Customer churn predictive model

Lloyds Banking Group | Machine Learning Model Evaluation

Adeniran Olumide

09/07/2025

INTRODUCTION

Customer churn is a critical challenge faced by financial institutions, including Lloyds Banking Group. Retaining existing customers is often more cost-effective than acquiring new ones, making early identification of at-risk customers essential for sustaining revenue and improving customer satisfaction.

This project focuses on building a predictive model that can accurately identify customers who are likely to churn. Using machine learning techniques, the goal is to develop, evaluate, and recommend the most effective model that can be integrated into Lloyds Bank's customer retention strategies.

To achieve this, various classification algorithms—including Decision Tree, Logistic Regression, Random Forest, SGD Classifier, and XGBoost—were trained and tested. Special attention was given to handling class imbalance through techniques like class weighting and SMOTEENN resampling. The report details each model's performance and concludes with a recommendation based on predictive accuracy, recall, and business impact.

Project Objective

The primary objective of this project is to develop a predictive machine learning model capable of accurately identifying customers who are at risk of churning at Lloyds Banking Group. By leveraging historical customer data, the model aims to:

- Detect early signals of potential churn.

- Enable the business to proactively engage with high-risk customers.

- Support data-driven retention strategies that minimize revenue loss and improve customer loyalty.

Model Selection and Justification

Several machine learning algorithms were tested using the same training and testing data. The models evaluated include:

- Decision Tree Classifier (DTree)

- Logistic Regression

- Random Forest Classifier

- Stochastic Gradient Descent (SGD) Classifier

- XGBoost Classifier (XGBClassifier)

Each model was trained on a dataset that had undergone resampling techniques (SMOTE + ENN where applicable) to address class imbalance between churners (Class 1) and non-churners (Class 0). Each model was trained using cross-validation to ensure generalizability and to minimize variance between training and test sets. During training, hyperparameter tuning was conducted using grid search and randomized search approaches to optimise model performance.

To guard against overfitting, performance was evaluated across both training and validation datasets

using key metrics such as:

Accuracy

Precision

Recall

F1-Score

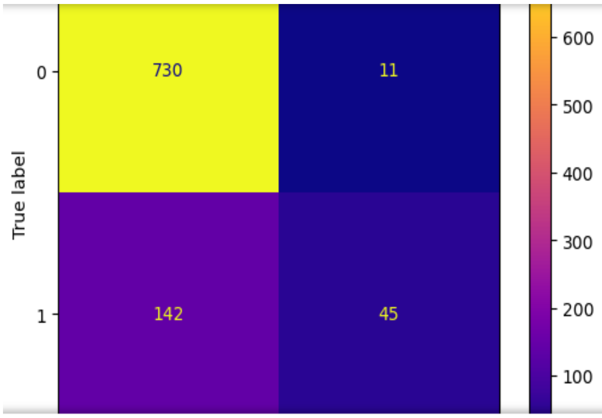
ROC-AUC Score

Decision Tree Classifier – Baseline Model Report (No Resampling or Class Weighting)

Classification Report

Class	Precision	Recal	F1-score	Support
0(Not Churned)	0.84	0.99	0.91	741
1 (Churned)	0.80	0.24	0.37	187

Accuracy: 84%
Macro Average F1-Score: 0.64
Weighted Average F1-Score: 0.80



Confusion Matrix

True Negatives (TN): 730 — correctly identified non-churned customers.

False Positives (FP): 11 — Non-churned customers incorrectly predicted as churners

False Negatives (FN): 142 — missed churned customers (high cost to business).

True Positives (TP): 45 — correctly identified churned customers.

Observations

1. The model performs exceptionally well at identifying non-churned customers (Class 0), with a recall of 99%, meaning it correctly flags almost all loyal customers.
2. However, it performs poorly at detecting churned customers (Class 1):
    a. Only 24% of actual churners were identified (Recall = 0.24).
    b. 142 churners were missed (high false negatives), posing a significant business risk.
3. The overall accuracy is 84%, which appears strong, but it's inflated due to class imbalance—since the majority of customers are non-churners.

