WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

Customer Churn Prediction Using Machine Learning: A Review of Techniques, Imbalance Handling, Explainability, and Personalized Retention

Mrs. Gorisetty. Nirosha[1], Saripalli Jabili[2]

[1] Assistant Professor & Term Paper Supervisor | [2] Student (JNTU No.: 24341A05J8), B. Tech (CSE) 5th Semester
Department of Computer Science and Engineering, GMRIT Deemed to be University, Rajam, Andhra Pradesh, India

ABSTRACT—

Customer churn prediction applies machine learning (ML) to identify customers who are likely to stop using a product or service, so that a business can act before they leave. This paper reviews fifteen studies published between 2013 and 2026 across banking, telecommunications, e-commerce, and industrial settings. It covers classical classifiers (Logistic Regression, Decision Tree, Random Forest, SVM), boosting and ensemble methods (XGBoost, weighted ensembles), deep models (Bi-LSTM, CNN/RNN, Transformer), social-network features, hard and soft data fusion, class-imbalance handling (SMOTE and density-regularized oversampling), explainable AI, and profit-based evaluation. The synthesis shows that preprocessing, imbalance handling, and the choice of evaluation metric influence outcomes as much as the choice of algorithm, and that most studies stop at predicting who will churn rather than explaining why and recommending what to do. To address this gap, the paper proposes a workflow that combines Logistic Regression, Random Forest, and XGBoost with SMOTE applied to the training data only, SHAP-based factor analysis, and rule-based personalized retention recommendations. The framework is a design proposal and is not experimentally evaluated in this paper; the paper closes with open challenges and future directions.

Keywords—*Customer Churn Prediction, Machine Learning, Customer Retention, Predictive Analytics, Classification Algorithms, Class Imbalance, Explainable AI, Personalized Retention Strategies.*

1. Introduction

1.1 Background and Context

Customer churn is the process by which customers stop using a company's products or services and move to a competitor. In competitive markets, retaining an existing customer is usually cheaper than acquiring a new one, so even a small improvement in retention has a direct effect on revenue [4], [13]. Churn prediction supports retention by ranking customers according to their likelihood of leaving, which allows a company to direct its retention budget to the customers where it matters most.

ML has become the standard tool for this task. Models learn patterns from demographic, service-related, and behavioural records and output a churn probability for each customer. Early work relied on Logistic Regression, Decision Trees, and SVMs [1], [7]. Later studies adopted Random Forest and XGBoost [3], [5], [10], deep sequence models [11], [15], and Transformer architectures [9]. In parallel, researchers have addressed three practical problems: churners are a small minority of customers (class imbalance) [1], [6], [8]; accuracy alone does not reflect business value [13]; and black-box predictions are hard for managers to act on [4], [8].

1.2 Churn Research Across Sectors

Churn research is spread across sectors that differ in the data they hold. Banks have transaction histories and sometimes customer feedback [1], [7]. Telecom operators have usage, billing, and call-network data [2], [3], [5], [6], [10]. E-commerce platforms have purchase and click sequences [11], [12], and industrial suppliers have a small number of high-value accounts [8]. A consolidated view therefore helps a practitioner choose methods that fit the data actually available.

1.3 Objectives and Contributions

This paper pursues four objectives:

- **Concept and drivers.** Explain what churn prediction is, why it matters, and which factors (service quality, satisfaction, pricing, usage patterns) are associated with churn.
- **Algorithm comparison.** Compare the ML techniques used in the reviewed studies, together with their data requirements, strengths, and limitations.
- **Imbalance, evaluation, and explainability.** Review how studies handle class imbalance, which metrics they report, and how they explain predictions.
- **Reasons and retention.** Propose a workflow that identifies probable churn reasons and links them to personalized retention actions.

The contributions are (i) a taxonomy that organizes the reviewed techniques into six families, (ii) a cross-study comparison of sector, approach, evaluation, and stated gaps, (iii) the identification of the separation between prediction and action as a gap in the reviewed work, and (iv) a design proposal that connects churn prediction, explanation, and retention action. Findings attributed to the reviewed studies and the proposed framework are kept in separate sections (Sections 3 and 5) so that they are not confused.

2. Review Methodology

This is a structured narrative review, not a full PRISMA systematic review. It covers fifteen studies selected for relevance to ML-based churn prediction: twelve IEEE journal articles [1], [2], [4]–[13], one *Procedia Computer Science* paper [3], one IEEE conference paper [14], and one article in *IEEE Communications Magazine* [15]. Publication years run from 2013 to 2026.

Inclusion criteria. A study was included if it (a) applies or evaluates ML or data-mining methods for predicting customer churn or churn-related behaviour, or (b) reviews such methods or proposes a way to evaluate them. Studies spanning banking, telecom, e-commerce, and industrial customers were retained to allow cross-sector comparison.

Data extracted. For each study, the sector, objective, data type, algorithms, imbalance treatment, evaluation measures, stated limitations, and research gaps were recorded and organized into the themes of Sections 3 to 6.

Scope note. The synthesis relies on each study's stated objectives and reported methodology. Numerical results are not compared across studies because datasets and protocols differ (see Section 7).

3. Related Work

3.1 Taxonomy of Techniques

The reviewed studies fall into six technique families (Table 1 and Fig. 1). Each family uses different input data and addresses a different part of the churn problem. Fig. 2 shows the customer data that these techniques convert into a churn-risk estimate.

