

Behavioral Member Churn Prediction Using Visit-Activity Data in a Club Membership System

Naura Gena Anasya¹, Handri Santoso²

^{1, 2} Master of Information Technology, Universitas Pradita, Scientia Business Park, Tangerang, Banten 15810, Indonesia
naura.gena@student.pradita.ac.id¹, handri.santoso@pradita.ac.id² (ORCID: ¹0009-0008-7270-5313, ²0000-0002-3265-8179)

Article Info

Article history:

Received 2026-09-08

Revised 2026-10-09

Accepted 2026-10-18

Keywords:

*Behavioral Churn Prediction,
Membership Retention,
Recency,
Primary Data,
Out-Of-Time Validation,
Data Anonymization.*

ABSTRACT

Membership businesses lose recurring revenue when members quietly stop attending, yet churn studies for the club-and-fitness sector in Indonesia are scarce and rarely use real behavioral data. This paper presents a behavioral churn-prediction case study on primary data extracted read-only from the operational database of a club membership information system; over one million visit records were aggregated into a cohort of 4,728 members, with personally identifiable and legally sensitive attributes excluded before analysis. To avoid label leakage, churn is defined temporally: the cohort is the members active in 2018, labeled churned if they neither visited nor paid during 2019—combining both signals corrects 445 visit-only mislabels—giving a 45% churn rate. Only pre-cutoff features are used. Four classifiers were compared against a majority baseline and a single-variable recency rule. Gradient Boosting reached test ROC-AUC 0.833 (Random Forest comparable), significantly above the recency-only rule (0.702; bootstrapped AUC gain +0.131, 95% CI 0.108–0.156, $p < 0.001$) and the 0.553 baseline, with PR-AUC 0.809. An out-of-time test (train 2017, test 2018) gives ROC-AUC 0.795, and a temporal-shift analysis links the drop to a change in median recency. We also report calibration (Brier 0.168), a data-quality audit, gender fairness, and an operational recency-threshold rule. The contribution is a leakage-free, verified, and statistically benchmarked churn pipeline for an under-studied domain, not a claim of novel drivers.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Subscription and membership businesses depend on recurring revenue, so the silent loss of members—customer churn—threatens financial sustainability, and empirical evidence has long held that retaining an existing customer is cheaper than acquiring a new one [20]. Early identification of at-risk members is therefore a high-value analytics task, increasingly one that must also explain its predictions [3]. While churn prediction is mature in telecommunications [1] and banking [2], the club, sports, and fitness membership sector in Indonesia has received little attention, and public datasets for this domain are effectively non-existent.

Two pitfalls shape this study. First, label leakage: when the churn label and every predictor are read from the same record at the same instant, a model reconstructs a present status rather than forecasting the future, and reports

implausibly high accuracy; a defensible task separates time. Second, benchmarking honesty: because recency of activity has been the strongest churn signal since the classic Recency-Frequency-Monetary (RFM) framework [19], any machine-learning result must be checked against a simple recency rule before value can be claimed. This paper adopts a temporal design on primary, anonymized data and tests the model against exactly that baseline.

The contributions are therefore methodological and evidential rather than a claim of new churn drivers. First, we build a leakage-free behavioral churn pipeline on primary Indonesian visit data, with the churn label verified against payment records. Second, we benchmark four classifiers against a majority baseline and a single-variable recency rule, and quantify how much the model actually adds. Third, we validate stability with an out-of-time test, and report data-

quality, calibration, and fairness checks so the result can be trusted rather than merely reported.

II. RELATED WORK

We read the literature in three groups whose gaps, taken together, define this study's position. The first is the mainstream of applied churn modelling, where tree-ensemble methods—Random Forest [4], gradient boosting [5]—reliably top linear baselines on tabular data [1], [2], [5], and recent work adds class-imbalance handling [6], [7] and SHAP- or permutation-based interpretability [3], [9], [13], [14], [15]. Its recurring weakness is evaluation: many studies report a single classifier's accuracy without a naive or heuristic baseline, so it is unclear how much the machinery actually buys. Reviews of resampling reinforce the point that method choices must be tested, not assumed [8], [10], [11], [12].

