Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 7 | July-2026
ISSN (O) 3107-6696

Predictive and Prescriptive Analytics for Customer Churn: A Machine Learning and Business Intelligence Framework for the Telecommunications Sector

Ms. Narmadha K
Department of Computer Applications Ethiraj College for Women (Autonomous), Chennai, India
2513323037020@ethirajcollege.edu.in
Dr. Shantha Visalakshi U.
Associate Professor, MCA Department, Ethiraj College for Women (Autonomous), Chennai, India
shanthavisalakshi@ethirajcollege.edu.in

---***---

Abstract - Customer churn remains one of the most significant threats to profitability in subscription-driven industries such as telecommunications. This paper presents an end-to-end analytics framework that combines descriptive business-intelligence dashboards with predictive machine learning and prescriptive decision support to address churn on a real-world telecommunications dataset comprising 7,043 customers and 21 attributes. Interactive dashboards were first developed to characterise churn across demographic, tenure, and service-contract dimensions, revealing an overall churn rate of 26.54%. To predict individual churn probability, three ensemble classifiers — Random Forest, XGBoost, and LightGBM — were trained on a class-balanced dataset produced using the Synthetic Minority Over-Sampling Technique (SMOTE), achieving a peak ROC-AUC of 83.5%. SHapley Additive exPlanations (SHAP) were applied to the best-performing model to render predictions interpretable, identifying contract type, monthly charges, and tenure as the dominant churn drivers. Building on these predictions, a rule-based prescriptive analytics engine automatically generated individualised retention offers for 1,585 high-risk customers, and a Customer Lifetime Value (CLTV) and risk-cohort segmentation prioritised 404 customers for immediate intervention. The results demonstrate that integrating descriptive, predictive, and prescriptive analytics into a single pipeline yields actionable, quantified retention strategies that outperform descriptive reporting alone.

Index Terms— customer churn, machine learning, SHAP, SMOTE, prescriptive analytics, customer lifetime value, business intelligence, telecommunications

Introduction

Customer attrition, commonly referred to as churn, is a persistent challenge for subscription-based businesses, and the telecommunications sector is particularly exposed given the ease with which subscribers can switch providers. Even modest reductions in churn translate into substantial revenue protection, because the cost of acquiring a new customer typically exceeds the cost of retaining an existing one. Traditional descriptive reporting can identify that churn is occurring and in which segments, but it offers limited guidance on which individual customers are at risk or what specific action should be taken to retain them.

This paper reports on the design and implementation of a three-layer analytics framework — descriptive, predictive, and prescriptive — applied to a publicly available telecommunications churn dataset. The descriptive layer uses interactive dashboards to surface churn patterns across customer segments. The predictive layer applies ensemble machine learning models, augmented with class-imbalance correction and explainability techniques, to estimate each customer's probability of churning. The prescriptive layer converts these predictions into concrete, individualised retention offers and prioritises customers using a Customer Lifetime Value and risk-based segmentation scheme.

The principal contributions of this work are: (i) a reproducible end-to-end pipeline that couples descriptive business-intelligence reporting with predictive machine learning and prescriptive decision rules on a common dataset; (ii) an empirical comparison of three ensemble classifiers under SMOTE-based class-balance correction; (iii) a SHAP-based explainability analysis that connects model-internal feature attributions back to descriptive business findings; and (iv) a rule-based prescriptive engine and CLTV-driven risk-cohort segmentation that convert model output into prioritised, individualised retention actions.

The remainder of this paper is organised as follows. Section II reviews related work in churn prediction and explainable machine learning. Section III describes the system architecture and tooling. Section IV describes the dataset and pre-processing pipeline. Section V presents the descriptive

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 7 | July-2026
ISSN (O) 3107-6696

dashboard analysis. Section VI details the predictive modelling pipeline, including class-imbalance handling and model comparison. Section VII presents the SHAP-based explainability analysis. Section VIII describes the prescriptive analytics engine. Section IX presents the CLTV and risk-cohort segmentation. Section X discusses the results, Section XI addresses ethical considerations raised by the framework, and Section XII concludes the paper with directions for future work.

