



Customer Churn Prediction in Waste Banks Using XGBoost and SMOTE

Indri Rahmayuni¹, Rahmi Putri Kurnia², Yance Sonatha³, Yulherniwati⁴, Afcha Arel Pratama⁵
^{1,2,3,4,5}Department of Information Technology, Politeknik Negeri Padang, Padang, Indonesia

*indri@pnp.ac.id

Abstract

Customer retention has become an important challenge in waste bank programs because declining member participation may reduce operational sustainability and weaken community-based waste management initiatives. However, churn prediction studies in non-commercial environmental programs such as waste banks remain limited. This study proposes a machine learning approach for customer churn prediction using operational transaction data from a waste bank managed by the Environmental Agency of Padang City, Indonesia. The dataset consisted of 34,188 transaction records representing 1,015 members collected between May 2024 and April 2026. Customer behavioral features were constructed from transaction history indicators, while class imbalance was handled using SMOTE and churn classification was performed using XGBoost under a leakage-aware customer-level train-test separation, ensuring a realistic evaluation on previously unseen members. Experimental results showed that the proposed model achieved an accuracy of 0.84, an F1-score of 0.73, and a ROC AUC value of 0.898. Feature analysis revealed that recency and transaction frequency were among the strongest predictors of churn behavior. The findings demonstrate the potential of machine learning to support participation monitoring and sustainability management in community-based waste bank systems.

Keywords: churn prediction; customer retention; SMOTE; waste; waste bank; XGBoost

How to Cite: I. Rahmayuni, R. P. Kurnia, Y. Sonatha, Y. Yulherniwati, and A. A. Pratama, "Customer Churn Prediction in Waste Banks Using XGBoost and SMOTE," *J. RESTI (Rekayasa Sist. Dan Teknol. Informasi)*, vol. 10, no. 4, pp. 1079–1088, 2026.

Permalink/DOI: <https://doi.org/10.29207/w8pcjx73>

Received: May 8, 2026

Accepted: August 5, 2026

Available Online: August 19, 2026

This is an open-access article under the [CC BY 4.0 License](#)

Published by [Ikatan Ahli Informatika Indonesia](#)

1. Introduction

Of 225 waste bank units established in Padang City by 2024, only 129 remained operationally active, indicating a participation decline of more than 40% in community-based waste management programs [1]. This condition highlights the importance of maintaining member participation and operational sustainability in waste bank systems. In this study, the term customer refers to registered waste-bank members.

Waste banks have been widely promoted in Indonesia as a community-based approach for reducing household waste and encouraging recycling [2], [3]. Residents deposit recyclable waste in exchange for monetary value or savings credits, which also fosters environmental awareness and household participation in sustainable waste management [4], [5].

However, maintaining long-term participation remains a major challenge, as waste bank sustainability strongly depends on continuous community engagement [6]. In practice, many registered members gradually become inactive, reducing transaction continuity and weakening

the effectiveness of community-based waste management.

In this context, member inactivity can be viewed as churn behavior, since it reflects the decline or cessation of participation. Identifying behavioral patterns associated with churn is therefore important for participation monitoring and retention in community-based waste management.

Customer churn prediction has been studied extensively in commercial sectors such as telecommunications, banking, and e-commerce using machine learning on historical behavioral data, where transaction and interaction patterns reliably signal disengagement and retention risk [7], [8]. Recent reviews confirm the growing adoption of machine learning techniques for churn prediction across multiple application domains [9], [10].

Among these methods, boosting-based models such as XGBoost perform strongly on structured tabular data [11]. Behavioral indicators derived from transaction history, particularly Recency, Frequency, and Monetary

(RFM) features, are widely used to represent customer engagement [12], [13] and have proven effective for churn classification [14].

Despite these advances, churn prediction in non-commercial environmental programs such as waste banks remains limited. Unlike commercial sectors, participation here is driven not only by financial incentives but also by environmental awareness, convenience, and community factors [5], [6], producing behavioral patterns that may differ from those in commercial churn studies.

A further issue in churn modeling is evaluation design. Random row-level splitting, when multiple records belong to the same customer, can cause information leakage and overly optimistic estimates [15], [16]. Customer-level separation instead provides a realistic scenario, requiring the model to predict churn for previously unseen members from historical behavior.

Accordingly, this study proposes a churn prediction framework for waste bank members. Customer behavior snapshots were constructed from historical transactions; RFM-based and waste-deposit features were generated; class imbalance was handled with SMOTE; and prediction used XGBoost under leakage-aware customer-level validation [11], [17].

The novelty of this study is twofold. First, it addresses a clear gap: while churn prediction is established in commercial sectors, no prior work has applied behavioral churn analytics to operational waste-bank participation data, where participation motivation differs fundamentally from profit-driven settings. Second, it contributes a leakage-aware design combining customer-level splitting with snapshot-based feature construction, yielding realistic estimates and interpretable indicators for participation monitoring.

2. Methods

This study developed a churn prediction framework using historical operational data, feature engineering, class imbalance handling, machine learning classification, and independent model evaluation. The overall workflow consisted of raw data auditing, dataset construction, preprocessing, model training, and predictive performance assessment. Figure 1 illustrates the overall research workflow used in this study, including data preparation, snapshot construction, class balancing, model development, and performance evaluation stages.

