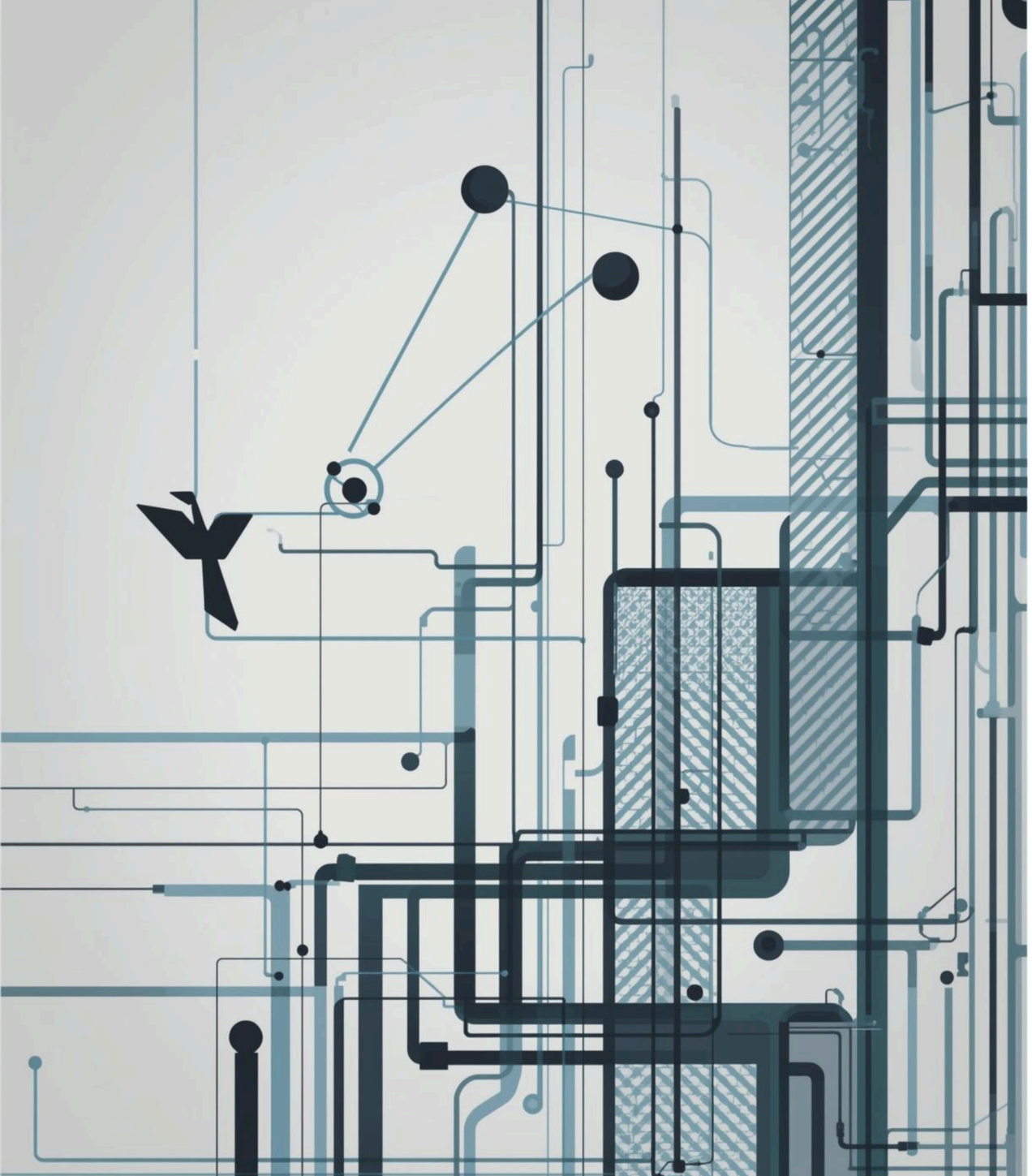


Telecom Insights

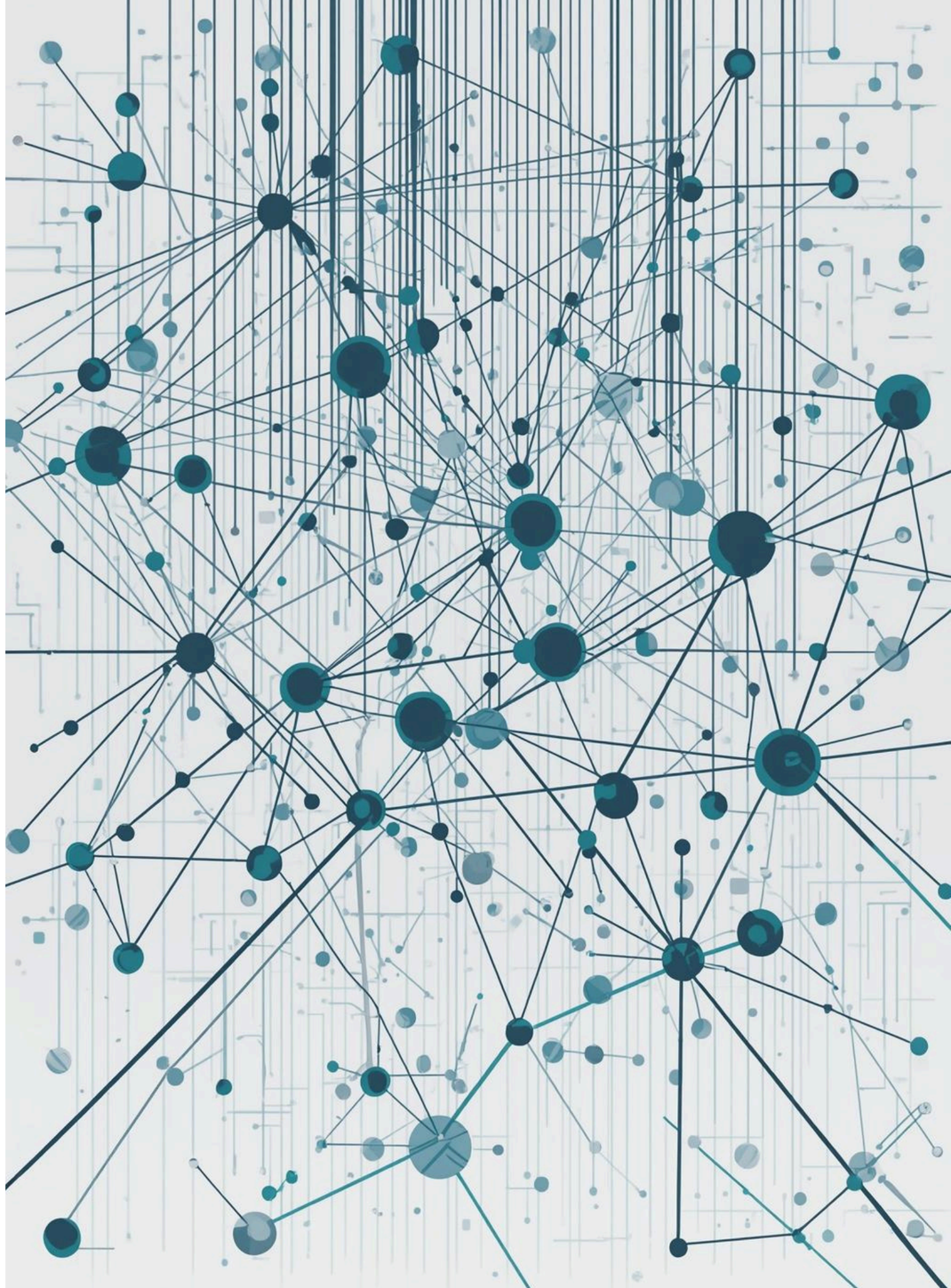
Predicting Churn, Preserving Value

Advanced Machine Learning and Explainable AI for Telecom Retention Strategies

PRESENTED BY

Audrey Chang

Table of Contents



Problem and Research Questions

- Problem: Telecom customer churn is costly; retention budgets are limited → need accurate and actionable risk ranking.
- Goal: Build an end-to-end churn prediction pipeline under realistic constraints (class imbalance, leakage control, reproducibility)

ENGINEERING RIGOR

Leakage-controlled preprocessing, consistent train/val/test splits, and reproducible artifact logging (configs, models, metrics)

MODEL COMPARISON

Benchmark a calibrated GLM baseline against tree/GBM models, deep tabular networks, and a stacking ensemble

EXPLAINABILITY IMPORTANCE

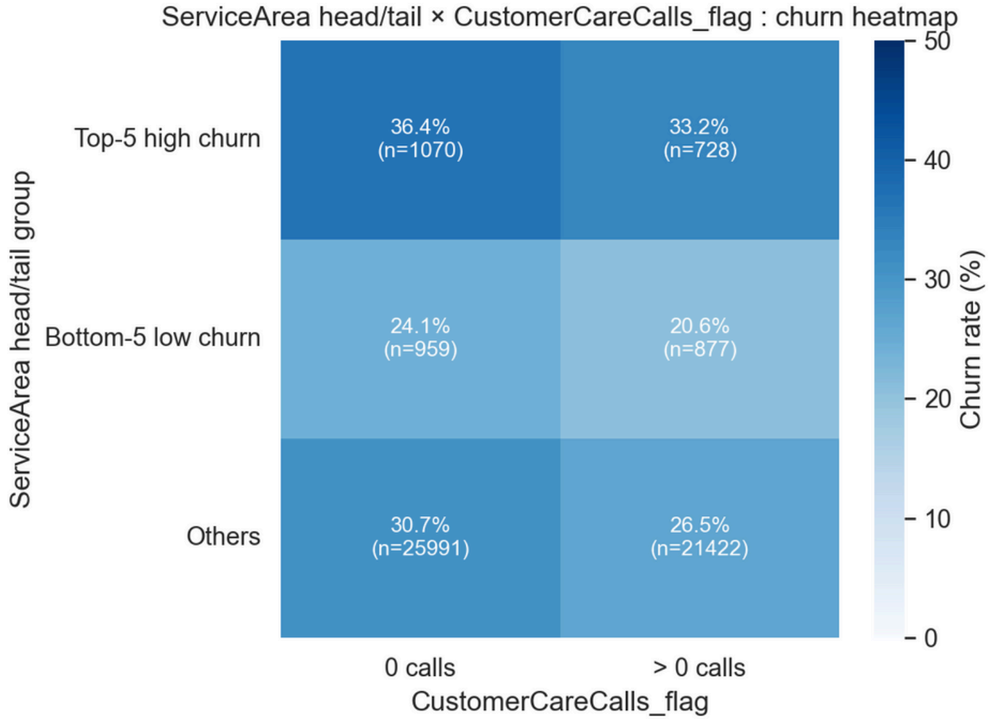
Translate interpretable drivers and failure modes (SHAP + error analysis) into retention playbooks and budget-aware targeting

