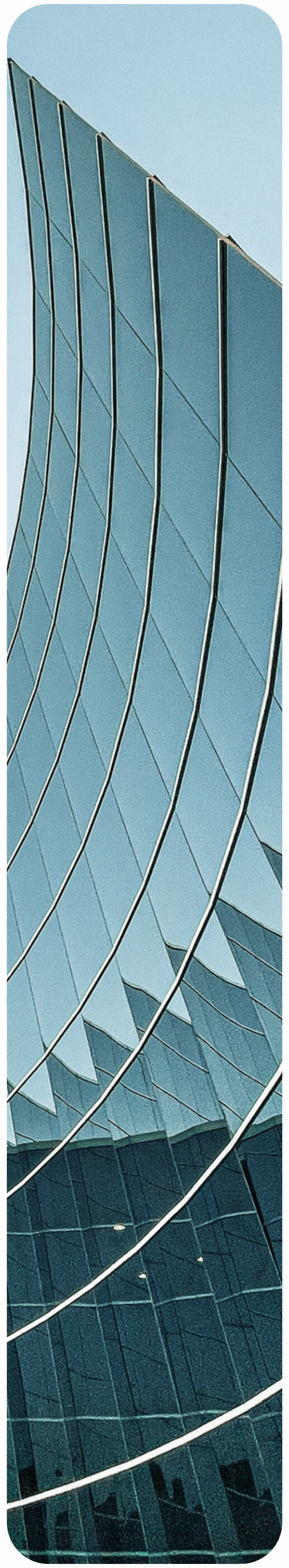


LLOYDS BANKING GROUP

Customer Churn Prediction:

Prediction Report

Date: May 17, 2025

Algorithm Description and Rationale

The goal of this project was to develop a machine learning model to predict customer churn based on various user feature, including demographics ,behavioral patterns and engagement history.

Algorithm Used:

Logistic Regression	Random Forest Classifier	XGBoost Classifier
Initially used for baseline performance and feature importance visualization	Provided higher accuracy and robust performance on the original and refined feature sets	Excelled in classification tasks and handled class imbalance and feature interaction effectively

Data Preparation:
Input Features:

Demographics	Transaction Date	Service Interaction	Digital Engagement
Gender	DaysSinceInteraction	InteractionType	LoginFrequency
MaritalStatus	AmountSpent	ResolutionStatus	ServiceUsage
IncomeLevel		InteractionLag	DaysSinceLogin
Age		DaysSinceInteraction	LoggedInLast30Days