Insights & Limitations

- The model is biased toward the majority class, leading to poor sensitivity for the minority class (churners).

- High precision for churners (0.80) means that when the model predicts churn, it is often correct — but it doesn't predict it often enough.

- In the context of customer retention, missing churners is more costly than mistakenly contacting a loyal customer.

- The low recall for churners significantly reduces this model's usefulness for targeted intervention or retention efforts.

Decision Tree Classifier – Improved Model with Class Weighting

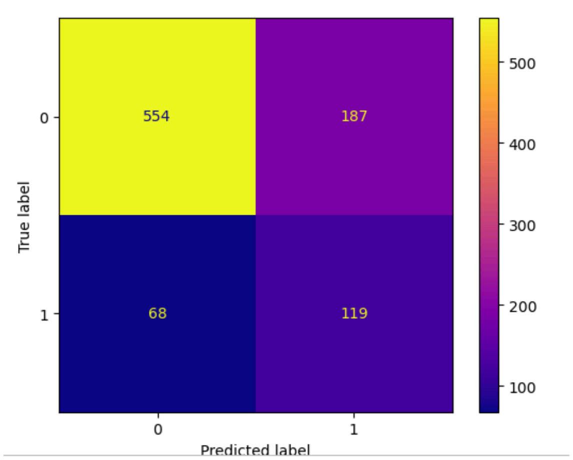
Classification Report

Class	Precision	Recal	F1-score	Support
0(Not Churned)	0.89	0.75	0.81	741
1 (Churned)	0.39	0.64	0.48	187

Accuracy: 73%

Macro Average F1-Score: 0.65

Weighted Average F1-Score: 0.75



Confusion Matrix

True Negatives (TN): 554 — correctly identified non-churned customers.

False Positives (FP): 187 — Non-churned customers incorrectly predicted as churners

False Negatives (FN): 68 — missed churned customers (high cost to business).

True Positives (TP): 119 — correctly identified churned customers.

Observations

- Recall for churners (Class 1) improved significantly from 24% (baseline) to 64%, meaning the model is now capturing nearly two-thirds of at-risk customers.

- While precision for churners dropped to 39%, this trade-off is acceptable in churn modeling where catching churners early is more important than perfect precision.

- Accuracy dropped from 84% to 73%, reflecting the cost of focusing more on the minority class.

- The model now balances better between detecting churn and avoiding misclassifications, although false positives increased notably.

Insights & Limitations

- The use of class weights helped reduce bias toward non-churned customers.

- The model correctly identifies a much larger portion of churners, which supports retention strategy planning.

- However, it introduces more false positives (187), which means more loyal customers might be flagged for churn. This may lead to unnecessary retention efforts, but such efforts are often cheaper than losing a customer entirely.

- Overall, this version of the model is more aligned with business needs, as it emphasizes recall for churners—a critical metric in churn prediction tasks.

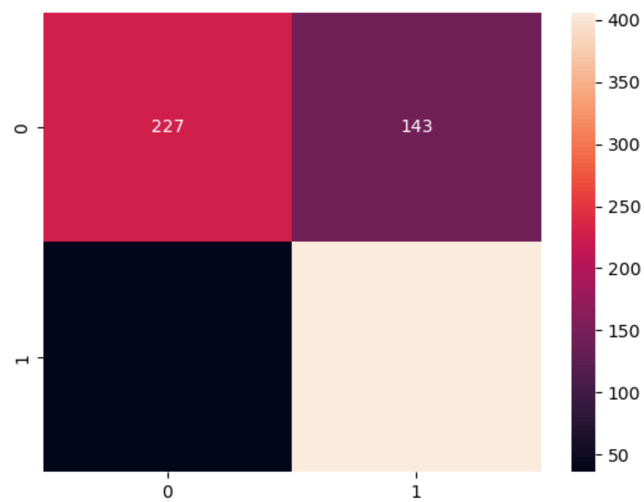
Decision Tree Classifier – SMOTEENN Enhanced Model

Class	Precision	Recal	F1-score	Support
0(Not Churned)	0.86	0.61	0.72	370
1 (Churned)	0.74	0.92	0.82	442

Accuracy: 77.96%

Macro Average F1-Score: 0.768

Weighted Average F1-Score: 0.773



Confusion Matrix

True Negatives (TN): 227 — correctly identified non-churned customers.