Table 1: Technique families in the reviewed studies

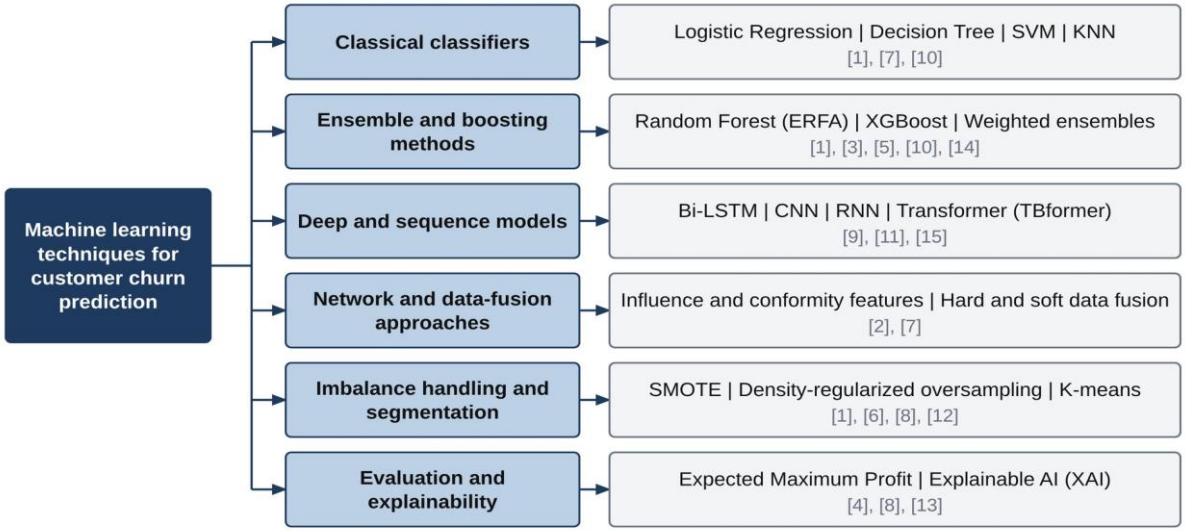


Fig. 1: Taxonomy of machine learning techniques used for customer churn prediction (authors' synthesis of the reviewed studies).

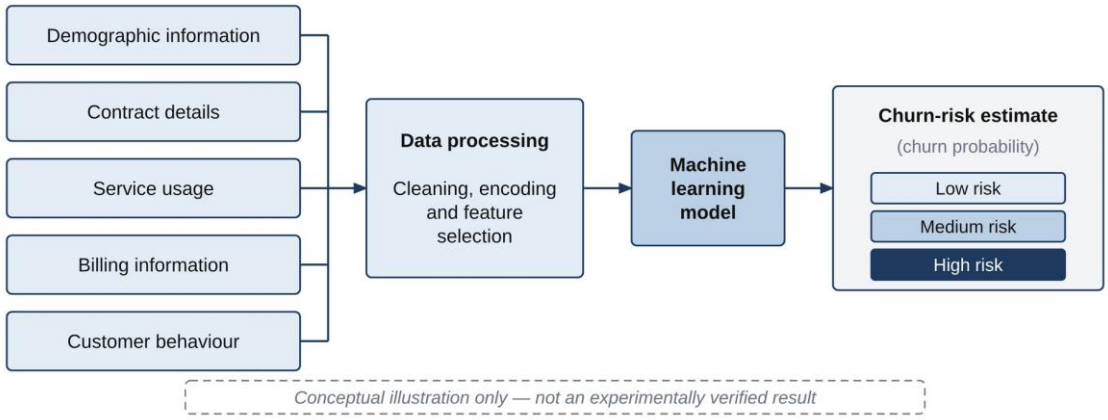


Fig. 2: Conceptual relationship between customer data and churn-risk prediction (conceptual illustration; not an experimental result).

3.2 Tree Ensembles and Boosting

Tree-based ensembles are the most common choice for tabular churn data. Ullah et al. apply Random Forest to telecom data and combine it with classification and clustering to identify the factors behind churn [5]. Al-Sultan and Al-Baltah propose an Improved Random Forest Algorithm (ERFA) for banking, tune its hyperparameters, and compare it with Logistic Regression, SVM, and a Neural Network on both imbalanced and balanced data [1]. Shrestha and Shakya use XGBoost on public and real Nepalese telecom data, with data cleaning, normalization, hyperparameter tuning, and 10-fold stratified cross-validation [3]. Afzal et al. compare Decision Tree, Random Forest, KNN, and XGBoost on telecom data and examine factors such as competitors' offers and the type of internet service [10]. Adiputra and Wanchai average the outputs of XGBoost and Random Forest and test the ensemble on telecom and insurance datasets, reporting F1-score and processing time [14]. Taken together, these studies treat boosting and bagging as strong default models, with ensemble gains depending on tuning and data preparation.

3.3 Social and Network Information

Customers do not decide in isolation. Amiri and Hosseini use phone-call networks to build *influence* and *conformity* features that describe how strongly a customer is affected by, and behaves like, their contacts, and they add these features to demographic and service features [2]. Their evaluation uses AUC, accuracy, precision, F1-score, MCC, Cohen's kappa, and Brier score. The approach requires relational data that many firms do not hold, which limits where it can be applied.

