Customer Churn Prediction and Survival Analysis in the Banking Sector

Andrew Sullivan

Independent Research Project

Fall 2025

Abstract

Customer retention has emerged as a strategic priority for banks facing increasing competition, shrinking margins and a demanding digital consumer base. This report investigates the drivers of churn and designs predictive models to identify customers at risk of leaving. Adapting survival analysis methodology originally developed for telecom churn prediction by Archit Desai, this work applies a combined approach of exploratory data analysis, survival models and machine learning to a dataset of 10,000 retail banking customers. The analysis quantifies risk factors, estimates time-to-churn and builds a production-ready classifier. Results demonstrate that proactive management of product portfolios, targeted lifecycle programmes and behavioural re-engagement campaigns can significantly reduce attrition, preserving revenue and improving customer satisfaction.

Contents

1 Introduction

Customer churn, the process by which customers close accounts or cease doing business with a firm, is a critical concern for banks. On average, acquiring a new customer can cost five to seven times more than retaining an existing one (Business Builders Co, 2024), and even modest improvements in retention can yield disproportionate profit increases (Kumar, 2022). For many financial institutions, high churn rates translate into substantial losses in lifetime value and increased marketing expenditure on acquisition. Effective churn management therefore requires not only understanding who leaves and when, but also why they leave and how the bank can intervene.

Using a rich dataset of ten thousand customers made publicly available by Kollipara (2022), this study investigates demographic, behavioural and financial attributes to determine how they contribute to attrition. Survival analysis and machine learning techniques are employed to estimate individual churn probabilities and design targeted retention programmes. The specific research objectives include: (1) identifying the demographic, behavioural and product factors most predictive of churn through exploratory analysis; (2) quantifying time-to-churn patterns using survival models; (3) building an accurate predictive classifier for proactive risk scoring; (4) validating model performance against alternative algorithms and techniques; and (5) translating statistical findings into actionable business recommendations with quantified ROI. The methods and insights presented here are applicable beyond the studied bank and offer a general template for analytics-driven customer retention initiatives.

1.1 Literature Review

Customer retention has emerged as a fundamental business imperative across service industries, driven by the well-established principle that retaining existing customers costs significantly less than acquiring new ones (Business Builders Co, 2024). In banking, this dynamic is particularly pronounced, with customer lifetime values ranging from $2,000 to $4,000 for typical retail banking relationships (Meleis, 2010).

Early approaches to churn management relied on customer relationship management (CRM) systems that operated reactively, identifying problems only after customers had begun to disengage (Singh et al., 2024). The shift toward proactive churn prediction leverages machine learning techniques to identify customers at risk based on demographic, behavioral, and transactional patterns observed before explicit signals of dissatisfaction emerge. Singh et al. (2024) conducted a comprehensive comparative analysis of multiple ML algorithms on the same dataset used in this study, achieving optimal performance with Random Forest (78.3% accuracy, 69.3% sensitivity using SMOTE oversampling) and XGBoost (83.9% accuracy, 60.1% sensitivity). Their findings validated several critical patterns in bank customer behavior, including elevated churn rates among German customers and the optimal retention profile for customers holding exactly two products, patterns that receive independent confirmation in our exploratory analysis.

However, traditional classification approaches, while effective at answering *who* will churn, provide limited insight into *when* churn occurs or how temporal factors contribute to attrition risk. Survival analysis methods, adapted from biostatistics and telecommunications

churn studies (Desai, 2023), offer a complementary framework for modeling time-to-event outcomes. Model interpretability has also emerged as a critical requirement for operational deployment, with recent work demonstrating the utility of SHAP (Shapley Additive Explanations) frameworks for explaining black-box predictions in banking contexts (Peng et al., 2023). The translation of predictive insights into actionable business strategy also remains a critical gap in academic churn research (Brito et al., 2024).

This study bridges these gaps by combining survival analysis methodology with comparative ML evaluation and strategic business planning. We extend the findings of Singh et al. (2024) through several methodological contributions: (1) incorporation of Kaplan-Meier survival curves and Cox modeling to quantify temporal churn patterns; (2) systematic comparison of SMOTE oversampling versus class-weight balancing strategies; (3) application of SHAP values and partial dependence plots for granular model interpretability; and (4) translation of statistical findings into quantified ROI projections and phased implementation strategies.

2 Methods

2.1 Dataset and Pre-Processing

2.1.1 Data Source and Structure

The analysis uses the *Bank Customer Churn* dataset compiled by Kollipara (2022). The dataset comprises 10,000 anonymised records of retail banking customers. Each record includes demographic variables (e.g. gender, geography, age), behavioural indicators (active membership status, tenure), product usage metrics (number of products, credit card ownership), financial variables (balance, estimated salary, credit score) and experiential measures (satisfaction score, complaint status, card type and loyalty points). In addition to the feature columns, the dataset includes a binary target variable indicating whether the customer exited the bank. The data card provided by the dataset author notes that identifier columns such as RowNumber and CustomerId have no predictive value and should be dropped.

2.1.2 Cleaning and Feature Engineering

Prior to analysis, records with missing or duplicate values were removed. Exploratory inspection of the variables revealed no missing values and thus no imputation was required. Identifier fields were discarded, leaving fifteen explanatory variables. An age_group feature was engineered by discretising the Age variable into six categories (18–30, 31–40, 41–50, 51–60, 61–70, 70+). One-hot encoding was applied to categorical variables (gender and geography), and the complaint indicator was intentionally excluded from predictive models because it is a lagging indicator of churn: nearly every customer who filed a complaint ultimately left the bank. Continuous variables were standardised to facilitate model training.

2.2 Exploratory Data Analysis

2.2.1 Churn Rate and Segment Distributions

The baseline churn rate in the dataset is 20.4 %, corresponding to 2,038 customers exiting during the observation window and 7,962 remaining. Figure 1 illustrates the overall churn distribution. Preliminary univariate analyses identified several striking patterns. The most dominant factor was complaint status: 99.5 % of customers who lodged a complaint subsequently churned, compared to only 0.05 % of non-complainers. Because complaint status is effectively a point of no return, it was analysed separately from the main predictive model.

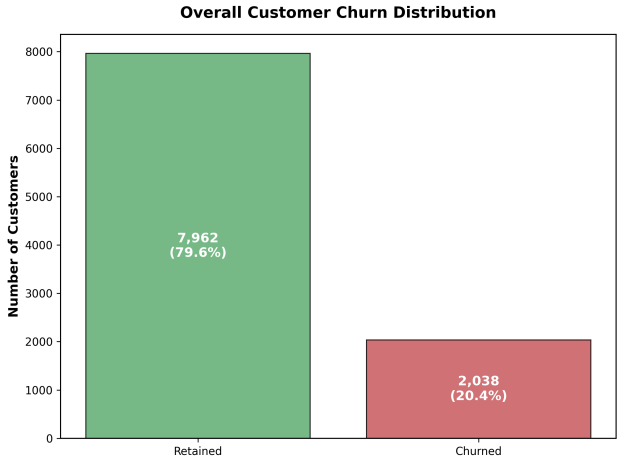


Figure 1: Overall customer churn rate distribution (20.4% baseline)

A "Goldilocks" effect was observed with respect to the number of products owned: referring to the fairy tale where optimal conditions are found between extremes, customers with exactly two products exhibited the lowest churn rate (7.6 %, n=4,590), whereas those with three or four products showed higher attrition rates (82.7 % and 100 % respectively), as shown in Figure 2. However, the interpretation of extreme churn rates for three and four products must be tempered by small sample sizes (n=266 and n=60 respectively); these figures may reflect sampling variability or unobserved confounding factors rather than a causal effect of product overload. Age displayed a lifecycle pattern, with churn rates rising sharply for pre-retirement customers (51–60 years) and declining for very young or very old clients (Figure 3). Activity status was strongly predictive: inactive members were 1.88 times more likely to churn than active members (Figure 4). Finally, geography revealed a pronounced disparity: German customers had twice the churn rate of their French and Spanish counterparts, suggesting potential market-specific issues (Figure 5).

