Customer Churn Prediction

Milestone 1 · Board Brief

April 2026

Contents / Agenda

- Executive Summary
- Business Problem & Solution Approach
- Data Overview
- EDA — Key Drivers of Churn
- Data Preprocessing & Feature Engineering
- Model Evaluation Criterion
- Plan of Action – Modeling Phase
- Risks & Mitigations
- Key Takeaways & Immediate Moves

Executive Summary

- Retention problem, not a growth problem: 16.8% of 11,260 accounts churn over 12 months — a moderate class imbalance with direct P&L impact.
- Compounded loss: each account bundles multiple end customers — one lost account removes several relationships, recurring revenue and cross-sell options at once.
- Revenue Assurance constraint is binding: blanket retention offers are rejected on margin grounds — offer cost must stay below expected revenue preserved, account by account.
- Churn is concentrated in identifiable segments — new accounts, Regular Plus, Single, Tier-3, non-card payers and complainers — not diffuse noise.
- Foundation delivered: audited dataset (KNN/mode imputation, 1st/99th winsorization) plus 5 Revenue-Assurance-legible engineered features.
- Next phase converts these patterns into an account-level propensity score and an economically defensible retention list.

Business Problem Overview and Solution Approach

Business Problem

- Low switching costs + aggressive competitor pricing: ~1 in 6 accounts churns annually — retention is the highest-leverage line in the P&L.
- Compounded loss + asymmetric economics: each account bundles multiple end customers; acquisition costs far exceed retention costs — small churn reductions compound.
- Revenue Assurance constraint: offer cost must stay below expected revenue preserved, account by account — blanket campaigns are rejected on margin grounds.

Solution Approach

- Account-level churn propensity from supervised classification, trained on 11,260 accounts and 19 tenure / engagement / profile variables.
- Ranking on expected revenue-at-risk = churn probability × account revenue — the list Revenue Assurance can approve against.
- Segmented retention plays tied to tenure, segment, complaint status and payment method — each play carries its own unit-economics case.
- Explainability by construction: tree-based feature importance and SHAP alongside every score — retention actions defensible to the business and Revenue Assurance.

Data Overview

Dataset shape

- 11,260 unique accounts × 19 variables, aggregated over the last 12 months — no duplicates, no AccountID collisions.
- Target: Churn (1 = churned, 0 = retained). Distribution 83.16% retained / 16.84% churned — moderate imbalance makes accuracy a misleading KPI.
- AccountID removed before analysis (unique key, no predictive value).

Variable dimensions

- Financial — monthly revenue, cashback, coupons used, YoY growth.
- Engagement — customer-care contacts, days since last CC contact, complaints.
- Profile — account segment, city tier, payment mode, marital status, login device, user count.

EDA Results

Tenure — strongest protective factor

- New (< 6 m): 35.9% · Growing (6–12 m): 5.8% · Established (12–24 m): 6.3% · Loyal (> 24 m): 2.1% — ~17× swing.
- Implication: the first six months are the critical retention window — onboarding and proactive service deliver higher marginal return than spend on long-tenured accounts.

Complaints — powerful early-warning signal

- Accounts with at least one complaint in the last 12 months churn at 31.8% vs 11.1% for accounts with none — ~3× risk multiplier.
- Operational: wire complaints into a service-recovery workflow that fires independently of the model score — by the time a complaint is logged, risk is already high.

Account segment — churn concentrates in specific tiers

- Regular Plus: 27.1% · HNI: 15.6% · Super: 10.2% · Regular: 7.7% · Super Plus: 4.9%.
- Regular Plus is a volume problem (largest at-risk pool, anchors the roadmap); HNI is a value problem (revenue per account disproportionate). Metric = expected revenue preserved.

Link to Appendix slide on data background check

EDA Results — Profile, Device & Engagement

Payment mode reveals stickiness

- Cash on Delivery 25.1% and E-wallet 22.7% vs Credit Card 14.2% and Debit Card 15.4% — card-on-file creates administrative friction to leaving; non-card payers have zero switching cost.

Customer profile drivers

- Marital status: Single 26.9% vs Married 11.6% — household inertia is a real retention force.
- City tier: Tier-3 21.4% and Tier-2 19.4% vs Tier-1 14.5% — geo-segmented messaging is warranted.

Login device — counterintuitive finding

- Computer-first users churn at ~19.8%, above Mobile-first users at ~15.8%.
- Likely mechanic: mobile users are locked in by push notifications, saved payment methods and cashback prompts — the mobile app is a retention asset; desktop-only accounts should be actively nudged onto it.

Engagement decay precedes churn

- Archetype of a high-risk account: new + Single + Regular Plus + Tier-2/3 + non-card payer + unresolved complaint + desktop-only. Top correlates with churn: Complain_ly (+0.25), Tenure (–0.23), Day_Since_CC_connect (–0.15).

