

Machine Learning Approaches for Understanding Customer Churn in Subscription-Based Digital Services

Dr. Carlo Mendoza

Assistant Professor, Institute of Subscription Analytics, Philippines

Abstract

Subscription-based digital services have become an important component of contemporary digital business models, providing continuous access to software applications, digital media, educational platforms, and other online resources. The financial sustainability of these services depends partly on maintaining relationships with existing subscribers and understanding the circumstances associated with subscription termination. Customer churn represents a significant analytical challenge because cancellation decisions may reflect changing usage behavior, perceived service value, payment difficulties, subscription costs, and dissatisfaction with service experiences. Traditional descriptive approaches provide useful information about customer attrition but may offer limited capacity to identify complex relationships among behavioral and transactional indicators. The present paper develops a hypothetical methodological framework for investigating machine learning approaches to customer churn prediction in subscription-based digital services. Cox Proportional Hazards Modeling, Logistic Regression, Random Forest, and Gradient Boosting. A fictional dataset containing 24,000 subscriber observations is introduced to demonstrate the intended research design. Customer engagement, subscription tenure, billing history, service interactions, and plan characteristics are considered potential explanatory variables.

The proposed methodology incorporates subscriber-level data preparation, behavioral feature engineering, chronological validation, classification modeling, and independent prediction assessment. Customer churn is operationalized as confirmed subscription termination or nonrenewal within a 30-day prediction period. Two illustrative analytical tables describe the hypothetical subscriber distribution and comparative classification outcomes. Precision, recall, and F1-score are selected as principal evaluation measures. All numerical values are manually constructed and do not represent collected customer information or completed machine learning experiments. The methodological discussion examines behavioral disengagement, billing-related attrition, class imbalance, temporal information leakage, probability calibration, and responsible interpretation of predictive factors. Particular attention is given to distinguishing churn prediction from causal explanations of customer behavior. The paper concludes that machine learning provides a relevant analytical foundation for understanding subscription attrition, although reliable conclusions require representative records, independent evaluation, appropriate privacy safeguards, and evidence-based validation.

Keywords: Customer Churn Prediction, Machine Learning, Subscription-Based Services, Digital Customer Behavior, Customer Retention, Logistic Regression, Random Forest, Gradient Boosting, Survival Analysis, Predictive Analytics.

1. Introduction

Subscription-based digital services have transformed the organization of customer relationships by providing continuing access to products and applications through recurring payments or periodic membership arrangements. Digital entertainment platforms, software services, electronic learning environments, and online professional tools frequently employ subscription models to maintain ongoing customer access. Unlike isolated purchases, subscription arrangements generate continuing interactions between service providers and users. These interactions create opportunities to observe engagement behavior, billing activity, service experiences, and changes in customer participation. Understanding the circumstances associated with subscription continuation and termination has consequently become an important area of digital business analytics and predictive information systems.

Customer churn generally refers to the termination or discontinuation of an existing customer relationship within a defined observation period. In subscription-based services, churn may occur through voluntary cancellation, expiration without renewal, or termination following unsuccessful payment recovery procedures. These circumstances may reflect different underlying mechanisms and should therefore be distinguished during analytical assessment. Temporary reductions in platform activity do not necessarily indicate subscription termination, while an unsuccessful payment attempt may not produce churn if the customer subsequently completes payment. Consequently, reliable prediction requires a clearly specified definition of the outcome and careful examination of the operational processes through which subscriber status changes.

Digital subscription platforms generate various forms of information that can contribute to churn analysis. These include subscription tenure, login frequency, session duration, feature utilization, payment history, account-plan characteristics, and customer support interactions. Changes in these variables may provide useful indicators of future subscription decisions. For example, declining service engagement can occur before cancellation, while repeated billing difficulties may increase the likelihood of interruption. However, these relationships are not necessarily uniform across subscription categories. Some subscribers maintain active memberships despite infrequent use, whereas highly engaged subscribers may discontinue a service because of changing requirements, financial considerations, or alternative offerings.