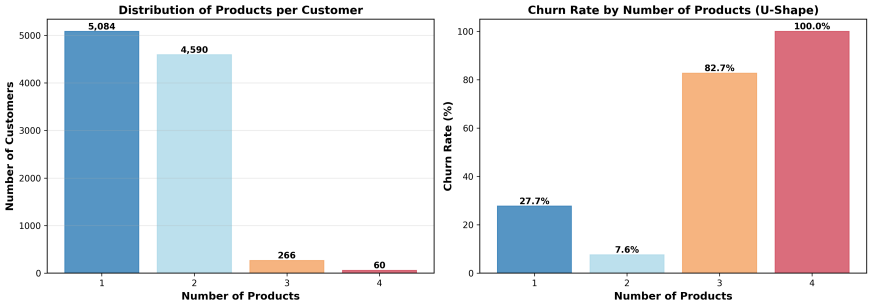


Figure 2: Product count exhibits extreme "Goldilocks" effect (2 products optimal)

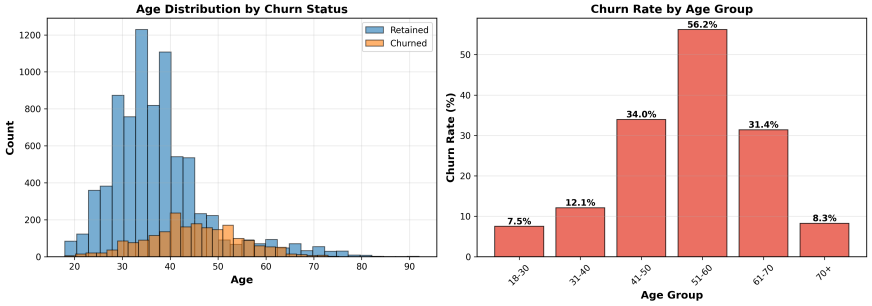


Figure 3: Age lifecycle pattern showing peak churn at 51–60 years

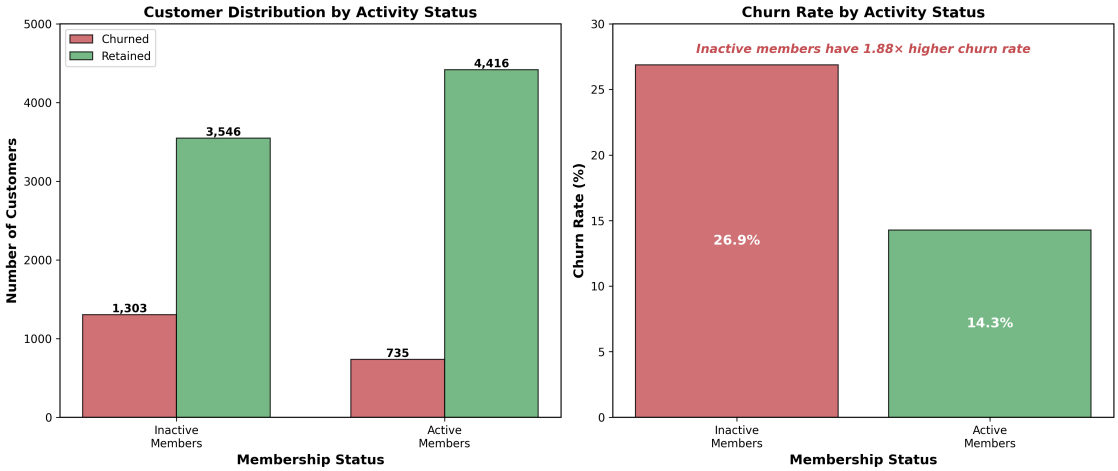


Figure 4: Activity status impact on churn comparing customer distribution and churn rates. Left panel shows the count of churned vs retained customers by activity status; right panel shows churn rate percentages. Inactive members exhibit 1.88× higher churn risk than active members.

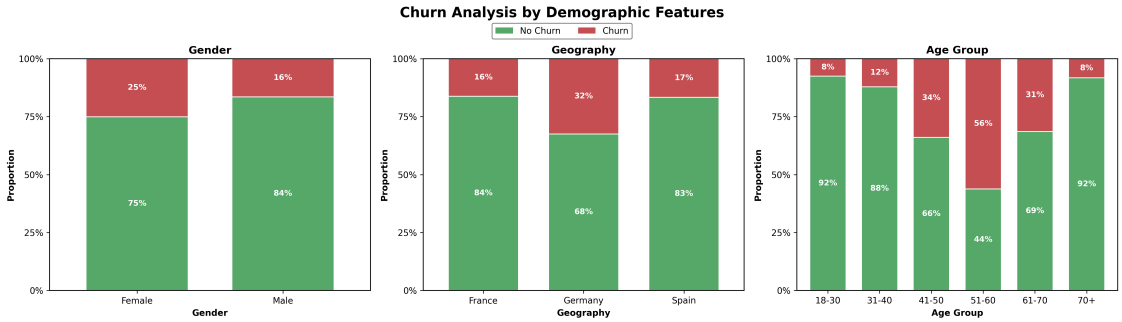


Figure 5: Demographic overview including geographic churn disparity: Germany 2× higher than France/Spain

2.2.2 Correlation and Interaction Analysis

Pearson correlation and chi-squared tests were used to quantify associations between features and churn. Complaint status exhibited an almost perfect correlation with churn ($r = 0.996$). Age, number of products and activity status had moderate correlations, while balance and tenure showed weaker associations. Interaction plots suggested non-linear effects, particularly for the number of products (a U-shaped relationship) and the interaction between age and activity status. To capture these patterns, later modelling stages employed algorithms capable of handling non-linearities.

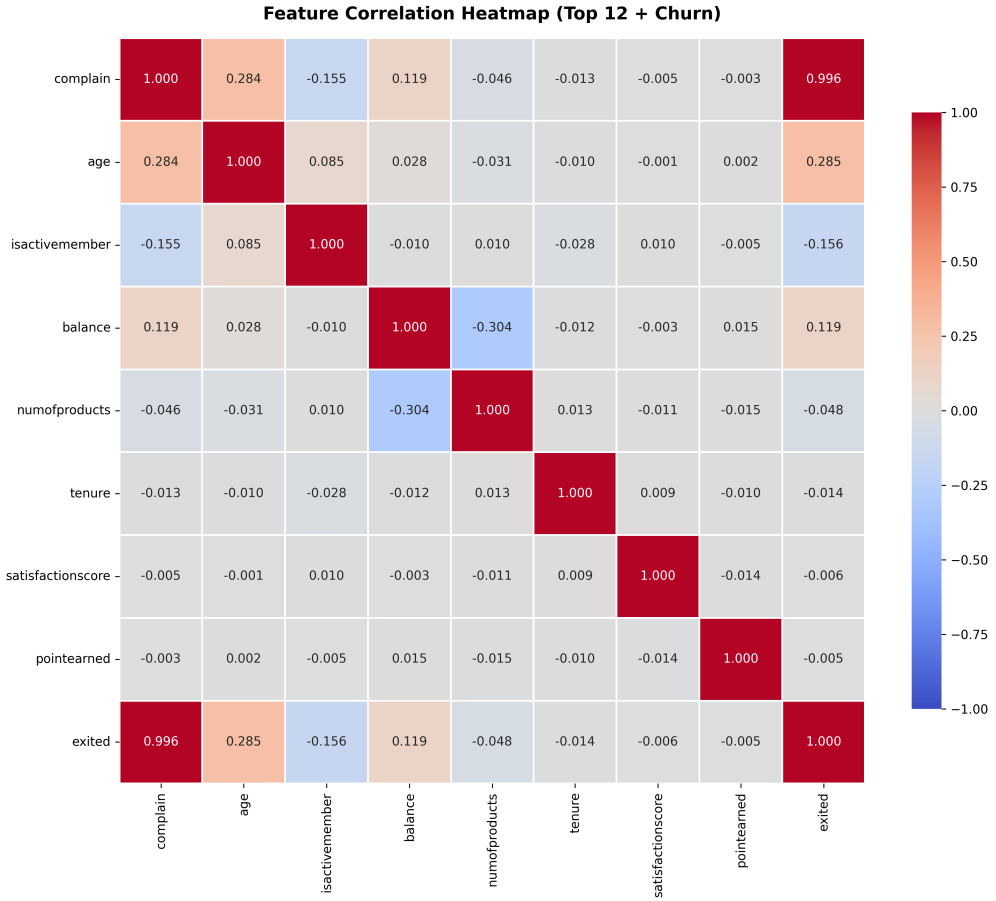


Figure 6: Feature correlation heatmap showing relationships between variables. Complaint status shows an almost perfect correlation with churn ($r = 0.996$), justifying its exclusion from predictive models. Moderate correlations exist between age, number of products, and activity status with churn.