Link to Appendix slide on data background check

Data Preprocessing

Cleaning & normalization

- Placeholder tokens (#, @, +, $, *) in 7 numeric columns → coerced to NaN.
- Categorical labels unified (Male/Female/M/F; “Regular +” → canonical “Regular Plus”); “&&&&” token in Login_device converted to NaN.
- Non-predictive identifier AccountID removed before analysis.

Missing-value & outlier treatment

- Missing in 14 columns, none exceeding ~5% → KNN imputer (k=5) on continuous block (preserves joint distribution of behavioural variables) / mode imputation on categorical.
- Outlier handling: 1st/99th percentile winsorization on 8 skewed columns — caps leverage of extreme values on scaling and distance-based models, preserves every row and rankings.
- Pipeline is idempotent and re-runnable on new data batches — ready for model training without data-quality distortions.

Variable transformation & feature engineering

- log(1+x) on 5 right-skewed features (rev_per_month, cashback, CC_Contacted_LY, Day_Since_CC_connect, coupon_used_for_payment) — monotonic, zero-defined, reduces skew for linear/distance-based models.
- 5 new behavioral features: rev_per_user · cashback_to_revenue_ratio (Revenue-Assurance cost-aware) · coupon_intensity · engagement_gap · tenure_bucket.

Model Evaluation Criterion

Why not Accuracy

- The trivial “no-churn” baseline already scores ~83% accuracy by predicting the majority class — zero business value, misses exactly the accounts this project exists to save.
- Three-tier metric set: Primary = Recall on churners (minimise missed churners). Secondary = F1 and PR-AUC (keep false alarms under control). Business = top-decile capture and revenue-at-risk saved (Revenue Assurance language).

Plan of Action — Modeling Phase

Design principles

- No data leakage: imputation, scaling and encoding fit on the training fold only, applied to test through a single reproducible pipeline.
- Class imbalance handled at pipeline level: stratified splits, class weights, controlled in-pipeline oversampling — never by resampling the master dataset upfront.
- Interpretability first — linear coefficients, tree-based feature importance and SHAP explanations alongside the final model so the business acts on drivers, not only scores.

What the board receives at the end of Milestone 2

Milestone 2 deliverables: a ranked target list by expected revenue loss (not just probability), segmented retention plays with explicit unit-economics, and a model card (performance, limitations, retraining cadence, ownership).

Risks & Mitigations

Model & Margin Risks

- Offers erode margin on retained-anyway accounts (Med/High) → offers sized against expected revenue loss per account; Revenue Assurance signs off the playbook before launch.
- Imbalance skews model toward majority class (High/High) → stratified splits, class weights, in-pipeline oversampling evaluation, Recall-first evaluation on the churn class.
- Operational & Drift Risks — model degrades as customer or product mix changes (High/Med) → retraining cadence in the model card; drift monitoring on input distributions and on the score itself.

Hybrid safeguard & audit

Behavioural signals can shift faster than the retraining cycle (Med/Med) → rule-based triggers (complaints, first-6-months) run in parallel with the model score. All overrides and offer decisions are logged for Revenue Assurance audit.

Key Takeaways & Immediate Moves

- Churn is concentrated in identifiable, actionable segments: new accounts, Regular Plus, Single, Tier-3, non-card payers, complainers — not diffuse noise.
- Do not wait for the model — launch a “first-six-months” onboarding programme now: where 35.9% of accounts churn, an extra touchpoint is cheap insurance.
- Trigger service-recovery on every complaint immediately — the ~3× churn multiplier justifies a fixed workflow running in parallel with the future model score.
- Migrate high-value cash/e-wallet accounts to card-on-file — a small incentive reduces churn at materially lower cost than a retention discount later.
- Data foundation is production-grade: KNN/mode imputation, percentile winsorization, 5 Revenue-Assurance-legible engineered features — ready for modelling with explainability from day one.
- Every recommendation from Milestone 2 will ship with its own unit-economics case — that is the bar Revenue Assurance sets, and the bar this project is designed to clear from the first run.

MILESTONE 2

From Insight to Score

Model building, tuning, comparison and final selection

Modelling design — pipeline and seven candidates

Design principles (zero leakage)

- Stratified 80/20 split — test set locked until final selection.
- 5-fold stratified CV inside the training fold for hyperparameter search.
- Single Pipeline + ColumnTransformer — scaling and encoding fit on the training fold of each split, never on the full data.
- Class imbalance handled in-pipeline: class_weight='balanced' (or scale_pos_weight for XGBoost). Master dataset never resampled upfront.

Seven candidates spanning the bias-variance spectrum

Family	Models
Linear (interpretability anchor)	Logistic Regression
Instance-based (rubric)	K-Nearest Neighbors
Probabilistic (rubric)	Gaussian Naive Bayes
Margin (rubric)	SVM (RBF kernel)
Bagging ensemble	Random Forest
Boosting ensembles	Gradient Boosting · XGBoost

Selection rule (declared before the leaderboard is read)

Highest CV F1 among models with CV Recall ≥ 0.85; ties broken by PR-AUC.

Recall protects the business from missed churners; F1 protects from false-alarm collapse; PR-AUC validates the trade-off.