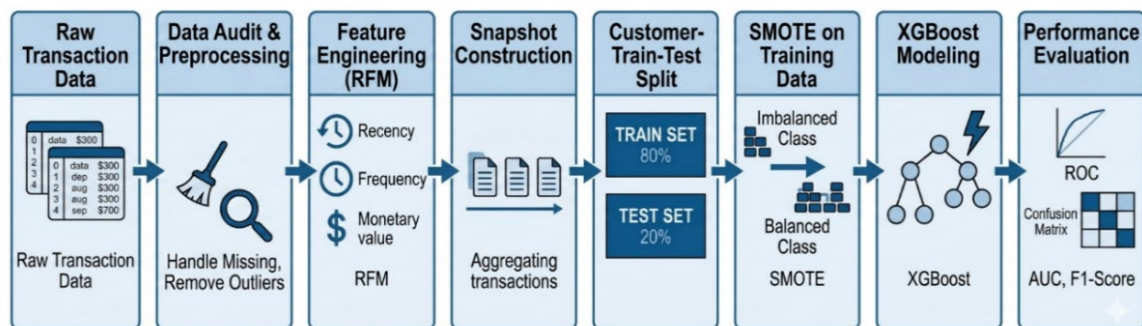


Figure 1. Overall research workflow

2.1 Raw Data Source

The research data were obtained from an operational spreadsheet-based waste bank information system managed by the Environmental Agency (Dinas Lingkungan Hidup) of Padang City. The dataset contains transaction and customer administration records collected from 15 May 2024 to 27 April 2026. The source workbook consisted of multiple worksheets containing transaction records, customer registration data, balance summaries, and supporting administrative information. The main transaction sheet contained 34,188 records and 13 attributes, while customer profile sheets contained more than 1,000 registered members.

Available variables included transaction timestamp, customer name, account number, deposited waste type, deposited weight, unit price, transaction value, gender, address, and organizational department.

2.2 Data Audit and Preprocessing

An initial audit was conducted to evaluate data completeness, consistency, and integration readiness. A small number of duplicated transaction records and incomplete customer entries were identified during the audit process. Additional issues included inconsistent customer identifiers, categorical label inconsistencies, and worksheets not directly relevant for predictive modeling.

Data cleaning was performed to standardize variables and improve dataset consistency. Customer identifier fields were reconciled into a unified account number key, while invalid and duplicated transaction entries were excluded. Date and numeric fields were converted into consistent formats to support temporal analysis and feature generation. Transaction records were then integrated with customer profile data using the customer account identifier.

Several attributes were excluded during preprocessing because they were either redundant or generated sparse categorical dimensions that did not improve predictive performance. In particular, detailed address information was removed from the final modeling stage.

2.3 Dataset Construction and Feature Engineering

To predict customer churn, transaction history was transformed into multiple customer-period observations using a snapshot approach. Each snapshot represents behavioral conditions at a specific observation date based on historical transaction behavior available up to that time. Snapshots were generated on a monthly basis from accumulated transaction records to simulate real-world monitoring conditions, where predictions are made using only historical information available before the prediction period.

For every snapshot, a binary target label named ChurnStatus was generated according to transaction activity in the subsequent observation period. Customers who became inactive were labeled as churn, while those who remained active were labeled as non-churn. This approach enables the model to learn temporal behavioral patterns associated with member inactivity rather than relying solely on static customer profiles.

In this study, a customer was labeled as churn if no waste deposit transaction was observed within the subsequent three-month observation period after the snapshot date. Customers who continued to perform transactions during the observation window were labeled as non-churn.

The three-month inactivity threshold was selected based on the operational transaction characteristics of the waste bank system, where active members generally perform waste deposit transactions within relatively regular monthly intervals.

To empirically justify the three-month inactivity window, the distribution of intervals between consecutive deposits was examined. The median inter-transaction interval was 7 days and the 90th percentile was 21 days, with 99.3% of all intervals shorter than 90 days. Active members therefore almost always transact again within three months, so an absence exceeding this window represents a clear departure from normal participation rather than a routine pause. The robustness of this choice was further verified by reconstructing the churn label under 30-day, 90-day, and 180-day windows; the results, reported in Section 3.4, show that model discrimination remains stable across all thresholds, confirming that the reported performance does not depend on an arbitrarily selected definition.

After construction, the final analytical dataset contained 8,607 observations representing 1,015 unique members. Predictor variables were derived from the RFM framework together with operational attributes commonly used in customer analytics [13], [14]. The final model therefore included six predictors:

transaction recency, transaction frequency, cumulative transaction value, deposited waste volume, gender, and organizational department.

2.4 Data Splitting and Class Balancing

To obtain a more realistic validation scenario, the dataset was divided into training and testing subsets using an 80:20 ratio at the customer level. This strategy ensured that the same customer could not appear in both subsets, thereby reducing the risk of information leakage during model evaluation. The final split produced 6,956 training observations and 1,651 testing observations.

The training data showed class imbalance, where non-churn observations were substantially more frequent than churn observations. Specifically, the training subset contained 5,098 non-churn observations and 1,858 churn observations. To improve minority-class learning, the Synthetic Minority Oversampling Technique (SMOTE) was applied only to the training data [17]. SMOTE was used to generate synthetic minority samples and improve class balance before model training [18].

2.5 Model Development and Evaluation

Categorical predictors were transformed using one-hot encoding prior to model training. The classification model used in this study was Extreme Gradient Boosting (XGBoost), a boosting-based ensemble algorithm proposed by Chen and Guestrin [11] that has demonstrated competitive performance on structured tabular classification tasks.

The XGBoost model was implemented using 250 estimators, a maximum tree depth of 4, and a learning rate of 0.05. To reduce overfitting and improve model generalization, subsample and colsample_bytree parameters were both set to 0.8. The logloss evaluation metric and a random state of 42 were also applied during model training. Preliminary parameter exploration was conducted empirically to obtain stable classification performance prior to final model evaluation.