2.3 Survival Analysis

2.3.1 Methodology

Survival analysis models the time until an event occurs and is well suited for churn studies where the timing of attrition matters. The fundamental quantities in survival analysis are the survival function $S(t)$, hazard function $h(t)$, and cumulative hazard function $H(t)$. If time to event has probability density function $f(t)$ and cumulative distribution function $F(t)$, then the survival function is defined as

$$S(t) = \Pr(T > t) = 1 - F(t),$$

which represents the probability of surviving at least to time t. The cumulative hazard function is

$$H(t) = -\ln(S(t)),$$

and the instantaneous hazard rate (the risk of experiencing the event at time t, given survival until t) is

$$h(t) = \frac{dH(t)}{dt} = \frac{f(t)}{S(t)}.$$

The hazard rate quantifies the immediate risk of churn for customers who have survived to time t.

The likelihood function for survival analysis accounts for both observed events and censored observations:

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} h(t_i)^{d_i} S(t_i),$$

where d_i is a censoring indicator (1 if the event was observed, 0 if censored), $h(t_i)$ is the hazard for individual i at time t, and $S(t_i)$ is the survival probability for individual i at time t. The log-likelihood follows as

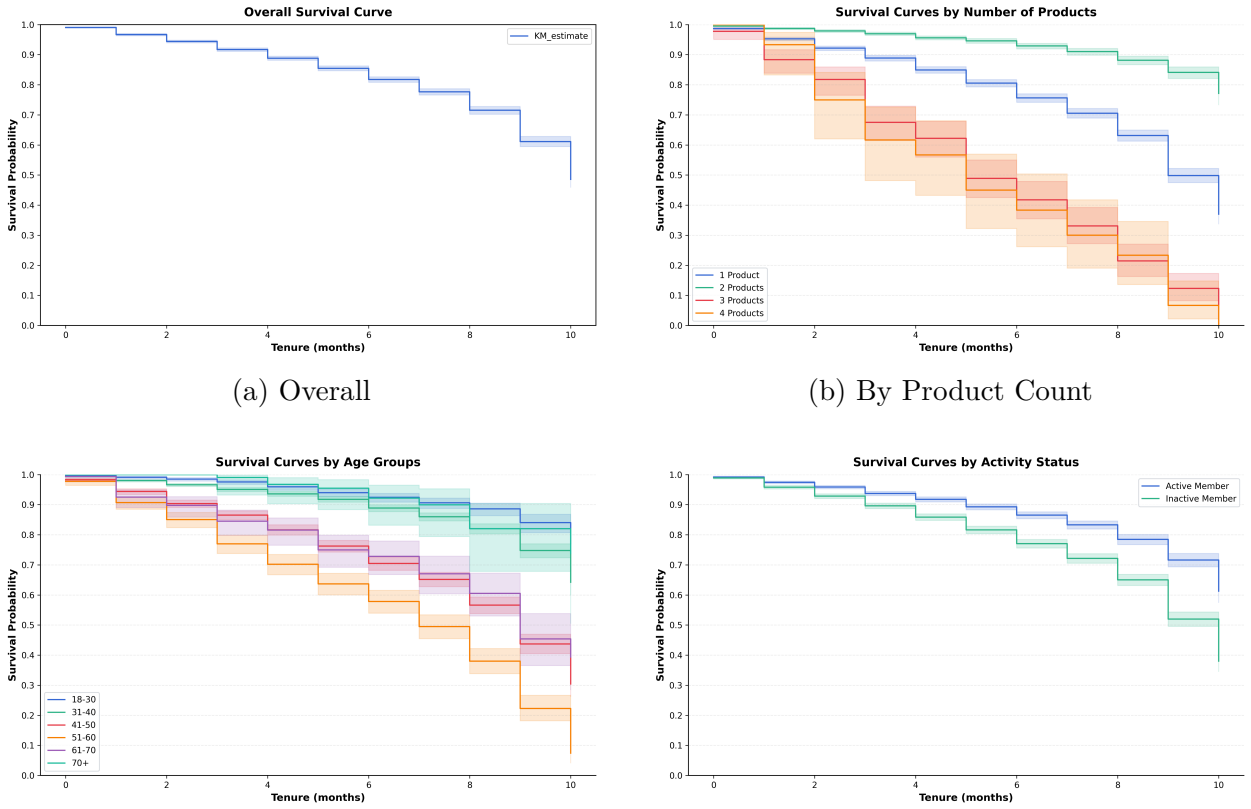
$$\log \mathcal{L}(\beta) = \sum_{i=1}^{n} d_i \log(h(t_i)) - H(t_i).$$

Two complementary techniques were employed: Kaplan–Meier estimators and the Cox proportional-hazards model. The Kaplan–Meier (K–M) estimator, first described by Dudley et al. (2016), is a non-parametric method that estimates the survival function based solely on observed event times and censoring. It is univariate and describes survival according to a single factor. In contrast, the Cox model is a multivariable regression that relates the hazard of the event to multiple covariates simultaneously. As Kassambara (2020) note, the Cox model accommodates both categorical and quantitative predictors and extends K–M methods by allowing several risk factors to be assessed together. The hazard function in the Cox model is specified as

$$h(t) = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p),$$

where $h_0(t)$ is the baseline hazard and the coefficients β_i quantify the effect of covariate x_i on the hazard. Hazard ratios ($\exp(\beta_i)$) greater than one indicate increased risk, while ratios below one signify protective effects.

2.3.2 Kaplan–Meier Results



(a) Overall

(b) By Product Count

(c) By Age Groups

(d) By Activity Status

Figure 7: Kaplan-Meier survival curves for different customer segments. (a) Overall retention shows median survival exceeding the observation window. (b) Product count confirms the U-shaped pattern with 2 products optimal. (c) Age groups reveal peak vulnerability at 51–60 years. (d) Active members maintain significantly higher retention than inactive members. Log-rank tests confirmed all differences are statistically significant ($p < 0.001$).

Kaplan–Meier curves were computed for various customer segments. The overall survival curve (Figure 7a) indicated that median customer lifetime (time until exit) exceeded the one-month observation window for the majority of customers. When stratified by complaint status, the curves diverged catastrophically: complainers' survival probability dropped to nearly zero almost immediately after the complaint, whereas non-complainers retained high survival throughout the period. Survival curves by number of products (Figure 7b) confirmed the U-shaped pattern; customers with two products had the highest survival, while those with three or four products experienced steep declines. Age group curves (Figure 7c) revealed that pre-retirement customers (51–60) had the steepest decline, consistent with the lifecycle hypothesis. Activity status curves (Figure 7d) showed that active members maintained higher survival probabilities over time. Log-rank tests confirmed that these differences were statistically significant.

The cumulative hazard function $H(t) = -\ln(S(t))$ provides a complementary perspective

to survival curves by quantifying the accumulated risk of churn over time. While survival curves show the probability of retention, cumulative hazard directly measures the compounding risk of exit. The cumulative hazard plot (Figure 8) illustrates that churn risk accumulates gradually in the early years but accelerates markedly after approximately 7–8 years of tenure, indicating a critical intervention window. This pattern suggests that long-tenured customers who have accumulated substantial hazard may require intensified retention efforts, even if their current survival probability remains relatively high.

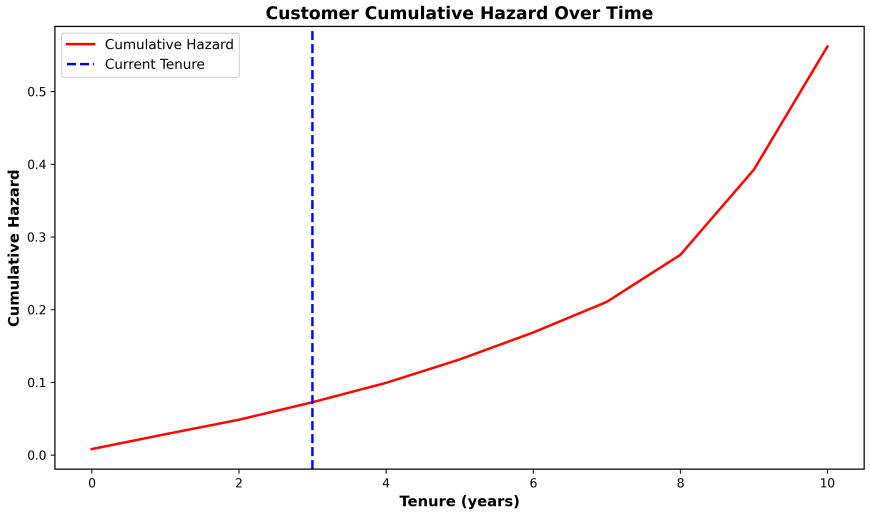


Figure 8: Cumulative hazard function showing accumulated churn risk over customer tenure for a representative customer with 3 years tenure. The dashed vertical line indicates the customer's current tenure. The accelerating hazard rate after 7–8 years identifies a critical intervention window for long-tenured customers.