Data & Evaluation

- **Dataset**: telecom churn dataset (tabular, mixed numeric + categorical, imbalanced target)
- **Target**: binary churn label (Churn01)
- **Split strategy:** train/validation/test split done before any threshold learning / encoding fitting
- **Leakage control**: exclude post-treatment / retention-intervention signals (e.g., retention call/accept variables)
- **Evaluation focus**: ranking + decision usefulness under imbalance

Feature Engineering

(Business-friendly + Leakage-safe)



Design principle: make features interpretable and stable; avoid high-cardinality explosion

Examples of engineered features (and why):

- CustomerCareCalls_flag = support/complaint signal (actionable)
- DroppedQuality_flag / DroppedBlockedCalls = network quality proxy
- HighUsage_flag = usage pressure (threshold learned on train)
- HandsetPrice_q = price binning + missing as category (robust)
- HighRiskServiceArea_flag = compress high-cardinality area into risk flag

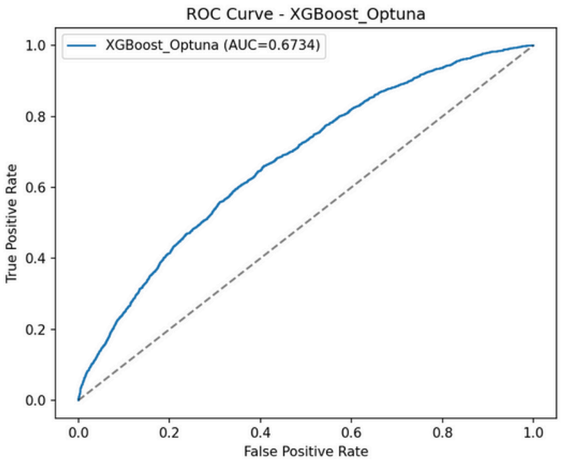
Models Compared

Four Families

- **Baseline**: Logistic Regression (interpretable, strong calibration baseline)
- **Tree/GBM**: Decision Tree, RandomForest, LightGBM, XGBoost (nonlinear + interactions)
- **Deep Learning**: Dense MLP, Embedding NN, Wide & Deep; imbalance strategies (strong weighting, focal loss)
- **Ensemble**: OOF Stacking over diverse base learners

Experimental Design & Metrics

How I Compare Fairly



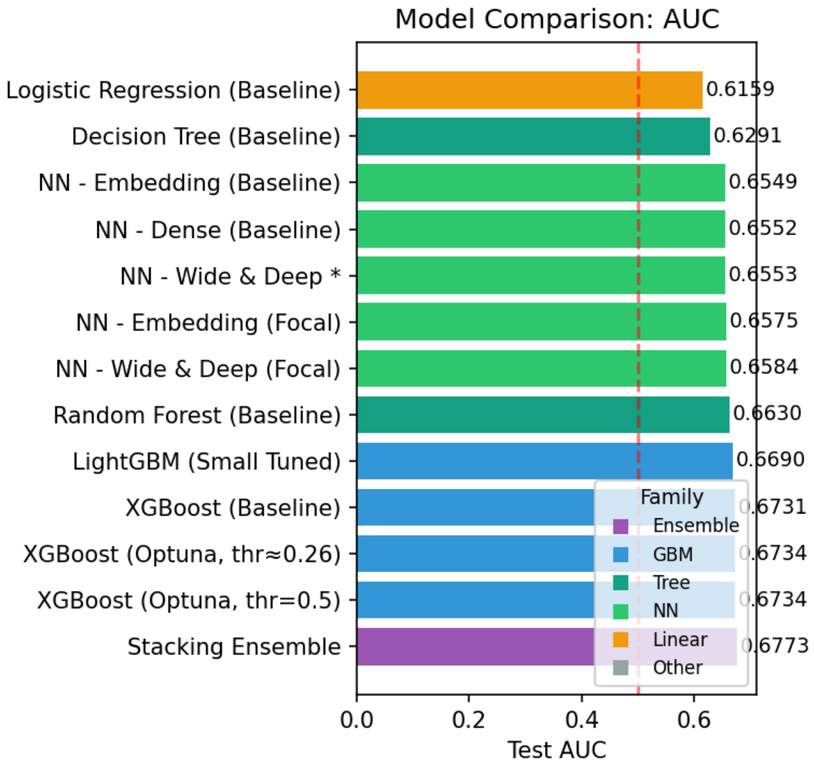
- Same train/val/test split across all models
- Hyperparameter tuning limited to validation (no test peeking)
- **Primary metrics**: ROC-AUC and PR-AUC (imbalance-aware)
- **Decision metrics**: F1 / Recall at tuned threshold; and threshold is chosen on validation

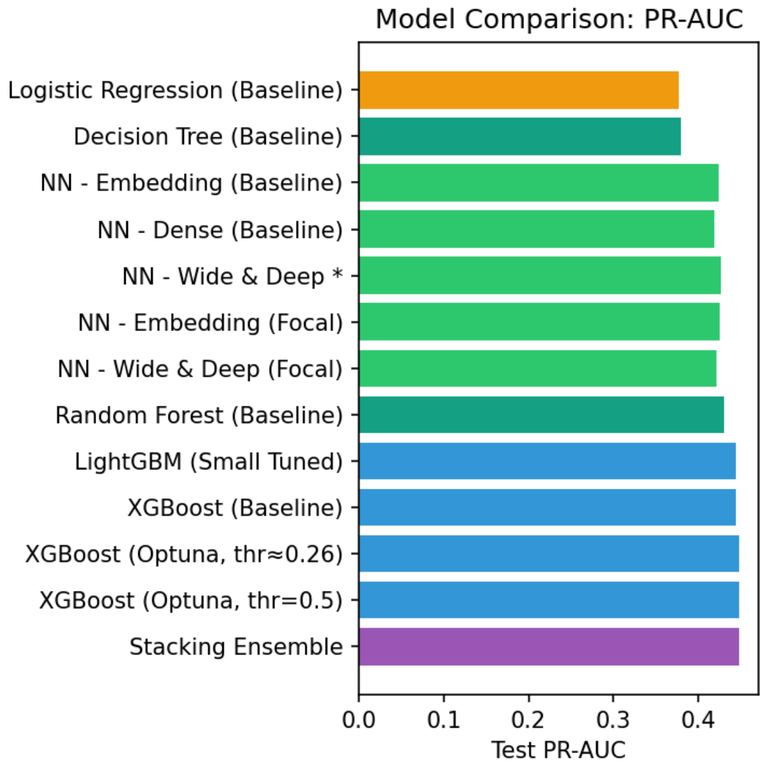
Metric	Why it matters
ROC-AUC	**Ranking quality**: how well the model orders churners above non-churners across all thresholds.
PR-AUC	**Positive-class focus**: more informative under class imbalance; emphasizes precision/recall on churners.
Confusion matrix	**Decision trade-off**: shows FP/FN at a chosen threshold → links directly to business costs and intervention policy.

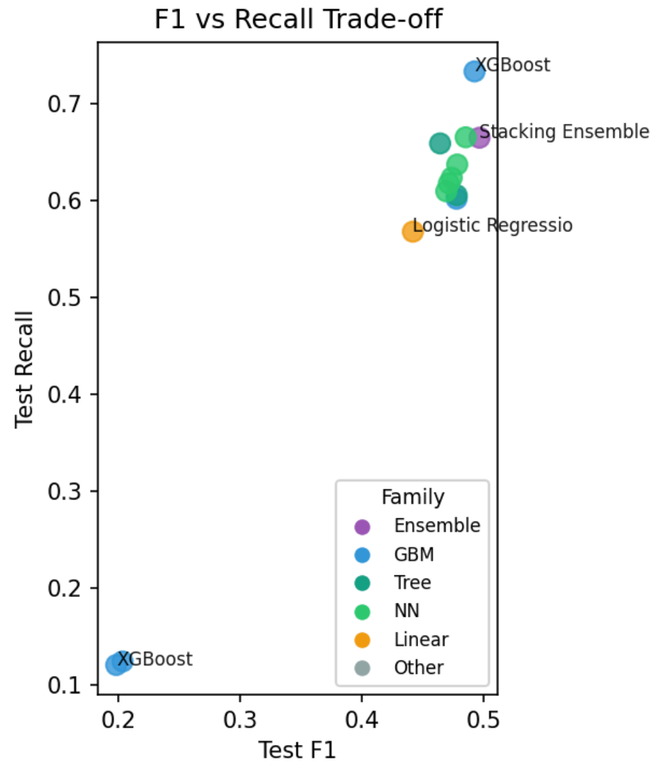
Main Results

Leaderboard + Key Takeaway

Model Leaderboard



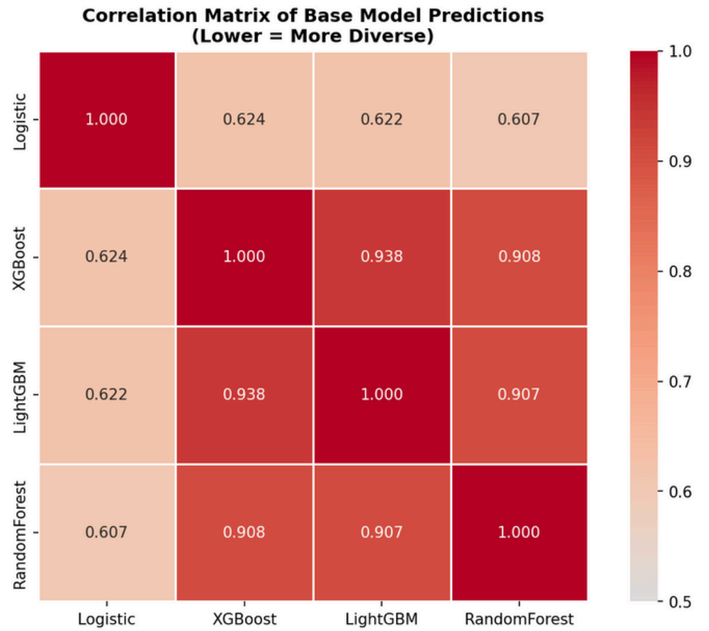




Advanced Modeling Beyond GBM

Tabular Deep Learning + Leakage-safe Stacking

- Controlled study: 7 NN variants under the same data split + training protocol(dense / embedding / wide&deep / deeper / focal / weighting)
- Best NN: Wide & Deep — Val AUC 0.6356, Test AUC 0.6553, PR-AUC 0.4259
- Imbalance handling: used class-weighted BCE (pos_weight); compared with stronger weighting and focal loss ablations
- Takeaway: Embeddings + wide component help, but gain is limited vs top GBMs on this tabular dataset