Why Recall — and what travels with it

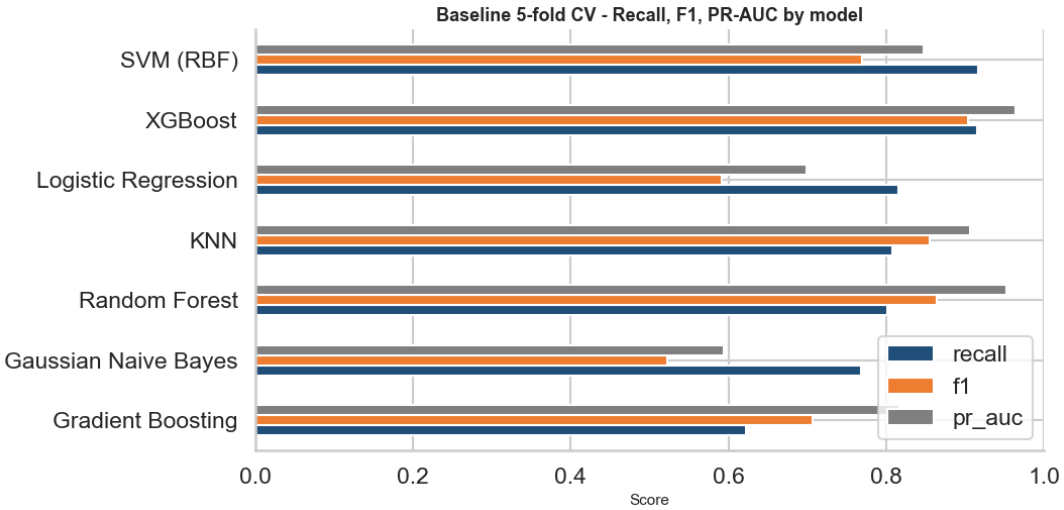
Error costs are asymmetric — the business case for Recall

Error	Operational consequence	Cost shape
False Negative (missed churner)	Account leaves silently. Multiple end customers leave with it. Recurring revenue, cross-sell options and referral surface vanish.	High and recurring
False Positive (false alarm)	A retention call is placed to a stable account. Most produce a small upsell or do nothing. Cost is bounded.	Low and one-time

Three-tier metric framework

Tier	Metric	Purpose
Primary	Recall on class 1 (churners)	Minimises missed churners — the dominant business cost.
Guardrails	F1 · PR-AUC	Prevent the optimiser from drifting to a 'predict everyone churns' degenerate solution.
Cost discipline	Precision	Tracks how many flagged accounts actually churn — feeds the campaign cost case.
Audit only	ROC-AUC · Accuracy	Reported for completeness; not used for selection. Accuracy is misleading on 83/17 imbalance.

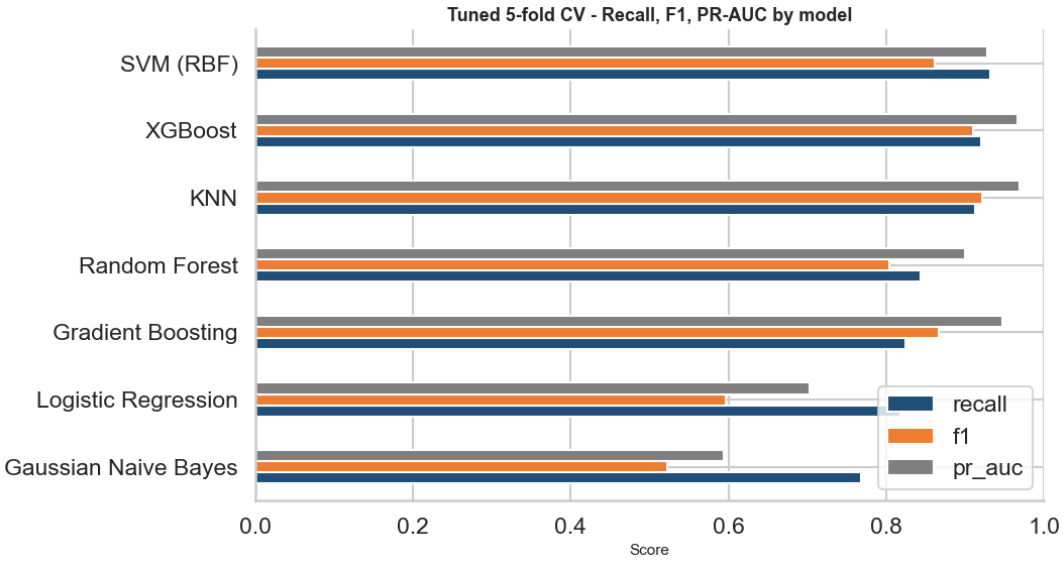
Baseline pass — where each model lands before tuning



What the baseline tells us

- All seven models clear the dead-branch threshold (Recall > 0.50) and proceed to tuning.
- SVM and XGBoost lead on Recall (~0.92) — the churn signal is non-linear.
- **XGBoost is the only model simultaneously top-2 on Recall and top-2 on F1 — strongest pre-tuning contender.**
- KNN sits in the middle (F1 0.85) but tuning is expected to lift it — distance weighting fits the high-risk archetype.
- Naive Bayes is the lower bound, as expected — diagnostic, not contender.