The second group is the RFM tradition [19] and the retention economics behind it [20], which established decades ago that recency is the strongest single churn signal. This is precisely why a modern model must be measured against a recency rule before any contribution can be claimed—a comparison the first group usually omits. The third group is domain-specific churn for membership, sports, and fitness businesses, which is comparatively thin and, in the Indonesian context, rarely built on real operational data rather than small public or synthetic sets.

This study sits at the intersection of the three: it brings the first group's ensemble modelling and interpretability to the third group's under-served domain, while adopting the second group's recency rule as an explicit, statistically-tested baseline and adding a leakage-free temporal design on primary, anonymized data. In our earlier work we compared algorithms and imbalance techniques for software defect prediction [16]; here the domain, the temporal behavioral target, and the baseline discipline differ fundamentally.

III. METHOD

A. Data Source, Ethics, and Anonymization

Data were collected as primary data through read-only (SELECT-only) queries against the operational PostgreSQL database of a club membership information system; no record was modified, and access was granted by the operator for academic analysis. Before analysis, all directly identifying information (name, national identity number, addresses, telephone, e-mail) and the internal member identifier were removed and rows aggregated so no individual can be re-identified. Attributes that are legally sensitive under Indonesia's Personal Data Protection Law (UU No. 27 of 2022)—religion, marital status, citizenship—were deliberately excluded to avoid regulatory risk and proxy discrimination.

B. Temporal Cohort, Churn Definition, and Label Verification

An observation cutoff of 31 December 2018 is fixed. The cohort is the 4,728 members with at least one recorded visit during 2018. A member is labeled churned if they neither visited nor made a payment during 2019, and retained otherwise. Using payment as a second activity signal is deliberate: a visit-only label would mark 445 members as churned who in fact paid dues in 2019, an error the combined label removes. The resulting churn rate is 45%. Aggregate monthly activity (Fig. 1) shows the visit series is dense during the 2018 cohort year and through mid-2019 before an operational/pandemic-related recording break; this motivates the payment cross-check and is stated openly as a limitation.

C. Features and Data Quality

Every predictor is computed only from data on or before the cutoff; each derived feature has an explicit, one-directional formula (Table I) so the design is fully reproducible and free of look-ahead leakage. Behavioral features summarise visits and settled payments; non-sensitive membership attributes (member type, gender, city, payment type, main-member flag, age) are also included. A data-quality audit (Table II) revealed that membership tenure was derivable for only half of the members; rather than impute a single constant, missing tenure is flagged and left to the model, and the audit is reported transparently.

TABLE I.
BEHAVIORAL AND DEMOGRAPHIC FEATURE DEFINITIONS.

Feature	Definition (all computed strictly before the cutoff)
recency_days	Days from the cutoff (31 Dec 2018) to the last pre-cutoff visit
visits_cy, visits_py	Visit count in the cutoff year and in the prior year
visits_total, active_months_cy	All-time visits up to the cutoff; distinct visit-months in the cutoff year
visit_tenure_days	Days from the member's first-ever visit to the cutoff
avg_visits_month, visit_trend	$\text{visits_cy} / \text{active_months_cy}$; $\text{visits_cy} - \text{visits_py}$
pay_total, pay_cy	Settled payments up to the cutoff; in the cutoff year
age_at_cut, tenure_at_cut_years	Age at cutoff; (cutoff – registration date), flagged when missing

D. Models, Baselines, and Protocol

Data are split 70/30 (stratified, seed 42); five-fold cross-validation is run on the training part with SMOTE fitted only inside each fold. Four classifiers—Logistic Regression, Decision Tree, Random Forest, and histogram-based Gradient Boosting—are compared in scikit-learn [17] at default settings. Two baselines anchor the comparison: a majority-class predictor (accuracy 0.553) and a single-variable rule that flags churn when days-since-last-visit exceed a threshold; that threshold, like every hyperparameter, is selected only on the training split so the comparison remains leakage-free. Models

are compared with the Friedman test over folds and a Wilcoxon test between the two best; the final model is refit on the training set and evaluated once on the untouched test set. An out-of-time test trains on the 2017 cohort and tests on the 2018 cohort.