Machine learning offers computational methods for examining complex relationships among customer characteristics and subsequent subscription outcomes. Logistic Regression provides an interpretable statistical baseline for binary classification, while Random Forest can model nonlinear relationships and interactions among explanatory variables. Gradient Boosting develops predictive models through sequential error reduction and can accommodate complicated combinations of behavioral features. Survival analysis offers a complementary perspective by examining the time until a subscription event occurs, while accounting for customers whose subscriptions remain active during observation. Nevertheless, methodological complexity does not automatically guarantee improved predictive reliability, and models must be evaluated under conditions reflecting genuine subscription behavior.

The present paper proposes a hypothetical comparative investigation of customer churn in subscription-based digital services. A fictional dataset containing 24,000 individual subscribers is considered, with each observation representing an active subscription at a defined prediction date. The proposed framework examines behavioral, transactional, and service-related characteristics associated with churn during the subsequent 30 days. Four predictive approaches are selected for comparison, including a survival-analysis configuration converted into a fixed-horizon churn probability. Two illustrative tables demonstrate subscriber distribution and comparative model performance. No actual subscription records were accessed or analyzed, and no predictive models were executed. The manuscript develops a reproducible methodological foundation for future empirical research.

2. Objectives of the Study

The primary objective of this study is to examine how machine learning approaches can be applied to understanding and predicting customer churn in subscription-based digital services. The investigation focuses on behavioral engagement, transactional characteristics, comparative predictive performance, and evaluation reliability. Three specific objectives establish the methodological direction of the proposed research framework.

2.1 Examination of Customer Engagement and Subscription Behavior

The first objective is to examine how digital engagement characteristics may provide information about subsequent subscription termination. Login frequency, active usage days, session duration, feature utilization, and changes in engagement may reflect different forms of customer participation. The proposed study considers how these variables could be represented through historical behavioral summaries and compared with future churn outcomes. Particular attention is given to detecting changes in engagement over time rather than assuming that customers with lower absolute usage are necessarily at greater risk of cancellation.

The investigation further aims to examine how engagement characteristics interact with subscription tenure and service-plan arrangements. New subscribers may exhibit usage behavior that differs from customers with established accounts, while annual and monthly subscription arrangements can involve different renewal opportunities. Consequently, the same activity pattern may carry different predictive information depending on account context. The proposed framework therefore emphasizes the interpretation of behavioral indicators alongside subscription characteristics and rejects the assumption that platform activity alone provides a complete explanation of customer retention.

2.2 Comparative Evaluation of Machine Learning Prediction Models

The second objective focuses on establishing a comparative framework for evaluating alternative churn prediction approaches. Cox Proportional Hazards Modeling, Logistic Regression, Random Forest, and Gradient Boosting are selected to represent different statistical and computational learning strategies. Their intended comparison examines the ability to identify customers whose subscriptions terminate during a common 30-day prediction period. Precision, recall, and F1-score provide principal measures of classification effectiveness, while probability calibration and ranking performance are considered supplementary evaluation dimensions.

A further objective involves examining whether flexible machine learning models offer useful advantages over conventional statistical approaches. Ensemble methods may identify nonlinear relationships among engagement, billing, and account characteristics, while Logistic Regression provides a relatively transparent representation of predictive associations. Survival modeling incorporates information about time to termination and right-censored observations. However, differences in data preparation, feature availability, and prediction thresholds can affect comparative outcomes. Consequently, the proposed framework emphasizes consistent forecasting horizons, independent testing, and transparent model configuration.

2.3 Assessment of Churn Prediction Reliability and Customer Management Constraints

The third objective is to identify methodological factors that may influence the reliability of subscription churn prediction. These include inconsistent termination labels, incomplete billing records, changes in customer engagement, unequal churn prevalence, and temporal information leakage. The proposed investigation considers how outcome definition, historical feature construction, and chronological evaluation could reduce selected risks. Particular attention is given to distinguishing information available before a prediction date from events occurring during the subsequent outcome period.

