algorithms

MDPI

Article

A Machine Learning-Driven CRM Approach for Identifying Member Churn in a Brazilian Agro-Industrial Cooperative: A Practical Case Study

Sergio Akio Tanaka [1,*], João Vitor da Costa Andrade [1], Alessandro Botelho Bovo [2], Attilio Converti [3], Danilo Sipoli Sanches [2] and Hugo Valadares Siqueira [1,*]

1 Department of Electrical Engineering, Federal University of Technology—Paraná (UTFPR), Ponta Grossa Campus, Ponta Grossa 84016-210, PR, Brazil; joaovitorandrade@alunos.utfpr.edu.br
2 Department of Electrical Engineering and Computer Science, Federal University of Technology—Paraná (UTFPR), Londrina Campus, Londrina 86036-370, PR, Brazil; alessandrobovo@utfpr.edu.br (A.B.B.); danilosanches@utfpr.edu.br (D.S.S.)
3 Department of Civil, Chemical and Environmental Engineering, University of Genoa, 16145 Genoa, Italy; converti@unige.it
* Correspondence: sergiotanaka@alunos.utfpr.edu.br (S.A.T.); hugosiqueira@utfpr.edu.br (H.V.S.)

Abstract

This study addresses member churn in a Brazilian agro-industrial cooperative by operationalizing a leakage-aware, governance-aligned machine-learning protocol within the organization's Customer Relationship Management (CRM) system. Using real-world CRM data under confidentiality constraints, we followed a KDD-based workflow. This workflow includes: (i) multi-source integration; (ii) targeted preprocessing with explicit handling of severe class imbalance via undersampling; (iii) a unified validation scheme with stratified cross-validation, hyperparameter search, and controlled AutoML benchmarking; (iv) comparison of tabular learners (Random Forest, XGBoost, and Support Vector Classifier) and a voting ensemble; and (v) SHAP-based explainability to support transparent decision-making. Class rebalancing substantially improved minority-class performance; for instance, the "Inactive" recall increased from 0.27 to 0.74 with SVC. Across ten folds, AutoML achieved competitive mean ROC-AUC (0.8844), followed by XGBoost (0.8690) and Random Forest (0.8660); global metrics supported operational feasibility (accuracy 0.79–0.80; ROC-AUC up to 0.8876), while the ensemble delivered comparable discrimination (ROC-AUC 0.8845) with a modest precision gain. SHAP analyses yielded business-coherent drivers and enabled actionable, instance-level communication in the CRM. The resulting microservices-based module exposes ranked churn propensities and explanations in dashboards for risk stratification and prioritization of retention actions. Overall, the work provides an interpretable, reproducible, and production-ready methodological blueprint for predictive CRM in seasonal cooperative environments under governance and confidentiality constraints.

Keywords: agricultural cooperatives; AutoML; class imbalance; CRM integration; customer churn prediction; ensemble methods; machine learning; SHAP; stratified cross-validation

Check for updates

Academic Editors: Carolina Del Valle Soto and Ramiro Velázquez

Received: 10 February 2026
Revised: 20 February 2026
Accepted: 25 February 2026
Published: 27 February 2026

1. Introduction

Brazilian agribusiness plays a central role in the national economy. In 2024, it accounted for approximately 23.5% of gross domestic product (GDP) [1]. In Paraná state, this prominence is reinforced by its leadership in soybean production [2] and in poultry production and tilapia aquaculture [3]. In this context, agro-industrial cooperatives coordinate producers, infrastructure, and services, integrating activities across the value chain [4]. However, this cooperative-based model is vulnerable to member churn—members exiting or becoming inactive—which erodes margins, service delivery, and investment capacity [5], underscoring the need for proactive monitoring and targeted retention interventions.