E. Evaluation Metrics

We report accuracy; precision, recall, and F1 in micro-, macro-, and weighted-averages; and ROC-AUC as a threshold-independent measure [18]. Feature influence uses permutation importance—unbiased, unlike impurity importance which favours high-cardinality features. Because a threshold decision is only meaningful on calibrated probabilities, we also report the Brier score and a reliability diagram.

IV. RESULTS AND DISCUSSION

A. Cohort, Label Verification, and Activity Series

The cohort is moderately imbalanced (2,115 churned versus 2,613 retained). The monthly activity series (Fig. 1) confirms dense recording during the cohort year and the horizon's first half, and the visit-plus-payment label removes 445 visit-only mislabels.

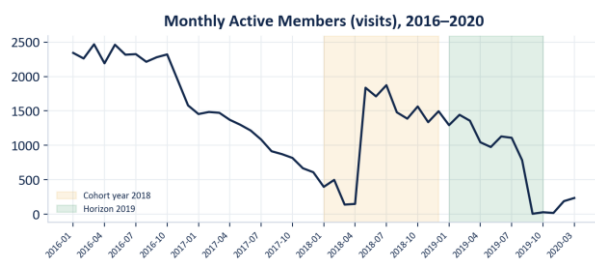


Fig. 1. Monthly active members (visits), 2016–2020, with cohort and horizon marked.

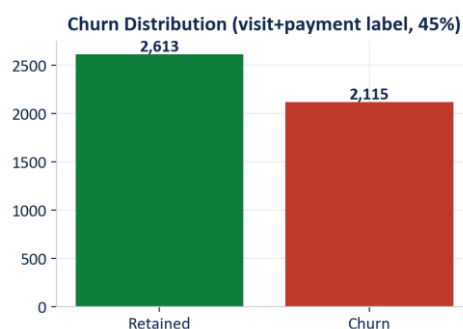


Fig. 2. Churn class distribution (visit-plus-payment label).

B. Data-Quality Audit

TABLE II.
DATA-QUALITY SUMMARY OF NUMERIC FEATURES.

Feature	Missing	% at mode	Min	Median	Max
visits_cy	0.0%	20.6%	1.0	4.0	416.0
visits_py	0.0%	68.4%	0.0	0.0	460.0
visits_total	0.0%	7.2%	1.0	28.0	5573.0
active_months_cy	0.0%	40.1%	1.0	2.0	12.0
recency_days	0.0%	2.9%	0.0	75.0	362.0

visit_tenure_days	0.0%	1.3%	0.0	2575.5	3516.0
pay_total	0.0%	41.6%	0.0	1.0	330.0
pay_cy	0.0%	58.9%	0.0	0.0	52.0
age_at_cut	7.4%	3.0%	5.0	39.0	85.0
tenure_at_cut_years	50.5%	1.7%	0.0	1.27	18.09

The audit surfaced that tenure_at_cut_years is missing for about half the cohort—an artefact of default registration dates in the source system. Reporting this openly, and flagging rather than imputing it, restores confidence that the remaining features were inspected.

C. Model Comparison and the Recency Baseline

TABLE III.
FIVE-FOLD CV PERFORMANCE (MEAN $\pm$ SD) WITH RECENCY AND MAJORITY BASELINES.

Model / baseline	ROC-AUC	PR-AUC	Macro-F1
Random Forest	0.838 $\pm$ 0.010	0.796	0.744
Gradient Boosting	0.833 $\pm$ 0.009	0.792	0.746
Logistic Regression	0.816 $\pm$ 0.009	0.771	0.723
Decision Tree	0.751 $\pm$ 0.016	0.662	0.712
Recency-only rule (test)	0.702	0.612	—
Majority baseline (test)	—	—	acc 0.553

Gradient Boosting and Random Forest lead and are statistically indistinguishable from each other (Wilcoxon $p=0.19$), while the Friedman test confirms the four models differ ($p=0.005$). Crucially, the single-variable recency rule reaches only ROC-AUC 0.702 on the test set, well below the model's 0.833. A paired bootstrap (2,000 resamples of the test set) puts the AUC improvement of Gradient Boosting over the recency rule at +0.131 (95% CI 0.108–0.156; $p<0.001$), and the gain also holds on precision–recall (PR-AUC 0.809 vs 0.612 at a 0.45 prevalence). This is the paper's central quantitative result: the model adds a real, statistically significant margin over the classic recency heuristic—the evidence needed before recommending machine learning at all.