3.4 Deep and Sequential Models

When behaviour is recorded over time, sequence models can use it directly. Wei applies a Bi-LSTM to e-commerce data in a personalized, contextualized way [11]. Li et al. propose TBformer, a multi-scale Transformer with time-behaviour attention that fuses static customer data with dynamic behaviour for internet service platforms [9]. Usman et al. compare classical ML with CNN and RNN architectures on the Cell2Cell and KDD Cup benchmark datasets, motivated by the manual feature selection that traditional methods require [15]. The reviewed papers list higher data and computation requirements as the cost of these models.

3.5 Data Fusion and Segmentation

Alizadeh et al. combine hard data (recorded transactions) with soft data (customer opinions and feedback) for banking churn, using Decision Tree, change mining, and K-means [7]. Wu et al. pair churn prediction with K-means segmentation so that each customer group can receive a targeted retention strategy [6]. Lee et al. define churn dynamically from customer behaviour rather than a fixed inactivity window, and segment customers into new, short-term, high-value, and churned groups before prediction [12]. These studies move beyond a single yes/no prediction toward action at the level of customer groups.

3.6 Class Imbalance

Churners are usually far fewer than non-churners, so a model can reach high accuracy while missing most churners. Studies respond at the data level: Wu et al. apply SMOTE [6]; Al-Sultan and Al-Baltah run experiments on balanced and unbalanced data [1]; Inkesar introduces density-regularized oversampling that creates new minority samples without simply duplicating records [8]; and Shrestha and Shakya cite the imbalance of telecom data as the motivation for testing XGBoost [3]. A methodological point applies to all of these: resampling must be done on the training split only, otherwise synthetic points leak into the test set and inflate scores.

Fig. 3 illustrates the imbalance with a hypothetical dataset. A model that labelled every customer in that dataset as non-churn would reach 90% accuracy while identifying none of the churners (0% recall). The figures are illustrative and are not results of this paper.

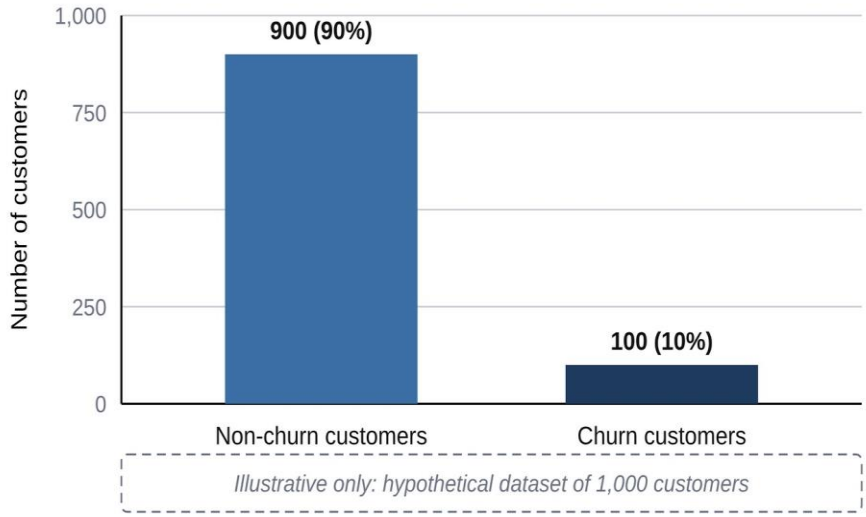


Fig. 3: Illustrative class imbalance in a hypothetical dataset of 1,000 customers (values are hypothetical).

3.7 Business-Value Evaluation and Explainability

Verbraken et al. argue that models should be judged by the profit of the retention campaign they enable, not by accuracy alone, and they introduce the Expected Maximum Profit criterion, which also indicates how many customers to target [13]. The review by Manzoor et al. of 212 articles from 2015 to 2023 identifies profit-based evaluation and explainable ML as key gaps [4]. Inkesar addresses the second gap by pairing oversampling with Explainable AI so that businesses can see why a customer is predicted to churn [8]. Fig. 4 is a schematic of how a SHAP-based explanation attributes one customer's predicted churn risk to individual factors; no SHAP values were computed for this paper.

4. Research Gap and Problem Statement

4.1 Recurring Challenges

Six challenges recur across the reviewed studies, and none is fully solved. Table 2 lists the evidence for each, the mitigation currently used, and the gap that remains.

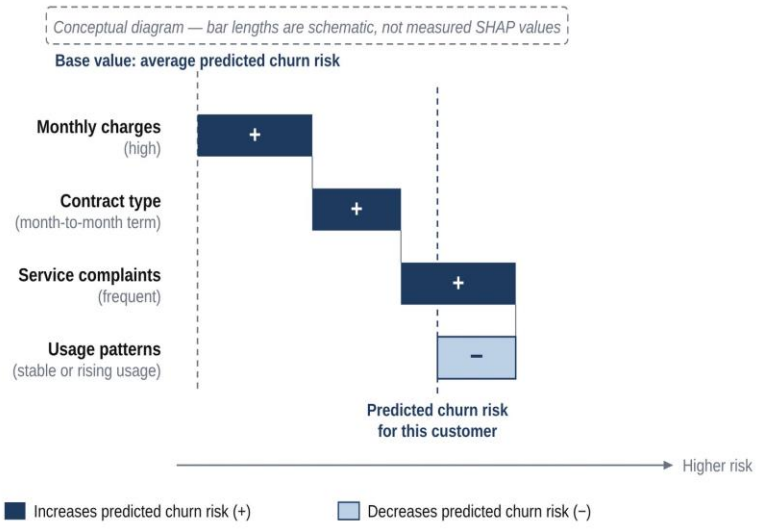


Fig. 4: Conceptual illustration of a SHAP-based churn-risk explanation (schematic; no SHAP values were computed).