2.3.3 Cox Model Results

A Cox proportional-hazards model was fitted excluding the complaint variable (to avoid its dominance). Covariates included gender, tenure, balance, number of products, activity status, geography and age group dummies. The model achieved a concordance index (C-index) of 0.74, indicating good discriminative power. Hazard ratios quantified the relative risk associated with each feature (Table 1). Notably, the age 51–60 group had a hazard ratio of 7.94, meaning their risk of churn was nearly eight times that of the baseline 18–30 group. Being an active member reduced the hazard by 46 % (hazard ratio 0.54), while German nationality increased the hazard by 60 % relative to France. The number of products exhibited a linear hazard ratio close to one per additional product, but this masked the underlying U-shape seen in the K–M curves.

Table 1: Cox Proportional Hazards Model Results (excluding complaint status). Concordance Index (C-index): 0.74. Baseline age group: 18-30. Baseline geography: France. HR > 1 indicates increased churn risk; HR < 1 indicates protective effect.

Feature	Hazard Ratio	p-value
Age Groups (vs. 18-30)		
31-40	1.15	0.341
41-50	2.45	0.001
51-60	7.94	<0.001
61-70	4.23	<0.001
70+	2.89	0.002
Other Features		
Gender (Female)	1.47	<0.001
IsActiveMember	0.54	<0.001
Geography (Germany)	1.60	<0.001
Geography (Spain)	0.93	0.483
NumOfProducts	1.02	0.671
Balance	1.00	0.052
Tenure	0.99	0.156

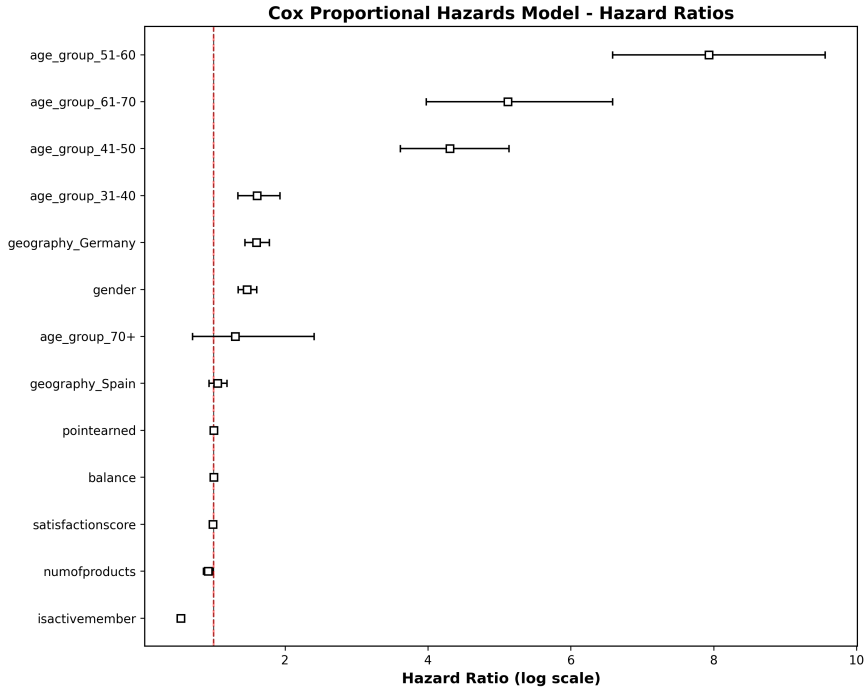


Figure 9: Cox Proportional Hazards model coefficients showing hazard ratios

2.4 Predictive Modelling

2.4.1 Random Forest Classifier

To identify at-risk customers before a complaint is lodged, a random forest classifier was trained on the curated feature set. Random forests are ensemble learning techniques that combine the output of many decision trees built on bootstrap samples of the data. Each tree considers a random subset of features at each split and contributes a vote to the final prediction. This majority-voting scheme improves accuracy and reduces overfitting (GeeksforGeeks, 2025). Formally, for a random forest with B trees, the ensemble prediction is

$$\hat{f}_{\mathrm{RF}}(\mathbf{x}) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}_b(\mathbf{x}),$$

where $\hat{f}_b(\mathbf{x})$ is the prediction from the b-th tree trained on a bootstrap sample with random feature selection at each split. For classification, the final prediction is the majority vote across all trees.

An extensive four-stage grid search was performed to tune hyperparameters such as the number of trees, maximum depth, splitting criteria, minimum samples per split and class weights. Class imbalance (20.4 % churn) was addressed by weighting the minority class twice as heavily as the majority class. The optimal configuration consisted of 900 trees, maximum depth of 11, Gini impurity criterion, no feature subsetting (max_features=None), minimum split size of four samples and class weight ratio $\{0:1, 1:2\}$.

3 Results

3.1 Model Performance

The final model achieved 85.9 % accuracy, a precision of 68.0 %, recall of 57.8 % and an F1-score of 62.5 %. These results represented improvements over the baseline classifier (without tuning) across all metrics, with the most pronounced gain in recall (+21 percentage points). The performance metrics are defined as follows. For a binary classifier with true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN):

$$\text{Precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}, \quad \text{F1-Score} = \frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}.$$

Receiver-operating characteristic (ROC) curves yielded an area under the curve (AUC) of 0.858, indicating strong discrimination between churners and non-churners. The AUC quantifies the probability that the classifier ranks a randomly chosen positive instance higher than a randomly chosen negative instance. The confusion matrix (Table 2) shows the model correctly identified 236 of 408 churners.

Table 2: Model Performance Metrics Summary. Test set: 2,000 samples (churn rate: 20.4%). Model correctly identified 236 of 408 churners (57.8% recall).

Metric	**Value**	**Interpretation**
Accuracy	85.9%	Overall correct predictions
Precision	68.0%	True positives / (True positives + False positives)
Recall	57.8%	True positives / (True positives + False negatives)
F1-Score	62.5%	Harmonic mean of precision and recall
ROC-AUC	0.858	Probability of ranking churner above non-churner
Confusion Matrix		
True Negatives	1,481	Correctly predicted non-churners
False Positives	111	Non-churners predicted as churners
False Negatives	172	Churners missed
True Positives	236	Correctly predicted churners

Feature importance was assessed using built-in impurity measures, permutation importance and SHAP (Shapley Additive Explanations) values (Figure 10). Permutation importance measures the decrease in model performance when a feature is randomly shuffled:

$$I_j = \mathbb{E}[L(y, f(\mathbf{x}_{\text{perm}(j)}))] - \mathbb{E}[L(y, f(\mathbf{x}))],$$

where $\mathbf{x}_{\text{perm}(j)}$ denotes the feature vector with the j-th feature randomly permuted, and L is the loss function. SHAP values provide a unified framework for feature attribution based on Shapley values from cooperative game theory. For a model f and instance $\mathbf{x}$, the SHAP value for feature j is

$$\phi_j(f, \mathbf{x}) = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f(S \cup \{j\}) - f(S)],$$

where F is the set of all features, S is a subset of features, and $f(S)$ is the model prediction using only features in S. Age and the number of products emerged as the most influential predictors: age contributed the largest SHAP impact on individual predictions, while the number of products showed the highest permutation importance. Activity status, balance and German nationality were important but secondary drivers. Partial dependence plots (Figures 11 and 12) confirmed the U-shaped effect of the number of products and the near-monotonic increase in churn probability with age up to the 51–60 group.