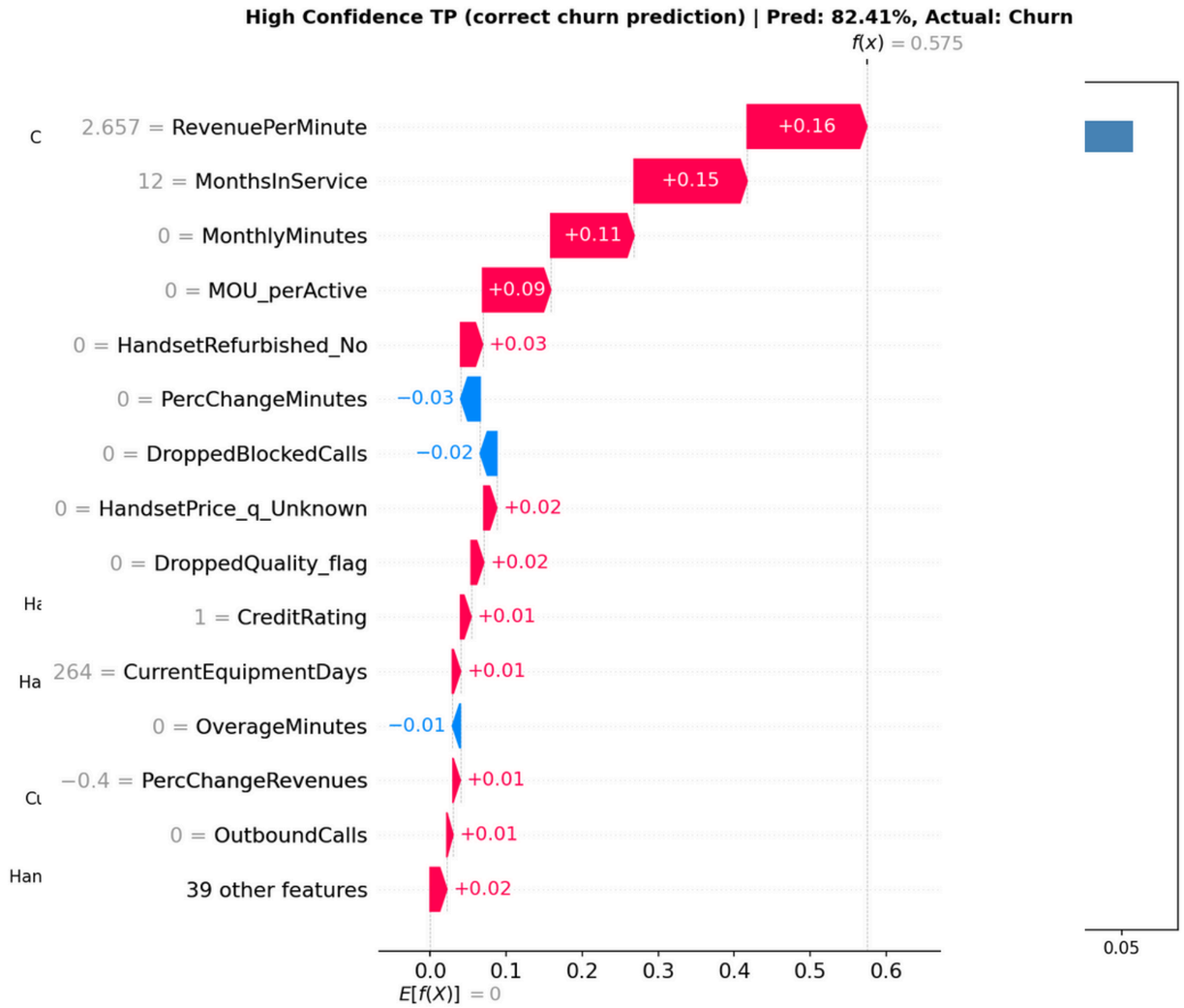


- **Why stacking:** combine complementary errors across model families
- Evidence of diversity: avg prediction correlation ≈ 0.768 (not identical signals)
- Negative result: naïve averaging did not beat the best single model (important sanity check)
- Leakage-safe method: meta-learner trained on out-of-fold (OOF) base predictions
- Result: Stacking Test AUC 0.6773 vs best single XGB 0.6734 (+0.58% relative lift)

Explainability: SHAP Insights

From “Black Box” to Action

- Global SHAP shows top churn drivers: **CurrentEquipmentDays (largest), PercChangeMinutes, MonthsInService, DroppedBlockedCalls, RevenuePerMinute, MonthlyMinutes, etc**
- Use SHAP dependence plots to show **nonlinear risk patterns** (e.g., device age / usage change effects)
- SHAP is **associational**, not causal → interventions should be validated with experiments



Business Case Study

Why calibrated probabilities matter

- Decision rule (expected value):
  - **Expected Loss = P(churn) × LTV**
  - Retain only if: **P(churn) × LTV > Cost(retention)**

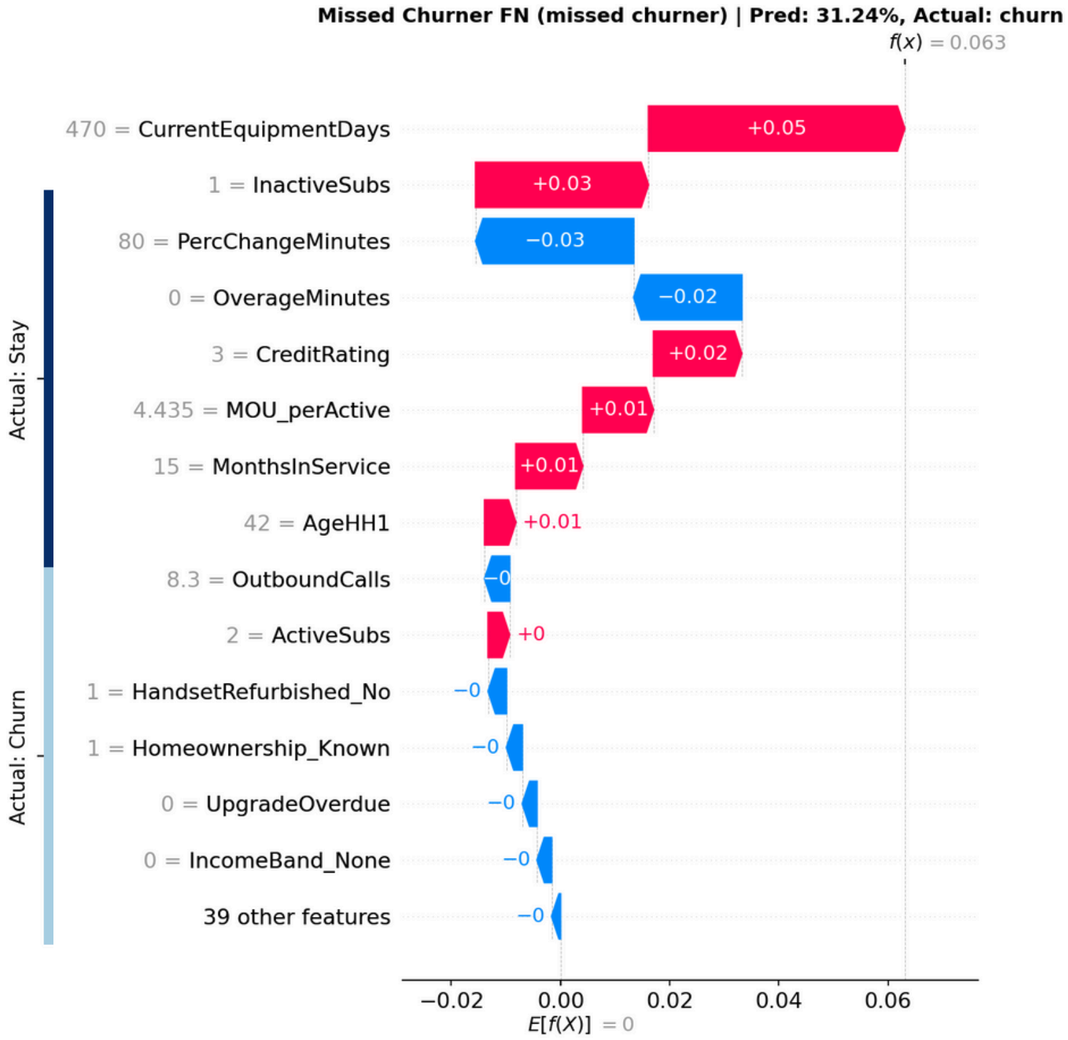
- Example assumption:
  - Cost = **$200**, LTV = **$1000** → break-even probability = **0.20**

- Why calibration matters:
  - If model says **0.80** but true churn rate is **0.10**, you over-spend and get negative ROI (you're "buying insurance" that's too expensive).