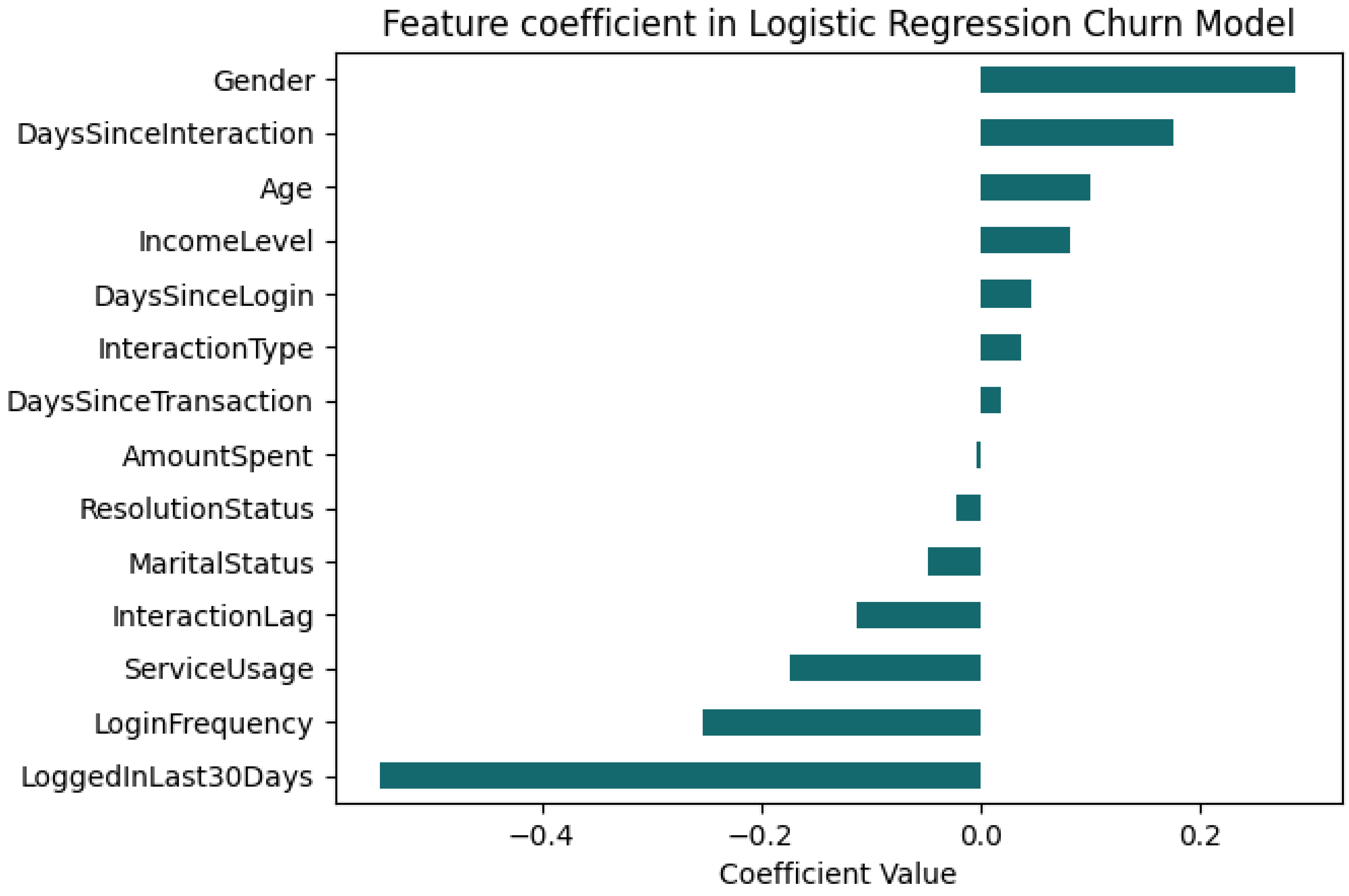
Target Variables:

Churn status

The dataset is split using train_test_split from sklearn , 30% of the data is allocated for testing and rest for training

Logistic Regression:

log loss = 0.505



Most Important Predictors

- LoggedInLast30Days
- LoginFrequency
- ServiceUsage
- Gender
- DaysSinceInteraction

Least important Predictors

- AmountSpent
- ResolutionStatus
- InteractionType

Logistic Regression with selected feature:

log loss = 0.504

Random Forest & XGBoost:

