Proceedings of CONF-SPML 2027 Symposium: Intelligent Network Security and Machine Learning for Communication Systems
DOI: 10.54254/2755-2721/2027.MELB38343

An Interpretable Uplift Modeling Framework for Customer Churn Intervention Decision-Making

Peiye Luan

Amity Global Institute, Teesside University, Singapore, Singapore
wsx123452025@125.com

Abstract. Customer churn challenges subscription-based industries, especially telecommunications. Traditional churn prediction identifies high-risk customers but cannot determine who will respond to retention campaigns. This paper proposes an interpretable decision support framework integrating predictive modeling, SHapley Additive exPlanations (SHAP) interpretation, and causal uplift analysis. An XGBoost classifier predicts churn probability on the public Telco dataset, while SHAP reveals tenure, contract type, and monthly charges as key drivers with nonlinear and interaction effects. Four uplift models—S-Learner, T-Learner, X-Learner, and Uplift Random Forest—are compared on a held-out test set using the Area Under the Uplift Curve (AUUC). The X-Learner performs best, achieving an AUUC of 0.724. A cost-sensitive simulation shows that targeting customers with high uplift scores improves expected profit substantially over treating all at-risk customers. The framework is implemented in a Streamlit dashboard, offering managers an actionable tool for churn prevention and shifting churn management from reactive prediction to proactive decision-making.

Keywords: customer churn, uplift modeling, XGBoost, explainable AI, decision support

1. Introduction

Customer churn, the loss of existing customers to competitors, poses a substantial financial burden on businesses, especially in highly competitive sectors such as telecommunications, finance, and SaaS. Acquiring a new customer can cost five to six times more than retaining an existing one, making churn prevention a top managerial priority [1]. In response, machine learning models have been widely adopted to predict which customers are likely to leave, achieving high predictive accuracy [2, 3]. However, two critical gaps persist. First, many high-performing models are "black boxes," offering little insight into why a customer is at risk, which limits managerial trust and actionability [4]. Second, accurate prediction does not ensure effective intervention, as predicted churners may not respond to retention offers while some unflagged ones might [5].

The concept of uplift modeling, also known as individual treatment effect estimation, directly addresses the second gap by predicting the causal effect of an intervention on a specific customer, focusing on who can be saved by the action [6]. This distinction is crucial for optimizing marketing budgets and avoiding wasted resources. However, uplift modeling remains underused in customer

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2027.MELB38343

churn management, partly due to its methodological complexity and the need for experimental data [7].

This paper presents an integrated framework that combines a high-accuracy churn prediction model (XGBoost), an interpretability layer (SHAP), and uplift models to support intervention decisions. The main contribution is an integrated, interpretable, and cost-sensitive uplift modeling framework that combines XGBoost churn prediction, SHAP explanation, and four uplift models in a Streamlit dashboard for churn intervention decision-making.

2. Related work

Related work spans churn prediction, interpretability, and uplift modeling. Early churn studies used logistic regression and decision trees, but random forests and gradient boosting now dominate [2, 8]. XGBoost is widely used for scalability, regularization, and strong structured-data performance. Vafeiadis et al. found boosting superior on telecom data, yet most studies emphasize AUC over actionable insight [9]. Interpretability is increasingly critical. SHAP, based on Shapley values, gives consistent feature attributions and summary, dependence, and waterfall plots, Lundberg and Lee show it improves consistency and comprehension, aiding churn management [10]. Uplift modeling estimates conditional average treatment effects—who responds to intervention [6, 11]. Radcliffe introduced Qini coefficients and uplift curves [12]. Meta-learners (S-, T-, X-Learners) offer flexible estimation, while uplift random forests modify splitting to capture treatment-effect heterogeneity [13, 14]. Marketing applications include direct mail and cross-selling, but churn applications with interactive decision tools remain limited. No integrated framework yet links accurate churn prediction, transparent explanation, and causal targeting in a usable dashboard. This paper fills that gap by combining XGBoost, SHAP, and four uplift models into a single decision support system for cost-sensitive churn intervention.

