

# Customer Churn Prediction Model Report

---

## 1. Introduction

This report presents the development and evaluation of a machine learning model for predicting customer churn at Lloyds Banking Group. The model aims to identify customers who are likely to churn, enabling proactive retention strategies.

## 2. Data and Preprocessing

The dataset used for this analysis included customer demographics, transaction history, customer service interactions, and engagement metrics. Data preprocessing steps included:

- **Handling Missing Values:** Imputation with median for numerical features and mode for categorical features.
- **Outlier Treatment:** Capping outliers at the 1st and 99th percentiles.
- **Feature Scaling:** Standardization of numerical features.
- **Encoding Categorical Variables:** One-hot encoding.
- **Date Feature Engineering:** Converting dates to numerical features (days since a reference date).
- **Addressing Class Imbalance:** Applying SMOTE (Synthetic Minority Over-sampling Technique) to balance the dataset.

## 3. Model Selection

Several models were evaluated, including Random Forest, Logistic Regression, SVM, XGBoost, and KNN. **The Random Forest model was selected as the best performing model based on its overall accuracy and ability to handle complex relationships in the data.**

### Rationale for Choosing Random Forest:

- **Handles Non-linearity:** Random Forest can capture complex, non-linear relationships between features and churn.
- **Robust to Outliers:** Less sensitive to outliers due to its ensemble nature.
- **Feature Importance:** Provides insights into the importance of different features for churn prediction.
- **Handles High Dimensionality:** Can effectively handle datasets with many features.
- **Good Performance:** Demonstrated high accuracy and balanced performance metrics.

## 4. Model Training and Evaluation

The Random Forest model was trained using the preprocessed data and optimized through hyperparameter tuning using GridSearchCV with 5-fold cross-validation. The best parameters were: `{'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200}`

### Performance Metrics:

The model's performance was evaluated using the following metrics:

- **Accuracy:** Overall correctness of predictions.
- **Precision:** Proportion of correctly identified churners out of all predicted churners.
- **Recall:** Proportion of correctly identified churners out of all actual churners.
- **F1-score:** Harmonic mean of precision and recall, balancing both metrics.
- **AUC:** Area Under the Curve of the ROC (Receiver Operating Characteristic) curve, measuring the model's ability to distinguish between classes.

Evaluation Results:

Classification Report for Random Forest:				
	precision	recall	f1-score	support
0.0	0.80	0.98	0.88	159
1.0	0.25	0.02	0.04	41
accuracy			0.79	200
macro avg	0.52	0.50	0.46	200
weighted avg	0.68	0.79	0.71	200

Confusion Matrix for Random Forest:  
[[156 3]  
[ 40 1]]  
Accuracy Score for Random Forest: 0.785

## 5. Business Applications and Recommendations

Utilizing the Model's Predictions:

- **Customer Segmentation:** Identify high-risk customers based on their churn probability scores.
- **Targeted Retention Campaigns:** Design personalized retention strategies for high-risk segments.
- **Proactive Customer Service:** Offer proactive support to customers with high churn likelihood.
- **Product and Service Improvements:** Understand factors driving churn and improve offerings accordingly.

Potential Areas for Improvement:

- **Explore More Features:** Include additional relevant data sources, such as customer feedback or website interactions.
- **Refine Feature Engineering:** Experiment with different feature transformations or combinations.
- **Ensemble Methods:** Consider combining the Random Forest model with other strong models for potentially better performance.
- **Continuous Monitoring:** Track model performance over time and retrain as needed.

## 6. Conclusion

The Random Forest model provides a valuable tool for predicting customer churn at Lloyds Banking Group. By leveraging its predictions, the business can implement data-driven strategies to reduce churn and enhance customer retention. However, continuous monitoring and refinement are crucial to ensure long-term effectiveness.

Thank You!!!