The hyperparameter configuration was obtained through a systematic randomized search rather than manual trial-and-error. A RandomizedSearchCV procedure (25 candidate configurations ($n_iter = 25$), three-fold stratified cross-validation, optimizing ROC AUC) was conducted over the following search space: number of estimators {150, 250, 350, 500}, maximum tree depth {3, 4, 5, 6}, learning rate {0.01, 0.03, 0.05, 0.1}, subsample and column-subsample ratios {0.7, 0.8, 0.9, 1.0}, minimum child weight {1, 3, 5}, and gamma {0, 0.1, 0.3}. To prevent information leakage, SMOTE was embedded within the cross-validation pipeline so that oversampling was applied only to the training folds. The search produced a slightly more complex configuration that marginally improved cross-validation AUC but did not improve held-out test performance; the configuration reported above was

therefore retained as it provides a more favorable bias-variance trade-off (Section 3.6).

To verify that the selected algorithm is appropriate rather than arbitrarily chosen, XGBoost was benchmarked against four baseline classifiers under an identical pipeline: Logistic Regression, Decision Tree, Random Forest, and LightGBM. All models used the same customer-level split, the same SMOTE-balanced training data, and identical one-hot encoding. Logistic Regression additionally used standardized features. The comparison is reported in Section 3.5.

To assess model stability during training, five-fold stratified cross-validation was performed on the resampled training data [19]. Stratified sampling was applied to preserve class distribution across validation folds and improve the reliability of performance estimation during model development.

The final model was evaluated using an independent testing dataset. Model performance was assessed using accuracy, precision, recall, F1-score, and Receiver Operating Characteristic Area Under Curve (ROC AUC), which are widely used metrics for classification evaluation [20], [21].

In addition to these threshold-based measures, and because class imbalance can render accuracy optimistic, the evaluation was complemented with the Precision-Recall Area Under Curve (PR-AUC), the Matthews Correlation Coefficient (MCC), Balanced Accuracy, and Cohen's Kappa, which are more informative for imbalanced classification problems [22], [23]. To confirm that the model does not overfit to the synthetic minority samples generated by SMOTE, performance

on the real pre-oversampling training data was compared against the independent test set, and a five-fold cross-validation with SMOTE applied inside each fold was performed (Section 3.7). Additional robustness analyses were also conducted, including statistical significance testing, probability calibration, permutation importance, temporal validation, subgroup analysis, multicollinearity assessment, and nested cross-validation (Sections 3.8 and 3.9).

3. Results and Discussions

This section presents the experimental results of the proposed churn prediction model and discusses the findings with supporting tables and figures to improve clarity and interpretability.

3.1 Class Distribution and SMOTE Effect

Before model training, the dataset exhibited class imbalance in which non-churn observations were substantially more dominant than churn observations. The training subset contained 5,098 non-churn instances and 1,858 churn instances, indicating that churn behavior represented a relatively smaller portion of the dataset.

Imbalanced data distributions may reduce model sensitivity toward minority classes and lead to biased classification performance, particularly in churn detection tasks [18]. To address this issue, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training data to generate synthetic minority-class samples and improve class balance before model training [17], [24].

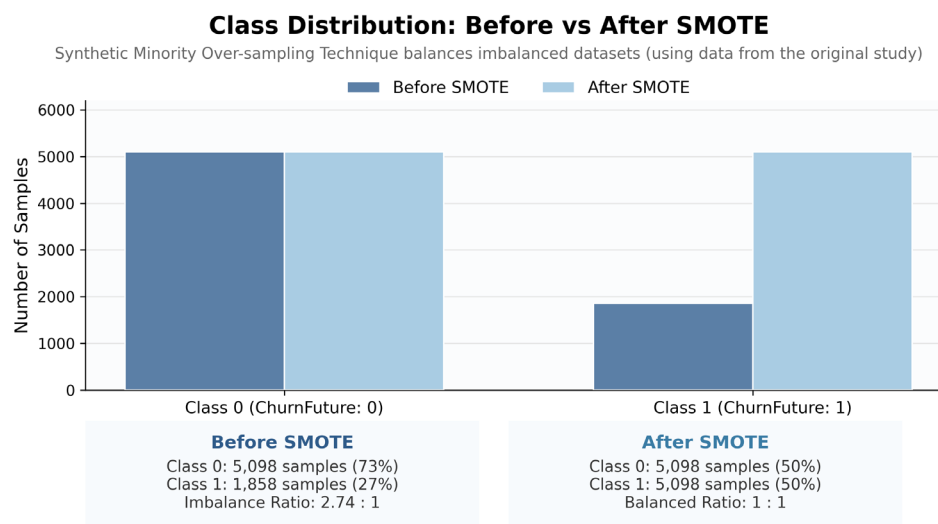


Figure 1. Training class distribution before and after SMOTE

Figure 2 illustrates the training class distribution before and after SMOTE application. The resampling process produced a more balanced class distribution, which may improve the model's ability to learn minority behavioral patterns during training.

3.2 Model Performance Evaluation

The performance of the proposed XGBoost model was evaluated using an independent testing dataset. Table 1 summarizes the classification performance metrics obtained from the independent testing dataset. The

model achieved an accuracy of 0.84, a precision of 0.71, a recall of 0.75, and an F1-score of 0.73. In addition, the ROC AUC value reached 0.898, indicating strong discrimination capability between churn and non-churn classes [24].