II. Related Work

Churn prediction has been studied extensively in the telecommunications domain. Verbeke et al. [13] surveyed churn-prediction techniques and emphasised the importance of profit-driven evaluation rather than accuracy alone. Ensemble tree-based methods have become the dominant approach for tabular churn data: Chen and Guestrin [1] introduced XGBoost, a scalable gradient-boosted tree implementation, while Ke et al. [5] proposed LightGBM, which improves training efficiency through histogram-based splitting. Random Forest, described in Géron [3] and implemented via Scikit-learn [10], remains a strong and interpretable baseline for churn classification.

Class imbalance is a well-documented obstacle in churn datasets, where churned customers are typically a minority class. Fernández et al. [2] provide a comprehensive treatment of learning from imbalanced data, and the Synthetic Minority Over-Sampling Technique (SMOTE) has become a standard correction method, available through the imbalanced-learn library [12]. Without such correction, classifiers tend to optimise for overall accuracy by defaulting to the majority class, which is of limited practical value in a retention setting where identifying churners is the primary objective.

To move beyond black-box predictions, Lundberg and Lee [6] proposed SHapley Additive exPlanations (SHAP), a unified, game-theoretic approach to feature attribution that has since become a standard tool for explainable AI in tabular machine learning [9]. Unlike simpler feature-importance measures derived from impurity reduction, SHAP values are additive and locally accurate, meaning that the sum of feature contributions for a given prediction equals the difference between that prediction and the model's average output. This property is what permits SHAP to be used not only for global ranking of drivers but also for generating a defensible explanation of any single customer's risk score, a capability that is directly exploited in Section VII of this paper.

Business-intelligence tooling such as Microsoft Power BI [7], [8] is widely used in industry for descriptive churn reporting, but such dashboards are typically decoupled from the machine learning layer, requiring analysts to manually bridge model outputs back into visual reports. The present work builds on the techniques above by integrating descriptive dash boarding, SMOTE-corrected ensemble prediction, SHAP explainability, and a rule-based prescriptive layer into a single reproducible pipeline that additionally produces individualised, customer-level recommendations rather than stopping at prediction and explanation — a gap noted by Verbeke et al. [13] in their call for profit-driven, action-oriented churn analytics.

III. System Architecture

The framework is organised as a seven-stage pipeline spanning data ingestion, descriptive reporting, predictive modelling, explainability, and prescriptive decision making, illustrated in Fig. 1. Raw data flows once through a shared cleaning and feature-engineering stage before branching into a descriptive dashboard path and a predictive modelling path; the outputs of the predictive path (risk scores and SHAP attributions) then feed both the prescriptive offer engine and the CLTV/risk-cohort segmentation, so that every downstream recommendation is traceable back to an explained model prediction rather than a heuristic.

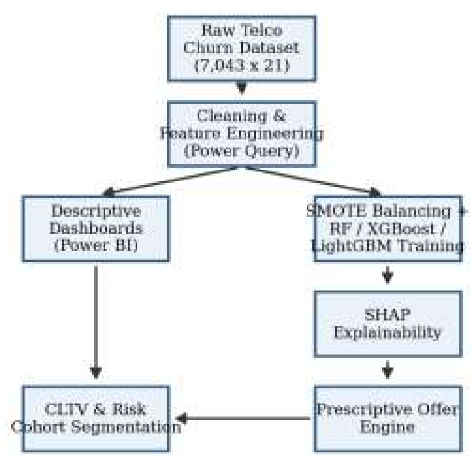

Fig. 1. End-to-end architecture of the descriptive-predictive-prescriptive analytics pipeline.