D. Test-Set Performance and Calibration

TABLE IV.
PER-CLASS REPORT ON THE HOLD-OUT TEST SET (GRADIENT BOOSTING).

Class	Precision	Recall	F1	Support
Retained	0.777	0.758	0.767	784
Churn	0.710	0.732	0.721	635

On the untouched test set the model attains accuracy 0.746 and ROC-AUC 0.833 (bootstrap 95% CI 0.812–0.854), with churn recall 0.732. The reliability diagram (Fig. 5) and a Brier score of 0.168 indicate moderate calibration: against a no-skill baseline that always predicts the prevalence (Brier 0.247), this is a Brier Skill Score of 0.32. Applying isotonic calibration lowers the Brier score further to 0.162 without harming ranking, so a threshold shift for higher recall rests on meaningful probabilities rather than arbitrary scores.

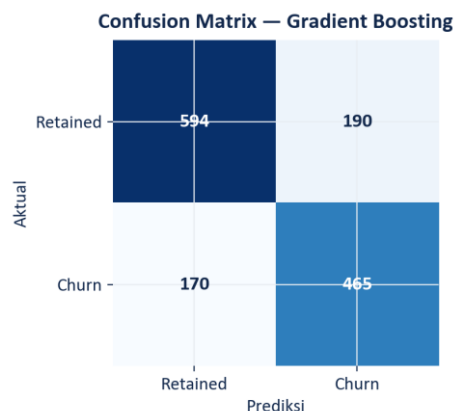


Fig. 3. Confusion matrix (Gradient Boosting).

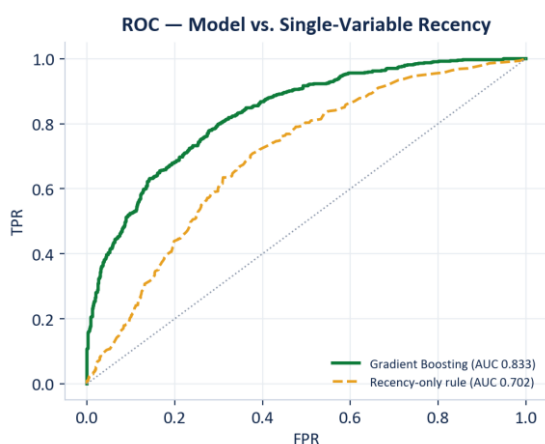


Fig. 4. ROC: Gradient Boosting vs. single-variable recency rule.

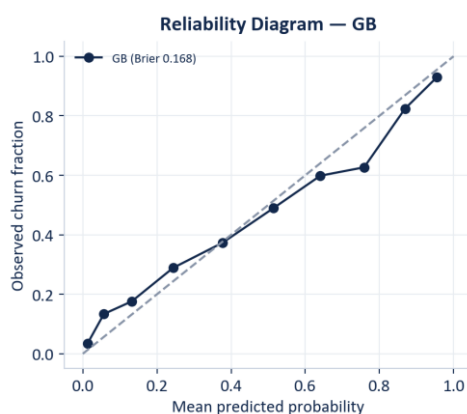


Fig. 5. Reliability diagram (Gradient Boosting).

E. Churn Drivers and Retention Insight

Permutation importance ranks visit recency first, followed by whether tenure is known, recent payment activity, and payment type (recency_days, tenure_missing, pay_cy, pay_total). We caution that the second-ranked feature, tenure_missing, is not a behavioral driver: it flags a migrated membership cohort whose registration date was lost, and that cohort happens to churn far less (29% vs 61%).