Table 1. Classification Performance Metrics

Metric	Value (95% CI)
Accuracy	0.84 (0.82–0.86)
Precision	0.71 (0.67–0.75)
Recall	0.75 (0.71–0.79)
F1-score	0.73 (0.70–0.76)
ROC AUC	0.898 (0.881–0.914)

Figure 3 presents the confusion matrix generated from the testing results. The model was able to correctly classify most non-churn observations while maintaining reasonable sensitivity in identifying churn cases. This result suggests that the proposed model can effectively detect inactive member patterns despite the behavioral variability present in the dataset.

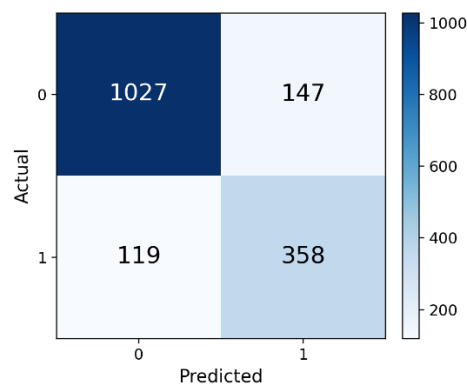


Figure 2. Confusion Matrix

Figure 4 illustrates the ROC curve of the proposed model. The curve demonstrates stable classification performance across multiple threshold settings and supports the high ROC AUC value obtained during evaluation. The obtained ROC AUC value is comparable with results reported in previous machine learning-based churn prediction studies, where ensemble and boosting approaches demonstrated strong classification performance on behavioral datasets [7], [8].

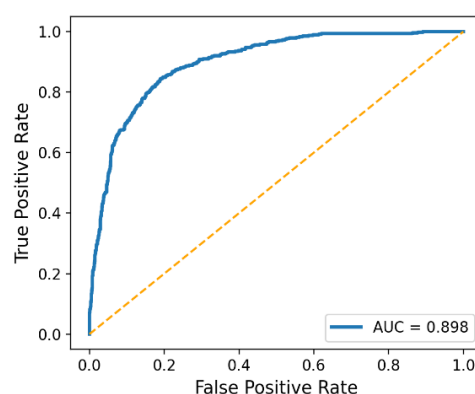


Figure 3. ROC curve of the XGBoost model

3.3 Behavioral Insight Analysis

To better understand churn behavior, an analysis of important predictive features was conducted. Figure 5 presents the feature importance generated by the XGBoost model. Among all evaluated variables, recency emerged as the most influential predictor, indicating that the duration since the last transaction plays a major role in identifying member inactivity patterns. Transaction frequency and several operational indicators also contributed to model performance, although with lower importance scores.

Several organizational-department variables also appeared among the top-ranked predictors; however, these mainly reflect local operational heterogeneity and are less generalizable than behavioral indicators such as recency and frequency. Gender-related effects should likewise be interpreted with caution, as they may reflect participation patterns specific to this dataset rather than causal behavioral differences.

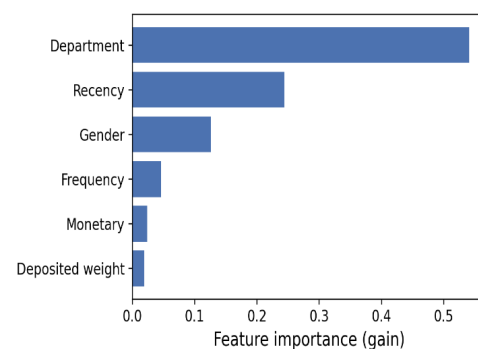


Figure 4. Feature importance of the XGBoost model

To further examine the influence of recency on churn behavior, Figure 6 illustrates the distribution of recency values between churn and non-churn members. The distribution shows that churned members generally exhibited substantially higher recency values compared with active members, indicating longer periods of inactivity before churn occurred.

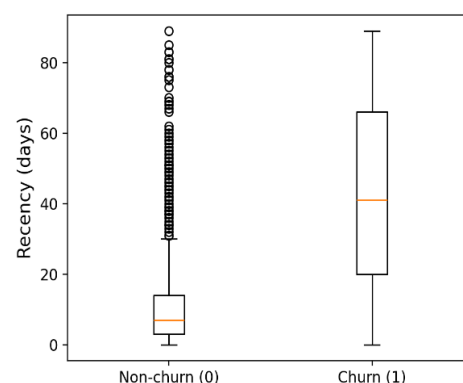


Figure 5. Recency distribution by churn status

The analysis also revealed that churned members tended to exhibit lower transaction frequency compared with non-churn members. Interestingly, several churned members still showed relatively high monetary values,

suggesting that churn behavior in waste bank programs is not solely associated with transaction volume. Instead, participation consistency appears to play a more important role in maintaining long-term engagement. These findings indicate that member inactivity is more strongly related to declining participation patterns than to transaction value alone. Therefore, recency and transaction frequency may serve as important behavioral indicators for identifying churn behavior in community-based waste management programs.

3.4 Churn Definition Validation

Figure 7 presents the distribution of intervals between consecutive deposits. The distribution is strongly right-skewed (median 7 days, 90th percentile 21 days), with 95.3%, 99.3%, and 99.8% of all intervals within 30, 90,

and 180 days. Active members therefore re-deposit well within three months, so inactivity beyond 90 days is a clear anomaly. Reconstructing the label under 30-, 90-, and 180-day windows (Figure 8) shows that a 30-day window inflates the churn rate to 34.9% by misclassifying short gaps, whereas 90 to 180 days changes it only marginally (27.6% to 25.4%). Crucially, ROC AUC stays stable across all definitions (0.891, 0.882, 0.879), confirming the result is driven by genuine behavioral signal, not an arbitrary threshold. The slightly lower values reported here arise because the sensitivity analysis uses datasets reconstructed under each alternative inactivity window with the behavioral numeric features, rather than the final model in Table 2; the comparison is intended only to demonstrate performance stability across thresholds.

