Telco Churn — Executive Summary

Objective

- **Decision use-case:** Predict customers at risk of churn next cycle to target retention spend efficiently (prediction > interpretation).
- **Business value:** Prioritise high-risk accounts, control offer costs with **Precision ≥ 0.80**, and use calibrated probabilities to size campaign volumes and forecast ROI.

Data & Scope

- **Source:** IBM Telco Customer Churn (Kaggle mirror).
- **Size/shape:** ~7k customers; >20 features across services, billing, contract and demographics.
- **Target:** Binary churn (Yes/No), prevalence ~26%.
- **Scope notes:** Snapshot modelling (no time-series features yet).

Method (concise)

- **Preparation:** Removed identifiers; fixed TotalCharges blanks; median/frequent imputation; one-hot encoding; numeric scaling.
- **Validation:** 80/20 stratified hold-out; model selection by **5-fold stratified CV**; threshold chosen to maximise recall subject to **Precision ≥ 0.80** on a validation split.
- **Miscalibration addressed:** Post-hoc calibration of predicted probabilities.

Models Compared (CV)

- Baseline majority-class.
- **Logistic Regression** (interpretable baseline).
- **Linear SVM + calibration** (margin-based).
- **HistGradientBoosting (HGBT)** (non-linear ensemble).

Selection metric: PR-AUC (robust to class imbalance).

Recommended Model

- **Champion: HGBT + isotonic calibration.**

    - **Hold-out:** ROC-AUC ≈ **0.83**; Average Precision ≈ **0.61** (well above 0.26 base rate).
    - **Calibration:** Raw model over-confident; isotonic brings reliability closer to the diagonal, improving decision quality.
- **Challenger (glass-box):** Logistic Regression (slightly lower lift, higher interpretability for policy sign-off).

What Drives Churn (action-oriented)

- **Contract:** Month-to-month customers churn significantly more than 1–2 year contracts.
- **Tenure:** Early-life customers (especially <12 months) are higher risk.
- **Charges:** Higher **MonthlyCharges** (and high **TotalCharges** at low tenure) increase risk.
- **Service:** Fibre-optic plans associate with higher churn vs DSL/none.
- **Payments/Billing:** Electronic cheque and paperless billing correlate with higher churn.

Actions We Recommend Now

- **Targeting:** Focus retention offers on **month-to-month, low-tenure, high-charge** customers first.
- **Commercial levers:** Incentivise **contract upgrade** and **payment-method change** away from electronic cheque.
- **Onboarding:** A 90-day early-life engagement programme to reduce first-year churn.
- **Operating point:** Deploy with **Precision ≥ 0.80**; scale recall to match budget.

Risks & Controls

- **Model drift:** Monitor class balance and PR-AUC monthly; re-calibrate quarterly.
- **Fairness:** Track precision/recall by key customer segments; investigate if gaps exceed 5–10%.
- **Governance:** Report Brier score and Expected Calibration Error alongside AUCs.

Next Steps

- **Data enrichment:** Add service quality and engagement (outages, support tickets, app logins), price-rise history, competitor events, NPS/complaints.

- **Features:** Tenure bands; interactions (**contract × charges**); consider monotonic constraints in GBDTs.
- **Experimentation:** A/B test calibrated thresholds tied to offer cost and expected margin; build precision–recall–cost curves to set budget.

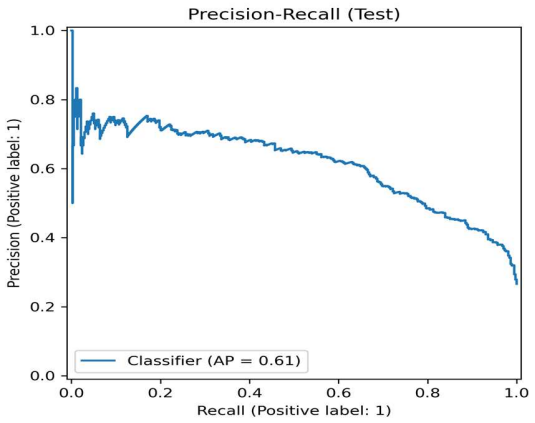

Figure 1. Precision–Recall (Test). Average Precision = **0.61** vs base rate ≈ **0.26**; operating policy targets **Precision ≥ 0.80** with recall tuned to budget.

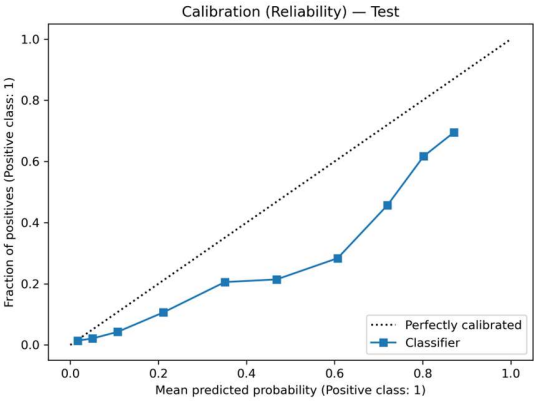

Figure 2. Calibration (Test). Raw scores are **over-confident** (below diagonal); post-hoc isotonic calibration applied before deployment to improve probability reliability.

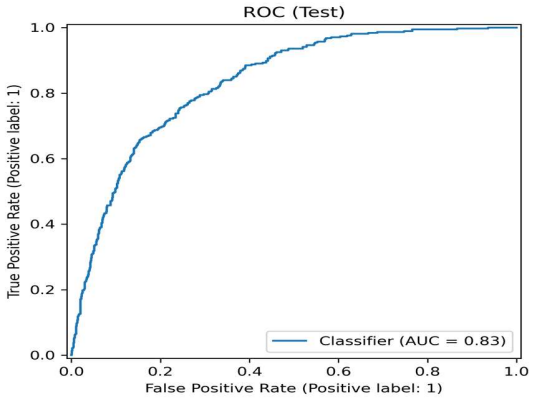

Figure 3. ROC (Test). AUC = **0.83**, indicating strong discrimination across thresholds.