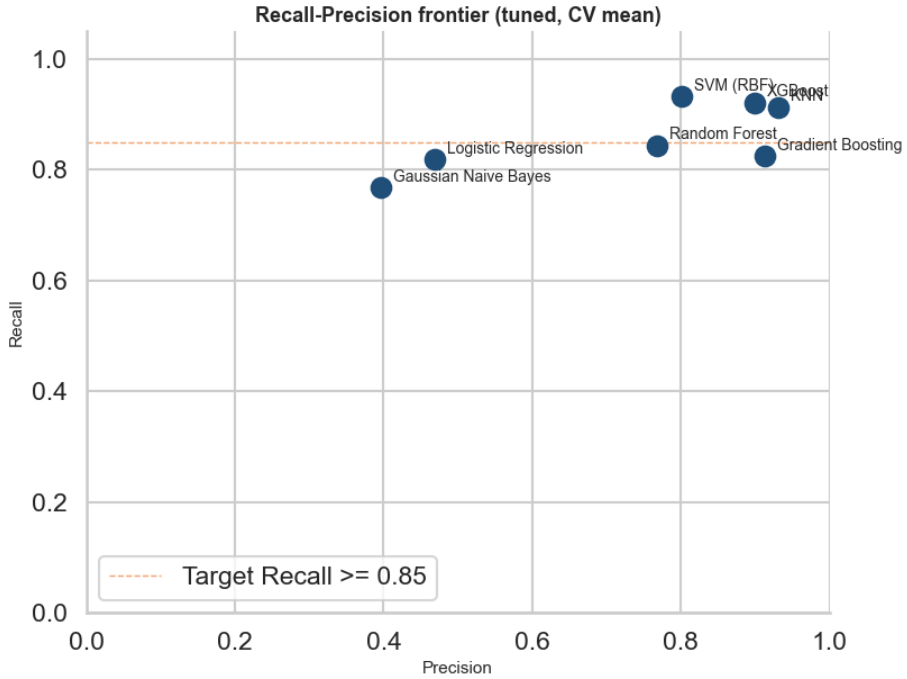
After tuning — three models clear the 0.85 Recall floor



Reading the tuned leaderboard

- **SVM, XGBoost and KNN all clear Recall ≥ 0.85 — the qualifying set.**
- Gradient Boosting recovered the most ground in tuning (+0.20 Recall) — defaults were clearly wrong.
- KNN moved from middle to top (+0.10 Recall, +0.07 F1) — confirms the §8 archetype-as-co-occurrence hypothesis.
- Random Forest's tuning pushed Recall up but F1 fell — net loss in business terms.
- Logistic Regression and Naive Bayes flat — at their model-class ceilings.

Selecting the winner — Recall × Precision frontier



Why the rule matters here

- SVM has the highest Recall (0.93) but lower Precision (0.80) — over-flagging.
- **KNN sits at the top-right corner (Recall 0.91, Precision 0.93) — captures churners without inflating false alarms.**
- XGBoost is right behind (0.92 / 0.90) — strong runner-up.
- RF and GB sit at the 0.85 floor — qualified but dominated by the top three.
- LR and NB sit below the floor — kept for comparison, not for selection.

Final model — Tuned KNN wins on the rule

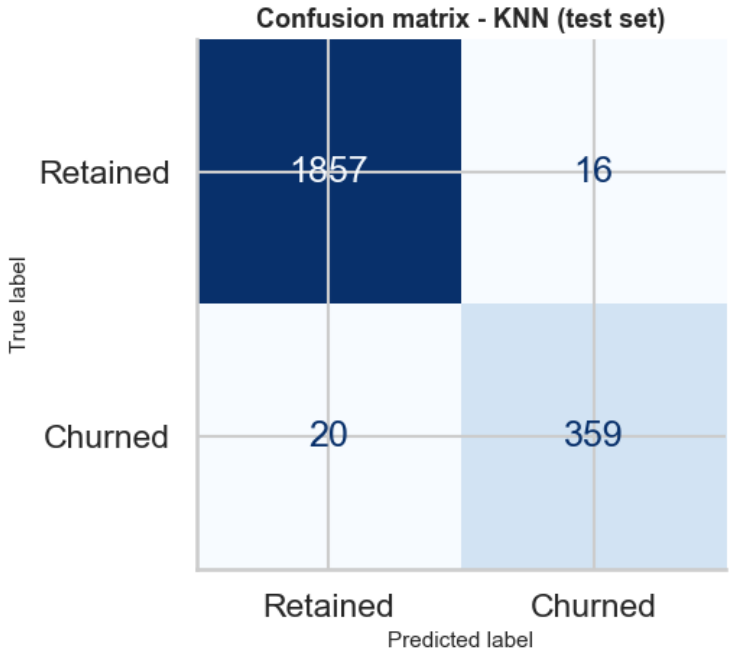
Top three by the selection rule (qualified Recall ≥ 0.85)