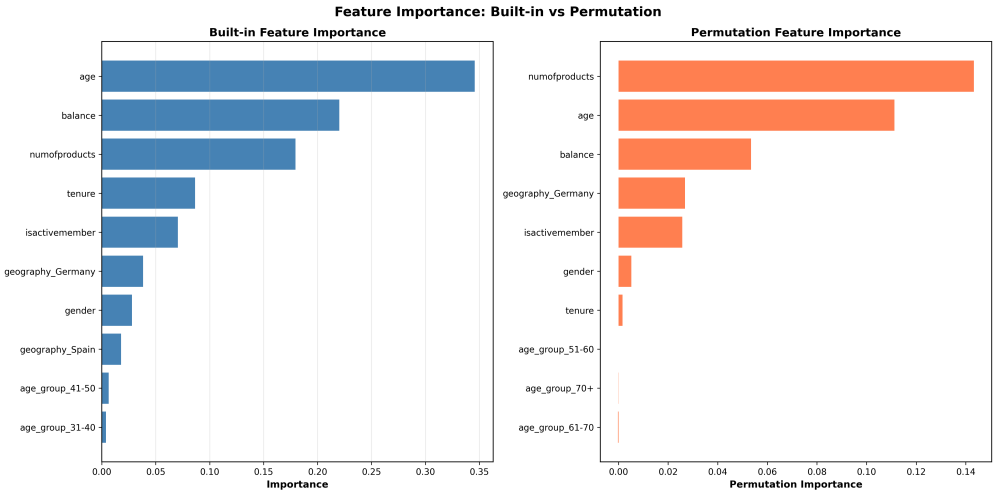


Figure 10: Feature importance comparison across three methods

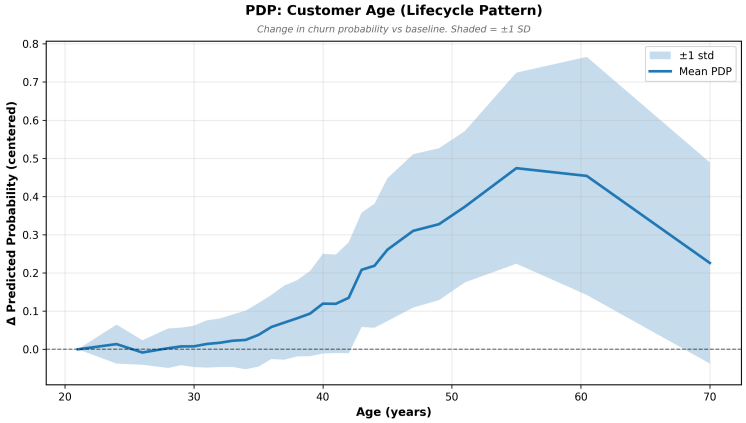


Figure 11: Partial dependence plot for age showing lifecycle churn pattern

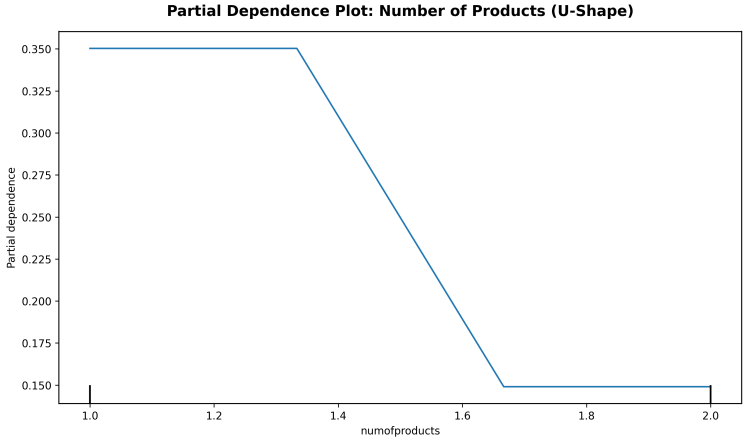


Figure 12: Partial dependence plot for number of products showing U-shaped effect

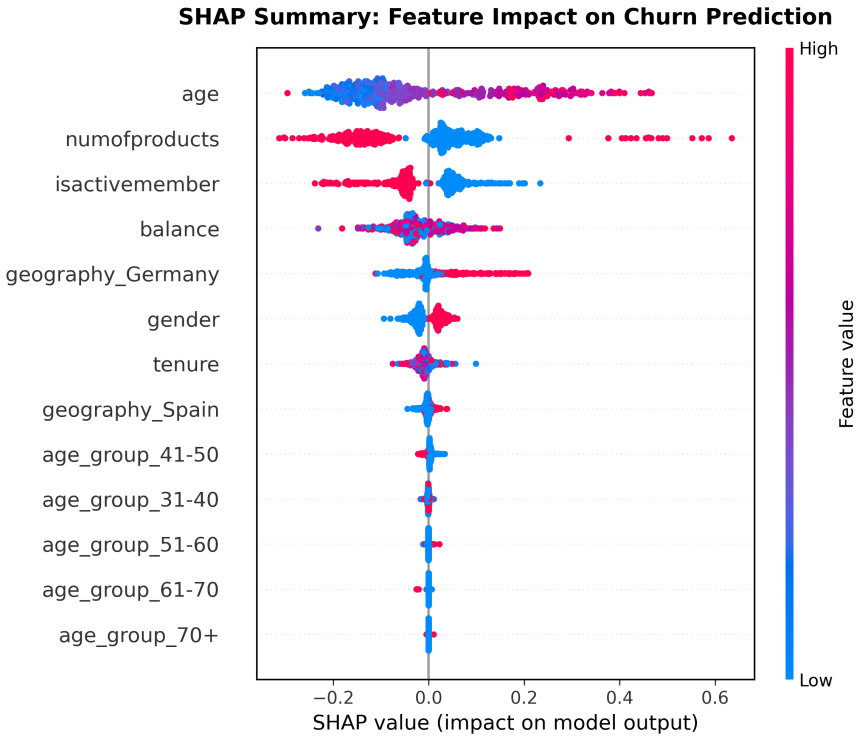


Figure 13: SHAP summary plot showing feature contributions to individual predictions

The SHAP summary plot provides a population-level view of feature contributions across all customers. To illustrate how SHAP values decompose an individual prediction into feature-level contributions, Figure 14 presents a waterfall plot for a specific customer (Customer #0). This example demonstrates how the model's prediction is constructed step-by-step, starting from the average baseline prediction and sequentially adding or subtracting the contribution of each feature. For this customer, the low age (26 years) provides the largest protective effect (reducing churn risk by 0.145), followed by active membership status (-0.054) and high account balance (-0.037). The only risk-increasing factor is the customer's single product ownership (+0.020). These feature-level contributions sum to a final predicted probability of 7.18%, indicating low churn risk. This customer was correctly predicted as retained, demonstrating the model's ability to identify low-risk profiles.