In this setting, CRM consolidates relational data—such as service records, purchases, credit, and technical assistance—providing the foundation for predictive churn analytics and the prioritization of retention actions. For structured CRM churn data, machine learning (ML) methods capture predictive signals [5–9], with widely used models including Random Forest (RF), XGBoost, and Support Vector Machine (SVM), as well as stacking ensembles that combine model predictions [10,11]. In line with contemporary practice, automated machine learning (AutoML) supports end-to-end modeling pipelines by automating preprocessing, addressing class imbalance, and tuning hyperparameters [12,13]. Additionally, SHAP-based interpretability provides model-agnostic local and global explanations grounded in Shapley value theory [14,15], with prior applications in churn scenarios [16].

Although churn prediction and predictive CRM have been extensively investigated in mature sectors such as telecommunications and banking, the systematic integration of AI-driven churn modeling within operational CRM systems for agro-industrial cooperatives remains insufficiently explored in the literature. Specifically, there is a lack of empirical studies reporting end-to-end, production-oriented CRM–ML pipelines that encompass data governance, preprocessing, model training, deployment on real cooperative data. This research gap is particularly critical in the context of agro-industrial cooperatives in Paraná, where seasonality, structural heterogeneity among members, and institutional governance constraints limit the direct transferability of models and practices developed for other industries.

Against this backdrop, a practical gap remains in churn prediction for agro-industrial cooperatives in Paraná [6,7]. To address this gap, we design and evaluate an integrated CRM–ML system that combines AutoML and explainable AI (XAI) via SHAP using real cooperative data to deliver consistent and interpretable churn risk estimates and support dashboard-driven managerial interventions.

The contributions of this study are methodological and operational, with an explicit emphasis on real-world evidence in an underexplored domain. Methodologically, we formalize a leakage-aware and governance-aligned protocol for churn prediction in seasonal, cooperative-based CRM systems, integrating class-imbalance mitigation, controlled AutoML benchmarking, ensemble learning, and post hoc explainability within realistic organizational constraints.

Operationally, we instantiate this protocol using real-world data from a Brazilian agro-industrial cooperative and integrate risk scores and explanations into CRM dashboards to support retention decision-making. This study does not introduce a novel machine learning algorithm; rather, it provides empirical evidence on the extent to which consolidated churn–CRM practices—predominantly validated in mature sectors such as telecommunications and banking—generalize to the cooperative agro-industrial context, where the academic evidence base remains limited and fragmented.

2. Related Work

Predictive CRM leverages historical relational data and supervised machine learning models to estimate churn risk and support proactive retention strategies [16,17]. Traditional rule-based approaches grounded in classical statistical frameworks (e.g., generalized linear models) are often limited in capturing the multivariate and nonlinear dynamics of member behavior, motivating the adoption of modern ML techniques, including bagging and boosting ensembles [9].

Related work shows that churn analytics has been predominantly developed and benchmarked in mature markets such as telecommunications and banking, frequently relying on ensemble learners and large-scale transactional data [5,8,11,18]. By comparison, studies explicitly centered on agro-industrial cooperatives remain scarce and are often reported as isolated contributions or in adjacent rural/cooperative settings [19,20]. This imbalance makes it important to examine whether consolidated churn–CRM practices from other domains generalize to the cooperative agro-industrial context, particularly when evaluated with real-world data and under practical organizational constraints that can affect both churn operationalization and model behavior.

Prior work demonstrates that CRM can transition from a reactive to a proactive approach when supervised classification models are adopted. When combined with explainability methods (e.g., SHAP), these models facilitate the identification of members prone to churn as well as the drivers underlying the predicted risk [21]. The incorporation of explainable artificial intelligence (XAI) enhances managerial trust in model-assisted decisions by increasing transparency and auditability [22]. Overall, the maturity of predictive CRM is closely tied to the convergence of big data analytics, statistical modeling, and machine learning, enabling preventive retention actions and value-oriented customer segmentation [16,17,23].