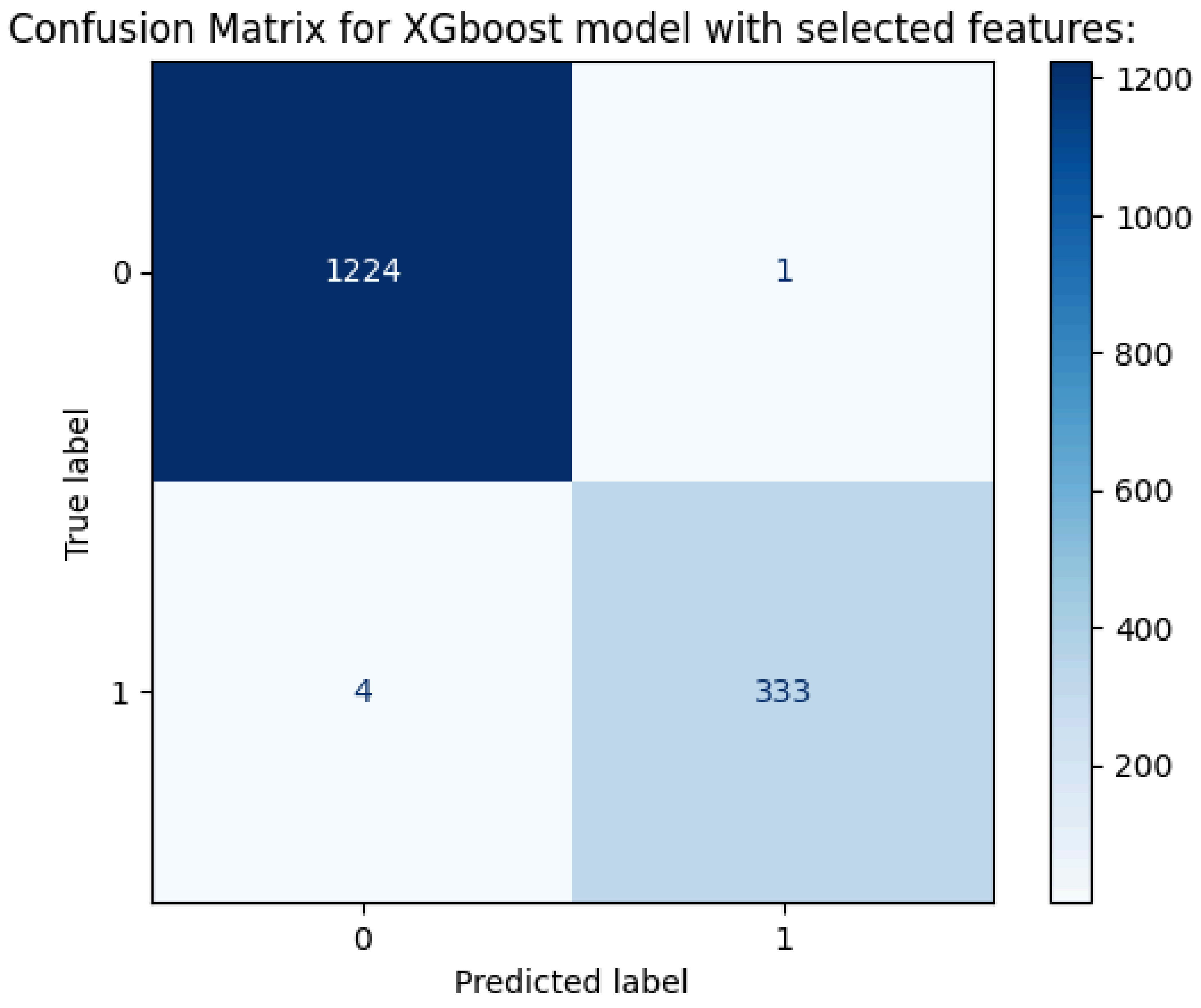
Model	Features	Accuracy	AUC Score
Random Forest	All	0.97	0.94
Random Forest	Selected	0.98	0.96
XGBoost	All	0.99	0.99
XGBoost	Selected	0.99	0.99

Cross validation:

Model	Features	CV Score
Random Forest	Selected	0.97
XGBoost	Selected	0.98

Selected Model: XGBoost with selected features

Confusion Matrix Analysis:



- **True Negatives (TN):** 1224 customers correctly predicted to not churn.
- **False Positives (FP):** 1 customer incorrectly predicted to churn.
- **False Negatives (FN):** 4 customers who actually churned but were not identified.
- **True Positives (TP):** 333 churners correctly identified.

Metrics from Confusion Matrix:

- **Accuracy** : 0.996
- **Precision** : 0.997
- **Recall(Churn Class)** : 0.988
- **F1 Score(Churn Class):** 0.992

Business Application:

Utilization of Predictions:

- Targeted Retention Campaigns: Spot your at-risk customers early and send them personalized offers or reach out through customer service.

- Churn Risk Dashboard: Put the model results into a live dashboard so managers can keep an eye on the chances of customers leaving.

- Customer Lifetime Value (CLV) Modeling: Use churn scores as part of your CLV models to plan your marketing and budgets.

Potential Areas for Improvement

- Feature Engineering: Dig into user habits and timing stuff, like session patterns and how engagement changes over time.
- Hyperparameter Tuning: Play around with Grid Search or Bayesian Optimization to get XGBoost and Random Forest working their best.
- Prediction Confidence: The model usually guesses churn with about 99% certainty. Looking at the tricky cases (around 45-55%) can help us tweak thresholds based on what's more important.
- Model Interpretation: Use SHAP values to break down individual predictions, so everyone trusts what the model's saying.
- Ensemble Blending: Mix together Logistic Regression, XGBoost, and Random Forest using soft voting to make the whole thing more reliable.