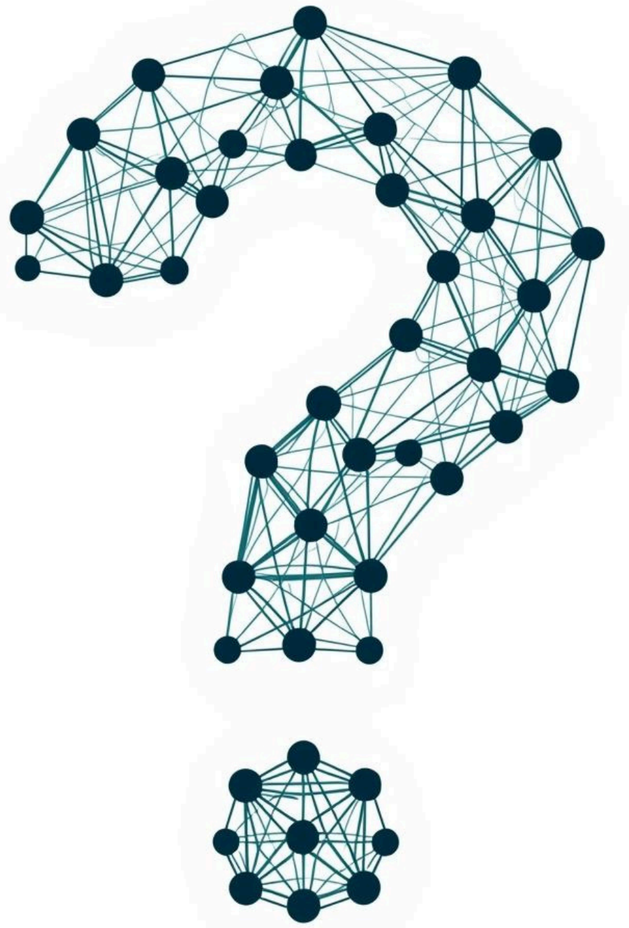
Error Analysis & Decisioning

Threshold + Confusion Matrix + Next Steps

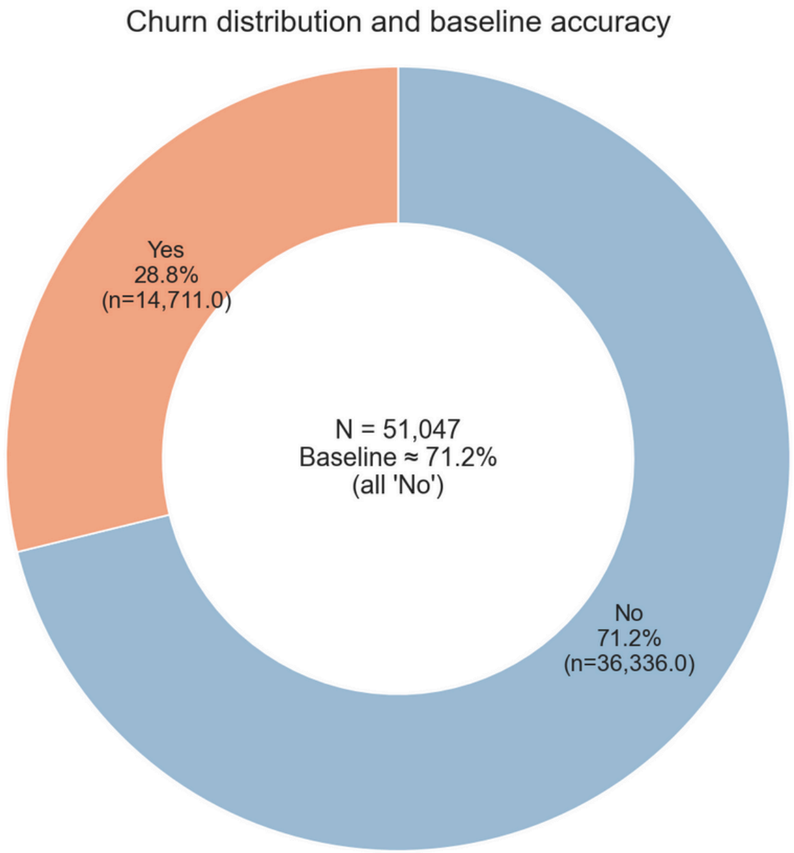
- Default threshold 0.5 can miss churners (example confusion matrix shows churn recall is low at 0.5)
- In my paper-style evaluation, I also tune threshold to hit business goals (e.g., threshold ≈ 0.2577 giving churn recall ≈ 0.734)
- Show one FN SHAP waterfall: competing signals + threshold too high→ action: adjust threshold, add features, or segment strategy
- Next steps: calibration check, uplift testing, monitoring drift by key driver