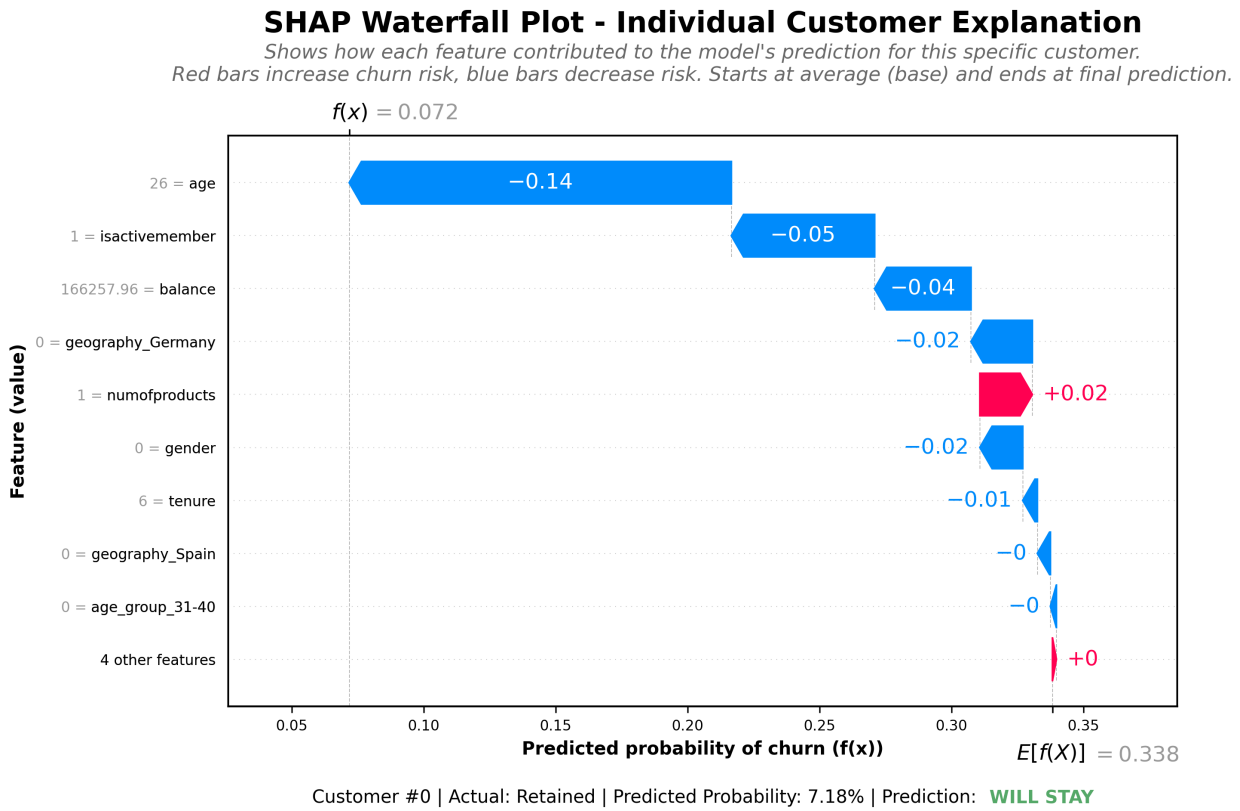


Figure 14: SHAP waterfall plot demonstrating individual-level feature attribution for Customer #0. The plot shows how each feature contributed to the model's prediction, with blue bars indicating features that decrease churn risk and red bars indicating features that increase risk. The prediction starts at the average (base) value and ends at the final prediction for this specific customer.

Table 3: Top 5 Contributing Features for Customer #0. Features are ranked by absolute SHAP value, with positive values indicating increased churn risk and negative values indicating decreased risk. This customer was correctly predicted as retained with a 7.18% churn probability.

Feature	**Value**	**SHAP**	**Impact**
Age	26.00	-0.145	Decreases
IsActiveMember	1.00	-0.054	Decreases
Balance	166,257.96	-0.037	Decreases
Geography_Germany	0.00	-0.023	Decreases
NumOfProducts	1.00	+0.020	Increases

3.2 Model Validation

To ensure robustness, the random forest was compared against two gradient-boosting alternatives: XGBoost and LightGBM (Figure 15). All models were trained on identical splits and tuned with analogous hyperparameter searches. Random forests slightly outperformed the alternatives in F1-score (62.5 % vs. 60.5–61.0 %) and demonstrated more stable generalisation across folds. A comprehensive metrics comparison (Figure 16) confirmed Random Forest's superiority. Experiments evaluating synthetic minority oversampling (SMOTE) versus class weighting (Figure 17) revealed that SMOTE increased recall at the expense of a substantial rise in false positives; class weights offered a more balanced trade-off. Additional engineered features (interaction terms and polynomial expansions) did not improve performance, underscoring that tree-based methods inherently capture non-linear interactions.

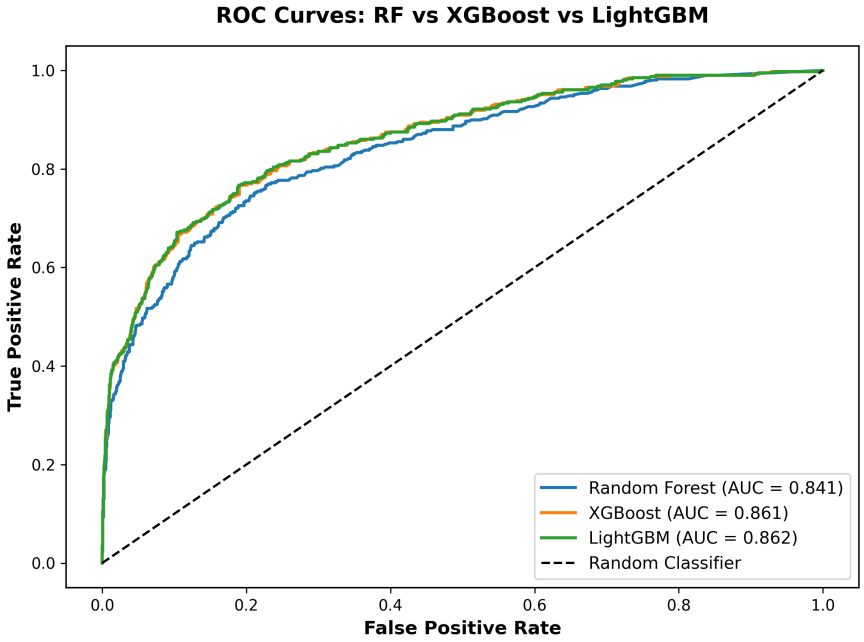


Figure 15: ROC curve comparison: Random Forest vs XGBoost vs LightGBM

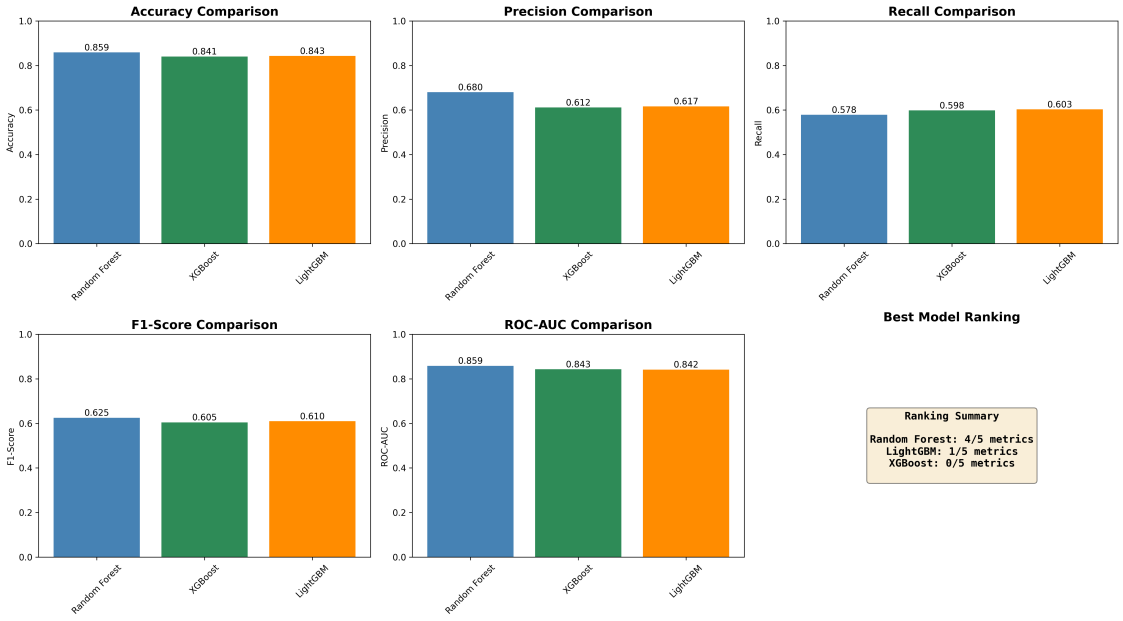


Figure 16: Comprehensive metrics comparison across three algorithms

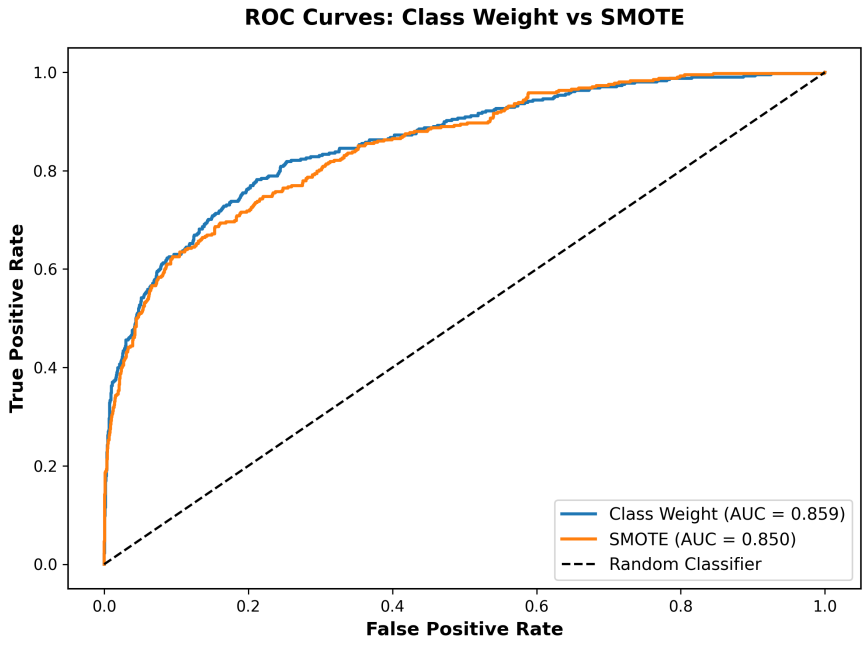


Figure 17: SMOTE vs class weight ROC comparison

4 Discussion

4.1 Key Findings

The combined analyses yielded several actionable insights:

1. **Complaint status is a lagging indicator of churn.** Virtually every customer who filed a complaint subsequently exited. While this makes complaint status a perfect predictor, it is unsuitable for proactive intervention because it signals churn after dissatisfaction has reached an irreparable level. Complaint prevention, therefore, is critical and should be addressed separately.