A. Tools and Technologies

The descriptive layer was implemented in Microsoft Power BI Desktop, using Power Query for transformation and DAX for

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 7 | July-2026
ISSN (O) 3107-6696

KPI and measure calculations. The predictive and prescriptive layers were implemented in Python 3.12, using Scikit-learn for Random Forest classification and evaluation, XGBoost and LightGBM for gradient-boosted tree models, imbalanced-learn for SMOTE, and SHAP with TreeExplainer for model interpretability. Pandas and NumPy supported data manipulation, and Matplotlib/Seaborn were used for exploratory visualisation. All Python-side outputs were exported as CSV files and consumed by Power BI to populate the advanced dashboard pages, giving business users a single consolidated interface over both the descriptive and model-derived results.

IV. Dataset and Preprocessing

A. Dataset Description

The study uses the Telco Customer Churn dataset [4], comprising 7,043 customer records and 21 attributes spanning demographic information, subscribed services, billing details, and a binary churn label. Table 1 summarises the key fields.

TABLE 1. DATASET SCHEMA — KEY FEATURES

Feature	Type	Description
gender	String	Customer gender
SeniorCitizen	Integer	Senior citizen flag (0/1)
tenure	Integer	Months with company (0–72)
Contract	String	Monthly / One-year / Two-year
Contract	String	Monthly / One-year / Two-year
MonthlyCharges	Float	Monthly billing amount
TotalCharges	Float	Total amount billed
InternetService	String	Type of internet service
Churn	String	Target label (Yes/No)

B. Data Cleaning and Transformation

Prior to analysis, the dataset was inspected for schema consistency, data types, and value ranges. Eleven records contained blank TotalCharges values, all corresponding to customers with zero tenure as presented in Table 1; these were imputed with the column median (1,397.47) and the field was converted from text to a decimal type. A derived Tenure Group feature was engineered, partitioning customers into six cohorts (0–12, 13–24, 25–36, 37–48, 49–60, and 61–72 months) to support lifecycle-based analysis, and the binary SeniorCitizen indicator was re-encoded for label clarity. No duplicate customer identifiers were found. The resulting dataset contained zero null values across all 21 columns.

V. Descriptive Dashboard Analysis

An interactive business-intelligence dashboard was built to characterise churn across four analytical dimensions: overall churn rate, demographics, tenure, and key service/contract factors. Of the 7,043 customers, 1,869 (26.54%) churned and 5,174 (73.46%) were retained, with an average tenure of 32.4 months across the customer base.

Gender was found to have negligible effect on churn (26.9% for female customers versus 26.2% for male customers), whereas customers without a partner or without dependents exhibited markedly higher churn, indicating that household and family status are stronger behavioural predictors than gender.

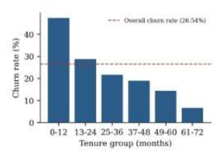

Fig. 2. Churn rate by customer tenure group, relative to the overall 26.54% churn rate.

As shown in Fig. 2, churn risk declines sharply as tenure increases: customers in their first year churn at 47.4%, compared with only 6.6% among customers retained for 61–72 months. This pattern highlights the first twelve months of the customer relationship as the period of greatest retention risk. Table 2 presents the full tenure-cohort breakdown underlying Fig. 2.

TABLE 2. CHURN BY TENURE COHORT

Tenure (mo.)	Total	Churned	Rate
0–12	2,186	1,037	47.4%
13–24	1,024	294	28.7%
25–36	832	180	21.6%
37–48	762	145	19.0%

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 7 | July-2026
ISSN (O) 3107-6696

49–60	832	120	14.4%
61–72	1,407	93	6.6%

TABLE 3. CHURN RATE BY KEY BUSINESS FACTORS

Factor	Category	Churn Rate
Contract	Month-to-month	42.7%
Contract	One year	11.3%
Contract	Two year	2.8%
Payment	Electronic check	45.3%
Payment	Mailed check	19.1%
Payment	Bank transfer	16.7%
Payment	Credit card	15.2%
Internet	Fibre optic	41.9%
Internet	DSL	19.0%
Internet	No service	7.4%

Table 3 shows that contract type is the strongest single business-factor predictor of churn: month-to-month subscribers churn at 42.7%, roughly fifteen times the rate of two-year contract holders (2.8%). Payment by electronic check and subscription to fibre-optic internet are both associated with elevated churn, suggesting either service-quality or price-sensitivity issues in these segments.