Table 2: Recurring challenges, current mitigation, and open gaps in the reviewed literature

Challenge	Evidence in the literature	Current mitigation	Open gap
Class imbalance	Churners are a small minority; models can miss them while scoring high accuracy [1], [3], [8]	SMOTE, density-regularized oversampling, balanced-data experiments [1], [6], [8]	Better minority recall; synthetic samples that reflect real churn behaviour [8], [14]
Data dependence	Models are tied to the features and datasets studied [3], [5], [6], [10]	Public plus real-world datasets; cross-validation [3]	External validation on new sectors and data [10]
Limited explainability	Managers need reasons, not only scores [4], [8]	XAI with oversampling [8]; factor analysis [5]	Explanations for deep and Transformer models [9], [11], [15]
Accuracy versus business value	The most accurate model is not always the most profitable [13]	Expected Maximum Profit [13]	Linking profit metrics to modern ML [13]
Complexity and cost	Transformers, fusion, and hybrid models are heavy [7], [9], [12]	Simpler ensemble baselines [10], [14]	Efficiency and real-time prediction [9], [12]
Data availability	Social-network and feedback data are often unavailable [2], [7]	Use available demographic and service data [3], [5]	Privacy-preserving use of richer signals [2], [7]

4.2 Identified Research Gap

Beyond the six challenges, the reviewed work as a whole shows a further gap: predicting churn and acting on it are treated as separate problems. Most studies end at the prediction, while the choice of retention strategy is left to the business [4], [5], [6]. The few studies that move toward action do so at the level of customer groups through segmentation [6], [12], or by explaining predictions without mapping them to specific interventions [8]. A customer-level link from the predicted risk, through the reasons behind it, to a concrete retention action is therefore not established in the reviewed studies.

4.3 Problem Statement

The problem addressed here is how to turn a churn score into a timely and targeted decision. A workable approach must (a) handle the class imbalance of churn data without leaking information into the evaluation, (b) report metrics that reflect the rarity of churners and, where cost data exist, business value, (c) explain the factors behind each customer's prediction, and (d) map those factors to retention actions that a business can review and adjust. Section 5 proposes a workflow for this purpose.

5. Proposed Framework

The proposed framework predicts churn probability, explains the main drivers for each customer, and maps those drivers to a retention action. It reuses the common pipeline found in the reviewed studies and adds the explanation and action stages that most of them leave out. It is a design proposal; no experimental results are reported in this paper.

5.1 Reference Pipeline

Despite the variety of models, the reviewed studies follow a common pipeline, which forms the base of the framework:

- **Data collection.** Customer demographic, service, transaction, and behavioural records; call or social-network data where available [2], [7].
- **Preprocessing.** Cleaning, handling missing values, encoding categorical features, and normalization [3], [10].
- **Feature selection and imbalance handling.** Selecting informative features and rebalancing classes with SMOTE or similar methods [6], [8].
- **Model training and prediction.** Training one or more classifiers to output a churn label or probability, with hyperparameter tuning and cross-validation [1], [3].
- **Evaluation.** Accuracy, precision, recall, F1-score, AUC, and in some studies profit-based or agreement metrics such as MCC and Cohen's kappa [2], [13].
- **Interpretation and action.** Identifying churn factors and linking them to retention decisions [4], [5], [8].

Fig. 5 shows this pipeline as a flowchart. SMOTE is applied to the training data only; the test set keeps its original class distribution.

5.2 Workflow

1. **Collect data.** Customer demographic, service, and behavioural data in CSV form, loaded in Python (Pandas, NumPy).
2. **Clean and encode.** Handle missing values, encode categorical variables, and prepare relevant features.
3. **Split first.** Divide the data into training and testing sets, stratified by the churn label.
4. **Rebalance the training set only.** Apply SMOTE (imbalanced-learn) to the training data so that the test set stays untouched [6]. When cross-validation is used, SMOTE is applied inside each training fold only.
5. **Select features.** Keep the informative features using importance scores computed on the training data.
6. **Train three models.** Logistic Regression as an interpretable baseline, Random Forest for non-linear patterns and feature importance [5], and XGBoost for accurate churn probabilities [3], with hyperparameter tuning and stratified cross-validation [3].
7. **Evaluate.** Report accuracy, precision, recall, F1-score, and ROC-AUC on the untouched test set, with recall and F1 emphasized because of imbalance [1], [14]. Where cost data exist, add a profit-based comparison [13].
8. **Explain.** Use feature importance and SHAP values to find the factors that push each customer's prediction toward churn [8].
9. **Recommend.** Assign each customer a risk tier from the predicted probability and map the top churn factors to a retention action.

Fig. 6 presents these steps as an architecture diagram, from data input to the final retention recommendation.