Moreover, recent empirical evidence reinforces these advances. The authors in [17] report that RF, SVM, logistic regression (LR), and Gradient Boosting (GB) provide robust decision-support frameworks that balance portfolio value, acquisition costs, and retention; for example, recall = 0.97 at the selected threshold and ROC-AUC = 0.80. From an organizational perspective, adopting predictive CRM reduces costs, increases loyalty, and enhances evidence-based decision-making, positioning CRM as a core component of data-driven management [21,24].

The literature tends to favor ensemble and hybrid models as state-of-the-art approaches for churn prediction, as they combine performance, robustness, and, in some settings, interpretability [5,11,18]. Under conditions of class imbalance, RF is often cited for its stability and generalization [8]. Effectiveness, however, should be assessed not only by predictive accuracy but also by computational cost and interpretability. Comparative analyses [25] examine XAI methods, such as TreeSHAP and KernelSHAP, and report a favorable trade-off between computational cost and explanatory depth and faithfulness. Evidence [26] shows that feature selection and resampling strategies (e.g., Synthetic Minority Over-sampling Technique (SMOTE), oversampling, undersampling) enhance performance—and, in some cases, improve interpretability—in severely imbalanced scenarios.

In this study, stratified k-fold cross-validation is employed as an internal independent and identically distributed (i.i.d.) evaluation protocol, providing a reproducible baseline for model comparison under the available organizational data. Nevertheless, for CRM decision support that unfolds over time and across branches or seasons, external validation is essential; therefore, temporal validation (e.g., season-based holdout) and unit-based validation (e.g., branch/filial holdout), when available, are defined as a priority direction for future work to strengthen external validity.

2.1. Ensemble Learning and Explainable AI (XAI)

Ensemble learning is a robust and well-established approach to churn prediction, leveraging complementary strengths of multiple models to improve accuracy, stability, and generalization [9]. In comparative analyses, XGBoost and RF often outperform traditional algorithms in terms of accuracy, precision, recall, and F1 score, with top features—domaindependent but often including contract duration, plan type, and monthly expenditure—reported in prior studies [18]. Class imbalance, a typical characteristic of churn datasets, can be mitigated using ratio-based data balancing that adaptively combines oversampling and undersampling. This approach can yield performance gains for XGBoost compared with SMOTE [7].

Building on these ensemble-based and XAI advances, recent approaches seek to combine high predictive performance with local and global explainability. XAI Churn TriBoost—built on XGBoost, CatBoost, and LightGBM—demonstrates strong performance (ROC-AUC, PR-AUC) and provides individual-level interpretations that support retention decision-making via SHAP and Local Interpretable Model-agnostic Explanations (LIME) [27]. However, feature importance is descriptive rather than prescriptive. Feasible counterfactual (recourse) explanations—as in Diverse Counterfactual Explanations for ML (DiCE-ML) formulated via mixed-integer linear programming (MILP)—identify minimal, feasible profile changes required to alter the model's prediction [28]. Computationally, TreeSHAP and KernelSHAP remain widely used choices for post hoc interpretability with tractable computational cost, which motivates hybrid, budget-aware strategies that selectively compute explanations for critical instances [25,26]. Taken together, these advances underscore practical trade-offs between performance, interpretability, and computational cost in churn modeling.

Beyond ensemble-based XAI, causal inference methods—e.g., the R-learner—can complement SHAP to broaden interpretive insight: the R-learner estimates putative causal effects of key variables on churn (under standard identification assumptions), whereas SHAP provides associational explanations (both local and global), jointly offering a more prescriptive basis for retention actions [23]. In addition, studies emphasize the importance of algorithmic fairness audits (e.g., disparate impact, equalized odds, calibration) for detecting and mitigating bias, thereby enhancing the reliability and ethical sustainability of AI within [22]. Finally, explainable ensemble methods have also been applied in adjacent domains, such as blockchain fraud detection, illustrating their relevance to complex, high-dimensional data settings [29].