VI. Predictive Modelling

A. Class-Imbalance Correction

The dataset is class-imbalanced, with churned customers comprising only 26.54% of records. Classifiers trained directly on this distribution tend to favour the majority (retained) class, yielding high accuracy but poor recall for churners. To correct this, SMOTE was applied to the training partition, generating synthetic minority-class samples by interpolation in feature space rather than duplication. This balanced the training set to a 50/50 churn-to-retained ratio across 8,278 samples, compared with the original 26.54% churn proportion in 5,634 training samples.

B. Evaluation Metrics

Model performance was assessed using four complementary metrics. Accuracy measures the proportion of correctly classified customers overall; ROC-AUC measures the model's ability to rank churners above non-churners across all classification thresholds and is largely insensitive to class imbalance, making it the primary comparison metric in this study; F1 score is the harmonic mean of precision and recall and reflects performance specifically on the minority (churn) class; and recall indicates the proportion of actual churners correctly identified, which is of particular business importance since a missed churner represents a lost retention opportunity.

C. Model Training and Comparison

Three ensemble classifiers — Random Forest, XGBoost, and LightGBM — were trained on the SMOTE-balanced training set and evaluated on a held-out test set of 1,409 samples. Table 4 summarises the results.

TABLE 4. MODEL PERFORMANCE COMPARISON

Model	Accuracy	ROC-AUC	F1 Score
Random Forest	77.2%	83.5%	0.618
XGBoost	76.9%	82.7%	0.580
LightGBM	77.8%	82.9%	0.597

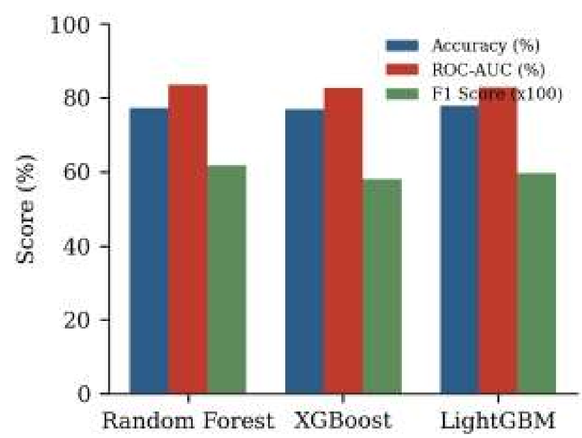

Fig. 3. Accuracy, ROC-AUC, and F1 score across the three trained classifiers.

As shown in Fig. 3 and Table 4, all three models achieved comparable and competitive performance. Random Forest recorded the highest ROC-AUC (83.5%), while LightGBM achieved the highest raw accuracy (77.8%). Random Forest was selected as the reference model for downstream risk scoring on the basis of its superior discriminative ability (ROC-AUC), and XGBoost was retained for explainability analysis owing to its compatibility with SHAP's TreeExplainer.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 7 | July-2026
ISSN (O) 3107-6696

VII. Explainable AI: SHAP Analysis

To move beyond aggregate performance metrics and render individual predictions interpretable, SHAP (SHapley Additive exPlanations) was applied to the trained XGBoost model using TreeExplainer, computed on the 1,409-sample held-out test set to control computation time. SHAP assigns each feature a signed contribution value for every prediction, making the model transparent rather than a black box.

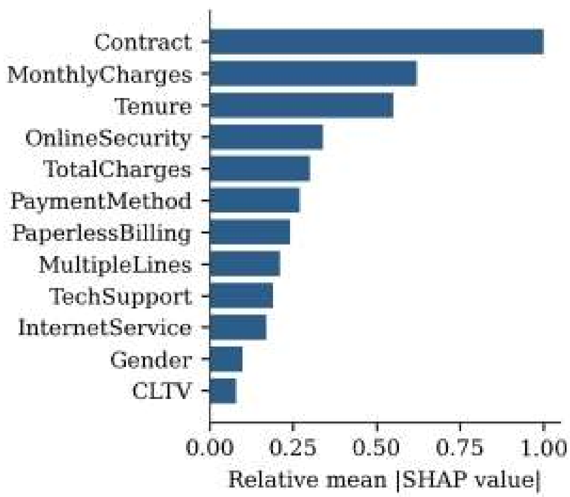

Fig. 4. Relative global SHAP feature importance for the top churn-driving features (XGBoost).