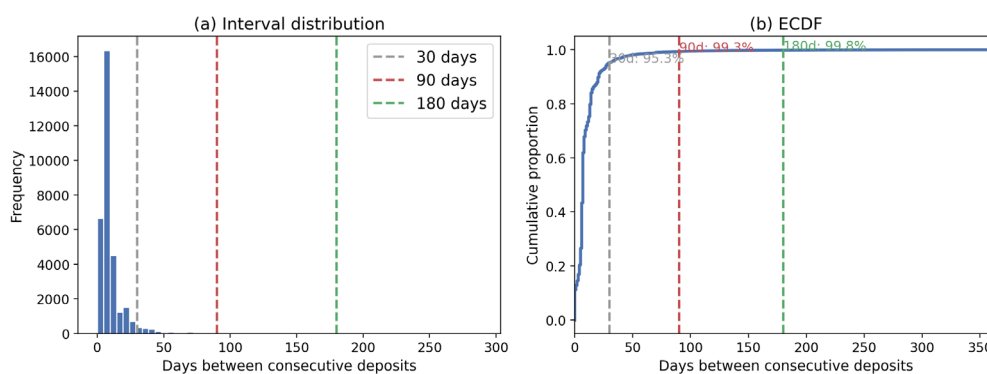


Figure 6. Distribution of intervals between consecutive deposits (histogram and ECDF) with the 30-, 90-, and 180-day thresholds marked.

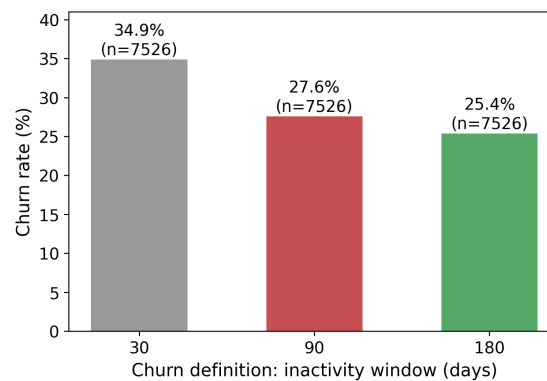


Figure 7. Churn rate under alternative inactivity windows (30, 90, and 180 days).

Table 2. Performance comparison of XGBoost against four baseline classifiers under an identical customer-level split and SMOTE setup.

Model	Acc.	Prec.	Recall	F1	ROC AUC	MCC	Bal. Acc.	Kappa
XGBoost	0.839	0.709	0.751	0.729	0.8985	0.615	0.813	0.615
LightGBM	0.835	0.702	0.746	0.724	0.8975	0.607	0.809	0.606
Random Forest	0.840	0.718	0.736	0.727	0.8963	0.614	0.809	0.614
Logistic Regression	0.843	0.750	0.686	0.716	0.8794	0.609	0.796	0.608
Decision Tree	0.826	0.670	0.782	0.721	0.8757	0.599	0.813	0.596

3.5 Comparison with Baseline Models

XGBoost was also compared against four baseline classifiers under an identical pipeline [25], [26]. As shown in Table 2, XGBoost attained the highest ROC AUC (0.8985), narrowly ahead of LightGBM (0.8975) and Random Forest (0.8963), together with the best F1

and MCC among the evaluated models; these boosting and ensemble methods clustered closely at the top and clearly outperformed the linear and single-tree baselines (Logistic Regression and Decision Tree, ROC AUC 0.879 and 0.876).

This consistent ranking confirms XGBoost is a well-justified and robust choice rather than dependent on a single algorithm. Therefore, XGBoost was retained because it achieved the best overall balance across ROC AUC, F1-score, and MCC. Statistical significance testing (Section 3.8) confirms that these performance differences are significant only against the linear and single-tree baselines.

3.6 Hyperparameter Optimization

Table 3 compares the configuration obtained from the randomized search against the configuration adopted in this study. Although the search identified a more complex configuration (maximum depth 5, learning rate 0.1, 500 estimators) that marginally increased the cross-validation AUC to 0.912, this configuration produced a lower test-set ROC AUC (0.889) and F1 (0.708) than the adopted configuration (ROC AUC 0.898, F1 0.729).

Table 3. Randomized-search configuration versus the adopted configuration on the held-out test set.

Configuration	CV ROC AUC	Test ROC AUC	Test F1
Randomized search	0.912	0.889	0.708
Adopted (this study)	0.904 $\pm$ 0.012	0.898	0.729

The five-fold cross-validation ROC AUC of the adopted configuration was 0.904 ± 0.012 . These results indicate that the reported hyperparameters already lie near the optimum and generalize better to unseen members, and that further tuning yields no practical benefit; the adopted configuration was therefore retained.

3.7 Extended Evaluation and Robustness

Because accuracy can be misleading under class imbalance, the model was further assessed using metrics designed for imbalanced problems. On the independent test set, XGBoost obtained a PR-AUC of 0.782, an MCC of 0.615, a Balanced Accuracy of 0.813, and a Cohen's Kappa of 0.615, all substantially above chance and consistent with the ROC-based results. Figure 9 shows the Precision-Recall curve, which remains well above the positive-class prevalence (0.289) across the operating range.