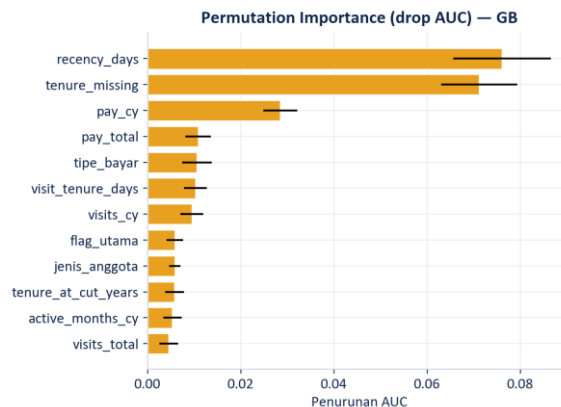


Fig. 6. Permutation importance (drop in test ROC-AUC).

We keep it for completeness but do not interpret it as a retention lever, and Section IV-I shows the model's advantage survives its removal. Recency leading is expected from RFM and is a confirmation, not a discovery; the operationally useful point is that recency is observable in real time, so the model can run as a live early-warning score. For retention practice this suggests triggering re-engagement once a member's visit gap crosses a threshold, and reviewing payment arrangements associated with drop-off. Because error costs are asymmetric—missing a churning forgoes revenue while a false alarm costs only a contact—the operating threshold should favour recall, a choice the calibrated ROC curve makes explicit.

F. From Prediction to an Operational Rule

TABLE V.
OPERATIONAL CONSEQUENCES OF A RECENCY-THRESHOLD EARLY-WARNING RULE (TEST SET).

Recency $\geq$ (days)	Members flagged	Recall (churn)	Precision	Lift	FPR
30	69.7%	0.846	0.543	1.21×	0.577
45	63.9%	0.803	0.562	1.26×	0.506
60	58.3%	0.756	0.580	1.3×	0.443
90	46.9%	0.639	0.611	1.37×	0.330
120	39.1%	0.551	0.631	1.41×	0.261
180	20.6%	0.298	0.647	1.45×	0.131

Table V translates the model into a deployment rule—flag any member whose days-since-last-visit exceeds a threshold—but we read its consequences honestly. At every threshold with usable recall the rule flags more than half the membership: a 60-day cut flags 58.3% of members to capture 76% of churners, a lift of only 1.3× over the 45% base rate. Precision rises with the threshold but recall collapses—at 180 days the rule reaches 1.45× lift yet misses 70% of churners. In other words this is coarse triage, not a precise early-warning system: it is useful only where contacting a large share of members is affordable, and a manager should set the operating point against an explicit contact-cost-versus-retained-value budget rather than assume a low threshold is free.

G. Out-of-Time Validation and Temporal Shift

Training on the 2017 cohort (3,508 members) and testing on the 2018 cohort (4,728) yields ROC-AUC 0.795, below the within-year 0.833 but still far above the recency baseline—so the learned pattern is reasonably stable across years. The modest drop is explained by a genuine distribution shift (Table VI): median recency roughly halves between the 2017 and 2018 cohorts while visit frequency and age stay flat, which is exactly the kind of covariate shift a recency-driven model is sensitive to.

TABLE VI.
DISTRIBUTION SHIFT BETWEEN THE 2017 AND 2018 COHORTS.

Cohort	n	Churn rate	Median recency	Median visits	Median age
2017	3,508	0.475	149.5	4.0	38.0
2018	4,728	0.447	75.0	4.0	39.0

H. Fairness Check

Test performance across gender groups is comparable (L: n=793, AUC 0.841, recall 0.722, precision 0.709; P: n=579, AUC 0.822, recall 0.732, precision 0.716; Unspecified: n=47, AUC 0.802, recall 0.880, precision 0.667). No sensitive attribute is a model input; the small Unspecified group is reported for transparency and, given its size, its figures should be read with caution—its blank gender field probably shares the same migrated-record origin as the missing tenure discussed above.

I. Robustness Checks

TABLE VII.
ABLATION AND ROBUSTNESS OF THE GRADIENT BOOSTING RESULT.