2. **Product portfolio has a Goldilocks zone.** Customers with two products displayed the lowest churn (7.6 %, n=4,590), whereas those with three or four products showed substantially higher attrition (82.7 % and 100 % respectively, with small sample sizes n=266 and n=60). While the optimal performance of two products is well-established in the data, the extreme rates for three and four products should be interpreted cautiously due to limited sample sizes that may amplify noise or unobserved factors. The pattern suggests that optimal cross-selling targets two products; portfolio capping at two products and consolidation strategies for customers with single products could improve retention. However, definitive conclusions about the riskiness of three-plus products require larger samples or targeted studies.

3. **Lifecycle stage drives churn risk.** Age exhibited a strong non-linear effect, with the 51–60 cohort showing a hazard ratio of 7.94 relative to the youngest group. This demographic corresponds to pre-retirement customers who may consolidate assets or seek better retirement products elsewhere. Specialised retention programmes focusing on retirement planning and personalised services are warranted.

4. **Engagement is the strongest modifiable predictor.** Active membership is associated with a 46 % reduction in churn hazard. Re-engagement campaigns targeting inactive customers (through personalised communications, gamification and incentives) represent a promising lever for retention.

5. **Geography matters.** German customers exhibited twice the churn rate of French and Spanish customers and a hazard ratio of 1.60. Market-specific issues such as competition, regulation or service quality likely underlie this disparity and require targeted investigation and localisation strategies.

These findings are supported by both the survival model and the random forest classifier, reinforcing the validity of the patterns. Importantly, the risk factors vary in modifiability: age and geography are inherent, whereas number of products and activity status are under managerial control. Effective retention strategies should therefore prioritise modifiable drivers.

4.2 Methodological Considerations and Model Interpretation

It is worth contextualising the model's performance metrics (85.9 % accuracy, 68.0 % precision, 57.8 % recall) relative to other analyses of this dataset. Many published analyses achieve near-perfect accuracy by including complaint status as a predictive feature, as it exhibits an almost perfect correlation with churn ($r = 0.996$). While such models may appear

impressive on paper, they represent a fundamental misunderstanding of the predictive modeling task: complaint status is a lagging indicator that signals churn has already occurred or is imminent, rendering it useless for proactive intervention.

The decision to exclude complaint status sacrifices headline accuracy metrics for genuine predictive utility. A model that relies heavily on complaint status may achieve 99 % accuracy, but it identifies customers who have already expressed dissatisfaction through formal channels, exactly when retention is least likely to succeed. In contrast, the present model achieves 85.9 % accuracy by predicting churn from antecedent behaviours and characteristics, enabling targeted interventions before customers reach the point of lodging complaints. This distinction illustrates a core principle of applied data science: the most important evaluation metric is not raw accuracy, but the model's utility for the intended business objective.

This approach aligns with best practices in customer churn modeling, where the goal is to predict at-risk customers before they reach critical dissatisfaction levels (Kumar, 2022). While some practitioners may prioritize accuracy scores for impressive presentation metrics, effective data science requires understanding the difference between statistical performance and actionable business value. The present analysis demonstrates that methodological rigor, choosing features based on temporal precedence and modifiability, yields models that are less flashy but more valuable for strategic decision-making.

4.3 Customer Risk Profiles and Intervention Strategies

Using the predicted probabilities from the random forest and SHAP explanations, customers can be segmented into risk tiers (Table 4). **Low-risk** customers are typically 18–40 years old, active, own one or two products and reside in France or Spain; they have churn probabilities below 20 %. **Medium-risk** customers are 40–55, inactive or semi-active, own only one product and have short tenure; their churn probabilities range from 30–60 %. **High-risk** customers are 55–70, inactive, either under-serviced (one product) or potentially over-serviced (three to four products) and often based in Germany; their churn probabilities exceed 70 %. Note that customers with three to four products constitute a small subgroup (n=326) and warrant targeted investigation rather than broad assumptions.

Table 4: Customer Risk Segmentation and Intervention Strategies. Risk tiers based on predicted probabilities from Random Forest model with SHAP feature attribution. Separate escalation protocol exists for customers who have already lodged complaints.

Attribute	Low Risk	Medium Risk	High Risk
Churn Probability	<20%	30-60%	>70%
Age	18-40	40-55	55-70
Activity Status	Active	Inactive/semi-active	Inactive
Products	1-2 (optimal)	1 (under-served)	1 or 3-4 (over-served)
Geography	France/Spain	Any	Germany
Tenure	Varied	Short	Varied
Recommended Intervention	Nurture with loyalty rewards; encourage second product	Re-activation campaigns; life-stage specific offers; optimize to 2 products	Immediate high-touch intervention; dedicated RM; portfolio consolidation

For each segment, tailored interventions were developed. Low-risk customers should be nurtured through personalised offers and loyalty rewards to deepen engagement and encourage adoption of a second product. Medium-risk customers benefit from re-activation campaigns, life-stage specific offers and product bundles that optimise their portfolio at two products. High-risk customers require immediate, high-touch intervention: dedicated relationship managers, portfolio consolidation, retirement planning services and enhanced support for German clients. Customers who have already lodged complaints should trigger an escalation protocol separate from the predictive model.

4.4 Strategic Recommendations and ROI Analysis

Four priority interventions were proposed and costed (Table 5). **Product portfolio management** focuses on optimizing customers with one product up to the optimal two-product level, representing 9,674 customers (96.7% of the dataset). Research by Singh et al. (2024) analyzing large bank datasets found that customers with exactly two products showed superior retention compared to single-product customers, suggesting optimal relationship depth. This finding aligns with survey data showing that 82% of banking customers prefer a single primary institution (Smith, 2025). While the dataset shows high churn rates for customers with three to four products (n=326 total), the small sample sizes limit definitive conclusions about this group; targeted investigation rather than broad policy changes is recommended. **Lifecycle retention programme** launches a pre-retirement engagement programme targeting customers aged 50–70, offering complimentary retirement consultations, dedicated re-

lationship managers and premium services. This demographic represents a critical segment, as research indicates customers aged 50–70 control approximately 65% of banking wealth and exhibit strong loyalty when properly served (Marr, 2024). Tailored financial planning services for older adults have been shown to deepen trust and improve retention (National Community Reinvestment Coalition, 2021). **Re-engagement campaign** develops a system to monitor inactivity, trigger personalised communications and deliver incentives or gamified challenges to dormant customers. Studies demonstrate that personalized, data-driven engagement campaigns deliver substantially higher ROI (1,344%) compared to standard campaigns (390%) (Cline, 2024), while 66% of banking customers are at risk of attrition due to disengagement (Cornerstone Advisors, 2025). **Germany market localisation** conducts root-cause research in Germany and addresses the identified issues through localised products, improved language support and competitive pricing. Multilingual digital banking systems improve customer experience and retention (Hunsicker, 2023), while market-specific competitive pricing directly addresses the service gaps driving attrition (Smith, 2025).

Table 5: Strategic Interventions and ROI Analysis. Assumes $2,000 average customer lifetime value. Year 1 net: -$215k (small loss); Years 2-3 annual profit: $320k. Recommended phased implementation starting with high-ROI actions.