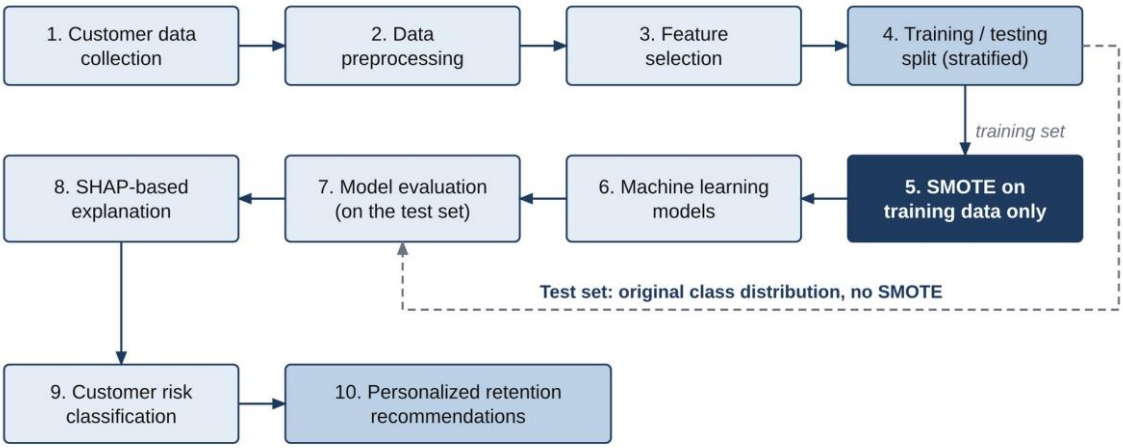


Fig. 5: Overall pipeline for ML-based customer churn prediction. The test set is not oversampled (authors' synthesis of the common pipeline in the reviewed studies).

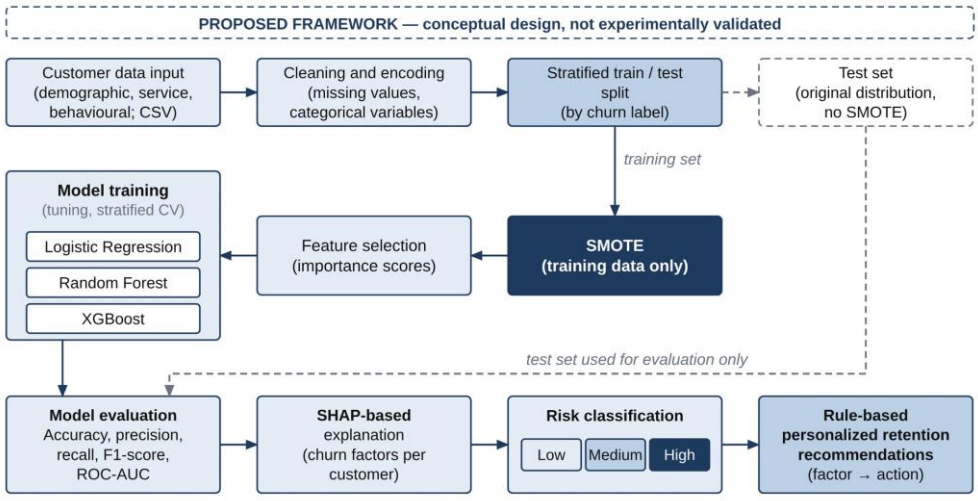


Fig. 6: Proposed customer churn prediction framework (authors' proposal; not experimentally validated).

5.3 Technique Rationale

The five techniques in the workflow play different roles. Table 3 summarizes each role together with a short technical note. The notes are standard background on the methods rather than findings of the reviewed studies.

Table 3: Role of each technique in the proposed workflow

Technique	Role	Technical note
Logistic Regression	Interpretable baseline	Models the log-odds of churn as a linear combination of features and converts it to a probability with the logistic function. Coefficients are directly readable, but non-linear effects and feature interactions are captured only if engineered.
Random Forest [5]	Non-linear patterns; feature importance	Averages many decision trees, each grown on a bootstrap sample with a random subset of features considered at each split. This reduces the variance of a single tree and yields feature-importance scores.
XGBoost [3]	Accurate churn probabilities	Builds trees sequentially, each new tree correcting the errors of the current ensemble, with regularization to limit overfitting. Widely used for tabular data.
SMOTE [6]	Rebalances the training set only	Creates synthetic minority samples by interpolating between a minority instance and one of its nearest minority neighbours. It can generate noisy samples where classes overlap and must never be applied to test or validation data.
SHAP [8]	Per-customer factor attribution	Assigns each feature a Shapley-value-based contribution to one prediction; the contributions and a base value sum to the model output (for tree-based classifiers, commonly expressed in log-odds). Positive contributions push toward churn.

5.4 From Churn Factors to Retention Actions

Table 4 shows the intended mapping logic. The factors and actions are illustrative examples consistent with the drivers discussed in [4], [5], and [10]; the actual rules must be derived from the SHAP output on the chosen dataset and agreed with the business. Fig. 7 presents the same mapping as a flowchart.