2.2. Machine Learning for CRM Churn Prediction

In agro-industrial cooperatives, churn prediction helps diagnose sources of dissatisfaction and prioritize data-driven retention actions that strengthen member–cooperative relationships and support organizational sustainability [17,21]. In mature domains such as telecommunications and banking, gradient-boosting models—particularly XGBoost—have demonstrated high accuracy, stability, and generalizability on structured tabular data with nonlinear relationships [9,23,30]. In these applications, predictive analytics helps anticipate needs and transactional behavior, personalize offers, and improve satisfaction and loyalty [31].

Beyond these sectors, empirical studies underscore the significance of ML in rural and cooperative settings. For example, Ref. [19] reports that RF outperforms random-effects logit for membership prediction, with out-of-sample, leak-free evaluation using a stratified holdout and out-of-bag estimation. From a technological standpoint, integrating edge computing and deep learning enables distributed, real-time inference, reducing latency and supporting predictive risk management in cooperative financial services [20]. Taken together, the combination of predictive accuracy, explainability, and algorithmic governance (e.g., fairness auditing and monitoring) increasingly positions customer retention analytics as a strategic component of modern CRM, aligning with XAI principles and supported by empirical evidence.

To synthesize these contributions across domains, Table 1 summarizes representative studies that combine ensemble learning and Explainable AI for churn prediction. Most works focus on telecom and banking datasets, relying on tree-based ensembles (RF, XGBoost, LightGBM, CatBoost) in conjunction with post hoc explanation methods, such as SHAP and Local Interpretable Model-agnostic Explanations (LIME). In contrast, relatively few studies address rural or cooperative contexts, which often emphasize predictive performance over structured explainability and governance. Overall, the reviewed evidence indicates that most AI + CRM churn studies are concentrated in telecom and banking, while agro-industrial cooperative CRM settings are rarely addressed. Moreover, the literature typically reports predictive performance but less frequently documents operational CRM integration, governance constraints, and deployment artifacts (e.g., microservices and dashboards) required for real-world adoption. Therefore, this study contributes by filling the AI–CRM–agro intersection with an end-to-end, deployable, and explainable pipeline grounded in real cooperative data.

Table 1. Studies on churn prediction with ensemble learning and Explainable AI.

Year	Reference	Dataset	Description of Methods and Results
2018	[32]	Telecom (3333 records)	Compared ten classifiers, including DT, LR, k-NN, Naïve Bayes, SVM, MLP, RF, AdaBoost, and stochastic gradient boosting. RF and AdaBoost achieved the highest accuracy (≈96%).
2022	[8]	Telecom (7043 records)	Applied XGBoosted DTs to customer churn prediction and reported higher accuracy than other evaluated learning models.
2023	[21]	Systematic review (2017–2022)	Proposed a proactive framework combining ML and XAI (LIME and SHAP), with real-time integration and explicit class-imbalance treatment.
2023	[16]	Telecom (50,137 customers; 55 variables)	Assessed multiple models using LIME and SHAP; RF and XGBoost delivered the best interpretable performance.
2024	[17]	–	Applied DT, RF, SVM, and GB. RF reported ROC-AUC = 0.878 and F1 = 0.766.
2024	[33]	Telecom	Evaluated ten classifiers, including LR, SVM, KNN, NB, RF, and XGBoost, with ten-fold cross-validation. RF achieved AUC = 0.85.
2024	[7]	–	Introduced Ratio-Based Data Balancing, improving the performance of XGBoost and GB.
2024	[5]	Telecom	Compared DT, RF, and boosted trees. RF achieved 91.66% accuracy and 82.2% precision and was interpreted using SHAP and LIME.
2024	[34]	–	Proposed an Ensemble Fusion model combining RF, XGBoost, and CatBoost, achieving AUC $>$ 0.97 and precision of 96%.
2024	[11]	Telecom	Evaluated DT, LR, RF, LightGBM, XGBoost, and an artificial neural network; XGBoost with oversampling achieved the highest accuracy (≈0.80), and tree-based ensembles outperformed the neural network.
2025	[24]	Banking	Developed a voting ensemble with random forest, XGBoost, and gradient boosting, achieving 94% accuracy with stability across samples.
2025	[23]	Bank credit	Combined XGBoost with SHAP and causal inference using the R-learner. Reported AUC = 0.97 and precision = 0.96.
2025	[25]	Telecom (7043 customers)	Compared LIME, SHAP, ELI5, and TreeSHAP; TreeSHAP was 331$\times$ more computationally efficient than ELI5.
2025	[26]	–	Used Boruta, genetic algorithms, and CNNs for feature selection, combined with SMOTE and undersampling. Accuracy improved by 12% over baseline models.
2025	[22]	Telecom (50,137 customers)	Evaluated eleven models, including LR, RF, XGBoost, LightGBM, and CatBoost. LightGBM achieved 73.08% accuracy, with SHAP and LIME used for model interpretation.
2024	[29]	Blockchain	Applied random forest, XGBoost, and LightGBM with SHAP and LIME, achieving 95% accuracy and providing evidence of transferability to churn-prediction contexts.
2025	[27]	–	Proposed XAI Churn TriBoost combining XGBoost, CatBoost, and LightGBM, achieving 96.44% accuracy and F1 = 0.90.