The investigation also considers the relationship between predictive accuracy and responsible customer management. A churn score may indicate elevated statistical risk without establishing why a particular subscriber intends to leave. Similarly, successful identification of likely churners does not demonstrate that a retention intervention would be effective or economically beneficial. Consequently, the proposed framework emphasizes transparent probability interpretation, privacy-aware data processing, and independent evaluation of intervention outcomes. These considerations provide a foundation for understanding subscription churn without equating statistical prediction with causal explanation.

3. Review of Literature

3.1 Customer Retention and Subscription Attrition Research

Van den Poel and Larivière (2004) investigated customer attrition using proportional hazard models and examined the relevance of longitudinal customer information for understanding relationship termination. Gupta et al. (2004) examined customer valuation and demonstrated the importance of retention within broader customer value assessment. Fader et al. (2005) developed an approach linking behavioral customer analysis with customer lifetime value. These contributions established important foundations for interpreting attrition in relation to customer relationships, behavioral information, and economic considerations rather than treating churn as an isolated classification event.

Neslin et al. (2006) examined methodological differences in customer churn prediction and demonstrated the importance of comparative model evaluation. Ahn et al. (2006) investigated customer churn determinants and partial defection within telecommunications services. Burez and Van den Poel (2007) examined analytical customer relationship management in a subscription-based pay-television environment. Together, these studies demonstrated that customer attrition is associated with different forms of behavioral and service-related information. Their findings also emphasized that predictive usefulness and retention effectiveness depend on the characteristics of the service environment.

3.2 Machine Learning and Customer Churn Classification

Lemmens and Croux (2006) investigated bagging and boosting classification trees for predicting customer churn and examined the importance of model selection and sampling procedures. Coussemont and Van den Poel (2008) evaluated support vector machines in a subscription-service setting and compared their predictive performance with Logistic Regression and Random Forest. Hadden et al. (2007) reviewed computational approaches to customer churn management. These contributions established important methodological foundations for applying machine learning to customer attrition and highlighted the relevance of model configuration and predictive evaluation.

Burez and Van den Poel (2009) investigated class imbalance in customer churn prediction and examined resampling and alternative learning strategies. Tsai and Lu (2009) developed hybrid neural network approaches for churn prediction. Verbeke et al. (2011) examined comprehensible classification models through advanced rule induction techniques. These studies demonstrate that model development must consider predictive performance, minority-class recognition, and the interpretation of customer-related features. Their contributions support comparing alternative algorithms through appropriate evaluation procedures rather than relying exclusively on aggregate classification accuracy.

3.3 Profit-Oriented Evaluation and Predictive Reliability

Glady et al. (2009) examined churn modeling through customer lifetime value and demonstrated the relevance of economic consequences when assessing customer attrition. Verbeke et al. (2012) developed a profit-driven approach to churn prediction in telecommunications, emphasizing the connection between predictive ranking and retention campaign decisions. Verbraken et al. (2013) introduced a profit-oriented classification performance measure for churn prediction models. These studies established that conventional prediction metrics do not automatically describe the financial consequences of customer management decisions.

Breiman (2001) established Random Forest as an ensemble learning method, while Friedman (2001) developed Gradient Boosting as a sequential predictive optimization approach. Niculescu-Mizil and Caruana (2005) investigated probability calibration, and He and Garcia (2009) examined learning under class imbalance. Cawley and Talbot (2010) reviewed model-selection bias, while Saito and Rehmsmeier (2015) emphasized precision–recall evaluation for unequal class distributions. These contributions establish methodological principles relevant to comparing churn models and assessing the reliability of their predictions.

4. Research Methodology

4.1 Research Design and Customer Churn Classification Framework

The present paper adopts a hypothetical comparative quantitative research design for investigating subscription churn prediction. Four configurations are considered: Cox Proportional Hazards Modeling, Logistic Regression, Random Forest, and Gradient Boosting. The intended outcome is the termination of an active subscription within 30 days after a designated prediction date. Termination includes confirmed cancellation resulting in loss of access and nonrenewal after any applicable payment recovery period. Temporary inactivity and recoverable payment failures are not independently classified as completed churn events.