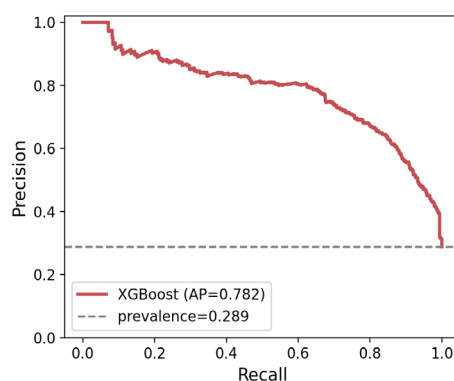


Figure 9. Precision-Recall curve of the XGBoost model on the independent test set.

To check for overfitting to the synthetic SMOTE samples, the real pre-oversampling training ROC AUC (0.927) was compared with the test value (0.898); the small gap of 0.028, together with a stable five-fold cross-validation ROC AUC of 0.904 ± 0.012 , confirms that the model generalizes to unseen members and was not inflated by oversampling.

3.8 Statistical Significance

To substantiate the performance claims, 95% confidence intervals (CIs) were estimated by bootstrap resampling with 2,000 replicates and are reported alongside the point estimates in Table 1. The proposed model attained a ROC AUC of 0.898 (95% CI 0.881 to 0.914) and an F1-score of 0.729 (95% CI 0.697 to 0.760). Pairwise comparisons using bootstrap AUC-difference tests and McNemar's test showed that XGBoost significantly outperformed Logistic Regression and Decision Tree (delta-AUC of 0.019 and 0.023, both $p < 0.001$), whereas the differences relative to LightGBM and Random Forest were not statistically significant ($p = 0.40$ and $p = 0.55$). XGBoost was therefore not statistically different from LightGBM and Random Forest, while remaining clearly superior to the linear and single-tree baselines. Its selection is justified by its consistently leading point estimates rather than by a claim of strict superiority over every model.

3.9 Robustness and Additional Analyses

Two further procedures assessed robustness. A temporal forward-chaining validation, in which the model was trained on earlier monthly snapshots and tested on the subsequent month, yielded a mean ROC AUC of 0.904 ± 0.031 across ten temporal folds, confirming stable performance under a realistic deployment scenario. A nested cross-validation (five outer folds with an inner randomized search) produced a ROC AUC of 0.916 ± 0.011 , slightly higher than the held-out test value of 0.898. Because the hyperparameter search used only the training data, this confirms that the reported results are not affected by optimistic bias. Because churn predictions are intended to support intervention decisions, the calibration of the predicted probabilities was examined.

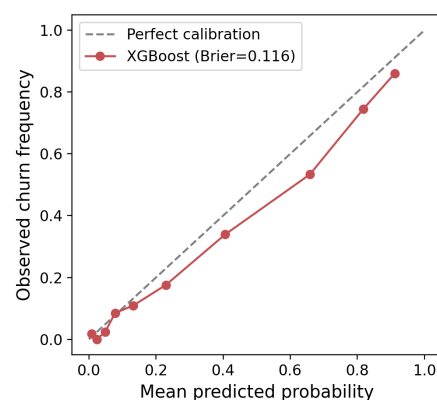


Figure 10. Calibration curve (reliability diagram) of the XGBoost model.

The model obtained a Brier score of 0.116, and the reliability diagram in Figure 10 shows that the predicted probabilities track the observed churn frequencies closely, indicating that the model is reasonably well calibrated for decision support.

To complement the gain-based feature importance, a model-agnostic permutation importance analysis was performed on the test set (Figure 11). Recency remained the dominant predictor by a wide margin, followed by organizational department and transaction frequency, whereas gender, monetary value, and deposited weight contributed marginally. This corroborates the behavioral interpretation that participation recency and consistency, rather than transaction value, drive churn. This ranking is consistent with the gain-based feature importance reported earlier (Figure 5).

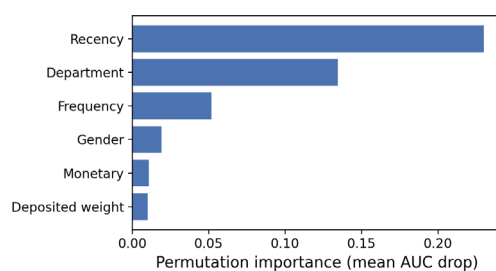


Figure 11. Permutation feature importance of the XGBoost model on the test set.

A subgroup analysis was conducted to check for demographic disparities. Discrimination performance was strong for both genders (ROC AUC 0.881 for male and 0.938 for female members), with no systematic performance degradation observed between the two groups, and generally consistent across the larger organizational departments, with wider variation only in small subgroups ($n < 30$) owing to limited sample sizes; gender should therefore be interpreted as a contextual rather than a causal factor. Multicollinearity among the numeric predictors was negligible, with all variance inflation factors below 3.5 (Recency 1.10, Frequency 1.64, Monetary 2.72, and deposited weight 3.42). Finally, the model is computationally lightweight: the final configuration comprises 250 trees with a maximum depth of 4, and on the Google Colaboratory CPU runtime training completed in 0.73 s on 10,196 balanced training samples while inference required 0.012 ms per member, making it suitable for routine operational deployment.

3.10 Discussion

These findings indicate that member engagement in waste banks exhibits behavioral patterns that can be effectively captured using ensemble learning. Although participation is driven by environmental and social motivations rather than commercial incentives, the resulting transaction behavior carries predictive signal comparable to that reported in commercial churn studies, where boosting approaches typically achieve ROC AUC values between 0.80 and 0.93 [7], [8], [14]. Behavioral churn analytics, traditionally applied in

profit-driven sectors, can therefore be meaningfully extended to community-based environmental programs.

