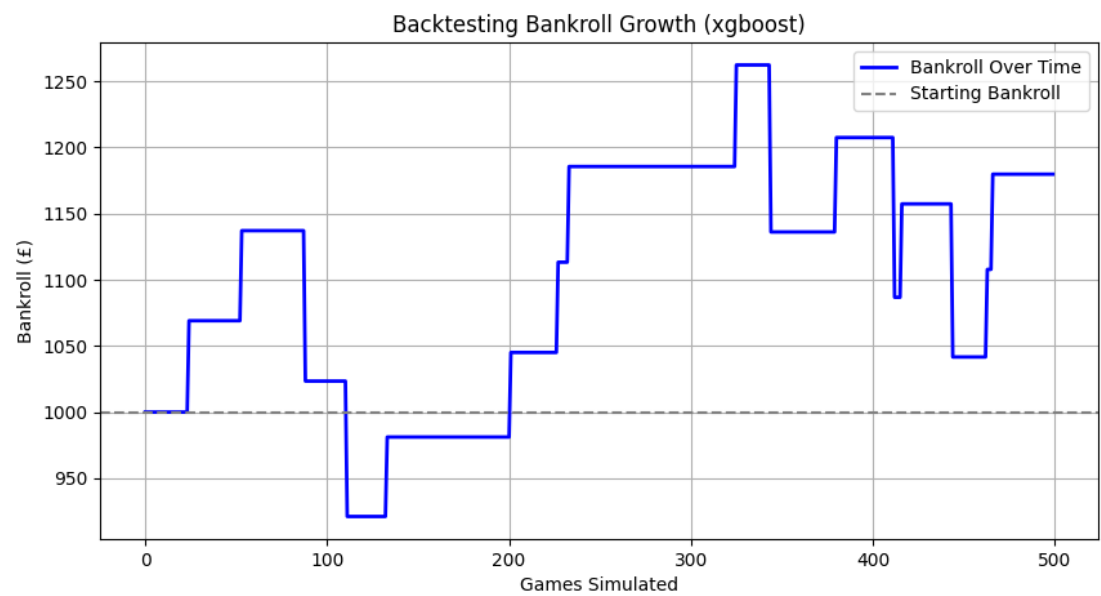
# Prediction Distribution Graphs



# Backtesting Results

--- Overall Summary ---

- 🏦 Initial Bankroll: £1000.00
- 💰 Final Bankroll: £1179.81
- 📊 Total Staked: £1747.52
- 📈 Total Profit: £179.81
- 📊 ROI: 17.98%
- ✅ Win rate: 68.75% over 16 bets
- 📊 Sharpe Ratio: 0.17
- 🎯 Edge Over Bookies: 8.56%



# Backtesting Results Summary