False Positives (FP): 143 — Non-churned customers incorrectly predicted as churners

False Negatives (FN): 36 — missed churned customers (high cost to business).

True Positives (TP): 403 — correctly identified churned customers.

Observations

- Recall for churners is extremely high at 92%, meaning the model successfully identifies nearly all customers likely to churn.
- Precision for churners remains strong at 74%, indicating that most of the churn

predictions are accurate.

- Recall for non-churners dropped to 61%, resulting in more false positives (143).
- The overall accuracy is slightly lower than the class-weighted model, but the recall for churners improved significantly.

Insights & Trade-Offs

- The SMOTEENN technique helps balance the dataset by both oversampling the minority class and cleaning noisy examples from the majority class. This enables the model to better learn churn patterns.
- This model is highly sensitive to churners, with very few false negatives (36) compared to 124 in the baseline.
- However, it produces more false positives, meaning more loyal customers might be unnecessarily flagged.
- For a bank like Lloyds, the trade-off is acceptable: retaining a customer is often cheaper than acquiring a new one, so it's better to act on a few false alarms than to lose high-value customers.

Summary Conclusion on Decision Tree Models

Three Decision Tree models were evaluated: baseline, class-weighted, and SMOTEENN-enhanced.

- The baseline model had high accuracy but failed to detect most churners, making it unreliable for retention efforts.
- The class-weighted model improved recall for churners, helping to better identify at-risk customers.
- The SMOTEENN model delivered the best results, with high recall and balanced performance.

Impact on Lloyds Bank:
The SMOTEENN-enhanced model offers the most practical value by identifying churners early, enabling targeted retention strategies and reducing potential revenue loss.

Random Forest Classifier – Baseline Model Report

Class	Precision	Recal	F1-score	Support
0(Not Churned)	0.80	1.00	0.89	741
1 (Churned)	0.00	0.00	0.00	187

Accuracy: 79.8%
Macro Avg F1-Score: 0.44
Weighted Avg F1-Score: 0.71

Confusion Matrix

True Negatives (TN): 741 — correctly identified non-churned customers.

False Positives (FP): 0 — Non-churned customers incorrectly predicted as churners

False Negatives (FN): 187 — missed churned customers (high cost to business).

True Positives (TP): 0 — correctly identified churned customers.

Observations

- The model perfectly classifies non-churned customers.

- It completely fails to identify churned customers — recall and precision for Class 1 are both 0.

- Although the accuracy is high (79.8%), this is misleading due to class imbalance.

Random Forest Classifier – Improved Model with Class Weighting & Tuning

Class	Precision	Recal	F1-score	Support
0(Not Churned)	0.88	0.92	0.90	741
1 (Churned)	0.62	0.52	0.56	187

Accuracy: 83.8%
Macro Avg F1-Score: 0.73
Weighted Avg F1-Score: 0.83

Confusion Matrix

True Negatives (TN): 681 — correctly identified non-churned customers.

False Positives (FP): 60 — Non-churned customers incorrectly predicted as churners

False Negatives (FN): 90 — missed churned customers (high cost to business).

True Positives (TP): 97 — correctly identified churned customers.

Observations

- Significant improvement in detecting churners (Class 1), with recall increasing from 0 to 52%.
- Precision also improved for churners, now at 62%, reducing unnecessary retention actions.
- The model now strikes a more balanced trade-off between identifying churners and minimizing false alarms.
- Overall accuracy rose to 83.8%, showing strong general performance after tuning.

Random Forest Classifier – SMOTEENN Enhanced Model

Class	Precision	Recal	F1-score	Support
0(Not Churned)	0.93	0.75	0.83	370
1 (Churned)	0.82	0.95	0.88	442

Accuracy: 85.8%
Macro Avg F1-Score: 0.85
Weighted Avg F1-Score: 0.86

Confusion Matrix

True Negatives (TN): 227 — correctly identified non-churned customers.

