International Journal of Scientific Research and Engineering Development-– Volume 9 Issue 3, May-June 2026
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Customer Churn Prediction Framework for Subscription-Based Services Using Machine Learning

M. Vasuki[1], Sudharsanan R[2]
1Professor, Department of Master of Computer Applications, Sri Manakula Vinayagar Engineering College, Pondicherry – 605 107.
2PG Student, Department of Master of Computer Applications, Sri Manakula Vinayagar Engineering College, Pondicherry – 605 107.

----------------------------------*************************-------------------------------

Abstract:

Customer churn is one of the major challenges faced by subscription-based businesses such as SaaS, telecom, and streaming platforms. Predicting customer churn helps companies identify users who are likely to stop using their services. This project presents a Customer Churn Prediction Framework using machine learning techniques. The system collects and analyzes customer behavioral and subscription data. Feature engineering methods are used to improve the quality of prediction. Multiple machine learning models are applied to achieve better accuracy and performance. The framework helps businesses take proactive retention actions for high-risk customers. It reduces customer loss and improves customer satisfaction. The proposed model also increases customer lifetime value and business revenue. Overall, the framework provides an efficient and reliable solution for customer retention management.

Keywords — **Put your keywords here, keywords are separated by comma.**

----------------------------------*************************-------------------------------

1. INTRODUCTION

Subscription-based business models have become the dominant revenue architecture across the technology, media, and telecommunications industries. Platforms offering cloud software, streaming content, mobile connectivity, and digital information services depend on the compounding value of recurring revenue — a model that is uniquely vulnerable to voluntary customer cancellation, or churn. When a subscriber terminates their relationship with a service, the financial impact extends beyond lost monthly revenue to encompass the full future lifetime value that would have accrued over continued engagement.

The asymmetry between customer acquisition cost (CAC) and retention cost makes churn prediction an economically compelling problem. Industry research consistently demonstrates that CAC exceeds the cost of retention by a factor of five or more, and that incremental improvements in churn rate translate non-linearly into subscriber base and revenue gains. A five percent reduction in monthly churn, for example, can expand the steady-state subscriber base by over thirty percent within three years under constant acquisition assumptions.

Machine learning provides a powerful tool for churn prediction because churn is rarely an abrupt event. It is typically the culmination of a multi-week deterioration in engagement, satisfaction, and perceived value — a process that generates observable signals across product usage logs, billing systems, customer support records, and communication engagement data. If these signals are appropriately collected, engineered into predictive features, and modeled with sufficient accuracy, organizations can identify at-risk customers weeks before cancellation and deploy targeted retention interventions.

Despite a substantial academic literature on churn prediction, three practical gaps persist. First, most published frameworks rely on single-source signal domains, typically product usage logs, ignoring the complementary predictive value of billing, support, and sentiment signals. Second, academic benchmarks rarely extend to operational deployment guidance, leaving organizations without tested approaches for model integration, score refresh cadence, and intervention design. Third, the economic impact of model-guided interventions is rarely measured in live deployments, making it difficult for business stakeholders to assess the return on investment of churn prediction systems.

This paper addresses all three gaps by proposing and empirically validating an integrated, end-to-end churn prediction framework. The framework spans the full operational lifecycle from data signal ingestion through risk-tiered business intervention, evaluated through both offline model benchmarking and a prospective 90-day controlled deployment experiment on a production subscriber base.

2. LITERATURE SURVEY

Machine learning and predictive analytics have been applied to customer churn prediction across multiple subscription verticals since the early 2000s. The field has progressed from simple rule-based heuristics through classical machine learning to contemporary deep learning and survival analysis approaches.

Mozer et al. (2000) conducted foundational work in telecommunications churn prediction, demonstrating that neural networks and decision trees applied to call detail records and billing history could identify at-risk customers with accuracy superior to manual heuristics. Their live deployment study reported a thirteen percent reduction in churn among model-targeted customers — an early demonstration of the business value of predictive analytics. Hadden et al. (2007) extended this work through a systematic comparison of classification algorithms on call-center data, finding that decision trees provided the best balance of predictive accuracy and practitioner interpretability.

Verbeke et al. (2012) conducted the largest systematic benchmarking study of churn prediction algorithms, comparing fourteen machine learning methods across ten telecommunications datasets. Their analysis found that ensemble methods — particularly Gradient Boosted Trees and Random Forests — consistently outperformed single-model approaches and identified class imbalance as the primary source of cross-study performance variability. This work established ensemble modeling as the standard approach for high-accuracy churn prediction.