The proposed analytical procedure consists of subscriber eligibility assessment, historical feature preparation, outcome labeling, chronological data separation, model development, and independent prediction evaluation. Every subscriber would contribute one eligible prediction snapshot to the analytical dataset. Features would be calculated using information available before the prediction date, while churn labels would be determined from subsequent subscription outcomes. Model parameters and classification thresholds would be selected using development information. The final test population would remain unavailable during optimization to support meaningful generalization assessment.

4.2 Hypothetical Subscriber Dataset and Behavioral Feature Engineering

For illustrative purposes, the study assumes a fictional dataset containing 24,000 unique subscribers to a digital subscription platform. Each subscriber is represented by one observation recorded while the corresponding subscription is active. The dataset contains 18,000 subscribers who remain active throughout the following 30 days and 6,000 whose subscriptions terminate during that period. This produces a hypothetical churn prevalence of 25%. These quantities are manually constructed and do not represent genuine customer records or observed platform statistics.

The proposed features would include subscription tenure, service-plan category, billed subscription amount, login frequency, active usage days, feature utilization, support interactions, and changes in engagement during a preceding 90-day observation window. Billing-related indicators would distinguish failed payment attempts that occurred before the prediction date from subsequent failures that form part of the outcome period.

4.3 Predictive Modeling and Independent Performance Evaluation

Logistic Regression would provide a transparent binary classification baseline, while Random Forest and Gradient Boosting would examine nonlinear associations and feature interactions. Cox Proportional Hazards Modeling would estimate subscription termination risk over time, with its survival estimates converted into 30-day churn probabilities for comparison. Event times would be measured from the prediction date, and subscribers who remained active throughout the complete 30-day period would be right-censored at day 30 for survival analysis. All four configurations would produce predictions for the same independent subscribers.

The principal evaluation would report minority-class precision, recall, and F1-score. Precision would measure the proportion of predicted churners who actually terminated their subscriptions, while recall would describe the proportion of churn events identified. F1-score would summarize the relationship between these measures. A completed investigation should also assess precision–recall area, ROC area, probability calibration, and expected retention value under explicitly defined intervention costs. Uncertainty intervals would be required for meaningful model comparison. No prediction algorithms were implemented for the present illustration.

5. Analysis

5.1 Illustrative Distribution of Subscription Continuation and Termination

Subscriber outcome distribution is an important consideration in churn prediction because termination events may occur less frequently than subscription continuation. Table 1 presents the hypothetical allocation of active and churned subscribers across the proposed evaluation partitions. The fictional

dataset contains a 75% continuation proportion and a 25% churn proportion. These values are selected exclusively to illustrate model evaluation and do not describe customer retention within any actual digital subscription platform.

The illustrative distribution also demonstrates why classification accuracy alone can provide incomplete information about churn detection. A model predicting that every subscriber will remain active would achieve 75% accuracy within this fictional dataset while identifying none of the customers who actually terminate their subscriptions. Consequently, minority-sensitive measures are necessary for meaningful assessment. The proposed framework preserves the original class proportions across independent evaluation partitions. This structure supports transparent comparison without artificially changing the distribution of churn outcomes during testing.

Table 1: Hypothetical Subscriber Outcome Distribution Across Evaluation Partitions

Subscription Outcome	Training Subscribers	Validation Subscribers	Test Subscribers	Total
Subscription Continued	10,800	3,600	3,600	18,000
Subscription Churned	3,600	1,200	1,200	6,000
Total	14,400	4,800	4,800	24,000

The constructed test population contains 1,200 churn events and 3,600 subscription continuations. These counts establish the mathematical basis for comparing the classification outcomes of the selected predictive models.