Configuration	Test ROC-AUC
Full model (all features)	0.833
Behavioral features only	0.780
Non-behavioral features only	0.748
Model without tenure features	0.800
Alternative label (visit-only)	0.819
Recency-only rule	0.702

Four checks probe the claim. First—and most important for the title—we separate behavioral from non-behavioral features. Behavioral features alone reach ROC-AUC 0.780 and non-behavioral (demographic and tenure) features alone 0.748; the full model (0.833) exceeds both. Behavioral signals are thus the single largest contributor and the model earns the label, but honesty requires stating that non-behavioral attributes carry real weight too—we therefore describe the study as behavioral-led rather than purely behavioral. Second, removing the artefact-affected tenure features lowers AUC only to 0.800, still beating the recency rule by +0.098 (0.800 vs 0.702); about a quarter of the headline +0.131 margin rests on that artefact and would not survive in a cleaner database, but roughly 0.10 of genuine improvement remains. Third, re-deriving churn from visits only still yields 0.819, so the result does not depend on the label definition. Fourth, SMOTE is near-inert on this now-balanced cohort (AUC 0.833 with vs

0.834 without); we retain it only for pipeline consistency and report the null effect openly. Isotonic calibration keeps ranking intact (0.168→0.162). The finding is stable.

J. Positioning and a Time-to-Event Perspective

Two points frame the contribution. First, relative to conventional churn studies that report a single classifier's accuracy, the value here is the verified, leakage-free pipeline and the explicit, statistically-tested comparison against a recency baseline: the study shows not only that a model works but by how much it beats the obvious heuristic, on primary data for a domain the literature has largely skipped. Second, because churn is intrinsically time-dependent, a survival-analysis view is a natural complementary benchmark. We adopt a fixed one-year horizon deliberately: it makes the label unambiguous and reproducible, matches the annual renewal cycle of the membership product, and keeps the comparison against the recency rule and the operational thresholds of Table V directly interpretable—indeed, that recency-threshold rule is itself a simple hazard-style operating policy. A discrete-time hazard model (logistic regression over member-month intervals) that predicts when—not just whether—a member lapses would refine the timing of interventions and is the concrete next step we set out for future work.

V. CONCLUSION

This paper built a leakage-free behavioral churn model on primary, anonymized visit data from a real Indonesian club membership system, with the churn label verified against payments. Gradient Boosting reached test ROC-AUC 0.83—comparable to Random Forest, clearly above both a majority baseline and, importantly, a single-variable recency rule (0.70)—and remained reasonably stable out-of-time (0.80). The contribution is a verified, honestly-benchmarked pipeline for an under-studied domain, not a claim that recency is a new driver. We reported openly that the two best models are statistically indistinguishable, that half the tenure values were an artefact and were flagged rather than imputed, and that the visit series is incomplete late in the horizon. We also quantify how much of the advantage is domain-specific: about a quarter of the +0.131 margin over the recency rule stems from a tenure data artefact, so the improvement expected to transfer to a cleaner database is the more conservative +0.10. On transferability we are careful to match the evidence: the out-of-time test shows the recency mechanism holds in direction and magnitude across years, but Table VI shows its calibration shifts—median recency halved between cohorts—so the decision threshold is period-sensitive and must be re-tuned for another period or organisation rather than assumed to carry over. Limitations include reliance on a single organisation and a 2018–2019 window; future work should confirm generalisation with newer data, a discrete-time hazard benchmark, and a controlled retention campaign.

DATA AVAILABILITY AND ETHICS STATEMENT

To make the reported numbers independently verifiable, the anonymized analysis cohort and the modelling code (feature construction, splitting, models, and evaluation) are released as supplementary material and deposited in a public repository (Zenodo, DOI: 10.5281/zenodo.22745576). The study uses genuine operational data, not synthetic or simulated data. Because the records are commercially confidential and concern real individuals, the source database cannot be shared; however, a fully anonymized, feature-level version of the analysis cohort—free of any direct identifier and of the excluded sensitive attributes—is provided as supplementary material so that the modelling results are reproducible. Data were accessed read-only with the operator's permission for academic use, and all preprocessing steps and formulas are specified in Section III to support independent verification.