Table 4: Illustrative mapping from SHAP-identified churn factors to retention actions

Dominant churn factor (from SHAP)	Probable reason	Example retention action
High monthly charges, price-sensitive segment	Pricing	Discounted plan or loyalty pricing
Short contract or month-to-month term	Low commitment	Incentive for a longer-term contract
Frequent service complaints or support calls	Service quality	Priority support and a service-recovery call
Falling usage over recent periods	Disengagement	Personalized usage offer or re-engagement message
Competitor offers or service type mismatch	Better alternatives elsewhere	Plan upgrade or matched offer [10]

Risk tiers (for example low, medium, and high) are set by probability thresholds that the business tunes to its campaign budget. Where profit data exist, the threshold can be chosen with the Expected Maximum Profit criterion [13].

5.5 Tools

The workflow uses Python with Pandas and NumPy for data handling, Matplotlib and Seaborn for plots, Scikit-learn and XGBoost for models, imbalanced-learn for SMOTE, and SHAP for explanation, in a Jupyter Notebook or Google Colab environment. A GPU is not required for these models, and a standard machine with 8 GB of RAM is sufficient for typical churn datasets.

6. Comparative Analysis and Discussion

6.1 Cross-Study Comparison

Table 5 summarizes the fifteen studies by year, sector, core approach, reported evaluation focus, and the gap associated with each. Entries are drawn from each study's stated aims and are not performance comparisons.

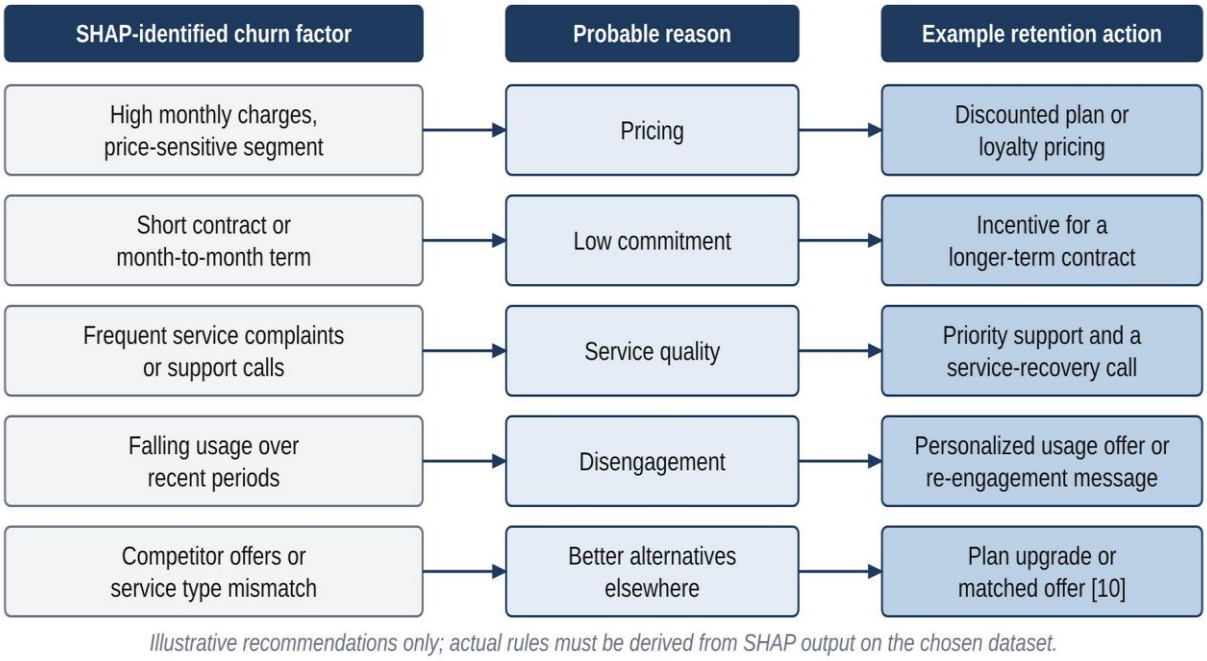


Fig. 7: Mapping predicted churn factors to personalized retention actions (authors' illustrative recommendations).

Table 5: Cross-study comparison of the reviewed literature (stated aims; not a performance comparison)

Ref.	Year	Sector	Core approach	Evaluation / methodology noted	Gap associated
[1]	2024	Banking	Improved Random Forest (ERFA) compared with Logistic Regression, SVM, and Neural Network	Balanced and unbalanced data experiments	Minority-class recall
[2]	2024	Telecom	Influence and conformity features from phone-call networks	AUC, accuracy, precision, F1, MCC, Cohen's kappa, Brier score	Explainability; richer data use
[3]	2022	Telecom (Nepal)	XGBoost on public and real data	Cleaning, normalization, tuning, 10-fold stratified CV	External validation
[4]	2024	Cross-sector review	Review of 212 articles (2015–2023); telecom, finance, online gaming	Profit-based evaluation and explainable ML identified as gaps	Profit-based evaluation; explainability
[5]	2019	Telecom	Random Forest with classification and clustering for factor identification	Factor analysis	Data dependence