Table 1. *Cont.*

Year	Reference	Dataset	Description of Methods and Results
2025	[28]	–	Introduced DiCE-ML based on MILP for recourse-oriented counterfactual explanations. Reported recall = 0.72.
2025	[18]	Multi-sector settings (telecom, finance, retail)	Compared XGBoost, gradient boosting, and random forest. XGBoost achieved accuracy = 0.93 and F1 = 0.91 across sectors.
2025	[9]	240 studies (bibliometrics) and 61 studies (qualitative analysis)	Showed that ensemble methods such as XGBoost, LightGBM, and CatBoost remain among the most effective approaches for structured churn data.
2023	[20]	Agricultural cooperatives in China	Integration of edge computing and deep learning in cooperative financial services increased operational efficiency, reduced latency, and improved predictive risk management.
2023	[19]	Agricultural cooperatives in Zambia	Used a random forest model to represent and explain dairy producers' decision-making, achieving 76.66% accuracy.
2025	[31]	30 clients and 20 managers of cooperative banks in India	Employed structured questionnaires and semi-structured interviews; SPSS version 29 analysis with correlation and regression ($R^2 = 0.56$) confirmed the Technology Acceptance Model. Perceived usefulness and ease of use affected satisfaction, while barriers included low digital literacy, privacy concerns, and limited language support.

Source: Authors' own elaboration.

3. Materials and Methods

This study presents the methodological framework for predicting member churn using real Customer Relationship Management (CRM) data from an agro-industrial cooperative in the state of Paraná, Brazil. The Knowledge Discovery in Databases (KDD) cycle was adopted in five stages: (i) Selection, (ii) Preprocessing, (iii) Transformation, (iv) Mining, and (v) Evaluation. The overall workflow is shown in Figure 1. In this study, KDD defines the end-to-end macro-workflow, whereas the Data Mining stage is operationalized using CRISP-DM as an internal, iterative protocol for model development and assessment, maintaining alignment with business and data understanding. Each stage specifies inputs, procedures, outputs, and validation criteria. The design of the system that operationalizes the analytical pipeline is then integrated through BPMN and UML.