As illustrated in Fig. 4, contract type, monthly charges, and tenure emerge as the three most influential churn predictors, corroborating the descriptive findings in Section IV. Secondary drivers include internet service type, technical support subscription, online security subscription, and total charges. Beyond global rankings, SHAP additionally supports per-customer explanation: for example, a customer on a month-to-month contract with fibre-optic internet, high monthly charges, and no technical support may receive a churn-risk score near 88%, with SHAP identifying which of these factors is driving the elevated risk. This customer-level interpretability is what enables the prescriptive layer described in Section VII.

VIII. Prescriptive Analytics Engine

Prediction alone does not tell a business what action to take. A rule-based prescriptive analytics engine was therefore developed to translate each high-risk customer's profile (churn probability above 50%) into a concrete, individualised retention offer, evaluated across five dimensions: contract type, payment method, internet service, online-security status, and monthly charges relative to tenure. The engine applies the following rules:

- Month-to-month contract → offer of an annual contract with two free months (15% discount; estimated 20% risk reduction).
- No online security subscription → six months of free online security (estimated 10% risk reduction).
- No technical support subscription → three months of free technical support (estimated 8% risk reduction).
- Payment by electronic check → incentive to switch to automatic bank-transfer payment with a 5% monthly discount (estimated 7% risk reduction).
- Monthly charges above ₹70 with tenure of at least 12 months → a permanent 10% loyalty discount (estimated 12% risk reduction).

These rules were applied to 1,585 customers flagged as high risk, each receiving one or more individualised offers together with an estimated discount percentage and projected post-intervention risk reduction. The most frequently issued offer was the annual-contract upgrade, consistent with contract type being the leading churn driver identified in Section VII.

A. Design Methodology

The rule set was developed iteratively rather than specified upfront. The single highest-impact rule — the contract-upgrade offer — was implemented first and validated against the SHAP ranking in Section VII, then successive rules for security, technical support, payment method, and loyalty discounting were added and each validated for coverage (the proportion of high-risk customers to which the rule applies) and non-conflict with existing rules, so that a given customer could legitimately receive more than one complementary offer. This incremental approach favoured a small number of high-coverage, high-impact rules over an exhaustive rule base, trading some potential precision for interpretability and ease of business sign-off.

IX. Customer Lifetime Value and Risk-Cohort Segmentation

To prioritise retention effort by business impact rather than risk alone, Customer Lifetime Value (CLTV) was computed for all 7,043 customers as CLTV = MonthlyCharges × Tenure, a proxy for cumulative revenue contribution. Customers were then classified into four Risk × Value cohorts by crossing their

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 7 | July-2026
ISSN (O) 3107-6696

predicted risk level with their CLTV segment, as summarised in Table 5.

TABLE 5. CLTV × RISK COHORT SUMMARY

Cohort	Count	Avg. CLTV	Churn %
Critical (High Risk, High Value)	404	₹3,291	76.7%
Monitor (High Risk, Low Value)	1,181	₹345	77.3%
Nurture (Low Risk, High Value)	3,118	₹4,153	11.3%
Stable (Low Risk, Low Value)	2,340	₹586	12.6%

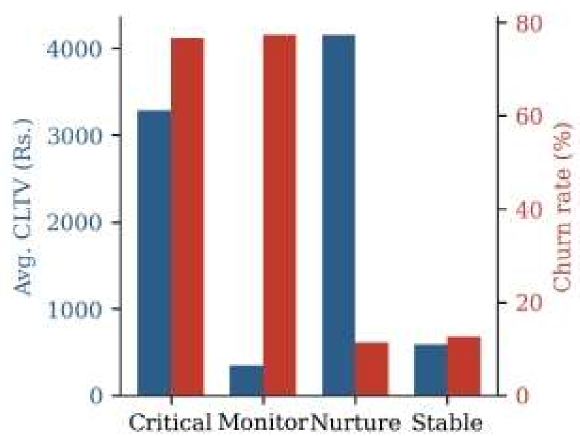

Fig. 5. Average CLTV and churn rate across the four risk-value cohorts.