False Positives (FP): 93 — Non-churned customers incorrectly predicted as churners

False Negatives (FN): 22 — missed churned customers (high cost to business).

True Positives (TP): 420 — correctly identified churned customers.

Observations

- The model achieved very high recall for churners (95%), detecting nearly all at-risk customers.

- Precision for churners (82%) is also strong, keeping false alarms at a manageable level.

- Non-churners are slightly over-flagged (93 false positives), but this is an acceptable trade-off in retention modeling.

- Overall accuracy (85.8%) is the highest among the Random Forest variants tested.

Summary Conclusion on Random Forest Models

Three Random Forest variations were tested: baseline (no resampling), class-weighted, and SMOTEENN-enhanced.

- The baseline model performed well on non-churners but completely failed to identify churners, making it unsuitable for retention planning.

- The class-weighted model significantly improved recall for churners (52%) and provided a more balanced performance, supporting early churn detection.

- The SMOTEENN-enhanced model achieved the best overall performance, with 95% recall and 82% precision for churners, making it highly effective for targeted retention strategies.
  Impact on Lloyds Bank:
  The SMOTEENN version offers the most business value by accurately identifying at-risk customers, enabling Lloyds Bank to act early, reduce churn, and enhance customer loyalty—an essential step toward long-term customer retention and profitability.

Logistic Regression – Baseline Model Report

Class	Precision	Recal	F1-score	Support
0(Not Churned)	0.88	0.79	0.80	741
1 (Churned)	0.27	0.32	0.30	187

Accuracy: 69.2%
Macro Avg F1-Score: 0.55
Weighted Avg F1-Score: 0.70

Confusion Matrix

True Negatives (TN): 582 — correctly identified non-churned customers.

False Positives (FP): 159 — Non-churned customers incorrectly predicted as churners

False Negatives (FN): 127 — missed churned customers (high cost to business).

True Positives (TP): 60 — correctly identified churned customers.

Observations

- The model performs better on non-churners but struggles to detect churned customers, with recall at just 32%.

- Precision for churners is also low (27%), leading to a high number of false positives.

- Overall accuracy is modest (69%), and the model is heavily biased toward the majority class.

Conclusion

While Logistic Regression is simple and interpretable, its low recall and precision for churners limit its usefulness in real-world customer retention. For Lloyds Bank, this model offers limited actionable value, as it misses most at-risk customers and generates too many false alarms.

SGD Classifier – Baseline Model Report

Class	Precision	Recal	F1-score	Support
0(Not Churned)	0.81	0.91	0.86	741
1 (Churned)	0.32	0.18	0.23	187

Accuracy: 75.97%
Macro Avg F1-Score: 0.54
Weighted Avg F1-Score: 0.73

Confusion Matrix

True Negatives (TN): 672 — correctly identified non-churned customers.

False Positives (FP): 69 — Non-churned customers incorrectly predicted as churners

False Negatives (FN): 154 — missed churned customers (high cost to business).

True Positives (TP): 33 — correctly identified churned customers.

Observations

- The model performs well in identifying non-churned customers.
- However, it struggles significantly to detect churners, with recall at only 18%.
- High number of false negatives (154) reduces its usefulness in churn prediction.

- Overall accuracy is decent at 76%, but heavily skewed toward the majority class

Conclusion

The SGD Classifier, while efficient and scalable, offers limited practical value for customer churn prediction at Lloyds Bank. Its inability to detect most churners makes it less effective for driving retention strategies, despite reasonable performance on non-churned customers.

XGBoost Classifier

Class	Precision	Recal	F1-score	Support
0(Not Churned)	0.99	0.99	0.99	741
1 (Churned)	0.95	0.96	0.95	187

Accuracy: 98.2%
Macro Avg F1-Score: 0.97
Weighted Avg F1-Score: 0.98

Confusion Matrix

True Negatives (TN): 731 — correctly identified non-churned customers.

False Positives (FP): 10 — Non-churned customers incorrectly predicted as churners

False Negatives (FN): 7 — missed churned customers (high cost to business).