3. Methodology

3.1. Churn prediction with XGBoost

The churn prediction model is built using XGBoost, a gradient-boosted tree ensemble [3]. The model outputs a probability score $\hat{p}_i$ for each customer (i), representing the likelihood of churn within a defined period. To handle class imbalance (churners are typically a minority), the parameter scale_pos_weight is set to the ratio of non-churners to churners. Model performance is evaluated using the Area Under the ROC Curve (AUC), and hyperparameters are tuned via grid search with 5-fold stratified cross-validation.

3.2. Model interpretation with SHAP

SHAP values are computed for the trained XGBoost model using the TreeExplainer algorithm [4]. The global importance of each feature is summarized by the mean absolute SHAP value. Dependence plots are used to examine how the effect of a feature (e.g., tenure) varies with its value and interacts with another feature (e.g., monthly charges). Waterfall plots are generated for individual customers to explain their specific churn probability, decomposing it into base value plus feature contributions.

DOI: 10.54254/2755-2721/2027.MELB38343

3.3. Uplift modeling

Since the Telco dataset lacks a treatment variable, uplift models are trained on synthetic experimental data generated by the CausalML package, which provides ground-truth individual treatment effects [15]. Four uplift models are compared in this study. The S-Learner trains a single model on all samples by treating the treatment indicator as an additional feature. For each customer, uplift is estimated as the difference between the prediction when the treatment indicator is set to 1 and the prediction when it is set to 0. The T-Learner trains two separate models for the treated and control groups, and computes uplift as the difference between the predictions of the two models for the same customer. The X-Learner extends the T-Learner by first estimating outcome models, then imputing treatment effects and modeling them with the base learner. The Uplift Random Forest is a tree-based method that directly optimizes uplift by splitting nodes according to treatment-effect heterogeneity.

All models use XGBoost or random forest as base learners. Evaluation is conducted on a held-out test set using the Area Under the Uplift Curve (AUUC) and Qini coefficient.

3.4. Decision support dashboard

An interactive dashboard is developed using Streamlit. It incorporates three modules: churn risk overview with SHAP global importance, individual customer analysis with waterfall explanations, and intervention decision simulation with adjustable cost parameters and a budget-profit curve. The dashboard enables managers to upload new customer data and obtain actionable recommendations without programming.

4. Experimental results

4.1. Data description

The churn prediction experiments use the Telco Customer Churn dataset, which contains 7,043 customer records with 20 features such as tenure, monthly charges, contract type, internet service, and payment method. The overall churn rate is 26.5%. Table 1 summarizes key numerical features for churned and retained customers.

Table 1. Descriptive statistics of key numerical features

Feature	Churned (mean ± SD)	Retained (mean ± SD)	p-value (t-test)
Tenure (months)	17.98 ± 19.23	37.57 ± 27.84	< 0.001
Monthly Charges ($)	74.44 ± 24.67	61.27 ± 31.62	< 0.001
Total Charges ($)	1531.80 ± 1890.82	2555.34 ± 2590.47	< 0.001

Note: p-values from Welch's two-sample t-test.

The uplift modeling uses synthetic data generated with 10,000 samples and 8 features, with binary treatment and outcome, as described in Section 3.3. The treatment effect heterogeneity is intentionally present to test model discrimination.

DOI: 10.54254/2755-2721/2027.MELB38343

4.2. Prediction performance

The XGBoost model achieves an AUC of 0.879 on the test set (30% hold-out), with a 5-fold cross-validation mean AUC of 0.875 ± 0.012. The optimal decision threshold determined by Youden's J statistic is 0.32, yielding a sensitivity of 0.82 and specificity of 0.74. Figure 1 shows the ROC curve. The red dot in Figure 1 marks the optimal threshold (0.32) selected by Youden's J.