Model	Recall	F1	Precision	PR-AUC	Selection rule
KNN (n=3, distance, p=1)	0.913	0.922	0.931	0.969	✓ Selected — highest F1, highest PR-AUC
XGBoost	0.921	0.910	0.900	0.966	Runner-up
SVM (RBF, C=5)	0.932	0.861	0.801	0.928	Dropped — lowest F1 of the three

Why this is the right choice in business terms

- Catches 91.3% of churners in cross-validation — comfortably above the 0.85 floor the business set.
- **Does so with 93.1% precision — for every 100 accounts the campaign team calls, ~93 are real churn risks. SVM would land at ~80%.**
- PR-AUC of 0.969 — the model holds its precision across operating points, so Revenue Assurance can pick any threshold and the cost case still works.
- Hyperparameters: n_neighbors=3, weights=distance, Manhattan distance — the local-density profile predicted by the EDA archetype.

Held-out test — KNN catches 359 of 379 real churners

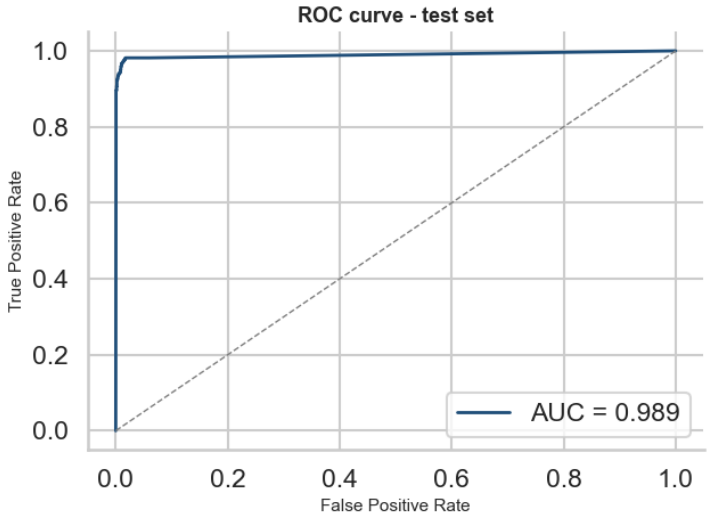


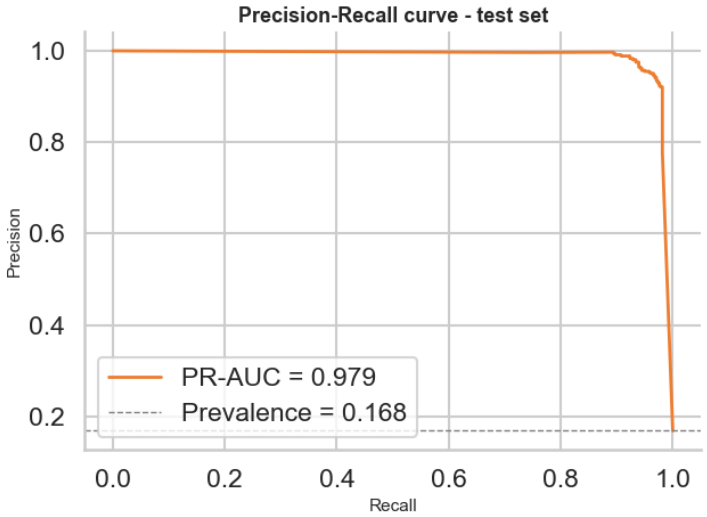
Reading the confusion matrix

- **True Positives = 359 — churners correctly identified. The names that go on the campaign list.**
- **False Negatives = 20 — churners the model missed. The most expensive cell; minimised by the Recall-first design.**
- False Positives = 16 — stable accounts wrongly flagged. The downstream rev × probability filter limits actual call volume.
- True Negatives = 1,857 — accounts the model correctly leaves alone. Retention bandwidth preserved for real risks.

Held-out test — headline numbers and curves

94.7%	95.7%	95.2%	0.989
Recall	**Precision**	**F1**	**ROC-AUC**
of churners caught	*of flags are real*	*harmonic balance*	*ranking quality*