5.2 Illustrative Comparison of Churn Prediction Models

Churn prediction models can be assessed by comparing their predicted cancellation risks with independently observed subscription outcomes. Precision describes the reliability of positive churn predictions, while recall measures the proportion of actual churn events identified. F1-score provides a combined measure of these two dimensions. Table 2 presents hypothetical prediction counts for four alternative models evaluated on the fictional independent test set. The values were manually constructed and do not represent actual machine learning predictions.

The illustrative comparison assigns the highest F1-score to Gradient Boosting, followed by Random Forest, Logistic Regression, and Cox Proportional Hazards Modeling. This pattern demonstrates how different approaches could be compared under a common 30-day outcome definition. However, the numerical rankings cannot establish actual superiority because no computational experiments were conducted. A genuine investigation would require trained models, measured test outcomes, confidence intervals, and comparison of expected retention value under realistic intervention assumptions.

Table 2: Hypothetical Classification Performance for 30-Day Subscription Churn Prediction

Predictive Model	True Positives	False Positives	Precision	Recall	F1-Score
Cox Proportional Hazards Model	660	330	0.667	0.550	0.603
Logistic Regression	708	306	0.698	0.590	0.640
Random Forest	780	360	0.684	0.650	0.667
Gradient Boosting	816	312	0.723	0.680	0.701

The illustrative comparison demonstrates potential differences in identifying customers likely to terminate their subscriptions. However, the preferred model would depend on calibrated prediction probabilities, intervention costs, available retention resources, and the consequences of incorrect classifications.

6. Discussion

6.1 Customer Engagement and Subscription Termination Patterns

Customer engagement characteristics can provide information about the relationship between users and subscription-based digital services. Login frequency, active usage days, and feature utilization describe different dimensions of platform participation. Changes in these indicators may reveal patterns associated with later subscription termination. However, low engagement does not necessarily indicate dissatisfaction because customers may maintain subscriptions for occasional or specialized use. Consequently, churn models should interpret behavioral characteristics within the context of subscription arrangements and customer requirements.

The proposed framework emphasizes the importance of examining changes in engagement rather than relying exclusively on absolute activity levels. A decline in usage may provide different predictive information for an established subscriber than for a recently registered customer. Similarly, billing difficulties and support interactions may be associated with termination through mechanisms that differ from reduced engagement. The hypothetical analysis does not identify actual customer behavior patterns because no platform records were examined. Its purpose is to establish relevant variables for future investigation.

6.2 Statistical Learning and Ensemble Classification Performance

Statistical and machine learning approaches differ in how they represent relationships among customer characteristics and churn outcomes. Logistic Regression estimates associations through a specified functional structure, while Random Forest captures nonlinear interactions through ensembles of decision trees. Gradient Boosting sequentially develops predictive models intended to reduce a defined loss function. Survival analysis provides an additional perspective by examining the timing of subscription termination. These methods therefore offer complementary analytical capabilities rather than interchangeable approaches with identical assumptions.

The hypothetical comparison assigns higher F1-scores to ensemble classifiers than to the selected statistical baselines. However, these values cannot establish empirical improvement or identify the characteristics responsible for any performance difference. A genuine investigation would need controlled feature comparisons, repeated validation, and assessment of model stability across different subscription periods. The computational requirements and interpretability of each approach would also influence selection. Predictive superiority should be established through independent evidence rather than assumed from the flexibility of a particular algorithm.

6.3 Retention Decisions and Probability-Based Customer Assessment

Churn prediction can support customer relationship management by identifying subscribers whose future continuation is uncertain. However, a high predicted churn probability does not establish that the subscriber would respond positively to a retention offer. Some customers may terminate their subscriptions regardless of intervention, while others may remain subscribed without receiving additional incentives. Consequently, predictive risk ranking and intervention effectiveness are distinct analytical problems. A retention strategy should not assume that all likely churners represent equally valuable or responsive intervention targets.