Ref.	Year	Sector	Core approach	Evaluation / methodology noted	Gap associated
[6]	2021	Telecom	Churn prediction with SMOTE and K-means segmentation	Group-level retention strategy	Data dependence
[7]	2023	Banking	Hard and soft data fusion with Decision Tree, change mining, K-means	Transactions combined with customer opinions	Data availability; complexity
[8]	2026	Industrial (B2B)	Density-regularized oversampling with Explainable AI	Few high-value accounts and few churn examples	Explainability; realistic synthetic samples
[9]	2025	Internet service platforms	TBformer: multi-scale Transformer with time-behaviour attention	Fuses static and dynamic behaviour data	Explainability; complexity
[10]	2024	Telecom	Decision Tree, Random Forest, KNN, XGBoost compared	Factors such as competitors' offers and internet service type	External validation
[11]	2025	E-commerce	Personalized, contextualized Bi-LSTM	Purchase and behaviour sequences	Explainability
[12]	2024	Online platforms	Dynamic churn definition with customer segmentation	Groups: new, short-term, high-value, churned	Efficiency; real-time prediction
[13]	2013	Cross-sector	Expected Maximum Profit criterion	Profit of the retention campaign; number of customers to target	Linking profit to modern ML
[14]	2023	Telecom and insurance	Weight average ensemble of XGBoost and Random Forest	F1-score and processing time	Minority-class recall
[15]	2021	Benchmark datasets (Cell2Cell, KDD Cup)	Classical ML compared with CNN and RNN architectures	Reduces manual feature selection	Explainability

Three patterns appear. First, telecom is the most studied sector (seven of the fifteen studies). Second, almost every study lists data dependence as a limitation, so conclusions need validation on new datasets. Third, the gaps named most often are explainability [2], [4], [8], [9], [11], [15], imbalance [1], [14], and external validation [3], [10].

6.2 Evaluation Metrics

Churn is a binary classification task, so the metrics derive from the confusion matrix (TP, FP, TN, FN, with churn as the positive class):

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

$$Precision = TP / (TP + FP) \quad (2)$$

$$Recall = TP / (TP + FN) \quad (3)$$

$$F1 = (2 \times Precision \times Recall) / (Precision + Recall) \quad (4)$$

Precision is the share of flagged customers who really churn; recall is the share of real churners that are caught. Because churners are rare, recall, F1, and AUC are more informative than accuracy [1], [14]. MCC, Cohen's kappa, and Brier score add chance-corrected agreement and probability calibration [2]. Business-oriented work goes further: a campaign has a contact cost and a retention benefit, and the best model is the one that maximizes expected profit [13].

6.3 Domain-Specific Applications

Table 6 groups the reviewed studies by application domain.

Table 6: Application domains in the reviewed studies

Domain	Studies	Focus in the reviewed work
Banking	[1], [7]	Predicting customers who may close accounts or stop using bank services; imbalanced data [1]; combining transaction records with customer opinions [7]
Telecommunications	[2], [3], [5], [6], [10]	Random Forest, XGBoost, and ensembles; social-network features from call data; prediction combined with segmentation; drivers such as competitors' offers and the type of internet service
E-commerce and online platforms	[9], [11], [12]	Sequence and behaviour models for customers about to stop purchasing or using a platform; dynamic churn definitions for businesses with irregular purchase cycles
Industrial and B2B	[8]	A small number of high-value accounts and few churn examples make imbalance handling and explanation central
Insurance	[14]	A weighted ensemble tested on an insurance customer dataset alongside telecom data
Cross-sector decision support	[4], [13]	Profit-based evaluation to decide how many customers to include in a campaign [13]; a review covering telecom, finance, and online gaming [4]

6.4 Discussion

Three conclusions follow from the comparison. First, the choice of algorithm is only one of several factors that shape results. Preprocessing, imbalance handling, and the evaluation metric matter as much, and a reported score is only meaningful when resampling is confined to the training data. Second, tree ensembles and boosting are strong defaults for tabular data, whereas deep and network-based models add value mainly when sequence or relational data exist and when their higher data and computation cost is acceptable. Third, accuracy and business value are different objectives: profit-based evaluation and explainability are the most frequently cited gaps, and both are needed before a prediction can support a retention decision.

The proposed framework in Section 5 responds to the gap identified in Section 4.2. It keeps the established pipeline, selects interpretable and widely used models, and adds per-customer explanation and an action layer. Its benefits remain untested, and validating them is the main next step.

7. Limitations of This Review

- **Selected sample.** Fifteen studies were reviewed. The review is narrative and does not claim to cover all churn literature.
- **No numerical comparison.** The studies use different datasets, preprocessing, and metrics, so reported scores are not directly comparable and are not tabulated here.
- **Stated aims, not full results.** The synthesis rests on each study's stated objectives, methods, limitations, and gaps; readers should consult the originals for exact figures.
- **Publication scope.** Most studies come from IEEE venues and are in English, which may under-represent work from other publishers and languages.
- **Proposal not evaluated.** The framework in Section 5 is a design; its benefits are untested.

8. Conclusion and Future Work

8.1 Conclusion

Machine learning is an effective way to identify customers who are likely to leave and to understand the factors behind their behaviour. The reviewed studies show that tree ensembles and boosting are strong choices for tabular churn data, that deep and network-based models add value when sequence or relational data exist, and that preprocessing, imbalance handling, and metric choice shape results as much as the algorithm itself. Profit-based evaluation and explainability are the most frequently cited gaps. The proposed workflow, which combines Logistic Regression, Random Forest, and XGBoost with

SMOTE on the training data, SHAP-based factor analysis, and mapped retention actions, aims to turn a churn score into a timely, targeted decision. Its validation on real datasets is the main next step.