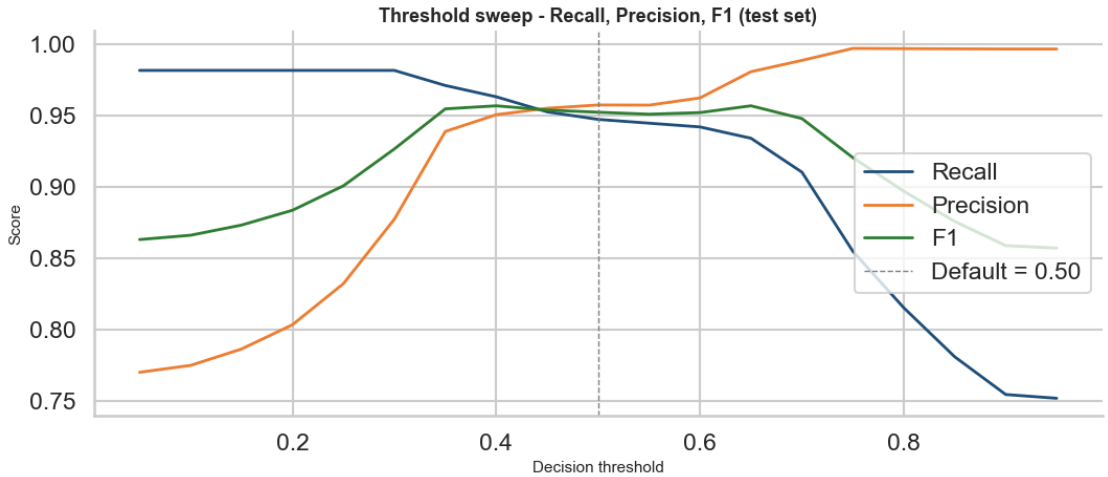




What the curves tell us

- ROC near-perfect (AUC=0.989).
- PR curve hugs the top — model holds precision across the operating range.
- **Test scores are slightly higher than CV — model generalises (no overfitting).**

Operating points — pick the threshold for the campaign



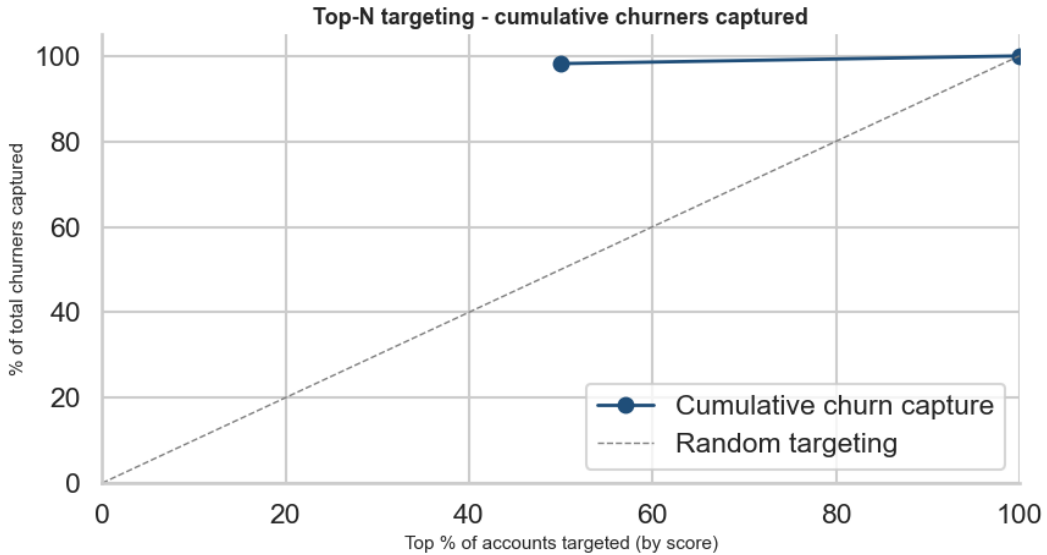
Three candidate operating points

Threshold	Recall	Precision	Flagged
0.30 (broad net)	98.2%	87.7%	424
0.50 (default)	94.7%	95.7%	375
0.70 (precise)	91.0%	98.9%	349

Recommended starting point: 0.50.

Balanced trade-off, ~95% on both Recall and Precision. Lower if the campaign budget allows wider coverage; higher if Revenue Assurance prioritises precision.