Research by Verbeke et al. (2012) and Verbraken et al. (2013) emphasizes the importance of connecting churn prediction with economic objectives. Within subscription-based digital services, retention decisions may depend on customer value, contact costs, incentive expenses, and expected incremental retention. However, predictive models alone cannot determine the causal effectiveness of retention measures. Controlled intervention studies would be necessary to establish whether specific actions actually improve subscriber continuation.

7. Challenges

7.1 Churn Label Ambiguity and Subscription Lifecycle Complexity

Customer churn is not always represented by a single clearly observable event within subscription-based digital services. A subscriber may cancel automatic renewal while retaining access until the end of a paid period. Another customer may experience a temporary payment failure before successfully completing renewal. Some services also provide grace periods, paused subscriptions, or alternative billing arrangements. Consequently, inconsistent treatment of these events can create inaccurate outcome labels and weaken the reliability of classification models.

Subscription lifecycle differences introduce additional challenges when customers use monthly, annual, or promotional plans. The opportunity for renewal-related churn may depend on the timing of the next billing event, making fixed-horizon prediction less informative for some accounts. The proposed framework therefore emphasizes a clearly defined termination event and consistent eligibility rules. A genuine investigation should document how cancellations, expired memberships, billing recovery periods, and reactivations are represented. These decisions are necessary for producing interpretable and comparable churn predictions.

7.2 Behavioral Data Leakage and Changing Customer Patterns

Temporal information leakage represents a major methodological risk when developing customer churn prediction models. Features calculated using information recorded after the prediction date can reveal

events associated with subsequent termination and artificially improve performance. For example, account status changes or final cancellation interactions should not be included as predictors when they occur during the outcome period. Consequently, each variable must be associated with a documented timestamp and evaluated according to its availability when the prediction would actually be generated.

Customer behavior may also change as digital services introduce new features, modify pricing arrangements, or alter subscription terms. Predictive relationships learned from earlier subscribers may become less reliable when applied to later customer populations. Such changes can affect engagement patterns, cancellation behavior, and billing outcomes. The proposed framework therefore emphasizes chronological testing and continued monitoring of model performance. Future empirical research should evaluate whether updated models improve reliability without introducing instability through excessive retraining.

7.3 Class Imbalance, Interpretability, and Customer Data Privacy

Customer churn datasets may contain substantially fewer termination events than subscription continuations. This imbalance can create misleading aggregate accuracy and complicate the detection of subscribers who are genuinely at risk of leaving. Resampling or class weighting may improve selected evaluation measures but can influence probability calibration and false-positive predictions. Consequently, model performance should be assessed through precision, recall, F1-score, and appropriate probability-based metrics. The preferred decision threshold should reflect the intended customer management application.

Customer data privacy introduces further challenges because subscription platforms may collect detailed information about usage behavior, payment history, and service interactions. Predictive modeling should therefore apply data minimization, access controls, and appropriate aggregation wherever practical. Sensitive personal characteristics should not be included without a legitimate and clearly justified analytical purpose. Interpretation also requires caution because statistical associations do not establish the motivations or intentions of individual customers. The proposed framework emphasizes responsible feature selection and transparent evaluation rather than intrusive profiling or unsupported behavioral judgments.

8. Conclusion

Machine learning provides a methodological foundation for examining customer churn within subscription-based digital services. The present paper developed a hypothetical comparative framework incorporating Cox Proportional Hazards Modeling, Logistic Regression, Random Forest, and Gradient Boosting. Its principal focus was the relationship among subscription characteristics, behavioral engagement, billing information, and subsequent customer termination. The proposed research design demonstrates how alternative prediction approaches could be evaluated through consistent outcome definitions, historical feature construction, and independent performance assessment.

The illustrative tables demonstrate possible procedures for reporting subscription outcomes and comparing churn classification performance. However, all numerical values were manually constructed and do not represent actual customer records or completed machine learning experiments. Establishing meaningful predictive findings would require authentic subscription datasets, appropriately implemented models, independent testing, and uncertainty assessment. Consequently, the hypothetical results cannot

establish that any particular algorithm provides superior churn prediction or that specific customer characteristics cause subscription termination.