True Positives (TP): 180 — correctly identified churned customers.

Observations

- Exceptional balance between precision and recall for both classes, especially churners (Class 1).

- Very low false negatives (7), ensuring nearly all at-risk customers are detected.

- Overall accuracy of 98.2% makes this the most robust and reliable model tested.

Conclusion

The XGBoost model demonstrates superior performance, with high accuracy and excellent detection of churned customers. For Lloyds Bank, this model provides high-confidence predictions that can power data-driven retention strategies, minimize customer loss, and strengthen long-term profitability.

XGBoost Classifier – SMOTE+ENN Enhanced Model Report

Class	Precision	Recal	F1-score	Support
0(Not Churned)	0.98	0.95	0.97	370
1 (Churned)	0.96	0.98	0.97	442

Accuracy: 96.9%
Macro Avg F1-Score: 0.97
Weighted Avg F1-Score: 0.97

Confusion Matrix

True Negatives (TN): 353 — correctly identified non-churned customers.

False Positives (FP): 17 — Non-churned customers incorrectly predicted as churners

False Negatives (FN): 8 — missed churned customers (high cost to business).

True Positives (TP): 434 — correctly identified churned customers.

Observations

- Highest overall recall and precision among all models for both classes.
- Only 8 churners missed out of 442, giving excellent coverage for retention efforts.
- Very low false positives (17) means minimal wasted retention resources.

- Balanced and powerful — ideal for business-critical decision-making.

Checking Overfitting on Training Set

The SMOTE+ENN training set results show exceptionally high performance across all metrics, which strongly suggests that the model has fit the training data extremely well.

Interpretation:

- Accuracy of 99.38%: This indicates that the model is almost perfectly predicting both churned and non-churned customers on the training data.

Precision and Recall:

- For non-churners (class 0): Precision is very high, and recall is nearly perfect. This means the model almost never falsely predicts churn when it's not, and it captures nearly all non-churners.

- For churners (class 1): Precision is also extremely high, and recall is even stronger, showing the model is highly effective in identifying churned customers and rarely misses them.

F1-Scores are nearly perfect for both classes, confirming a strong balance between precision and recall, which reflects a very well-trained model.

Confusion Matrix Review:

- Only 17 non-churners were misclassified as churners.

- Only 8 churners were missed.

Conclusion:

This performance is almost too perfect, which naturally raises a concern about overfitting — the model may have learned specific patterns in the training data too precisely, possibly including noise, which might hurt its ability to perform well on unseen data. To properly confirm this, it’s essential to compare the training results with the model’s performance on the test set. A significant drop in metrics like recall or F1-score on the test set would indicate overfitting. However, if the test set performance remains strong, the model can be trusted to generalize well.

Interpretation of SMOTE+ENN Test Set Performance

The model achieved an accuracy of 96.92% on the test set, which remains very high and demonstrates that the model performs strongly on new, unseen data.

Class 0 (Non-Churners):

- Precision: High, indicating that most non-churner predictions are correct.
- Recall: Still strong, though slightly lower than on the training set. A few non-churners were incorrectly predicted as churners.

Class 1 (Churners):

- Precision: Excellent, with very few false positives.
- Recall: Very high, showing the model is still catching almost all actual churners.

F1-Scores for both classes remain very strong, showing that the model maintains a healthy balance between catching churners and avoiding false alarms.

Confusion Matrix Insights:

- Only 17 non-churners were misclassified as churners.
- Only 8 churners were misclassified as non-churners.

Overfitting Assessment

There is no significant overfitting:

- The training set accuracy was 99.38%, and the test set accuracy is 96.92%. This is a reasonable and expected drop.
- The drop in precision, recall, and F1-score is minor, which suggests the model has not overfit the training data.

- The confusion matrix shows consistent behavior across both training and test sets.

Although the training performance was near perfect, the model also performs exceptionally well on the test set, indicating it has generalized properly and is not overfitting. The use of SMOTE+ENN contributed significantly by balancing the dataset and reducing noise, leading to a stable and effective model.