8.2 Future Research Directions

1. **Experimental validation of the proposed framework.** Run the Section 5 workflow on public churn datasets and, where possible, on real company data, and report all metrics with cross-validation.
2. **Better minority-class learning.** Combine oversampling methods such as SMOTE and density-regularized sampling with cost-sensitive learning and compare them on recall and F1 [1], [8], [14].
3. **Explainability for advanced models.** Extend XAI methods to sequence and Transformer models so that deep models can also give actionable reasons [9], [11], [15].
4. **Profit-aware modelling.** Integrate Expected Maximum Profit and similar criteria into the training and threshold selection of modern ML models [13].
5. **Richer and dynamic signals.** Use social-network, feedback, and temporal behaviour data where available, and move toward real-time prediction with dynamic churn definitions [2], [7], [12].
6. **Closed-loop retention.** Measure whether recommended actions actually reduce churn, for example through controlled retention campaigns, and feed the outcome back into the model.
7. **Privacy and fairness.** Study how to use richer customer data while protecting privacy and avoiding biased treatment of customer groups.

References

[1] S. Y. Al-Sultan and I. A. Al-Baltah, “An Improved Random Forest Algorithm (ERFA) Utilizing an Unbalanced and Balanced Dataset to Predict Customer Churn in the Banking Sector,” *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3395542.

[2] B. Amiri and S. H. Hosseini, “Unveiling the Power of Social Influence: A Machine Learning Framework for Churn Prediction With Network Analysis,” *IEEE Access*, vol. 12, pp. 71271–71285, 2024, doi: 10.1109/ACCESS.2024.3402684.

[3] S. M. Shrestha and A. Shakya, “A Customer Churn Prediction Model using XGBoost for the Telecommunication Industry in Nepal,” *Procedia Computer Science*, vol. 215, pp. 652–661, 2022, doi: 10.1016/j.procs.2022.12.067.

[4] A. Manzoor, M. A. Qureshi, E. Kidney and L. Longo, “A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners,” *IEEE Access*, vol. 12, pp. 70434–70463, 2024, doi: 10.1109/ACCESS.2024.3402092.

[5] I. Ullah, B. Raza, A. K. Malik, M. Imran, S. U. Islam and S. W. Kim, “A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector,” *IEEE Access*, vol. 7, pp. 60134–60149, 2019, doi: 10.1109/ACCESS.2019.2914999.

[6] S. Wu, W.-C. Yau, T.-S. Ong and S.-C. Chong, “Integrated Churn Prediction and Customer Segmentation Framework for Telco Business,” *IEEE Access*, vol. 9, pp. 62118–62136, 2021, doi: 10.1109/ACCESS.2021.3073776.

[7] M. Alizadeh, D. S. Zadeh, B. Moshiri and A. Montazeri, “Development of a Customer Churn Model for Banking Industry Based on Hard and Soft Data Fusion,” *IEEE Access*, vol. 11, pp. 29759–29768, 2023, doi: 10.1109/ACCESS.2023.3257352.

[8] A. Inkesar, “An Explainable Machine Learning Approach With Density-Regularized Oversampling for Industrial Customer Churn Prediction,” *IEEE Access*, vol. 14, pp. 119355–119383, 2026, doi: 10.1109/ACCESS.2026.3718205.

[9] Y. Li, Y. Tao, M. Zhu, Z. Chen, Z. Wen and B. Zhu, “TBformer: Multi-Scale Transformer With Time-Behavior Attention for Multi-Modal Customer Churn Prediction,” *IEEE Trans. Consum. Electron.*, vol. 71, no. 2, pp. 3192–3203, May 2025, doi: 10.1109/TCE.2025.3563905.

[10] M. Afzal, S. Rahman, D. Singh and A. Imran, “Cross-Sector Application of Machine Learning in Telecommunications: Enhancing Customer Retention Through Comparative Analysis of Ensemble Methods,” *IEEE Access*, vol. 12, pp. 115256–115267, 2024, doi: 10.1109/ACCESS.2024.3445281.

[11] Q. Wei, “Personalized and Contextualized Data Analysis for E-Commerce Customer Retention Improvement With Bi-LSTM Churn Prediction,” *IEEE Trans. Consum. Electron.*, vol. 71, no. 2, pp. 4406–4414, May 2025, doi: 10.1109/TCE.2024.3376672.

[12] N.-T. Lee, H.-C. Lee, J. Hsin and S.-H. Fang, “Prediction of Customer Behavior Changing via a Hybrid Approach,” *IEEE Open J. Comput. Soc.*, vol. 5, pp. 27–38, 2024, doi: 10.1109/OJCS.2023.3336904.

[13] T. Verbraken, W. Verbeke and B. Baesens, “A Novel Profit Maximizing Metric for Measuring Classification Performance of Customer Churn Prediction Models,” *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 961–973, May 2013, doi: 10.1109/TKDE.2012.50.

[14] I. N. M. Adiputra and P. Wanchai, "Customer Churn Prediction Using Weight Average Ensemble Machine Learning Model," in *Proc. 20th Int. Joint Conf. Computer Science and Software Engineering (JCSSE)*, 2023, pp. 90–94, doi: 10.1109/JCSSE58229.2023.10202105.

[15] M. Usman, W. Ahmad and A. Fong, "Design and Implementation of a System for Comparative Analysis of Learning Architectures for Churn Prediction," *IEEE Commun. Mag.*, vol. 59, no. 9, pp. 86–90, Sep. 2021, doi: 10.1109/MCOM.110.2100145.