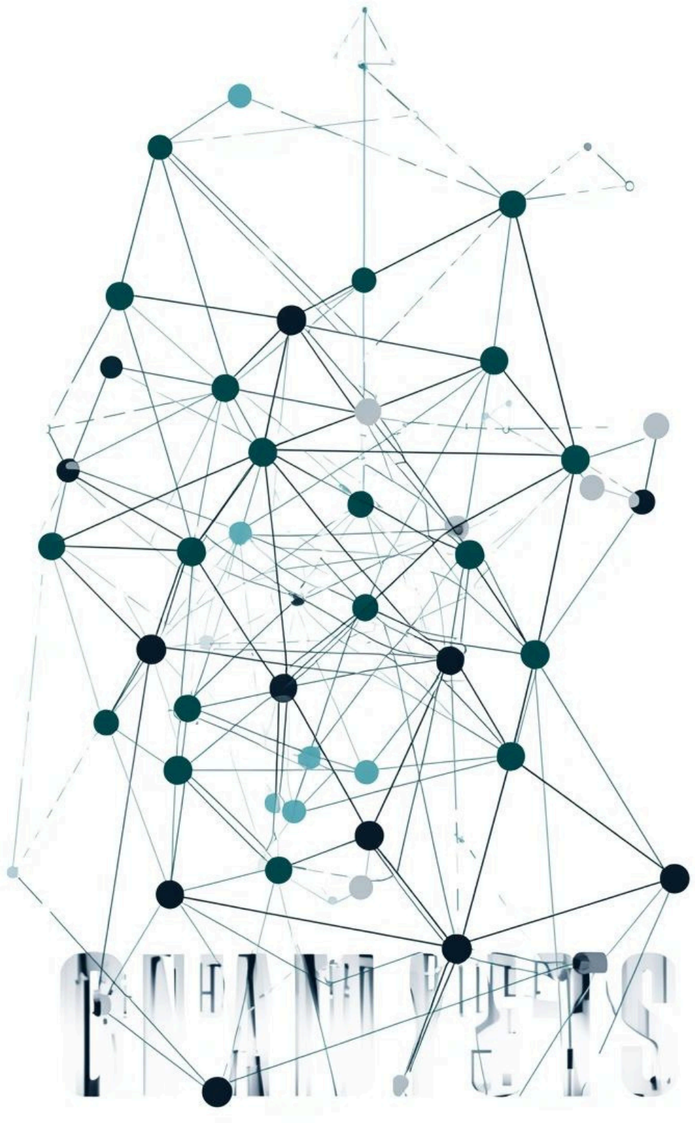


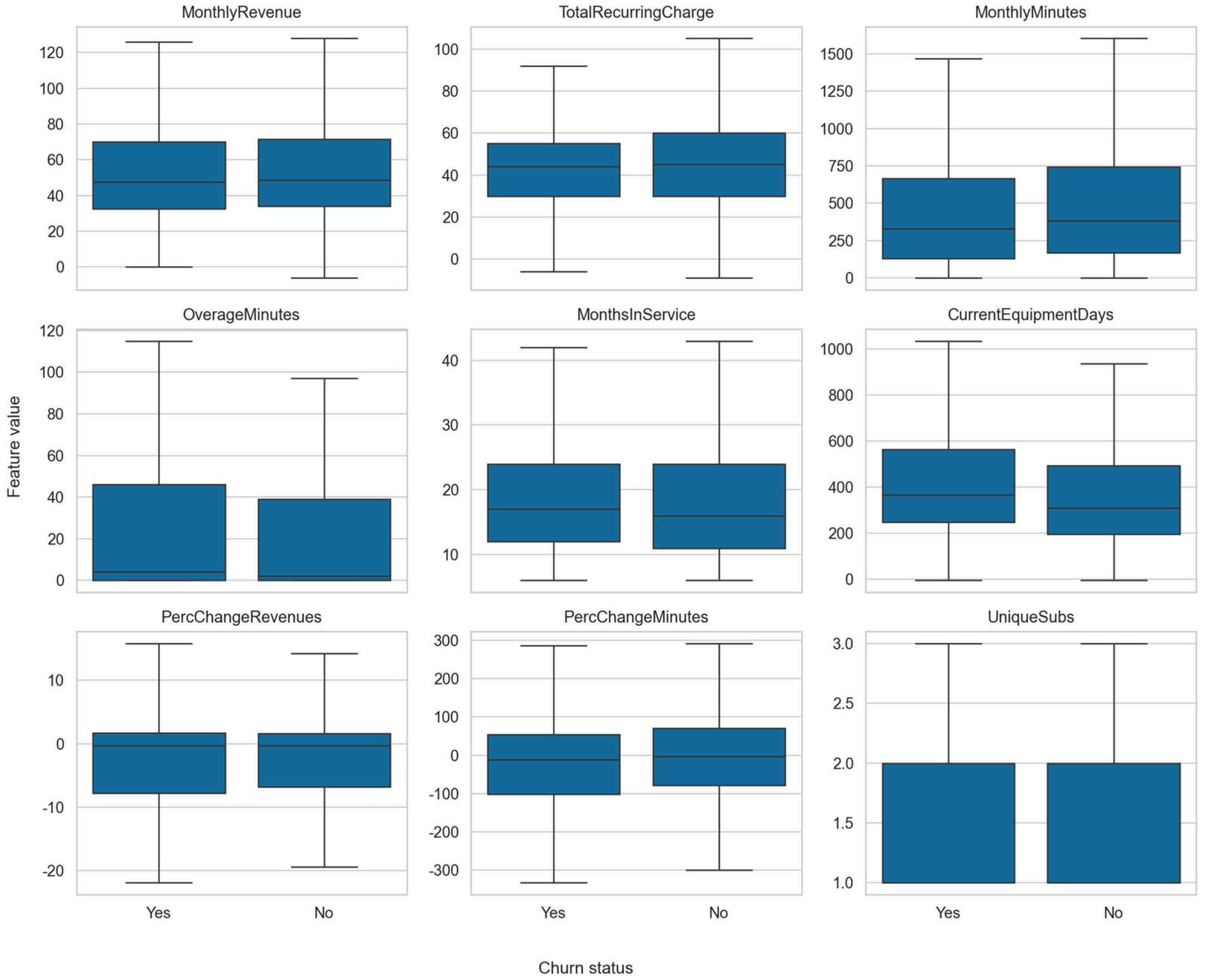
Q&A

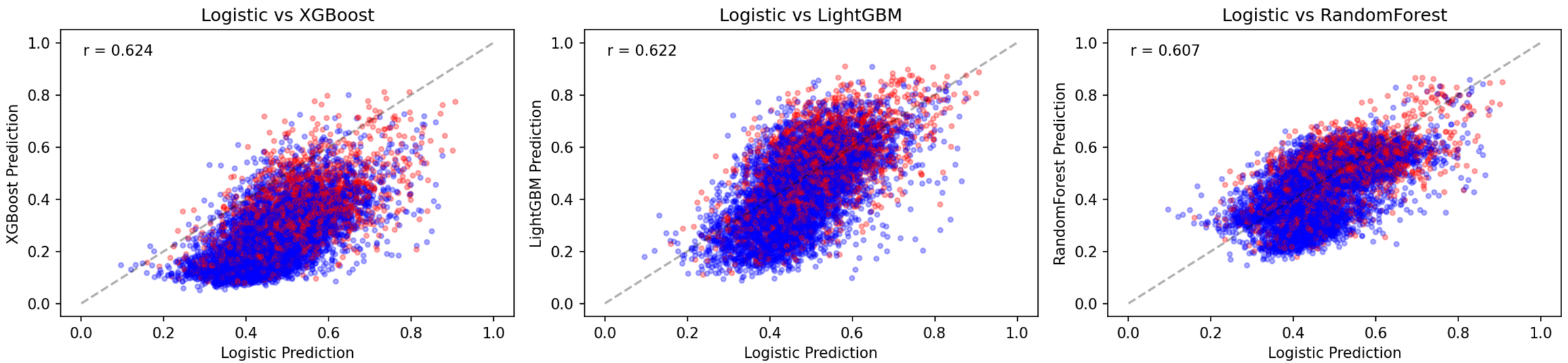


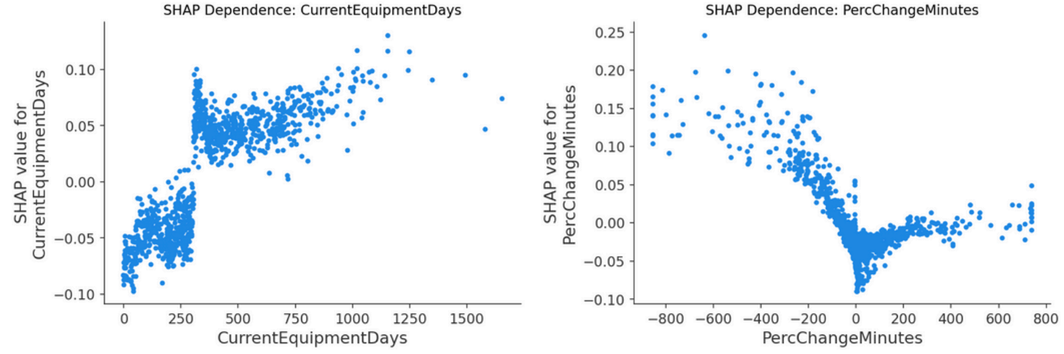
Backup Slide Menu



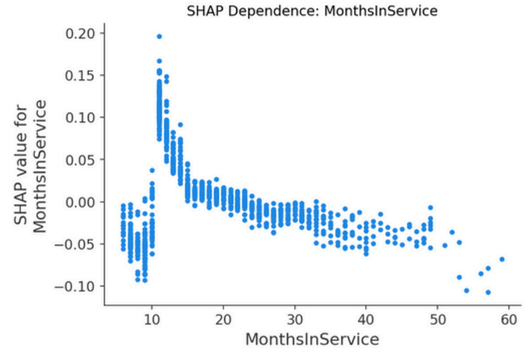


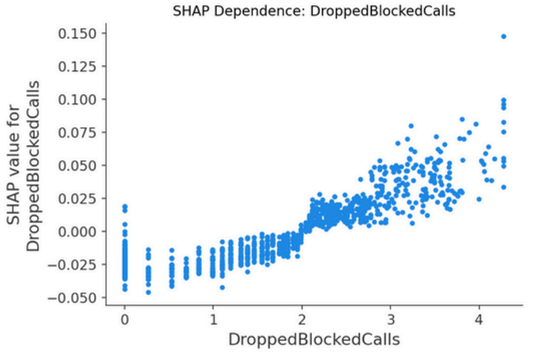

Distribution of key numeric features by Churn
MonthlyRevenue
TotalRecurringCharge
MonthlyMinutes
OverageMinutes
MonthsInService
CurrentEquipmentDays
PercChangeRevenues
PercChangeMinutes
UniqueSubs
Feature value
Yes
No
Churn status

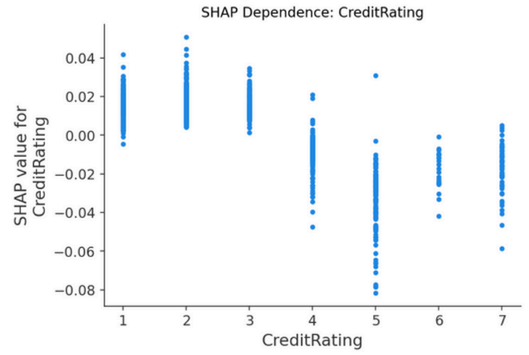

Base Model Predictions (Red=Churn, Blue=Stay)
Logistic vs XGBoost
r = 0.624
XGBoost Prediction
Logistic Prediction
Logistic vs LightGBM
r = 0.622
LightGBM Prediction
Logistic Prediction
Logistic vs RandomForest
r = 0.607
RandomForest Prediction
Logistic Prediction
0.0
0.2
0.4
0.6
0.8
1.0

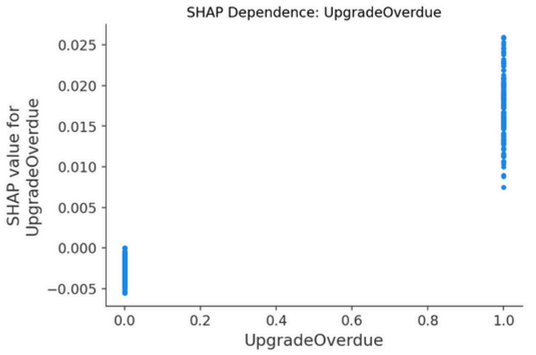

SHAP Dependence: CurrentEquipmentDays
SHAP value for CurrentEquipmentDays
CurrentEquipmentDays

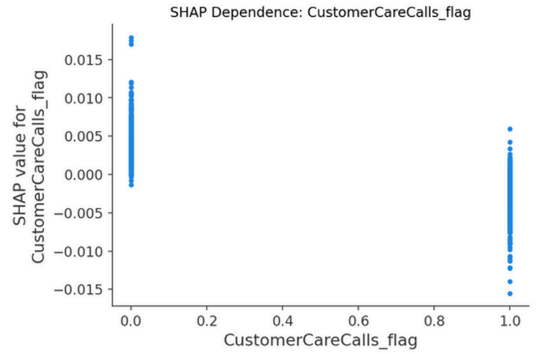
SHAP Dependence: PercChangeMinutes
SHAP value for PercChangeMinutes
PercChangeMinutes

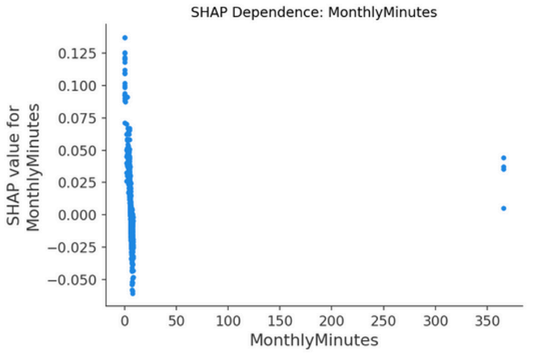

SHAP Dependence: MonthsInService
SHAP value for MonthsInService
MonthsInService

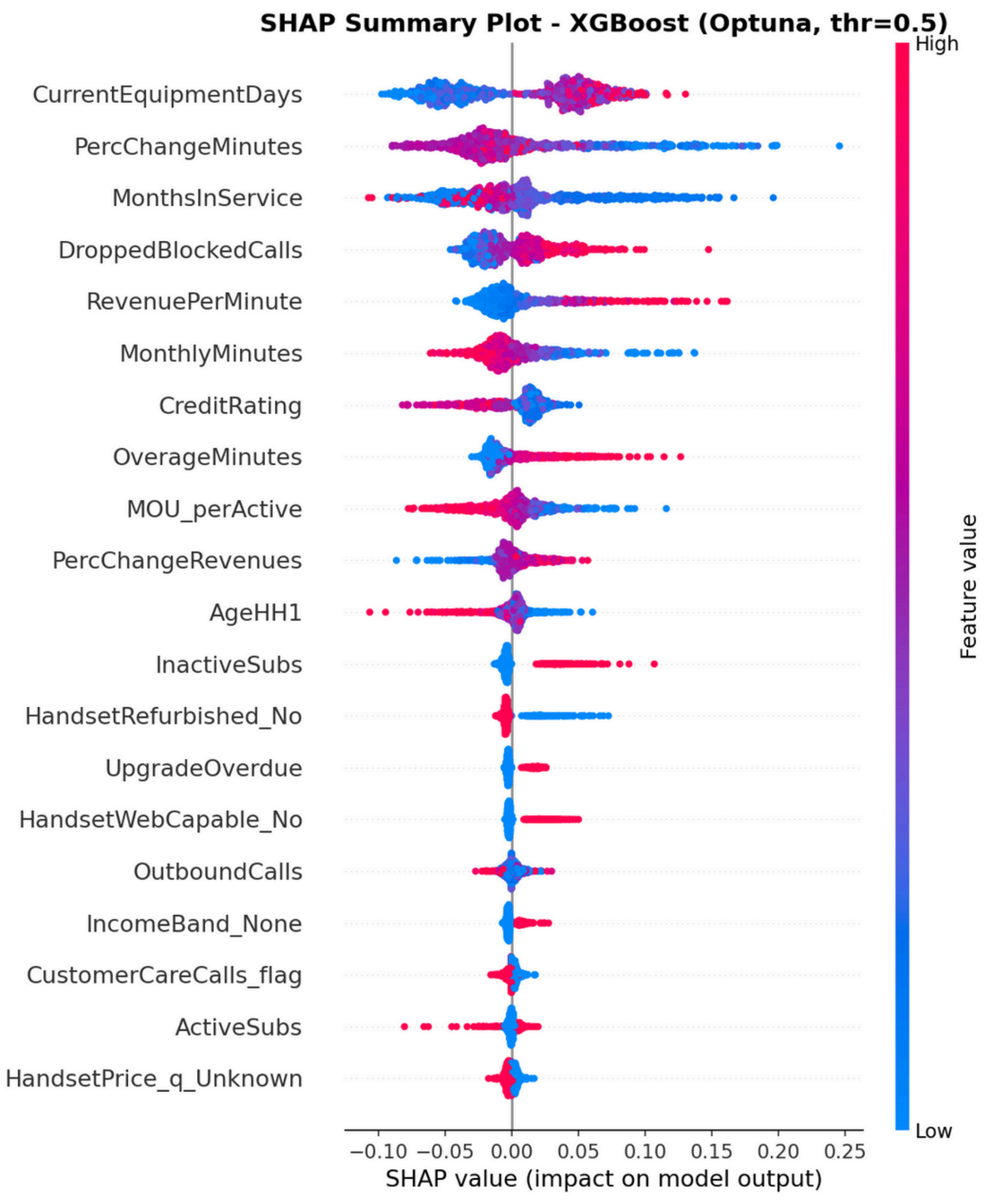

SHAP Dependence: DroppedBlockedCalls
SHAP value for DroppedBlockedCalls
DroppedBlockedCalls


SHAP Dependence: CreditRating
SHAP value for CreditRating
CreditRating


SHAP Dependence: UpgradeOverdue
SHAP value for UpgradeOverdue
UpgradeOverdue


SHAP Dependence: CustomerCareCalls_flag
SHAP value for CustomerCareCalls_flag
CustomerCareCalls_flag