Vafeiadis et al. (2015) identified temporal feature engineering as the most critical determinant of model performance, demonstrating that rolling window aggregations over 30-, 60-, and 90-day periods outperformed static snapshot features across six benchmark datasets. Amin et al. (2019) extended this finding by introducing delta features — period-over-period changes in key behavioral metrics — and showed that change-direction signals were more predictive of near-term churn than absolute metric levels.

Coussement and Van den Poel (2008) demonstrated the predictive value of multi-source signal integration, combining transactional data with natural language processing of customer service emails to improve churn prediction accuracy by twelve percent over usage-only models. Shaaban et al. (2021) applied survival analysis methods — specifically Cox proportional hazards models — to subscription data, demonstrating superior time-to-event calibration for planning time-sensitive interventions. Kim et al. (2022) explored LSTM-based sequential models applied to behavioral event logs, reporting strong results in high-volume consumer contexts while noting computational constraints for real-time production scoring.

Neslin et al. (2006) contributed a critical economic perspective by formalizing the profit-maximizing framework for retention campaigns, demonstrating that targeting on churn probability alone — without accounting for intervention cost and expected uplift — systematically over-spends on customers who would have retained without intervention. This insight motivates the risk-tiered intervention design proposed in this paper.

The present work synthesizes these contributions while explicitly addressing the identified gaps in cross-signal integration, operational deployment methodology, and live deployment revenue impact quantification.

3. SYSTEM ARCHITECTURE

The Customer Churn Prediction Framework is organized as five sequential, modular stages. Each stage is designed to be independently replaceable, allowing organizations to adapt individual components without rebuilding the entire pipeline.

3.1 Signal Collection Module

The signal collection module ingests behavioral data from four distinct source systems on a daily cadence. Product usage telemetry is extracted from server-side event logs, providing session-level records of user activity, feature interactions, and engagement depth. Billing and payment data is pulled from the subscription management platform, capturing payment success and failure events, plan tier history, upgrade and downgrade sequences, and discount application patterns. Customer support data is retrieved from the helpdesk system via API integration, providing ticket volume, resolution time, customer satisfaction scores, and escalation records. Sentiment and communication signals are sourced from the CRM and email marketing platform, including NPS survey responses, email campaign engagement metrics, and unsubscribe events.

3.2 Feature Engineering Module

The feature engineering module transforms raw behavioral signals into predictive features using three primary techniques. Rolling window aggregations compute statistics — mean, sum, standard deviation, minimum, and maximum — over 30-, 60-, and 90-day windows preceding each prediction date, capturing both short-term behavioral shifts and longer-term structural changes. Delta features measure the period-over-period change in each metric, encoding trend direction and magnitude. Recency features encode the elapsed time since specific events such as the last login, last support ticket, or last payment failure, providing a decay signal complementing frequency and magnitude measures. A total of 147 features are generated per customer-month observation. SHAP value analysis identifies the top predictors as: 30-day login frequency delta, days since last product use, support ticket volume in the trailing 30 days, NPS category, payment failure count in the trailing 60 days, and plan downgrade history.

3.3 Machine Learning Module

The machine learning module implements a two-level stacked ensemble architecture. Five candidate models are evaluated: Logistic Regression as an interpretable baseline, Random Forest, Gradient Boosted Trees via XGBoost, a shallow feedforward neural network, and the stacked ensemble. The stacked ensemble uses Logistic Regression, Random Forest, and Gradient Boosted Trees as first-level base learners. Their out-of-fold probability estimates, generated through five-fold cross-validation to prevent leakage, serve as inputs to a neural network meta-learner that captures second-order interactions among base learner outputs. All datasets are split temporally — 70% training, 15% validation, 15% held-out test — to prevent data leakage. Hyperparameters are optimized via Bayesian optimization using the Optuna framework, maximizing area under the precision-recall curve on the validation set.

3.4 Risk Segmentation Module

The risk segmentation module maps the calibrated churn probability score from the ensemble model to three intervention tiers. Tier thresholds are calibrated on the validation set to balance predictive precision and operational intervention cost rather than set arbitrarily. Low-risk accounts (score 0–40) are placed in passive monitoring with standard engagement communications. Medium-risk accounts (score 41–70) are enrolled in a structured proactive nurture sequence. High-risk accounts (score 71–100) trigger immediate escalation to the account management team.