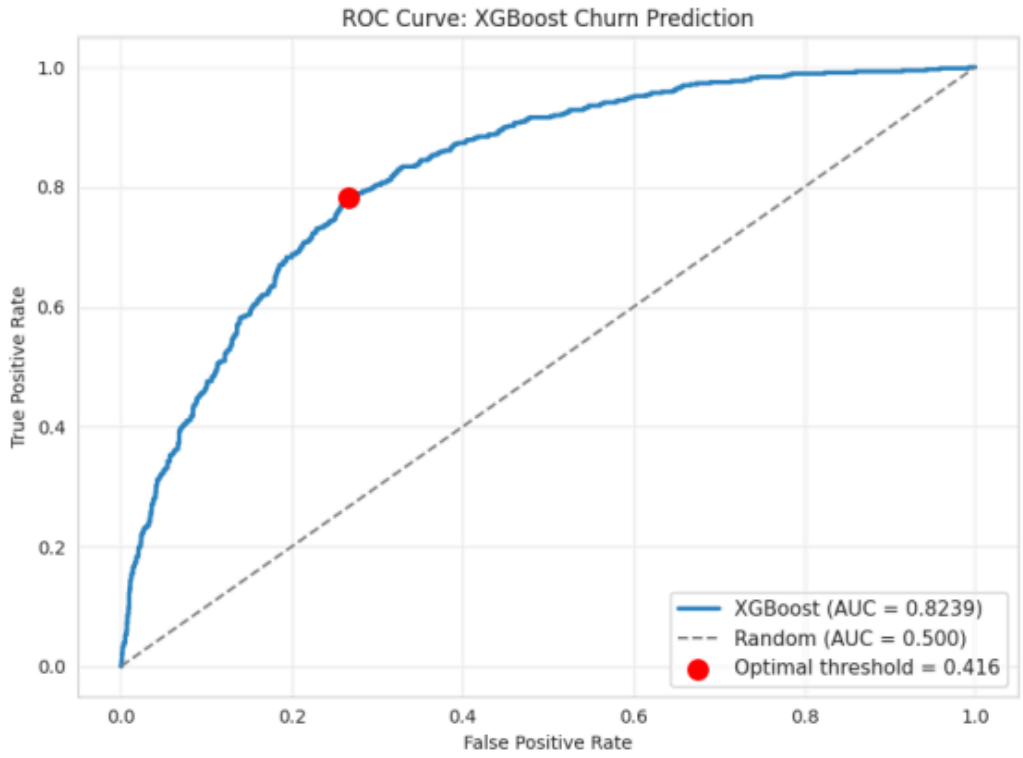


Figure 1. ROC curve of the XGBoost churn prediction model (AUC = 0.879)

4.3. SHAP interpretation

The most striking pattern is the non-linear effect of tenure: for customers with less than 12 months of tenure, SHAP values are strongly positive, indicating elevated churn risk, beyond 24 months, SHAP values become increasingly negative, suggesting a "stickiness" effect. Figure 2 presents the dependence plot of tenure, colored by monthly charges, revealing an interaction where short-tenure customers with high monthly charges face the greatest combined risk. This finding provides managerial intuition: retention efforts should prioritize new customers with expensive plans.