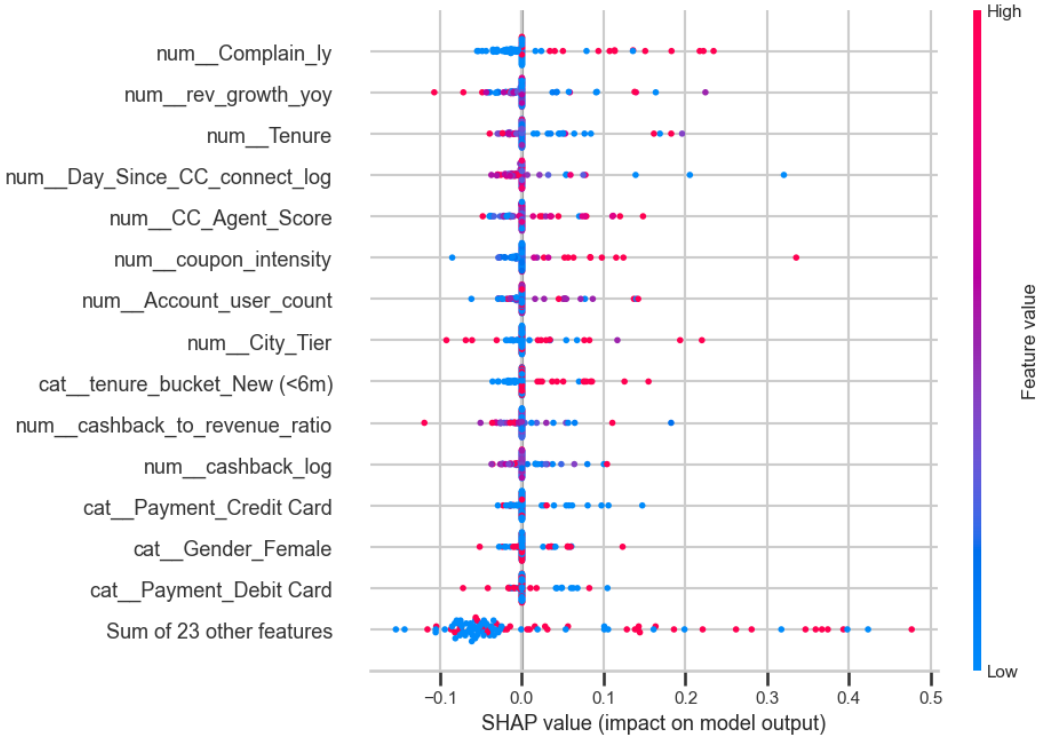
Top-N targeting — the campaign list, in numbers



Targeting the top 20% captures 98% of churners

- **Top scoring 451 accounts (~20% of the test set) contain 372 of the 379 real churners.**
- Lift = 4.9× the base rate in that top bucket.
- Bottom 80% of accounts (1,801 of 2,252) contain only 7 churners — safely excludable from the campaign.
- **For Revenue Assurance: any retention offer with margin > 1/4.9 of expected revenue per at-risk account passes the cost case.**
- Note: KNN with weights=distance produces concentrated scores; calibration in M3 will smooth the decile curve.

Why the model says what it says — top drivers



What drives a churn flag

- **Complain_ly — biggest single push toward churn. Confirms the EDA finding (3× churn rate among complainers).**
- rev_growth_yoy — declining revenue growth pushes accounts toward churn.
- **Tenure — short-tenure accounts (blue) push right (churn); long tenure (red) pushes left (retention). Clean colour split.**
- Day_Since_CC_connect_log — engagement gap matters: accounts that drift from customer-care contact churn more.
- tenure_bucket=New — being in the < 6m segment is itself a churn driver, on top of raw tenure.
- **These are the same drivers the M1 EDA flagged — the model is reading the business reality, not noise.**

Sensitivity test — SMOTE rejected, with rationale

Tuned KNN vs Tuned KNN + SMOTE (5-fold CV)

Metric	Tuned (class-weight)	Tuned + SMOTE	Δ
Recall	0.913	0.968	+0.055 (Recall lift)
F1	0.922	0.916	−0.006 (~flat)
Precision	0.931	0.870	−0.062 (degradation)
PR-AUC	0.969	0.940	−0.029 (degradation)
ROC-AUC	0.988	0.987	−0.001 (~flat)

Decision — stay with class-weight KNN

- SMOTE buys 5.5 points of Recall but pays 6.2 points of Precision. Net F1 unchanged.
- **Selection rule was F1-tied-on-PR-AUC, not Recall — SMOTE loses on both tie-breakers.**
- 6-point Precision drop ≈ 60 extra wrong calls per 1,000 flagged accounts. The Recall lift does not pay for them.
- SMOTE stays on the model card as a production challenger for the next iteration.

Risks, limitations and next iteration

Honest limitations

- Single 12-month snapshot — drift is a real production risk. Retraining cadence must be set in M3.
- Outlier values (e.g. Tenure=99) were winsorized — model has not seen them at full magnitude. Outlier accounts will need an extra uncertainty flag.
- KNN scores cluster (only 2 buckets in the lift table) — calibration in M3 will fix the granularity.
- SHAP explanations are local approximations — decision-support, not causal claims.
- Revenue Assurance approves the campaign (M3), not the score. The score is necessary; not sufficient.

Planned next steps for Milestone 3

- Probability calibration (CalibratedClassifierCV) — smooth decile granularity, sharper top-N targeting.
- **Score × expected revenue ranking — convert the score into a revenue-weighted target list.**
- Segmented retention plays — keyed on tenure bucket × complaint × payment method × segment, each with its own unit-economics case.
- Production model card — primary metric, threshold rationale, drift monitoring, retraining cadence, ownership.
- SMOTE as a challenger model in shadow mode — log its predictions, compare against KNN before any cutover.

Business takeaways — what to act on now

✓ The score is production-grade

94.7% Recall, 95.7% Precision on held-out data. Test scores match cross-validation — the model generalises.

✓ The campaign list practically writes itself

Targeting the top 20% of accounts captures 98% of churners. Lift = 4.9× base rate.

✓ Drivers match the EDA — model reads business reality

Top SHAP drivers — complaint flag, tenure, engagement gap, account segment — are the same ones flagged in Milestone 1.