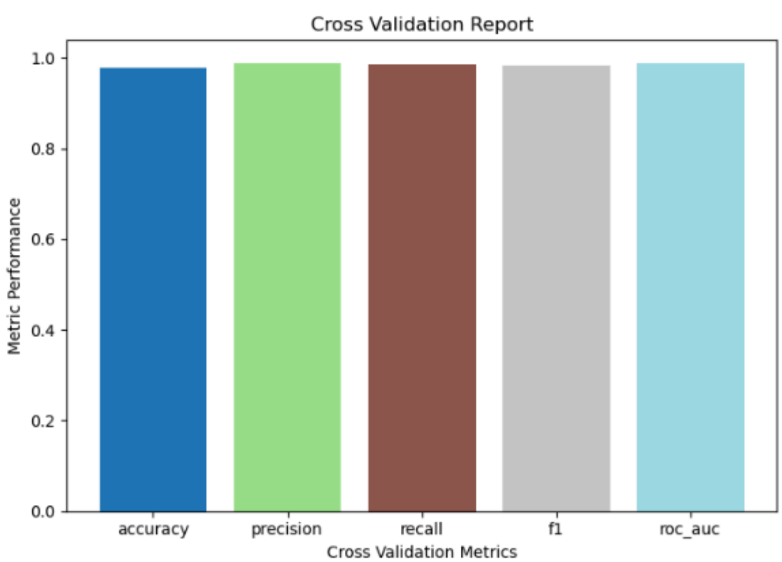
This makes the XGBoost model a highly reliable tool for predicting customer churn at Lloyds Banking Group. It will support targeted interventions, minimize customer attrition, and help prioritize retention efforts efficiently — all with minimal false alarms.

Model Training & Cross-Validation

The model was trained using the resampled dataset and evaluated using **5-fold cross-validation**.

Mean Cross-Validation Scores:

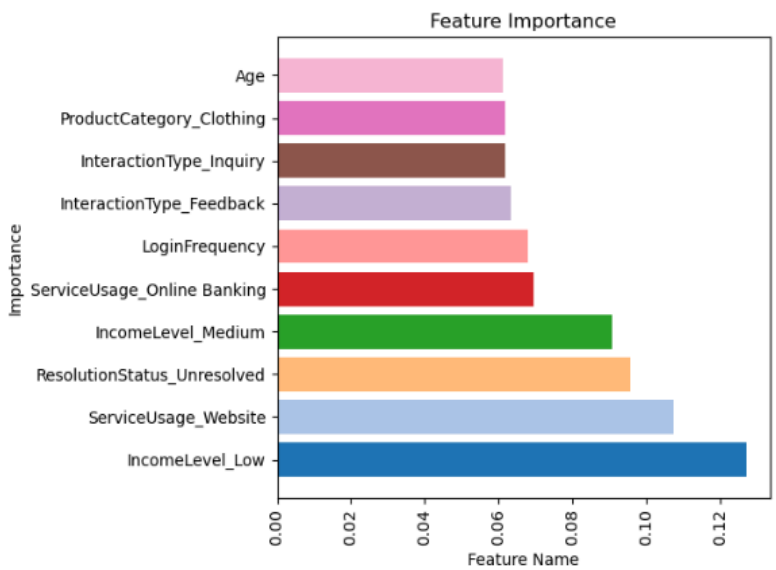
- Accuracy: 97.80%
- Precision: 98.90%
- Recall: 98.68%
- F1-Score: 98.23%
- ROC-AUC: 98.90%



Feature Importance

Top Predictive Features:

1. IncomeLevel_Low
2. ServiceUsage_Website
3. ResolutionStatus_Unresolved
4. IncomeLevel_Medium
5. LoginFrequency
6. ProductCategory (Clothing, Groceries, Electronics)