Although participation motivations differ from commercial sectors, customer engagement is still reflected in observable transaction patterns such as recency and frequency. Ensemble and boosting methods capture the complex nonlinear relationships in such structured behavioral data [27], [28], and their effectiveness for churn prediction has been reported across domains [29].

Notably, monetary value was not consistently associated with retention: several churned members had previously shown high transaction values. Retention thus appears governed by behavioral consistency and routine participation rather than transaction amount [5], [14].

These patterns can be read from a community-participation perspective. The dominance of recency suggests participation is largely habit-driven: once the routine of depositing is interrupted, resumption becomes unlikely, mirroring habit-discontinuation effects in pro-environmental behavior. The role of frequency is consistent with regular participation reflecting sustained environmental engagement. The weak link between monetary value and retention is especially informative: unlike commercial settings, a member may deposit a large quantity once yet still disengage, because participation is driven by routine, convenience, and intrinsic motivation rather than economic value. This helps explain why commercial churn models may not transfer directly to community-based programs.

Methodologically, customer-level train-test separation with snapshot-based construction prevents records from the same member appearing in both training and testing, reducing information leakage [15], [16]. The evaluation therefore reflects realistic operation, where churn must be predicted for previously unseen members from historical data.

Practically, the framework shows the value of predictive analytics in operational waste banks: indicators such as recency and frequency can help administrators identify members with declining engagement before prolonged inactivity, supporting more data-driven and sustainable participation management.

In practice, the framework can be embedded in the existing spreadsheet-based waste-bank information system with minimal overhead. A monthly batch job would recompute each member's behavioral features and churn probability, and a simple dashboard could flag members using tiered, recency-based alerts (for example, low, medium, and high risk). Administrators could then trigger graduated interventions, such as reminder messages, home pick-up offers, or community outreach, prioritizing high-risk members before prolonged inactivity occurs. Because the model is lightweight and uses only routinely recorded

transaction fields, it can be retrained periodically without specialized infrastructure, enabling continuous participation monitoring.

Several external factors were not represented in the model and warrant acknowledgement: fluctuations in recyclable-material prices, government policies and education campaigns, household economic conditions and relocation, and seasonal effects such as holidays or rainy-season disruptions may all influence participation. Because the model relies solely on internal transaction records, these drivers are not captured and may explain part of the unexplained variation. The findings should therefore be interpreted with caution, as they derive from a single municipal system; external validation across multiple waste banks in different regions is required before broad transferability can be claimed.

The temporal forward-chaining validation reported in Section 3.9 partially mitigates this concern by confirming stable performance across time; however, it does not substitute for external validation on independent waste-bank systems, which remains an essential direction for future work.

Future studies may extend this work by incorporating explainable artificial intelligence approaches, temporal behavioral modeling, rule-based risk interpretation techniques, and multi-source behavioral or socio-economic data integration to improve model interpretability and generalizability.

4. Conclusions

This study proposed a machine learning-based approach for customer churn prediction in waste bank programs using historical transaction data from the Environmental Agency of Padang City. A snapshot-based dataset construction combined with customer-level train-test separation was applied to provide a more realistic evaluation scenario while reducing the risk of information leakage.

The experimental results showed that the XGBoost model combined with SMOTE achieved strong predictive performance, with an accuracy of 0.84 and a ROC AUC value of 0.898. The findings also revealed that recency and transaction frequency were among the most influential behavioral indicators associated with churn behavior, suggesting that declining participation consistency is more strongly related to inactivity than transaction value alone.

This study extends churn prediction analytics into the waste bank domain, an area that remains largely underexplored compared with commercial sectors. The results demonstrate that behavioral transaction data can support participation monitoring and data-driven decision-making in community-based waste management systems.

The contribution of this study is threefold. Methodologically, it introduces a leakage-aware,

customer-level evaluation design with snapshot-based feature construction that avoids the over-optimistic estimates of row-level splitting. Practically, it provides a participation-monitoring framework in which routine indicators such as recency and frequency help administrators flag members with declining engagement before prolonged inactivity. From a domain perspective, it characterizes behavioral churn patterns specific to community-based programs, most notably the decoupling of monetary value from retention. Together these extend churn analytics into a previously underexplored non-commercial domain rather than merely applying existing commercial models. These findings demonstrate that leakage-aware machine learning can support operational decision-making in community-based environmental programs beyond conventional commercial churn applications.

However, this study was limited to a single waste bank environment and did not incorporate external behavioral or socio-economic factors. Future studies may explore explainable artificial intelligence and rule-based risk interpretation to strengthen operational decision support.

Data and Code Availability. The analysis code and a de-identified dataset that reproduce all results reported in this study are publicly available at <https://github.com/Indri-TI/waste-bank-churn-prediction>. The raw transaction records contain personal information and are owned by the Environmental Agency of Padang City; they therefore cannot be shared publicly, and requests for access may be directed to the corresponding author subject to approval by the Agency.