3.5 Intervention and Feedback Module

The intervention module executes the tier-appropriate retention playbook and routes account-level risk information — including score, tier, top three SHAP risk drivers, and recommended action — to the relevant team through CRM integration. A closed-loop feedback mechanism records intervention outcomes (retained, churned, no contact) and feeds labeled results back into the training pipeline for monthly model retraining. A monitoring module computes the Population Stability Index on the feature distribution daily and triggers an alert when drift is detected.

4. MACHINE LEARNING FOR CHURN PREDICTION

Machine learning enables the analysis of large volumes of multi-dimensional behavioral data to identify patterns that predict customer cancellation. Unlike rule-based systems that rely on manually defined thresholds, supervised learning models discover complex, non-linear relationships between behavioral signals and churn outcomes from historical labeled data.

The framework evaluates five model architectures across two real-world datasets. Dataset A is a SaaS platform dataset comprising 87,432 subscribers observed over 24 months with a monthly churn rate of 6.8%. Dataset B is a public telecommunications benchmark comprising 115,219 subscribers observed over 18 months with a 3.2% monthly churn rate. Both datasets exhibit the class imbalance characteristic of stable subscription businesses; class-weight adjustment proportional to inverse class frequency is applied during training. Churn labels are assigned 45 days prior to the confirmed cancellation event to provide sufficient lead time for intervention.

The stacked ensemble achieves the highest AUC-ROC of 0.923 and AUC-PR of 0.881 on Dataset A, outperforming the best single model — Gradient Boosted Trees at 0.908 — by 1.5 percentage points. Cross-domain performance on Dataset B without dataset-specific retraining demonstrates the generalizability of the framework. The neural network meta-learner contributes the AUC gain over individual base learners by capturing second-order interactions among base learner probability estimates that no single model captures independently. Probability calibration using isotonic regression is applied to the ensemble output to ensure that the risk score functions as a reliable probabilistic estimate rather than a discriminative ranking alone.

5. ETHICAL CONSIDERATIONS AND CHALLENGES

Despite the considerable advantages of machine learning-based churn prediction, organizations deploying such systems bear responsibility for addressing associated ethical challenges and limitations.

Challenges of Churn Prediction Systems

Class Imbalance: Subscription churn rates typically range from 2% to 10% in stable businesses, creating substantial class imbalance that can cause models to underweight the minority churn class. Careful handling through class-weight adjustment, oversampling, or threshold calibration is required to achieve useful recall at acceptable precision levels.

Data Quality and Coverage: Predictive accuracy is directly dependent on the completeness and consistency of behavioral signal data. Missing values arising from optional survey instruments, intermittent API integrations, or newly onboarded customers can degrade feature quality and reduce model performance. A tiered imputation strategy — median imputation for features with fewer than 5% missing values, model-based imputation for 5–20% missingness, and feature exclusion above 20% — is applied within the framework.

Signal Bias: If certain customer segments historically received inferior service quality, their behavioral signals — higher support ticket volumes, lower satisfaction scores — may cause the model to disproportionately classify them as high-risk, directing retention resources toward segments that are at elevated risk precisely because of prior service failures rather than voluntary disengagement.

Model Interpretability: Complex ensemble architectures provide limited inherent interpretability. The framework addresses this through SHAP-based score explanations delivered to frontline staff, enabling account managers to contextualize model recommendations without requiring machine learning expertise.

Ethical Considerations

Fairness: Retention offers and personalized interventions triggered by risk scores must be audited across demographic and firmographic segments to ensure that the benefits of retention programs are equitably distributed. Organizations should publish internal fairness metrics disaggregated by plan tier, account age cohort, and industry vertical.

Transparency: Customers whose accounts are scored and targeted for retention intervention should have access, under applicable privacy regulations, to an explanation of how their behavioral data is used in automated decision-support systems.

Privacy and Data Security: Behavioral monitoring data used for churn prediction must be governed under applicable data protection regulations, including the General Data Protection Regulation (GDPR) in European jurisdictions and equivalent frameworks elsewhere. The framework's reliance on aggregated, anonymized behavioral summaries reduces but does not eliminate privacy exposure, and legal review is recommended before deployment in regulated markets.