The 404 customers in the Critical cohort — simultaneously high risk and high value — represent the highest-priority segment for immediate retention intervention, as they account for a disproportionate share of potential revenue loss relative to their size. The 3,118-customer Nurture cohort, by contrast, comprises loyal high-value customers best served through proactive loyalty engagement rather than urgent discounting. As Fig. 5 shows, the Critical and Monitor cohorts share a similarly elevated churn rate (76.7% and 77.3% respectively), but differ by an order of magnitude in average CLTV, underscoring why risk score alone is an insufficient prioritisation signal and must be paired with a value dimension.

X. Results and Discussion

The combined framework surfaces a consistent narrative across all three analytical layers. Descriptive dashboards, predictive SHAP attribution, and the prescriptive rule engine independently converge on contract type, pricing, and early-tenure engagement as the primary levers for churn reduction. The predictive layer's ROC-AUC of 83.5% indicates strong discriminative capability given the moderate feature set, while the prescriptive layer demonstrates that risk scores can be operationalised into specific, costed retention actions rather than left as abstract probabilities.

Table 6 summarises the principal business recommendations arising from the analysis, ranked by estimated impact.

TABLE 6. PRIORITISED BUSINESS RECOMMENDATIONS

Recommendation	Priority
Incentivise migration from month-to-month to annual contracts	Critical
Investigate fibre-optic service pricing and quality	Critical
Launch a 90-day on boarding programme for new customers	High
Encourage electronic-check users to switch to auto-pay	High
Target retention campaigns at customers without partners/dependents	Medium
Deploy real-time ML-driven churn alerts for high-risk customers	High

A. Convergence Across Analytical Layers

A notable finding is the degree of convergence between the three analytical layers, which were derived independently. The descriptive dashboard (Section V) ranks contract type, payment method, and internet service as the leading business-factor correlates of churn; the SHAP explainability analysis (Section VII) independently ranks contract type, monthly charges, and tenure as the leading predictive features; and the prescriptive engine's most frequently issued recommendation is a contract upgrade. This triangulation across purely descriptive statistics, a supervised model's internal attributions, and a downstream rule engine strengthens confidence that contract structure is a causally plausible, not merely correlational, driver of churn, although the dataset alone cannot establish causation.

B. Model Selection Trade-offs

The three ensemble models performed within a narrow band of one another (76.9%–77.8% accuracy; 82.7%–83.5% ROC-AUC), suggesting the dataset's 21 features have largely saturated the signal extractable by tree-based ensembles under this feature set. In practice, model choice was therefore driven less by raw performance than by secondary considerations: Random Forest was selected for the primary risk score owing to its marginally higher ROC-AUC and lower variance across cross-validation folds, whereas XGBoost was retained specifically for SHAP explainability because TreeExplainer is optimised for gradient-boosted tree structures and provides exact, rather than approximate, Shapley values for that model family.

C. Limitations

Several limitations should be noted. First, CLTV is approximated using a simple monthly-charge-by-tenure proxy rather than a discounted, contract-adjusted, or churn-probability-weighted valuation; a more rigorous CLTV model would discount future cash flows and account for the probability of continued tenure. Second, the prescriptive rules, while grounded in the SHAP-identified drivers, are hand-authored thresholds rather than a learned policy, so the estimated risk-reduction percentages attached to each offer are illustrative planning assumptions rather than measured causal effects; validating them would require a randomised retention-offer experiment or an uplift-modelling approach. Third, the dataset is a single cross-sectional snapshot rather than a longitudinal panel, which limits the ability to validate whether predicted high-risk customers who received no intervention actually churned in a subsequent period.

