Customer Churn Prediction Report

Overview

In this project, a machine learning model was developed to predict customer churn, focusing on the likelihood of a customer leaving Lloyds Banking Group. Predicting churn enables the company to proactively identify at-risk customers and implement retention strategies to improve long-term customer loyalty.

Selected Algorithm and Rationale

After evaluating various algorithms, including Random Forest, XGBoost, and Neural Network models, the XGBoost algorithm was selected for the following reasons:

- Performance: XGBoost provided the highest F1-score across both churned and non-churned classes, achieving a balanced performance suitable for our imbalanced dataset.
- Interpretability: XGBoost offers feature importance, enabling us to identify the most influential factors contributing to churn, which can guide targeted intervention strategies.
- Robustness and Scalability: XGBoost performs well on large datasets with minimal parameter tuning, making it an efficient choice for long-term use.

Model Training and Performance Metrics

The model was trained on the preprocessed data, which included techniques for handling class imbalance (using SMOTE oversampling), feature scaling, and encoding of categorical variables. A 5-fold cross-validation approach was implemented to assess model generalizability.

Performance Metrics:
- Precision: High precision in both churned and non-churned classes, indicating low false-positive rates.

- Recall: High recall for both classes, demonstrating the model's effectiveness in identifying churned customers accurately.

- F1-score: Balanced F1-score for both classes, confirming the model's reliability in handling imbalanced data.

- ROC-AUC Score: The model achieved a ROC-AUC score of 0.994, showcasing strong discriminatory power in distinguishing between churned and non-churned customers.

- Confusion Matrix:

    – True Negatives: 1,045
    – True Positives: 248
    – False Positives: 5
    – False Negatives: 8

The model's high accuracy and low rate of misclassification indicate that it is well-suited for accurately predicting churn, enabling targeted intervention.

Model Utilization and Business Implications

The predictions from this model can provide actionable insights for Lloyds Banking Group to enhance customer retention by:

- Identifying High-Risk Customers: By flagging customers with a high likelihood of churning, the model helps focus retention efforts on at-risk individuals, enabling personalized retention campaigns.
- Personalized Interventions: With information on key drivers (e.g., low engagement, specific transaction types), the model can inform customized offers or service improvements tailored to customer needs.
- Resource Allocation: The model enables efficient allocation of resources toward retention strategies, focusing on high-risk segments to maximize return on investment.

Suggested Areas for Improvement

While the XGBoost model demonstrated strong performance, further enhancements could improve its accuracy and applicability:

- Additional Features: Incorporating more customer interaction data or social engagement metrics could further refine the model's predictions.
- Regular Model Updates: Retraining the model periodically with new data would ensure it adapts to shifting customer behavior trends.
- Model Interpretability Tools: Integrating tools like SHAP (SHapley Additive exPlanations) would provide deeper insights into feature contributions, enhancing transparency and trust among stakeholders.