SHAP Dependence: MonthlyMinutes
SHAP value for MonthlyMinutes
MonthlyMinutes

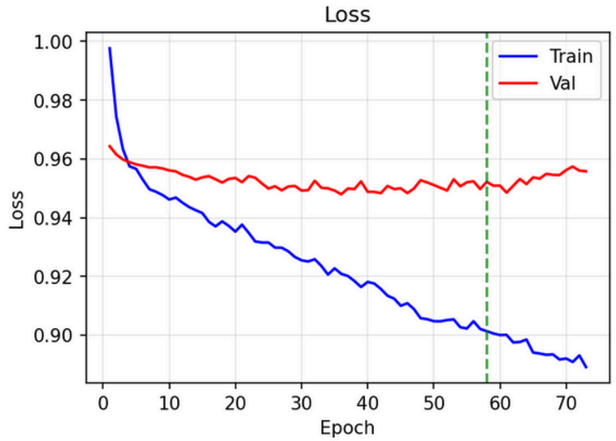

SHAP Summary Plot - XGBoost (Optuna, thr=0.5)
CurrentEquipmentDays
PercChangeMinutes
MonthsInService
DroppedBlockedCalls
RevenuePerMinute
MonthlyMinutes
CreditRating
OverageMinutes
MOU_perActive
PercChangeRevenues
AgeHH1
InactiveSubs
HandsetRefurbished_No
UpgradeOverdue
HandsetWebCapable_No
OutboundCalls
IncomeBand_None
CustomerCareCalls_flag
ActiveSubs
HandsetPrice_q_Unknown
−0.10 −0.05 0.00 0.05 0.10 0.15 0.20 0.25
SHAP value (impact on model output)
High
Low
Feature value

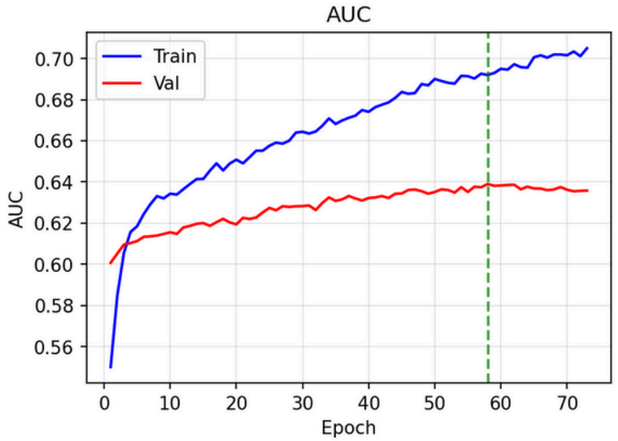
Model family	Examples in your project	Interpretability	Nonlinearity captured	Deployment complexity
Model family	Logistic Regression	**High** (coefficients, odds ratios)	**Low** (mostly linear unless manual interactions)	**Low** (fast, stable, easy to monitor)
Tree (single)	Decision Tree	**Medium** (rules, but unstable)	**Medium** (piecewise splits)	**Low–Medium** (simple but can overfit)
Tree ensemble (bagging)	Random Forest	**Medium–Low** (feature importance ok, less transparent)	**High**	**Medium** (bigger model, slower inference)
Gradient boosting	XGBoost / LightGBM	**Medium** (SHAP works well, but model is complex)	**High**	**Medium** (needs careful tuning + monitoring)
Deep learning (tabular)	Dense / Embedding / Wide&Deep (+ focal/weights)	**Low** (harder to explain; SHAP possible but heavier)	**High**	**High** (training/inference stack, reproducibility, drift)
Ensemble (stacking)	Stacking (meta-learner over base models)	**Medium–Low** (explain via meta-weights + SHAP per base)	**High** (combines diverse signals)	**High** (more components, OOF logic, versioning)

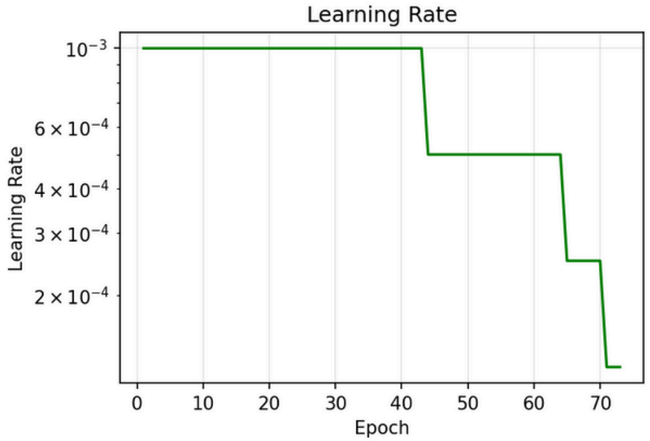
🏆 Model Performance Leaderboard (Test Set)

	Model	Family	Test AUC	Test PR-AUC	Test F1	Test Recall	Threshold
1	Stacking Ensemble	Ensemble	0.6773	0.4486	0.4965	0.6647	0.500
2	XGBoost (Optuna, thr=0.5)	GBM	0.6734	0.4484	0.1979	0.1205	0.500
3	XGBoost (Optuna, thr≈0.26)	GBM	0.6734	0.4484	0.4925	0.7331	0.258
4	XGBoost (Baseline)	GBM	0.6731	0.4440	0.2031	0.1242	0.500
5	LightGBM (Small Tuned)	GBM	0.6690	0.4448	0.4777	0.6017	0.500
6	Random Forest (Baseline)	Tree	0.6630	0.4306	0.4778	0.6053	0.500
7	NN - Wide & Deep (Focal)	NN	0.6584	0.4215	0.4852	0.6652	0.500
8	NN - Embedding (Focal)	NN	0.6575	0.4252	0.4782	0.6371	0.500
9	NN - Wide & Deep *	NN	0.6553	0.4259	0.4692	0.6094	0.500
10	NN - Dense (Baseline)	NN	0.6552	0.4192	0.4714	0.6176	0.500
11	NN - Embedding (Baseline)	NN	0.6549	0.4240	0.4737	0.6235	0.500
12	Decision Tree (Baseline)	Tree	0.6291	0.3794	0.4642	0.6588	0.500
13	Logistic Regression (Baseline)	Linear	0.6159	0.3774	0.4418	0.5677	0.500