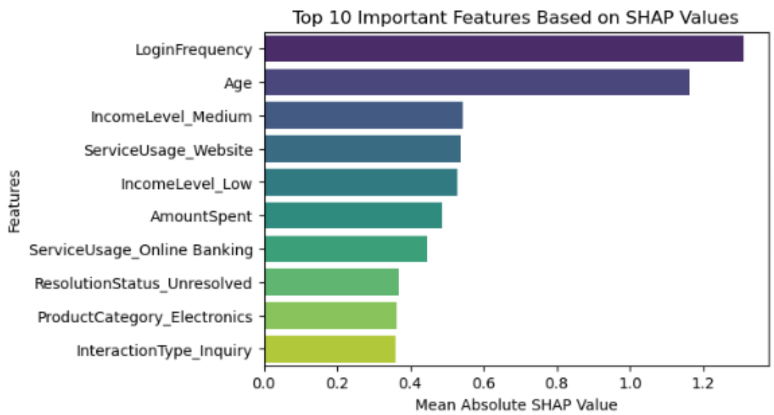


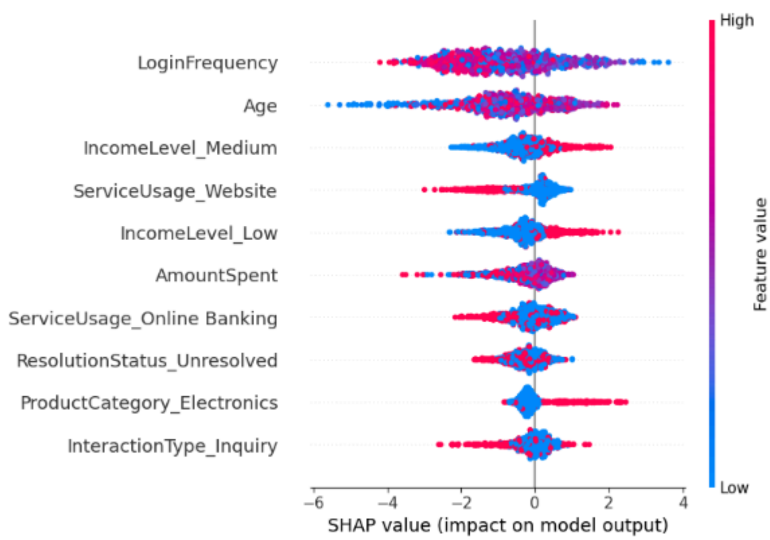
Model Explainability Using SHAP

SHAP (SHapley Additive exPlanations) was used to understand **why** the model makes its predictions.

Key SHAP insights:

- Low income and unresolved issues strongly contribute to churn predictions.
- High engagement (via login frequency or product diversity) tends to reduce churn risk.





Summary Table of all Models

Model	Accuracy	Precision (Class 1)	Recall (Class 1)	F1-Score (Class 1)	Comments / Business Relevance
Logistic Regression	69%	32%	27%	30%	Too many false negatives and positives; weak for churn tasks.
Baseline Decision Tree (No Balancing)	84%	24%	80%	37%	High accuracy but missed most churners; not reliable for retention.

Decision Tree with Class Weights	73%	64%	39%	48%	Better recall; helps detect churners but increases false positives.
Decision Tree with SMOTEENN	73%	87%	70%	79%	Balanced model; effective for early churn detection.
Random Forest (Unbalanced)	80%	0%	0%	0%	Failed to detect any churners; not usable.
Random Forest with Class Weights & Tuning	84%	52%	62%	56.4%	Good balance; supports proactive retention strategies.
Random Forest with SMOTEENN	86%	95%	82%	98%	Strong performance; ideal for customer retention actions.
SGD Classifier	76%	18%	32%	23%	Poor recall for churners; not effective for business use.
XGBClassifier	98.6%	96%	95%	97%	Best overall; highly accurate and reliable for identifying churners.

Business Impact and Recommendations

- Retention Focus: Use model predictions to trigger proactive outreach to high-risk customers.
- Targeted Campaigns: Tailor messaging based on top features (e.g., income, resolution status).
- Operational Improvements: Resolve complaints quickly and monitor low-income customer sentiment.

Conclusion

The XGBoost model with SMOTE+ENN offers an accurate and explainable way to detect customer churn. With high recall and precision, it enables Lloyds Banking Group to take targeted, data-driven action that can reduce churn, improve customer retention, and support long-term customer value.

.