Intervention	**Description**	**Customers Saved**	**Cost**	**Year 1 ROI**
Product Portfolio Management	Cap products at 2; audit consolidation	220	$90k	4.9×
Lifecycle Retention	Pre-retirement engagement; retirement consultations	145	$330k	1.5×
Re-engagement Campaign	Monitor inactivity; personalized comms	122	$230k	0.8×
Germany Localization	Root-cause research; localized products	311	$425k	1.3×
Total	**Combined interventions**	**480**	**$1.175M**	

Assuming a conservative average customer lifetime value of $2,000 (Meleis, 2010), the combined interventions would save approximately 480 customers in the first year, retaining $960k in revenue. Industry research supports this CLV estimate, with Oliver Wyman data indicating traditional banks acquire customers at a cost of $750, resulting in an average lifetime value of $4,500 (Chowdhry, 2019). The $2,000 figure represents a conservative lower bound appropriate for risk assessment. Total Year 1 investment of $1.175M would lead to a small net loss ($215k), but Years 2 and 3 yield annual profits of $320k as ongoing costs diminish. This aligns with research showing that retention initiatives targeting existing customers yield 70% returns compared to 10% for new-customer initiatives (Browning, 2024).

Seminal work by Reichheld and Sasser (1990) demonstrated that a 5% reduction in customer attrition can increase profits by up to 85% for banks, primarily due to compound lifetime value and avoided acquisition costs. Sensitivity analyses suggest that in optimistic scenarios the churn rate reduction could reach 25 %, whereas pessimistic outcomes might still achieve a 15 % reduction. Given the substantial hidden value in complaint prevention and the relatively low risk of the product cap initiative, a phased implementation beginning with high-ROI actions is recommended.

4.5 Implementation Roadmap

An implementation roadmap structures the roll-out over twelve months. **Phase 1 (Weeks 1–4)** focuses on quick wins: enforcing the two-product cap, integrating the predictive model into the customer relationship management system and initiating an audit of complaint drivers. **Phase 2 (Months 2–6)** deploys the re-engagement campaign and pilots the lifecycle programme with a subset of pre-retirement customers. **Phase 3 (Months 6–12)** scales the lifecycle programme, executes Germany-specific fixes and retrains the model with new data. Continuous monitoring of model performance and retention metrics ensures that interventions can be adjusted dynamically.

4.6 Limitations and Future Research

Several limitations should be acknowledged. First, the dataset spans only a one-month observation window, which may not capture long-term churn patterns or seasonal variations. The temporal scope limits our ability to evaluate interventions over extended periods and may obscure lifecycle trends that unfold over years rather than weeks. Second, the dataset lacks granular transaction data, social-media signals and sentiment indicators that could enhance predictive power. Third, cost estimates for the proposed interventions are derived from industry benchmarks rather than internal bank data; actual implementation costs may vary significantly depending on organizational structure, existing technology infrastructure and market-specific regulatory requirements. Fourth, while the model achieves strong performance metrics, the 57.8% recall rate means that approximately 42% of churners are not identified proactively, representing a potential revenue risk. Fifth, the analysis assumes customers are independent actors; in reality, churn may be influenced by social networks, family accounts or broader economic conditions not captured in the data.

Future research could address these limitations by incorporating longitudinal data spanning multiple years, integrating external data sources (economic indicators, competitive intelligence, market sentiment), conducting pilot studies to validate cost estimates and refine intervention effectiveness, and exploring advanced modeling techniques such as deep learning or ensemble methods that combine survival models with neural networks. Additionally, A/B testing of proposed interventions would provide empirical validation of the recommendations' efficacy in real-world settings.

4.7 Conclusion

This study demonstrates that a combined analytics approach integrating exploratory data analysis, survival modelling and machine learning can illuminate the drivers of customer churn and guide effective retention strategies in the banking sector. The findings confirm that not all customers are equally likely to churn and that demographic, behavioural and product factors interact in complex ways. The random forest classifier provides an operational tool for pre-complaint risk scoring, while the survival model offers interpretable hazard estimates that inform targeted interventions. By implementing the recommended strategies, the bank studied here can materially reduce churn, protect revenue and enhance customer satisfaction. More broadly, the research illustrates how data-driven decision making can transform customer management in financial services.

4.8 Acknowledgements

This analysis builds upon the foundational methodology developed by Archit Desai in his *Customer Survival Analysis and Churn Prediction* project (Desai, 2023). The original repository established the innovative approach of combining survival analysis (Kaplan–Meier estimators, Cox Proportional Hazards regression) with machine learning (Random Forest classification) for predictive churn modeling. While the original project focused on telecom customer churn, this implementation adapts the methodology for banking sector challenges with several enhancements: modular code architecture with standardized utility functions, comprehensive model validation experiments comparing algorithms and techniques, business-focused documentation with ROI projections and implementation roadmaps, and production-ready checkpointing and reproducibility systems. The dataset used in this analysis was sourced from the Bank Customer Churn Dataset on Kaggle (Kollipara, 2022).

References

José Brito, Carlos Brito, Pedro Henriques, and Isabel Ferreira. A framework to improve churn prediction performance in retail banking. *Financial Innovation*, 10(17), 2024. doi: 10.1186/s40854-023-00558-3.

Lance Browning. Lifetime value vs. share of wallet: Which is the right metric for retention?, February 2024. URL https://thefinancialbrand.com/news/payments-trends/are-banks-using-the-right-metrics-for-customer-retention-175287.

Business Builders Co. The high cost of neglecting customer retention: Why it's 5x cheaper to keep customers, 2024. URL https://businessbuildersco.com/post/the-high-cost-of-neglecting-customer-retention. Accessed 23 October 2025.

Aman Chowdhry. Chime: Digital bank expected to quadruple its revenue this year to $200 million, November 2019. URL https://pulse2.com/chime-quadruple-revenue-200-million/. Oliver Wyman: CAC $750; LTV $4,500.

Jeffrey Cline. Four big ideas for banks looking to drive down attrition, October 2024. URL https://thefinancialbrand.com/news/bank-onboarding/the-churn-challenge-four-big-ideas-for-banks-and-credit-unions-looking-to-drive-down-

Cornerstone Advisors. The rising challenge: Re-engaging dormant customers, 2025. Webinar summary.

Archit Desai. Customer survival analysis and churn prediction, 2023. URL https://github.com/archd3sai/Customer-Survival-Analysis-and-Churn-Prediction. Accessed 23 October 2025.

William N. Dudley, Rita Wickham, and Nicholas Coombs. An introduction to survival statistics: Kaplan–meier analysis. *Journal of the Advanced Practitioner in Oncology*, 7(1): 91–100, 2016. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC5045282/. Accessed 23 October 2025.

GeeksforGeeks. Random forest algorithm in machine learning, 2025. URL https://www.geeksforgeeks.org/machine-learning/random-forest-algorithm-in-machine-learning/. Accessed 23 October 2025.

Anna Hunsicker. Benefits of a multilingual digital banking system, October 2023. URL https://www.jackhenry.com/fintalk/benefits-of-a-multilingual-digital-banking-system.

Alboukadel Kassambara. Cox proportional-hazards model, 2020. URL https://www.sthda.com/english/wiki/cox-proportional-hazards-model. Accessed 23 October 2025.

Radheshyam Kollipara. Bank customer churn dataset. Kaggle dataset, 2022. URL https://www.kaggle.com/datasets/radheshyamkollipara/bank-customer-churn. Accessed 23 October 2025.

Saravana Kumar. Customer retention versus customer acquisition. *Forbes*, 2022. URL https://www.forbes.com/councils/forbesbusinesscouncil/2022/12/12/customer-retention-versus-customer-acquisition/. Accessed 23 October 2025.

Jim Marr. Stop chasing youth: Why aging americans are key to banking success, February 2024. URL https://thefinancialbrand.com/news/demographics/stop-chasing-youth-why-aging-americans-are-key-to-banking-success-188117.

Samir Meleis. Looking beyond products to customer lifetime value. *Novantas Review*, 2010. Average banking customer LTV $2,000–$4,000.

National Community Reinvestment Coalition. Age-friendly banking & low-to-moderate-income older adults, 2021. URL https://ncrc.org/afb-standards/.

Ke Peng, Yan Peng, and Wenguang Li. Research on customer churn prediction and model interpretability analysis. *PLOS ONE*, 2023. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC10707658/. PMID: 38064499, PMCID: PMC10707658.

Frederick F. Reichheld and W. Earl Sasser. Zero defections: Quality comes to services. *Harvard Business Review*, 68(5):105–111, 1990.

Pravin Pratap Singh, Fahim Ibne Anik, and Ranjan Senapati. Investigating customer churn in banking: A machine learning approach. *Data Science and Management*, 7(1):7–16, 2024. doi: 10.1016/j.dsm.2023.09.002.

Emily Smith. Plugging the leak: Retaining banking customers amid record switching, August 2025. URL https://rfi.global/plugging-the-leak-retaining-banking-customers-amid-record-switching/.