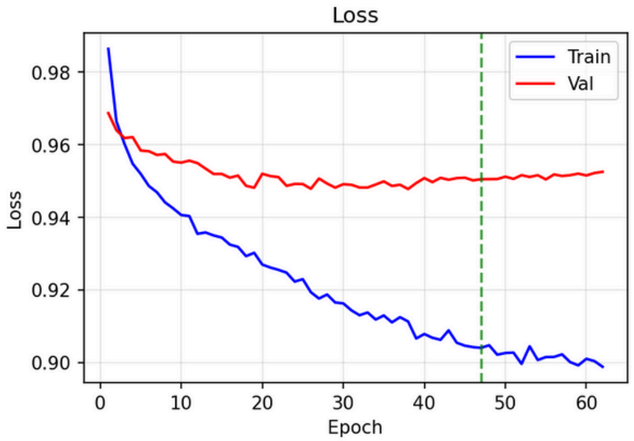
Training Curves: embedding_deep

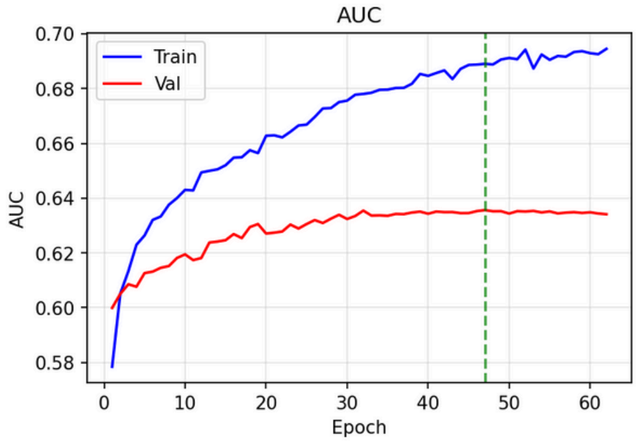


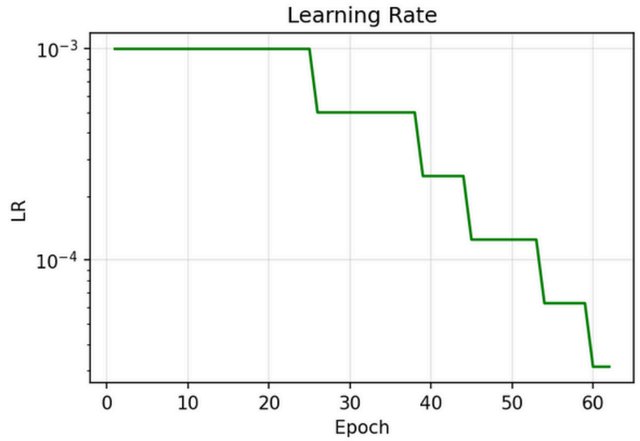


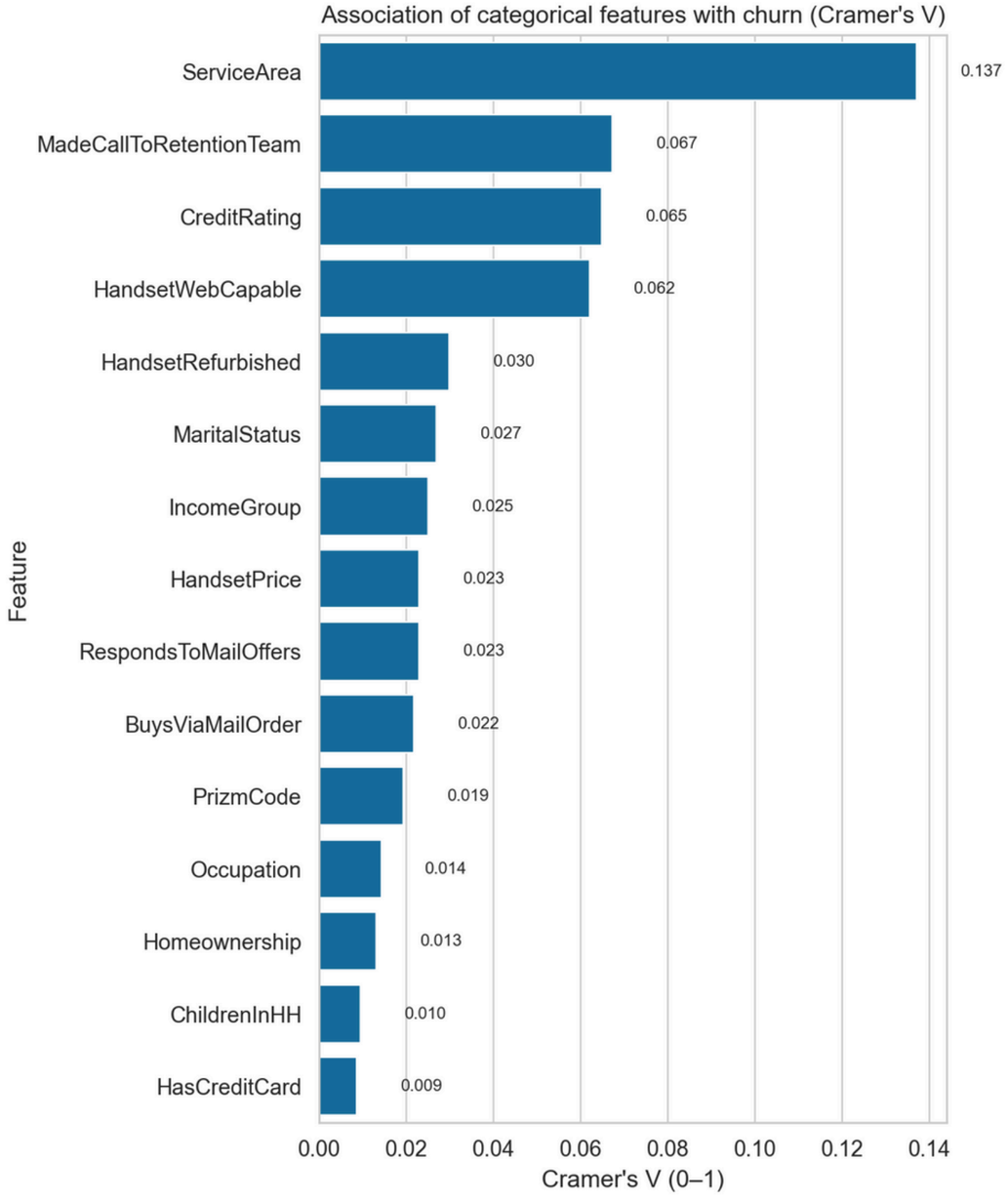


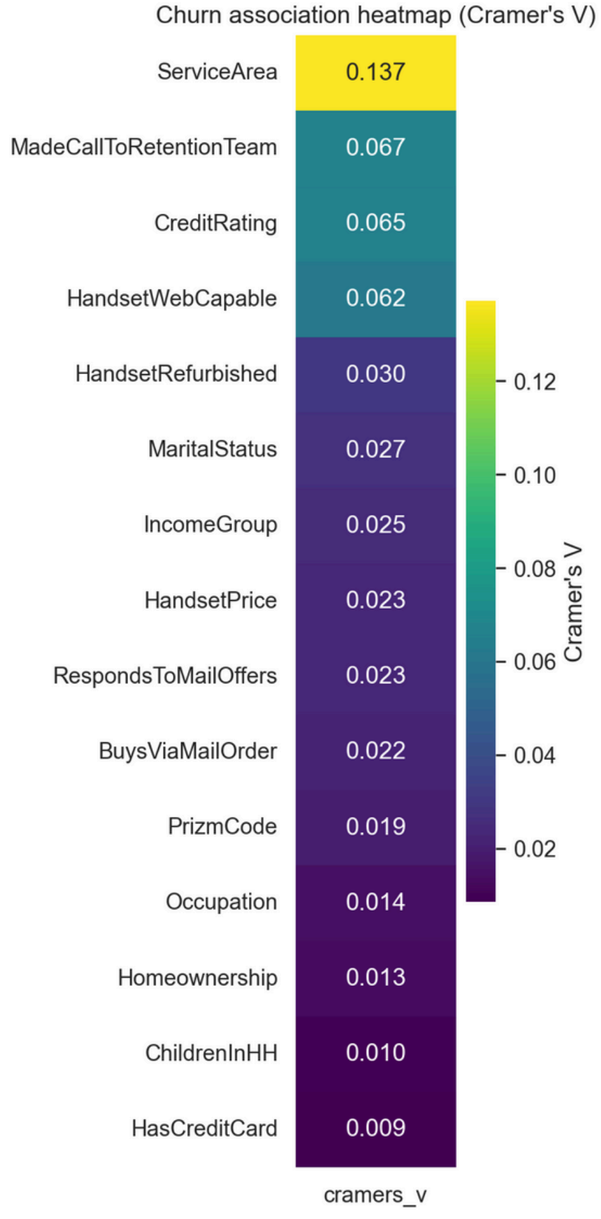
Training Curves: wide_and_deep

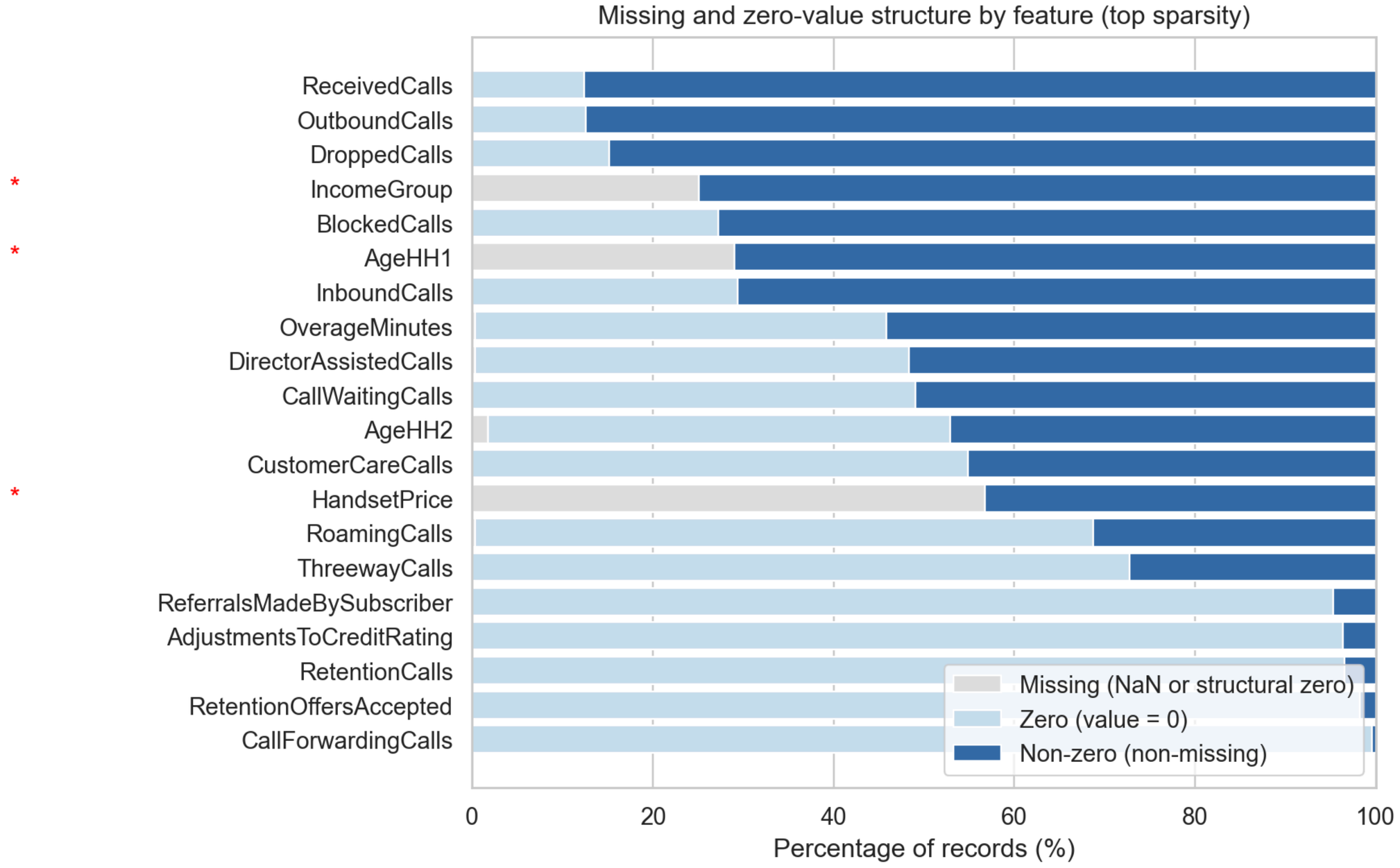




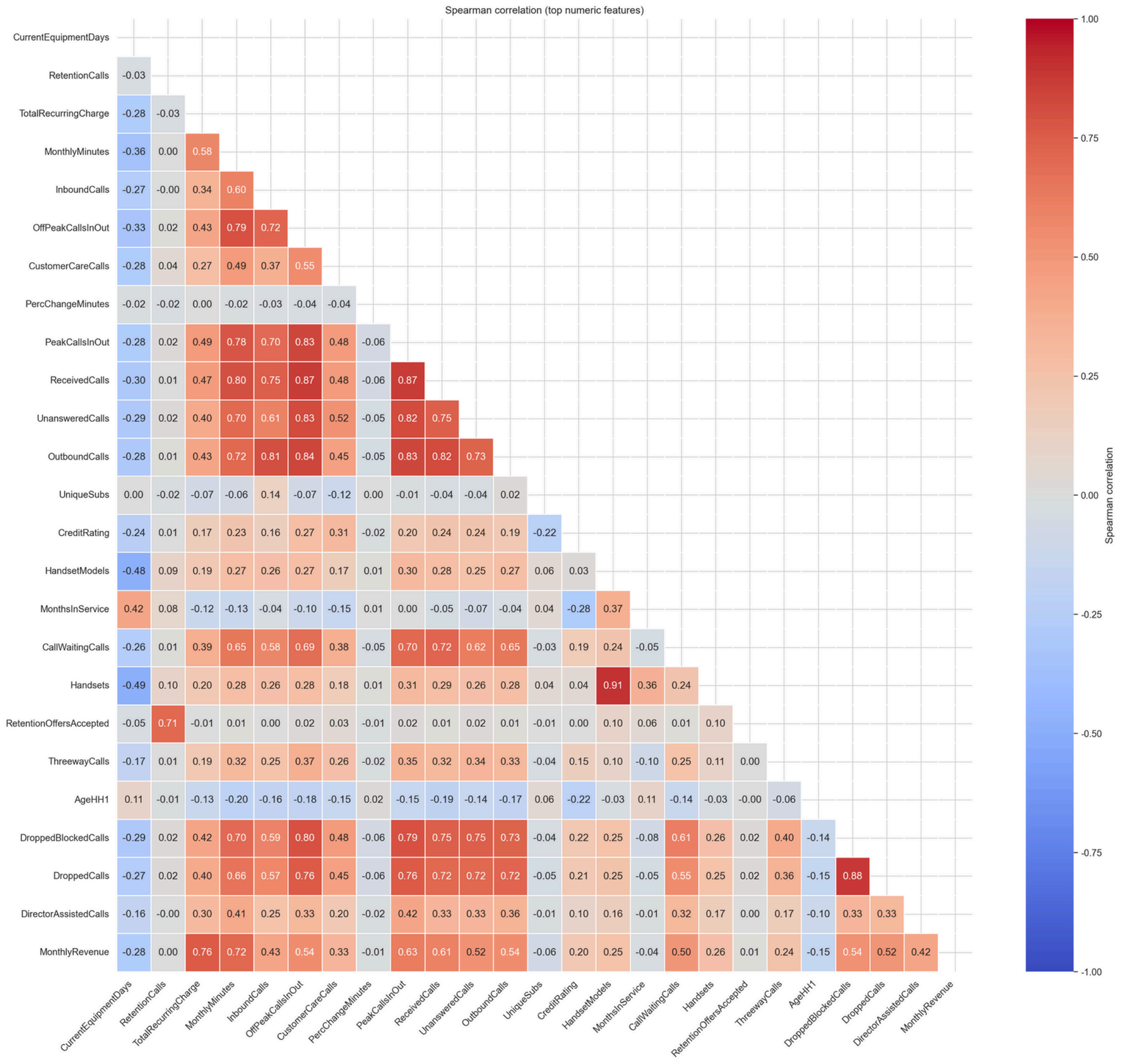


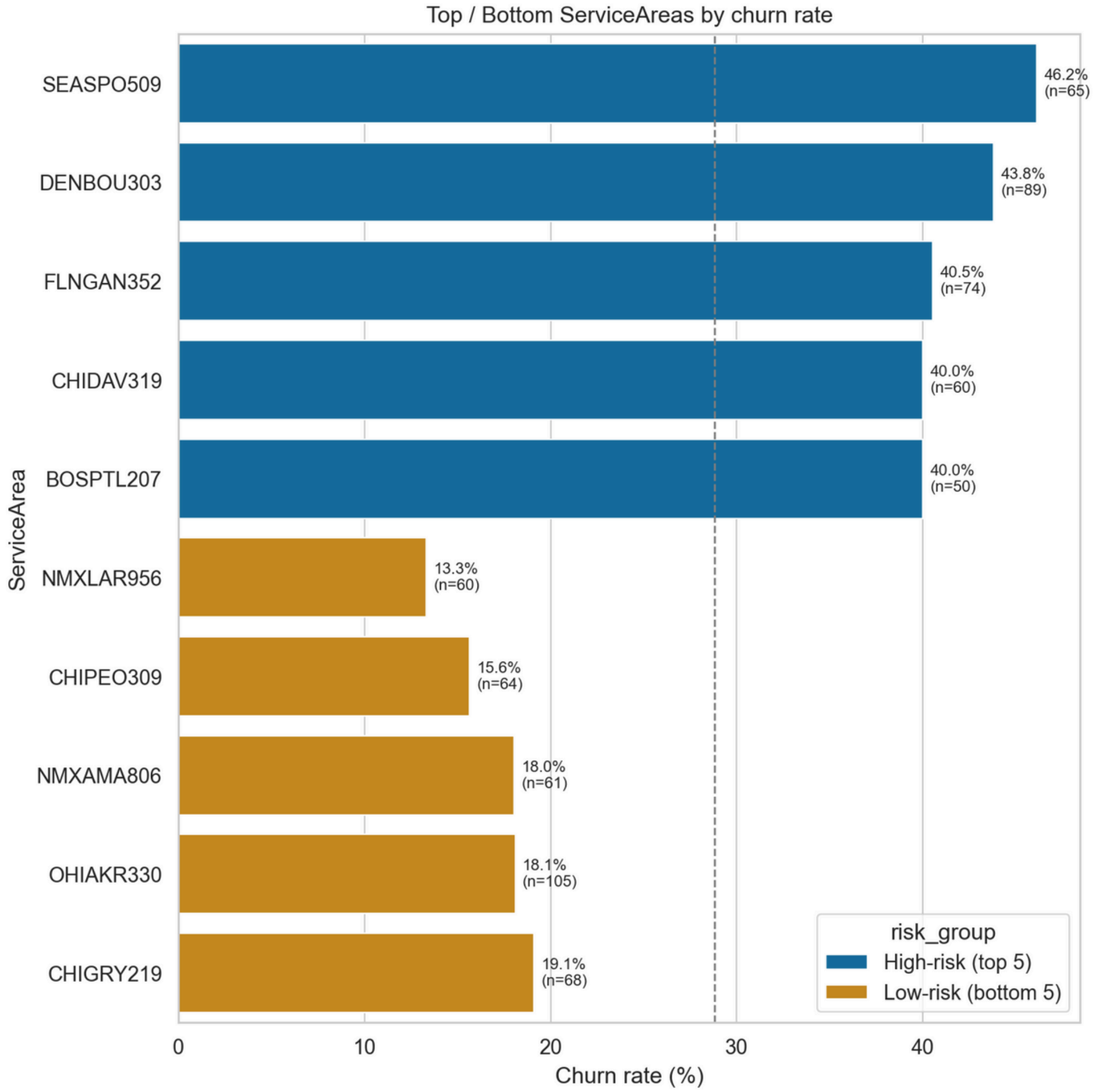

Association of categorical features with churn (Cramer's V)
ServiceArea
MadeCallToRetentionTeam
CreditRating
HandsetWebCapable
HandsetRefurbished
MaritalStatus
IncomeGroup
HandsetPrice
RespondsToMailOffers
BuysViaMailOrder
PrizmCode
Occupation
Homeownership
ChildrenInHH
HasCreditCard
0.137
0.067
0.065
0.062
0.030
0.027
0.025
0.023
0.023
0.022
0.019
0.014
0.013
0.010
0.009
Feature
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Cramer's V (0–1)


Churn association heatmap (Cramer's V)
ServiceArea
MadeCallToRetentionTeam
CreditRating
HandsetWebCapable
HandsetRefurbished
MaritalStatus
IncomeGroup
HandsetPrice
RespondsToMailOffers
BuysViaMailOrder
PrizmCode
Occupation
Homeownership
ChildrenInHH
HasCreditCard
0.137
0.067
0.065
0.062
0.030
0.027
0.025
0.023
0.023
0.022
0.019
0.014
0.013
0.010
0.009
cramers_v
Cramer's V
0.12
0.10
0.08
0.06
0.04
0.02



* 0 is treated as missing for this feature


Spearman correlation (top numeric features)
CurrentEquipmentDays
RetentionCalls -0.03
TotalRecurringCharge -0.28 -0.03
MonthlyMinutes -0.36 0.00 0.58
InboundCalls -0.27 -0.00 0.34 0.60
OffPeakCallsInOut -0.33 0.02 0.43 0.79 0.72
CustomerCareCalls -0.28 0.04 0.27 0.49 0.37 0.55
PercChangeMinutes -0.02 -0.02 0.00 -0.02 -0.03 -0.04 -0.04
PeakCallsInOut -0.28 0.02 0.49 0.78 0.70 0.83 0.48 -0.06
ReceivedCalls -0.30 0.01 0.47 0.80 0.75 0.87 0.48 -0.06 0.87
UnansweredCalls -0.29 0.02 0.40 0.70 0.61 0.83 0.52 -0.05 0.82 0.75
OutboundCalls -0.28 0.01 0.43 0.72 0.81 0.84 0.45 -0.05 0.83 0.82 0.73
UniqueSubs 0.00 -0.02 -0.07 -0.06 0.14 -0.07 -0.12 0.00 -0.01 -0.04 -0.04 0.02
CreditRating -0.24 0.01 0.17 0.23 0.16 0.27 0.31 -0.02 0.20 0.24 0.24 0.19 -0.22
