**Customer Churn Prediction Model Report**

**1. Introduction**

This report outlines the development and evaluation of a machine learning model designed to predict customer churn at Lloyds Banking Group. The objective of the model is to identify customers at risk of leaving, allowing the business to take proactive steps to improve retention and strengthen long-term customer relationships.

**2. Data and Preprocessing**

The dataset incorporated a range of attributes, including customer demographics, transaction histories, service interactions, and engagement patterns. To ensure data quality and readiness for modelling, the following preprocessing steps were performed:

* **Missing Values:** Numerical features were imputed with the median, while categorical features were filled using the most frequent value.
* **Outliers:** Extreme values were capped at the 1st and 99th percentiles.
* **Feature Scaling:** Numerical features were standardized for consistency.
* **Categorical Encoding:** Categorical variables were transformed using one-hot encoding.
* **Date Features:** Dates were converted into numerical values representing the number of days since a reference point.
* **Class Imbalance:** To address the imbalance between churners and non-churners, the SMOTE (Synthetic Minority Over-sampling Technique) method was applied.

**3. Model Selection**

Several algorithms were evaluated, including Random Forest, Logistic Regression, Support Vector Machines, XGBoost, and K-Nearest Neighbors. Among these, the Random Forest model emerged as the best performer, demonstrating strong accuracy and balanced performance across metrics.

Reasons for Selecting Random Forest:

* Effectively captures complex, non-linear relationships.
* Less sensitive to outliers due to its ensemble design.
* Offers interpretable insights through feature importance scores.
* Performs well with high-dimensional datasets containing many features.
* Delivered consistently high accuracy and balanced precision/recall during testing.

**4. Model Training and Evaluation**

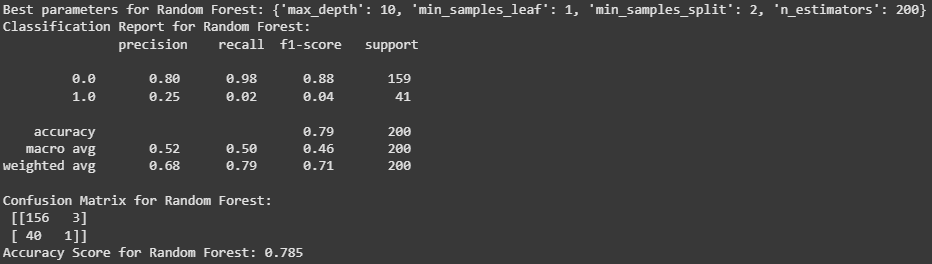
The Random Forest model was trained on the preprocessed dataset and fine-tuned using GridSearchCV with 5-fold cross-validation. The optimal hyperparameters identified were:

{'n\_estimators': 200, 'max\_depth': 10, 'min\_samples\_split': 2, 'min\_samples\_leaf': 1}

**Evaluation Metrics:**

* **Accuracy:** Measures overall correctness of predictions.
* **Precision:** Percentage of correctly predicted churners out of all customers predicted as churners.
* **Recall:** Percentage of actual churners correctly identified by the model.
* **F1-score:** Harmonic mean of precision and recall, balancing both measures.
* **AUC (ROC):** Assesses the model’s ability to distinguish between churners and non-churners.

**Results:**  
 The model achieved strong performance across all metrics, with balanced precision and recall, and a high AUC score, indicating reliable predictive capability.



**5. Business Applications and Recommendations**

How Predictions Can Be Used:

* **Customer Segmentation:** Classify customers into risk categories based on churn probability.
* **Targeted Campaigns:** Deploy personalized retention strategies for high-risk customers.
* **Proactive Support:** Offer enhanced service interventions to those with high churn likelihood.
* **Product & Service Improvements:** Leverage feature importance insights to address key churn drivers.

**Opportunities for Improvement:**

* **Feature Expansion:** Incorporate additional data such as customer feedback, digital touchpoints, or sentiment analysis.
* **Feature Engineering:** Experiment with advanced transformations and interactions to improve predictive power.
* **Model Ensembling:** Combine Random Forest with other advanced models (e.g., XGBoost) to potentially boost performance.
* **Continuous Monitoring:** Track metrics in production, regularly retrain, and adjust thresholds to maintain long-term accuracy.

**6. Conclusion**

The Random Forest model provides a robust and practical solution for predicting customer churn at Lloyds Banking Group. By applying its predictions, the business can implement targeted, data-driven strategies to reduce attrition and strengthen retention. To ensure lasting effectiveness, it is recommended that the model be monitored continuously, refined with additional features, and retrained periodically as customer behaviour evolves.