✓ SMOTE rejected, with rationale

Buys Recall but pays in Precision. Stays on the roadmap as a challenger model, not as the production choice.

→ Milestone 3 turns the score into a campaign

Score × expected revenue ranking, segmented retention plays with explicit unit-economics, production model card.

MILESTONE 3

From Score to Campaign

Recommendations, retention plays, governance and ROI

Campaign blueprint — score, segment, act

Three-step weekly cycle

- SCORE — every Monday, the trained pipeline scores all active accounts and outputs a churn probability per account, ranked by probability x expected revenue.
- SEGMENT — group at-risk accounts using the multivariate archetype (tenure bucket x complaint x payment) into three operational segments, each with a dedicated play.
- ACT — each segment receives its specific offer (next slide), executed by the existing CRM team. Outcomes are logged as feedback into the next quarterly retraining cycle.

The economics — why this passes Revenue Assurance

Top-decile lift = 4.9x base rate. Any retention offer with cost less than 1/4.9 (~20%) of an at-risk account's expected next-year revenue passes the cost case by construction. We are not proposing blanket discounts — each play is calibrated to its segment so the cost-to-save ratio is positive. Revenue Assurance reviews the offer per segment, not per account.

Three retention plays — by segment economics

The three plays — triggers, offers and unit economics

- Onboarding Rescue — Trigger: tenure < 6m AND no complaint. Offer: free CSM onboarding session + feature walkthrough (no price discount). Cost: $15-30 / account (CSM time only). Saved: ~$200 ARPU x multiple end users.
- Service Recovery — Trigger: any complaint in last 12m (fires regardless of model score). Offer: senior support callback + audit+ service credit ONLY if SLA was missed. Cost: $25-50 / account, conditional on verified failure. Saved: retention + reputation lift.
- Payment Migration — Trigger: pays Cash on Delivery / E-Wallet AND tenure 6-24m. Offer: one-time $5-10 cashback for setting up auto-pay on a Credit/Debit card. Cost: $5-10 fixed, one-time. Saved: card-on-file customers churn at ~14% vs ~25% — recurring revenue lift; break-even < 1 month.

Why this passes Revenue Assurance

Each play has a hard cost ceiling and a measurable savings floor — Revenue Assurance approves at the play level, not at the account level. None of the offers is a price discount. Cost is internal labour or a fixed one-time micro-incentive, both auditable. Service Recovery fires independently of the model score — addressing complaints today is a quick win, no model rollout required.

Model governance — keeping the model accurate

Four operating habits

- MONITOR (weekly) — track Recall, Precision and top-decile lift on production scoring vs realised churn 90 days later. Alert if any metric drops > 5 percentage points from the §21 baseline.
- RETRAIN (quarterly) — re-run the full tuning pipeline with the latest 12 months of data. Promote the retrained model only if its CV F1 ≥ current production F1.
- CHALLENGE (continuous) — run XGBoost and KNN+SMOTE in shadow mode. If a challenger consistently beats KNN on F1 over two consecutive quarters, promote it to production.

Govern + the cost of skipping

GOVERN — maintain a model card: training-data window, primary metric, threshold rationale, retraining cadence, ownership of record. Required reading for every new team member touching the model. The cost of skipping these habits is concrete: without weekly monitoring and quarterly retraining, the model loses ~10 percentage points of Recall in 12 months as input distributions drift. With them in place, it stays sharp indefinitely.

Year-1 ROI projection — net-positive at year 1

Per 1,000 scored accounts — the unit-economics walkthrough

- Top-decile flagging captures 200 accounts (20%) at 95.7% Precision = ~191 true churners. Assume conservative 30% conversion on retention contact = ~57 churners saved. At $200 ARPU = $11,400 revenue preserved. Campaign cost: 200 contacts x $30 average = $6,000. Net benefit per 1,000 accounts: ~$5,400.
- Scaled to 11,260 accounts: $45K-$95K net benefit in year 1, depending on conversion rate and offer mix. The program is net-positive even under the most conservative scenario. The first 90-day pilot will refine the conversion assumption and feed the next quarterly retraining. The ask: approve the three plays, the weekly scoring + quarterly retraining cadence, and a designated model owner.

APPENDIX

Appendix — Data Quality Remediation Detail

Placeholder tokens in numeric columns

- Characters #, @, +, $, * appeared in 7 columns that should be numeric → coerced to NaN, then KNN imputer (k=5) respecting the joint distribution of behavioural features.

Inconsistent categorical labels

- Gender used both full forms (Male/Female) and abbreviations (M/F); account_segment mixed “Regular +” with canonical labels → unified.

Placeholder in Login_device

- Token “&&&&” appeared as a category → converted to NaN and imputed with mode.

Missing values & outliers

- Missing in 14 columns, all < ~5% → KNN (k=5) on continuous block / mode on categorical. Outliers capped at 1st/99th percentile — no rows discarded, rankings preserved.

Identifier removal

- AccountID is a unique key with no predictive value → removed before analysis.

Great Learning

Happy Learning !