HandsetModels -0.48 0.09 0.19 0.27 0.26 0.27 0.17 0.01 0.30 0.28 0.25 0.27 0.06 0.03
MonthsInService 0.42 0.08 -0.12 -0.13 -0.04 -0.10 -0.15 0.01 0.00 -0.05 -0.07 -0.04 0.04 -0.28 0.37
CallWaitingCalls -0.26 0.01 0.39 0.65 0.58 0.69 0.38 -0.05 0.70 0.72 0.62 0.65 -0.03 0.19 0.24 -0.05
Handsets -0.49 0.10 0.20 0.28 0.26 0.28 0.18 0.01 0.31 0.29 0.26 0.28 0.04 0.04 0.91 0.36 0.24
RetentionOffersAccepted -0.05 0.71 -0.01 0.01 0.00 0.02 0.03 -0.01 0.02 0.01 0.02 0.01 -0.01 0.00 0.10 0.06 0.01 0.10
ThreewayCalls -0.17 0.01 0.19 0.32 0.25 0.37 0.26 -0.02 0.35 0.32 0.34 0.33 -0.04 0.15 0.10 -0.10 0.25 0.11 0.00
AgeHH1 0.11 -0.01 -0.13 -0.20 -0.16 -0.18 -0.15 0.02 -0.15 -0.19 -0.14 -0.17 0.06 -0.22 -0.03 0.11 -0.14 -0.03 -0.00 -0.06
DroppedBlockedCalls -0.29 0.02 0.42 0.70 0.59 0.80 0.48 -0.06 0.79 0.75 0.75 0.73 -0.04 0.22 0.25 -0.08 0.61 0.26 0.02 0.40 -0.14
DroppedCalls -0.27 0.02 0.40 0.66 0.57 0.76 0.45 -0.06 0.76 0.72 0.72 0.72 -0.05 0.21 0.25 -0.05 0.55 0.25 0.02 0.36 -0.15 0.88
DirectorAssistedCalls -0.16 -0.00 0.30 0.41 0.25 0.33 0.20 -0.02 0.42 0.33 0.33 0.36 -0.01 0.10 0.16 -0.01 0.32 0.17 0.00 0.17 -0.10 0.33 0.33
MonthlyRevenue -0.28 0.00 0.76 0.72 0.43 0.54 0.33 -0.01 0.63 0.61 0.52 0.54 -0.06 0.20 0.25 -0.04 0.50 0.26 0.01 0.24 -0.15 0.54 0.52 0.42
CurrentEquipmentDays
RetentionCalls
TotalRecurringCharge
MonthlyMinutes
InboundCalls
OffPeakCallsInOut
CustomerCareCalls
PercChangeMinutes
PeakCallsInOut
ReceivedCalls
UnansweredCalls
OutboundCalls
UniqueSubs
CreditRating
HandsetModels
MonthsInService
CallWaitingCalls
Handsets
RetentionOffersAccepted
ThreewayCalls
AgeHH1
DroppedBlockedCalls
DroppedCalls
DirectorAssistedCalls
MonthlyRevenue
Spearman correlation
1.00
0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
-1.00


Top / Bottom ServiceAreas by churn rate
ServiceArea
SEASPO509
DENBOU303
FLNGAN352
CHIDAV319
BOSPTL207
NMXLAR956
CHIPEO309
NMXAMA806
OHIAKR330
CHIGRY219
46.2%
(n=65)
43.8%
(n=89)
40.5%
(n=74)
40.0%
(n=60)
40.0%
(n=50)
13.3%
(n=60)
15.6%
(n=64)
18.0%
(n=61)
18.1%
(n=105)
19.1%
(n=68)
risk_group
High-risk (top 5)
Low-risk (bottom 5)
0
10
20
30
40
Churn rate (%)