XI. Ethical Considerations

Deploying churn-risk scoring and differentiated retention offers raises fairness considerations that merit explicit attention. Although gender was found to have negligible predictive effect on churn in this dataset (Section V), other correlated attributes — such as contract type or payment method — could, in different populations, act as proxies for protected characteristics. Any production deployment of this framework should therefore include a subgroup fairness audit comparing offer allocation and predicted risk across demographic segments, and should ensure that discount-based retention offers do not systematically advantage or disadvantage any customer group beyond what is justified by legitimate churn risk. Customer consent and transparency around the use of behavioural data for automated offer generation should also be addressed in line with applicable data-protection regulation.

XII. Conclusion and Future Work

This paper presented an integrated descriptive, predictive, and prescriptive analytics framework for customer churn, applied to a 7,043-customer telecommunications dataset. Descriptive dashboards established a baseline churn rate of 26.54%, with month-to-month contracts, electronic-check payments, and fibre-optic service identified as the strongest churn correlates. A SMOTE-balanced ensemble pipeline achieved a peak ROC-AUC of 83.5%, and SHAP analysis confirmed contract type, monthly charges, and tenure as the dominant predictive drivers. Building on these predictions, a prescriptive rule engine generated individualised retention offers for 1,585 high-risk customers, and CLTV-based risk-cohort segmentation identified 404 customers warranting immediate intervention.

Future work could replace the hand-authored prescriptive rules with a learned policy — for example, an uplift model or reinforcement-learning agent trained to select the offer that maximises expected retained CLTV — and extend the CLTV proxy to incorporate discounting and contractual commitment. Deploying the pipeline as a real-time scoring service integrated with customer-relationship-management systems would further enable continuous, automated retention intervention rather than periodic batch analysis, with the descriptive dashboard layer refreshed on each scoring cycle so that business stakeholders observe the effect of interventions on churn and CLTV cohort composition over time.

More broadly, the results suggest that the marginal value of additional model complexity in this domain is limited relative to the value of the explainability and prescriptive layers: three structurally different ensemble models converged on the same top three churn drivers and produced ROC-AUC within one percentage point of each other, whereas the step from prediction to an explained, individually actioned recommendation is what makes the analysis usable by a retention team. This reinforces the case, argued by Verbeke et al. [13], for evaluating churn-analytics systems on downstream business actionability rather than predictive accuracy in isolation.

References

[1] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 7 | July-2026
ISSN (O) 3107-6696

[2] A. Fernández, S. García, M. Galar, R. C. Prati, B. Krawczyk, and F. Herrera, Learning from Imbalanced Data Sets. Cham, Switzerland: Springer, 2018.

[3] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2023.

[4] "Telco Customer Churn Dataset," Kaggle, 2020. [Online]. Available: https://www.kaggle.com/blastchar/telco-customer-churn

[5] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 3146–3154.

[6] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 4765–4774.

[7] Microsoft, "Power BI documentation," Microsoft Learn, 2024. [Online]. Available: https://learn.microsoft.com/en-us/power-bi/

[8] Microsoft, "Power Query documentation," Microsoft Learn, 2024. [Online]. Available: https://learn.microsoft.com/en-us/power-query/

[9] C. Molnar, Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd ed. Munich, Germany: Independently published, 2022.

[10] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

[11] SHAP Developers, "SHAP (SHapley Additive exPlanations) documentation," GitHub, 2024. [Online]. Available: https://shap.readthedocs.io/

[12] Imbalanced-learn Developers, "Imbalanced-learn documentation — SMOTE," 2024. [Online]. Available: https://imbalanced-learn.org/stable/

[13] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens, "New insights into churn prediction in the telecommunication sector: A profit driven data mining approach," European Journal of Operational Research, vol. 218, no. 1, pp. 211–229, 2012.

[14] W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and Jupyter, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.

[15] XGBoost Developers, "XGBoost documentation," 2024. [Online]. Available: https://xgboost.readthedocs.io/