Overall, the study identifies churn definition, behavioral variability, billing complexity, temporal information leakage, class imbalance, probability calibration, and data privacy as central considerations in subscription analytics. Future empirical research should examine these factors through representative datasets and reproducible evaluation procedures. Machine learning may support customer retention planning when predictions are interpreted alongside operational constraints and independently assessed intervention outcomes. Nevertheless, reliable conclusions must be based on verified prediction performance and defensible evidence rather than illustrative comparisons or assumptions about customer behavior.

References

1. Ahn, J. H., Han, S. P., & Lee, Y. S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy*, 30(10–11), 552–568. <https://doi.org/10.1016/j.telpol.2006.09.006>
2. Burez, J., & Van den Poel, D. (2009). Handling class imbalance in customer churn prediction. *Expert Systems with Applications*, 36(3), 4626–4636. <https://doi.org/10.1016/j.eswa.2008.05.027>
3. Coussement, K., & Van den Poel, D. (2008). Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1), 313–327. <https://doi.org/10.1016/j.eswa.2006.09.038>
4. Coussement, K., & Van den Poel, D. (2009). Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. *Expert Systems with Applications*, 36(3), 6127–6134.
5. Glady, N., Baesens, B., & Croux, C. (2009). Modeling churn using customer lifetime value. *European Journal of Operational Research*, 197(1), 402–411.
6. Hung, S. Y., Yen, D. C., & Wang, H. Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3), 515–524.
7. Idris, A., Rizwan, M., & Khan, A. (2012). Churn prediction in telecom using Random Forest and PSO based data balancing in combination with various feature selection strategies. *Computers & Electrical Engineering*, 38(6), 1808–1819.
8. Keramati, A., & Ardabili, S. M. (2011). Churn analysis for an Iranian mobile operator. *Telecommunications Policy*, 35(4), 344–356.
9. Khattak, A., Mehak, Z., Ahmad, H., Asghar, M. U., Asghar, M. Z., & Khan, A. (2023). Customer churn prediction using composite deep learning technique. *Scientific Reports*, 13, Article 17294. <https://doi.org/10.1038/s41598-023-44396-w>
10. Lemmens, A., & Croux, C. (2006). Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(2), 276–286.
11. Manzoor, A., Qureshi, M. A., Kidney, E., & Longo, L. (2024). A review on machine learning methods for customer churn prediction and recommendations for business practitioners. *IEEE Access*, 12, 70434–70463. <https://doi.org/10.1109/ACCESS.2024.3402092>

12. 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. <https://doi.org/10.1509/jmkr.43.2.204>
13. Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592–2602. <https://doi.org/10.1016/j.eswa.2008.02.021>
14. Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*, 38(12), 15273–15285.
15. Rathi, S., Puranik, A., Pophale, V., Kutwal, P., Kulkarni, V., Pratham, S., & Maral, V. (2023). A survey and implementation of machine learning algorithms for customer churn prediction. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(10). <https://doi.org/10.17762/ijritcc.v11i10.8625>
16. 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.
17. Wagh, S. K., Andhale, A. A., Wagh, K. S., Pansare, J. R., Ambadekar, S. P., & Gawande, S. H. (2024). Customer churn prediction in telecom sector using machine learning techniques. *Results in Control and Optimization*, 14, Article 100342. <https://doi.org/10.1016/j.rico.2023.100342>
18. Wei, C. P., & Chiu, I. T. (2002). Turning telecommunications call details to churn prediction: A data mining approach. *Expert Systems with Applications*, 23(2), 103–112.
19. Xie, Y., Li, X., Ngai, E. W. T., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36(3), 5445–5449.
20. Amin, A., Anwar, S., Adnan, A., Nawaz, M., Howard, N., Qadir, J., Hawalah, A., & Hussain, A. (2017). Comparing oversampling techniques to handle the class imbalance problem: A customer churn prediction case study. *IEEE Access*, 4, 7940–7957.