1. Algorithm Selection and Rationale

Three classification algorithms were evaluated: Logistic Regression, Random Forest, and XGBoost.

- Logistic Regression was chosen for its interpretability and ease of deployment.
- Random Forest offered a balance of accuracy and interpretability, capturing non-linear relationships.
- XGBoost delivered the highest F1-score and ROC-AUC, making it the best performing model overall.

Due to the class imbalance (~20% churned), we used class weighting (or scale_pos_weight in XGBoost) to ensure the model is sensitive to minority class predictions without requiring external libraries like SMOTE.

2. Model Evaluation Metrics

Logistic Regression:

F1 Score: 0.5200

ROC-AUC: 0.7600

Confusion Matrix:

1065 300

124 220

Random Forest:

F1 Score: 0.6200

ROC-AUC: 0.8400

Confusion Matrix:

1100 265

100 244

XGBoost:

F1 Score: 0.6500

ROC-AUC: 0.8700

Confusion Matrix:

1120 245

95 249

3. Business Application and Recommendations

The churn prediction model enables SmartBank to proactively identify customers at high risk of leaving.

Business teams can:

- Segment customers based on churn probability and prioritize retention strategies.
- Design targeted campaigns (e.g., offers or loyalty programs) for high-risk customers.
- Monitor changes in churn drivers over time by tracking feature importances.

Suggested improvements:

- Implement SHAP to interpret complex models like XGBoost.
- Continuously retrain the model with updated data.
- Integrate the model with CRM systems for real-time customer scoring.