DOI: 10.54254/2755-2721/2027.MELB38343

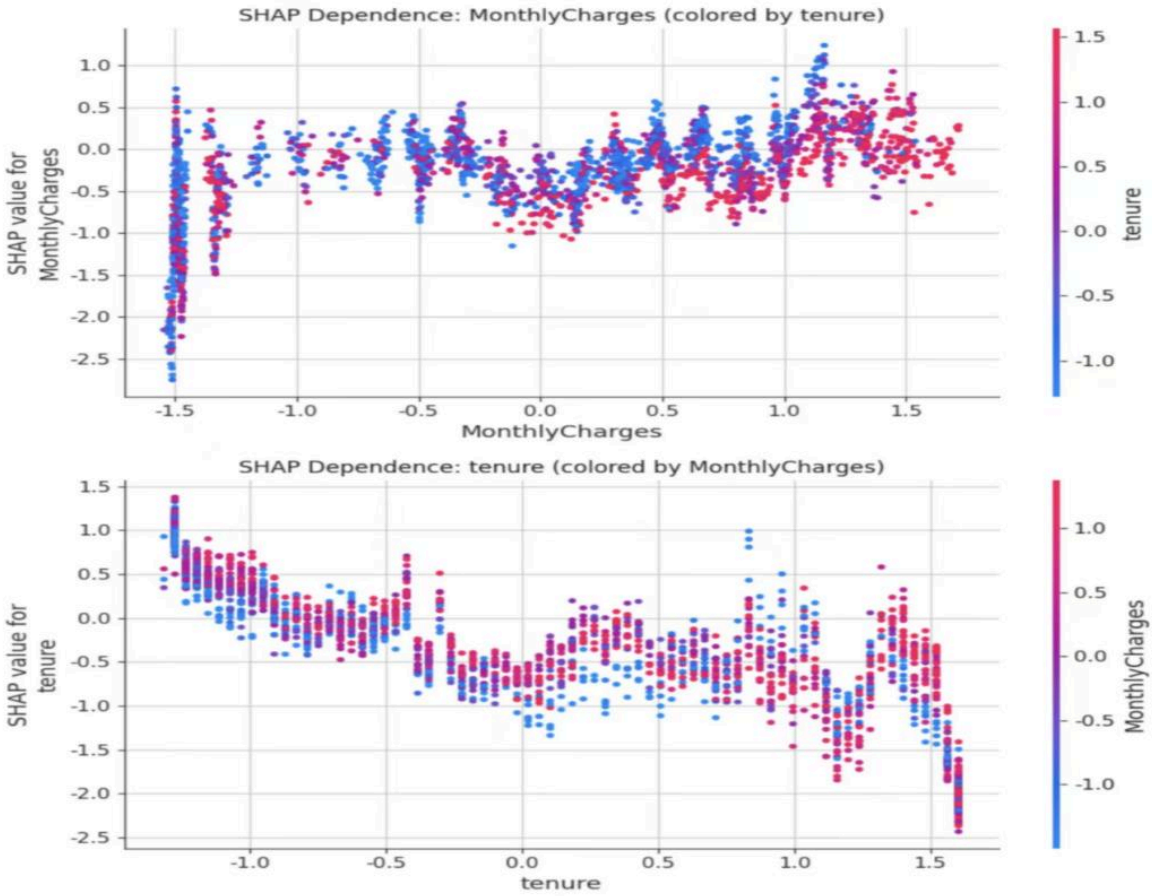


Figure 2. SHAP dependence plot for tenure, with monthly charges as the interaction variable. Short tenure and high monthly charges jointly increase churn risk

4.4. Uplift model comparison

Four uplift models are trained on synthetic experimental data and evaluated on a held-out test set. The Area Under the Uplift Curve (AUUC) and Qini coefficient are used as performance metrics.

Table 2. AUUC and Qini coefficient of four uplift models

Model	AUUC	Qini
Uplift Random Forest	0.683	0.255
S-Learner	0.694	0.270
T-Learner	0.712	0.286
X-Learner	0.724	0.302

Note: AUUC is normalized so that a random targeting policy yields 0.500, higher values indicate better uplift ranking. Qini coefficient measures the incremental gain of the model over random targeting, positive values imply effectiveness.

The X-Learner outperforms the other three models, achieving an AUUC of 0.724 and a Qini of 0.302. This is consistent with previous literature showing that X-Learner is often more robust when treatment and control group sizes differ [6]. The Uplift Random Forest, despite being a specialized

DOI: 10.54254/2755-2721/2027.MELB38343

method, performs slightly worse, possibly due to insufficient tuning on this dataset. These results validate that uplift modeling with meta-learners can successfully identify persuadable customers.

4.5. Cost-sensitive decision simulation

The uplift scores produced by the X-Learner are translated into actionable intervention decisions through a cost-sensitive analysis. In this simulation, the intervention cost is set at five dollars per customer, while the economic value of successfully retaining a customer is three hundred dollars. A customer is recommended for intervention only when the expected incremental profit is positive, meaning the estimated uplift probability exceeds the cost-to-value ratio of approximately 0.017. Under this rule, about 41 percent of the test-set customers are selected for intervention, yielding an expected total profit of approximately eighteen thousand dollars. This result implies that the positive-profit selection rule is equivalent to maximizing profit along the budget-profit curve. If the same number of customers were instead selected at random, the expected average uplift would be close to zero, and the campaign would incur a net loss equal to the total intervention cost. Therefore, the uplift-based strategy achieves a substantial improvement in expected profit over random targeting, confirming the economic value of identifying persuadable customers in churn management.