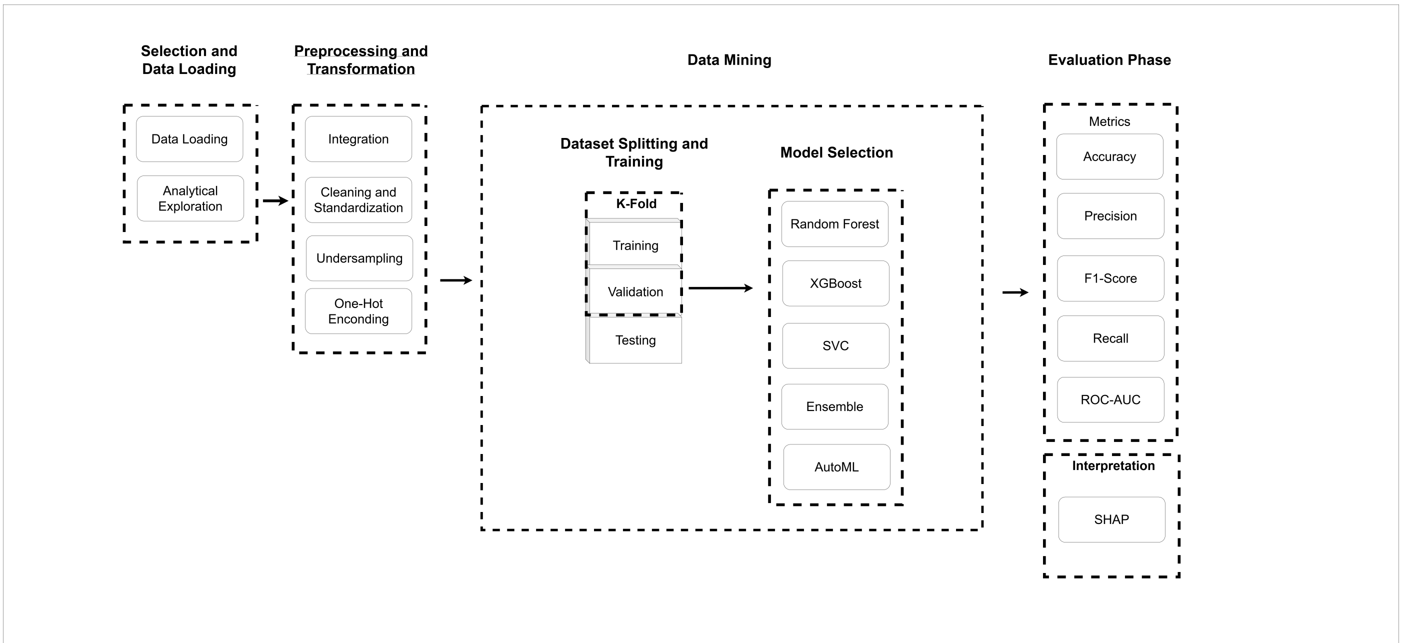


Figure 1. Methodological workflow based on the KDD cycle.

3.1. Selection

Data selection was conducted by the cooperative, which, following governance and privacy policies, provided a de-identified dataset with sensitive fields anonymized for safe sharing and reuse. The dataset metadata are publicly listed on Zenodo; however, file access is restricted and granted only upon reasonable request, subject to the cooperative's authorization and data-use conditions. The cooperative authorized the use of these secondary CRM records for research under its data-governance framework. Prior to access, direct identifiers were removed and only a minimized set of variables was shared.

The information is distributed across two datasets, totaling 15,034 members (9028 active), 112 columns, and 524,005 rows:

- `anonimized_data`: registry data for identification and profiling, 81 columns and 139,587 rows;
- `classif_prod_crm`: financial transactions with historical records of economic participation, 31 columns and 384,418 rows.

Selected Dataset

Table 2 presents the glossary of variables used in the modeling pipeline, concisely describing each column's role, data type, and operational meaning in the CRM context. For confidentiality reasons, some columns may have been removed and certain names may have been altered; therefore, the information reported here approximates the original analytical function and may not be identical to production labels. This process preserves attribute semantics and statistical usefulness, though small differences in naming or granularity may occur.

Table 2. Column glossary.