REFERENCES

- [1] A. Bhatnagar and S. Srivastava, "Customer churn prediction: A machine learning approach with data balancing for the telecom industry," *International Journal of Computing*, vol. 24, no. 1, pp. 9–18, 2025, doi: 10.47839/ijc.24.1.3873.
- [2] Y. Li and K. Yan, "Prediction of bank credit customers churn based on machine learning and interpretability analysis," *Data Science in Finance and Economics*, vol. 5, no. 1, pp. 19–34, 2025, doi: 10.3934/DSFE.2025002.
- [3] A. El Attar and M. El-Hajj, "Explainable AI-driven customer churn prediction: A multi-model ensemble approach with SHAP-based feature analysis," *Frontiers in Artificial Intelligence*, vol. 9, art. 1748799, 2026, doi: 10.3389/frai.2026.1748799.
- [4] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [5] I. AlShourbaji, N. Helian, Y. Sun, A. G. Hussien, L. Abualigah, and B. Elnam, "An efficient churn prediction model using gradient boosting machine and metaheuristic optimization," *Scientific Reports*, vol. 13, art. 14441, 2023, doi: 10.1038/s41598-023-41093-6.
- [6] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [7] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, no. 17, pp. 1–5, 2017.
- [8] Hairani, T. Widiyaningtyas, and D. D. Prasetya, "Addressing class imbalance of health data: A systematic literature review on modified SMOTE strategies," *JOIV: International Journal on Informatics Visualization*, vol. 8, no. 3, pp. 1310–1318, 2024, doi: 10.62527/joiv.8.3.2283.
- [9] M. Imani, M. Joudaki, A. Beikmohammadi, and H. R. Arabnia, "Customer churn prediction: A systematic review of recent advances, trends, and challenges in machine learning and deep learning," *Machine Learning and Knowledge Extraction*, vol. 7, no. 3, art. 105, 2025, doi: 10.3390/make7030105.
- [10] N. A. Azhar, M. S. Mohd Pozi, A. Mohamed Din, and A. Jatowt, "An investigation of SMOTE based methods for imbalanced datasets with data complexity analysis," *IEEE Transactions on Knowledge and Data Engineering*, 2022, doi: 10.1109/TKDE.2022.3179381.
- [11] T. Wongvorachan, S. He, and O. Bulut, "A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining," *Information*, vol. 14, no. 1, art. 54, 2023, doi: 10.3390/info14010054.
- [12] J. H. Joloudari, A. Marefat, M. A. Nematollahi, S. S. Oyelere, and S. Hussain, "Effective class-imbalance learning based on SMOTE and convolutional neural networks," *Applied Sciences*, vol. 13, no. 6, art. 4006, 2023, doi: 10.3390/app13064006.
- [13] F. A. Abdalla et al., "Comparative analysis of data augmentation methods for enhancing the performance of churn prediction models," *International Journal of Analysis and Applications*, vol. 23, art. 296, 2025, doi: 10.28924/2291-8639-23-2025-296.
- [14] H. A. Rasool, K. K. Aljawaheri, and A. A. M. Karram, "Explainable customer churn prediction in telecom using ensemble learning and SHAP analysis," *Mathematical Modelling of Engineering Problems*, vol. 12, no. 12, art. 121226, 2025, doi: 10.18280/mmep.121226.
- [15] D. A. Ardhani and K. D. Tania, "Knowledge discovery on e-commerce customer churn using interpretable machine learning: A comparative study of SHAP-based classifiers," *Journal of Applied Informatics and Computing*, vol. 9, no. 5, pp. 2695–2702, 2025.
- [16] N. G. Anasya and T. H. Rochadiani, "A multi-dataset comparison of class-imbalance techniques and machine learning algorithms for software defect prediction," manuscript submitted for publication, Universitas Pradita, 2026.
- [17] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [18] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.
- [19] P. S. Fader, B. G. S. Hardie, and K. L. Lee, "RFM and CLV: Using iso-value curves for customer base analysis," *Journal of Marketing Research*, vol. 42, no. 4, pp. 415–430, 2005, doi: 10.1509/jmkr.2005.42.4.415.
- [20] F. F. Reichheld and W. E. Sasser, "Zero defections: Quality comes to services," *Harvard Business Review*, vol. 68, no. 5, pp. 105–111, 1990.