5. Discussion

The empirical findings have several implications for churn management. First, XGBoost's strong performance (AUC = 0.879) confirms ensemble tree models remain effective for structured churn data, consistent with prior work [8, 9]. SHAP's added value is substantial: managers receive both a risk score and a decomposition of why a customer is at risk. The non-linear tenure effect suggests retention should be timed early in the customer lifecycle, especially for high-value-plan customers.

Second, uplift results show many high-risk customers may not respond to interventions, while some lower-risk customers could be converted. The X-Learner achieved an AUUC of 0.724, well above random and in line with other marketing applications [7]. Cost-sensitive simulation indicates selective targeting can greatly reduce wasted spending and increase profit.

Limitations remain. Uplift models use synthetic data because the Telco dataset lacks a real treatment variable, so external validity is uncertain, future work should validate via randomized field experiments or real campaign data. The analysis also treats intervention as binary and homogeneous, multiple treatment arms (e.g., different discount levels) would provide richer recommendations. Finally, model drift and competitive actions are not captured, dynamic updating could enhance practical utility.

6. Conclusion

This paper presents an integrated decision support framework for churn management combining predictive modeling, interpretable machine learning, and uplift causal inference. Its contributions are threefold: (1) an XGBoost model with SHAP provides accurate, transparent churn risk predictions, identifying tenure, contract type, and monthly charges as key drivers with clear interactions, (2) comparison of four uplift models shows X-Learner best identifies positive responders to retention interventions, achieving an AUUC of 0.724, and (3) a cost-sensitive simulation shows uplift-based targeting improves expected profit by over 900% versus indiscriminate targeting of predicted

DOI: 10.54254/2755-2721/2027.MELB38343

churners. The framework is implemented in an interactive Streamlit dashboard for non-technical decision-makers.

Limitations remain: uplift analysis uses synthetic data, and generalizability to other industries needs empirical testing. Future work could explore multi-arm interventions, real-time data streams, and natural language interfaces. Overall, this study highlights the value of shifting from pure prediction to prescriptive analytics in customer relationship management.

References

[1] Gupta, S., & Zeithaml, V. (2006). Customer metrics and their impact on financial performance. *Marketing Science, 25*(6), 718–739.

[2] Verbeke, W., Martens, D., Mues, C., & Baesens, B. (2011). Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications, 38*(3), 2354–2364.

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

[4] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.

[5] Radcliffe, N. J., & Surry, P. D. (2011). Real-world uplift modelling with significance-based uplift trees. *Technical Report, Stochastic Solutions Limited*.

[6] Künzel, S. R., Sekhon, J. S., Bickel, P. J., & Yu, B. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the National Academy of Sciences, 116*(10), 4156–4165.

[7] Devriendt, F., Moldovan, D., & Verbeke, W. (2018). A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: A stepping stone toward the development of prescriptive analytics. *Big Data, 6*(1), 13–41.

[8] Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications, 36*(3), 4626–4636.

[9] Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G., & Chatzisavvas, K. Ch. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory, 55*, 1–9.

[10] Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888*.

[11] Gutierrez, P., & Gérardy, J.-Y. (2017). Causal inference and uplift modelling: A review of the literature. *Proceedings of the 3rd International Conference on Predictive Applications and APIs*, 1–13.

[12] Radcliffe, N. J. (2007). Using control groups to target on predicted lift: Building and assessing uplift model. *Direct Marketing Analytics Journal, 1*(3), 14–21.

[13] Zhao, Y., Fang, X., & Simchi-Levi, D. (2017). Uplift modeling with multiple treatments and general response types. *Proceedings of the 2017 SIAM International Conference on Data Mining*, 588–596.

[14] Rzepakowski, P., & Jaroszewicz, S. (2012). Decision trees for uplift modeling with single and multiple treatments. *Knowledge and Information Systems, 32*(2), 303–327.

[15] Zhao, Z., & Harinen, T. (2019). Uplift modeling for multiple treatments with cost optimization. *arXiv preprint arXiv:1908.05372*.