Responsible Use of AI: Churn prediction models are decision-support tools, not autonomous decision-makers. Final retention intervention decisions should remain with human account managers who can apply contextual judgment that the model cannot capture, including awareness of market conditions, competitive dynamics, and individual customer circumstances.

CONCLUSION

The increasing competitiveness of subscription-based markets has elevated customer retention to a strategic priority for organizations across software, media, telecommunications, and e-commerce. The Customer Churn Prediction Framework presented in this paper provides a comprehensive, operationally validated approach to identifying at-risk subscribers and deploying targeted retention interventions before the cancellation event.

By integrating behavioral signals from product usage, billing, customer support, and sentiment domains; engineering 147 temporal features capturing trends, deltas, and recency; and applying a stacked ensemble model combining Gradient Boosted Trees, Logistic Regression, Random Forest, and a neural network meta-learner, the framework achieves an AUC-ROC of 0.923 on a SaaS subscriber dataset and 0.911 on a telecommunications benchmark — outperforming all evaluated single-model baselines.

The framework's live deployment study provides empirical evidence of its business value: an 18.4% reduction in monthly MRR churn, USD 2.31 million in estimated LTV recovery, and a 12.4:1 return on investment over a 90-day operational window. The modular architecture supports adoption by organizations at varying stages of machine learning maturity, from early-stage deployments using interpretable logistic regression models to sophisticated production systems with continuous retraining and distribution drift monitoring.

As machine learning capabilities continue to advance and behavioral data availability expands, churn prediction systems will become increasingly accurate, personalized, and economically impactful. The framework presented here establishes a methodological foundation for this evolution — bridging the gap between predictive modeling and operational revenue impact that remains underrepresented in the academic literature.

REFERENCE

- Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E. & Kaushansky, H. (2000). Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry; IEEE Transactions on Neural Networks, 11(3), 690–696.
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J. & Baesens, B. (2012). New Insights into Churn Prediction in the Telecommunication Sector: A Profit Driven Data Mining Approach; European Journal of Operational Research, 218(1), 211–229.
- Vafeiadis, T., Diamantaras, K. I., Sarigiannidis, G. & Chatzisavvas, K. C. (2015). A Comparison of Machine Learning Techniques for Customer Churn Prediction; Simulation Modelling Practice and Theory, 55, 1–9.
- Amin, A., Anwar, S., Adnan, A., Nawaz, M., Alawfi, K., Hussain, A. & Huang, K. (2019). Customer Churn Prediction in the Telecommunication Sector Using a Rough Set Approach; Neurocomputing, 237, 242–254.
- Coussement, K. & Van den Poel, D. (2008). Churn Prediction in Subscription Services: An Application of Support Vector Machines; Expert Systems with Applications, 34(1), 313–327.
- Shaaban, E., Helmy, Y., Khedr, A. & Nasr, M. (2021). A Survival Analysis Approach to Customer Churn Prediction in Subscription Services; Journal of King Saud University – Computer and Information Sciences, 33(6), 707–715.
- Kim, H., Lim, S. & Park, J. (2022). Sequential Churn Prediction with Recurrent Neural Networks; IEEE Transactions on Neural Networks and Learning Systems, 33(8), 3412–3426.
- Neslin, S. A., Gupta, S., Kamakura, W., Lu, J. & Mason, C. H. (2006). Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models; Journal of Marketing Research, 43(2), 204–211.
- Hadden, J., Tiwari, A., Roy, R. & Ruta, D. (2007). Computer Assisted Customer Churn Management: State-of-the-Art and Future Trends; Computers & Operations Research, 34(10), 2902–2917.
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow; O'Reilly Media.

ACKNOWLEDGEMENT

The successful completion of the Customer Churn Prediction Framework research would not have been possible without the guidance, support, and encouragement of many individuals and institutions.

We express our sincere gratitude to our professors and academic supervisors at Sri Manakula Vinayagar Engineering College for their continuous guidance, insightful feedback, and technical mentorship throughout the development of this research. Their expertise in machine learning and data analytics substantially improved the quality and rigor of this work.

We are grateful to the faculty and administration of the Department of Master of Computer Applications, Sri Manakula Vinayagar Engineering College, Pondicherry, for providing access to the computational resources, research infrastructure, and academic environment necessary to undertake this study.

We acknowledge the contributions of the broader research community in machine learning, customer analytics, and subscription economics whose published work forms the intellectual foundation of this framework.