References

- [1] Dinas Lingkungan Hidup Kota Padang, Dokumen Informasi Kinerja Pengelolaan Lingkungan Hidup Daerah (DIKPLHD) Kota Padang 2025 [Environmental Performance Information Document of Padang City 2025]. Padang, Indonesia, 2025.
- [2] S. Dhokhikah, Y. Trihadiningrum, and S. Sunaryo, "Community participation in household solid waste reduction in Surabaya, Indonesia," *Resources, Conservation and Recycling*, vol. 102, pp. 153–162, 2015. <https://doi.org/10.1016/j.resconrec.2015.06.013>.
- [3] A. Asteria and H. Heruman, "Bank sampah sebagai alternatif strategi pengelolaan sampah berbasis masyarakat di Tasikmalaya [Waste banks as an alternative strategy for community-based waste management in Tasikmalaya]," *Jurnal Manusia dan Lingkungan*, vol. 23, no. 1, pp. 136–141, 2016. <https://doi.org/10.22146/jml.18783>.
- [4] A. Kaza, L. Yao, P. Bhada-Tata, and F. Van Woerden, *What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050*. Washington, DC, USA: World Bank, 2018. <https://doi.org/10.1596/978-1-4648-1329-0>.
- [5] N. Ferronato and V. Torretta, "Waste mismanagement in developing countries: A review of global issues," *International Journal of Environmental Research and Public Health*, vol. 16, no. 6, p. 1060, 2019. <https://doi.org/10.3390/ijerph16061060>.
- [6] I. Rachman, "Community participation on waste bank sustainability for household solid waste management," *Journal of Environmental Science and Sustainable Development*, vol. 4, no. 2, 2021. <https://doi.org/10.7454/jessd.v4i2.1123>.
- [7] W. Verbeke, D. Martens, C. Mues, and B. Baesens, "Building comprehensible customer churn prediction models with advanced rule induction techniques," *Expert Systems with*

- Applications, vol. 38, no. 3, pp. 2354–2364, 2011. <https://doi.org/10.1016/j.eswa.2010.08.023>.
- [8] A. Idris, A. Khan, and Y. S. Lee, “Intelligent churn prediction in telecom: employing mRMR feature selection and RotBoost based ensemble classification,” *Applied Intelligence*, vol. 39, no. 3, pp. 659–672, 2013. <https://doi.org/10.1007/s10489-013-0440-x>.
- [9] 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. <https://doi.org/10.1109/ACCESS.2024.3402092>.
- [10] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. Chatzisavvas, “A comparison of machine learning techniques for customer churn prediction,” *Simulation Modelling Practice and Theory*, vol. 55, pp. 1–9, 2015. <https://doi.org/10.1016/j.simpat.2015.03.003>.
- [11] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
- [12] A. Ahmad, A. Jafar, and K. Aljoumaa, “Customer churn prediction in telecom using machine learning in big data platform,” *Journal of Big Data*, vol. 6, no. 1, 2019. <https://doi.org/10.1186/s40537-019-0191-6>.
- [13] A. M. Hughes, *Strategic Database Marketing: The Masterplan for Starting and Managing a Profitable Customer-Based Marketing Program*, 4th ed. New York, NY, USA: McGraw-Hill, 2011.
- [14] A. Keramati, R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, B. Mozaffari, and U. Abbasi, “Improved churn prediction in telecommunication industry using data mining techniques,” *Applied Soft Computing*, vol. 24, pp. 994–1012, 2014. <https://doi.org/10.1016/j.asoc.2014.08.041>.
- [15] S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, “Leakage in data mining: Formulation, detection, and avoidance,” *ACM Transactions on Knowledge Discovery from Data*, vol. 6, no. 4, pp. 1–21, 2012. <https://doi.org/10.1145/2382577.2382579>.
- [16] S. Kapoor and A. Narayanan, “Leakage and the reproducibility crisis in machine-learning-based science,” *Patterns*, vol. 4, no. 1, p. 100804, 2023. <https://doi.org/10.1016/j.patter.2023.100804>.
- [17] 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. <https://doi.org/10.1613/jair.953>.
- [18] H. He and E. A. Garcia, “Learning from imbalanced data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009. <https://doi.org/10.1109/TKDE.2008.239>.
- [19] S. Arlot and A. Celisse, “A survey of cross-validation procedures for model selection,” *Statistics Surveys*, vol. 4, pp. 40–79, 2010. <https://doi.org/10.1214/09-SS054>.
- [20] 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. <https://doi.org/10.1016/j.ipm.2009.03.002>.
- [21] T. Fawcett, “An introduction to ROC analysis,” *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006. <https://doi.org/10.1016/j.patrec.2005.10.010>.
- [22] S. J. Haddadi, A. Farshidvard, F. dos S. Silva, J. C. dos Reis, and M. dos S. Reis, “Customer churn prediction in imbalanced datasets with resampling methods: A comparative study,” *Expert Systems with Applications*, vol. 246, p. 123086, 2024. <https://doi.org/10.1016/j.eswa.2023.123086>.
- [23] 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, p. 54, 2023. <https://doi.org/10.3390/info14010054>.
- [24] B. Krawczyk, “Learning from imbalanced data: open challenges and future directions,” *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, 2016. <https://doi.org/10.1007/s13748-016-0094-0>.
- [25] V. Chang, K. Hall, Q. A. Xu, F. O. Amao, M. A. Ganatra, and V. Benson, “Prediction of customer churn behavior in the telecommunication industry using machine learning models,” *Algorithms*, vol. 17, no. 6, p. 231, 2024. <https://doi.org/10.3390/a17060231>.
- [26] T. E. Yimer, “Mitigating class imbalance in churn prediction with ensemble methods and SMOTE,” *Scientific Reports*, vol. 15, art. no. 16256, 2025. <https://doi.org/10.1038/s41598-025-01031-0>.
- [27] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. <https://doi.org/10.1023/A:1010933404324>.
- [28] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001. <https://doi.org/10.1214/aos/1013203451>.
- [29] G. Zhu, B. Wang, X. Liu, B. Song, and Z. Wang, “Customer churn prediction using improved balanced random forests,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 5445–5449, 2009. <https://doi.org/10.1016/j.eswa.2008.06.121>.