Column	Type	Description
`client_id`	key	Member ID in the CRM.
`producer_code`	key	Alternative producer ID.
`family_group_id`	key	Family/group ID in the CRM.
`group_id`	derived	Group ID used in modeling (derived from `family_group_id` or `producer_code`).
`deactivation_datetime`	datetime	Deactivation date and time.
`activation_datetime`	datetime	Activation date and time.
`reference_flag`	categorical	Reference indicator used for standardization.
`contract_events_flag`	boolean	Indicator of contractual events.
`check_flag`	boolean	Check indicator.
`protest_flag`	boolean	Protest indicator.
`lawsuit_flag`	boolean	Legal action indicator.
`inss_withholding_active`	boolean	INSS withholding is active.
`focus_client`	boolean	Focus customer marker.
`uses_member_portal`	boolean	Uses the member/cooperative portal.
`pet_segment_flag`	boolean	Pet segment indicator.
`chose_declaration`	boolean	Opted for declaration.
`classification_valid_date`	date	Classification validity date.
`own_property_flag`	boolean	Owns property.
`printed_signed_doc`	boolean	Printed/signed document present.
`producer_dap_flag`	boolean	Producer DAP/AF registration.
`activity_segment`	categorical	Activity segment.
`structure_size`	categorical	Structure or size/scale.
`producer_type`	categorical	Producer type.
`fiscal_regime`	categorical	Tax/fiscal regime.
`inss_withholding_type`	categorical	INSS withholding type.
`producer_classification`	categorical	Producer classification.
`registration_date`	date	Registration date.

Source: Authors' elaboration based on the original dataset; some fields may be removed or renamed for confidentiality.

Methodologically, the data dictionary establishes a data contract that guides preprocessing: keys must be treated as non-predictive identifiers to avoid leakage; boolean variables can be used as-is; categorical features require suitable encoding and, when pertinent, semantic ordering; temporal attributes demand granularity normalization and potential derivation of seasonal features. Due to confidentiality, some operational examples were generalized and certain fields may have been aggregated or binarized to reduce re-identification risk; these adaptations do not compromise statistical interpretation but should be considered when comparing results across environments. By making explicit the derived nature of `group_id` and the operational meaning of indicators, the table supports reproducibility, feature-engineering auditability, and comparability across experiments, ensuring transparency and consistency throughout the analytical pipeline even under confidentiality constraints.

3.2. Preprocessing

The process began with an integrity diagnosis to establish a baseline and assess the quality of the raw data. A minimal model was executed without imputation or feature selection, with the sole purpose of verifying the need for a rigorous data cleaning protocol. The following were observed: (i) severe target imbalance $\sim$ 87.2% active members; (ii) high rates of missingness, for example `deactivation_datetime` $\sim$ 99.9% and `activation_datetime` $\sim$ 80%; and (iii) constant columns. Artificially elevated metrics, with AUC values near 1.0, indicated potential data leakage. To confirm this, a control experiment with shuffled labels was performed and produced random performance, which supported the need for the preprocessing protocol.

Procedures:

I. **Integration**: joining `anonimized_data` to `classif_prod_crm` through `client_id/producer_code`, with priority given to `classif_prod_crm`. Members directly linked to the cooperative were filtered using `client_id` $\leq$ 50,000, and null or out-of-range identifiers were removed.

II. **Temporal organization**: chronological ordering of crop seasons and definition of activity status based on the two most recent seasons of the family group. A member was considered active if there was participation in at least one of these seasons, otherwise inactive. This season-based definition is aligned with the cooperative's operational CRM monitoring and underpins the subsequent churn-propensity indicator defined over consecutive crop seasons. A `group_id` was created, prioritizing `family_group_id` and, if absent, `producer_code`.

III. **Cleaning and standardization**: removal of redundant, low-information, or highly missing attributes; normalization of `reference_flag`; standardization of labels in `structure_size`; binarization of the activity indicator. The process yielded a set of **20 attributes** (categorical or boolean): `contract_events_flag`, `check_flag`, `protest_flag`, `lawsuit_flag`, `inss_withholding_active`, `focus_client`, `uses_member_portal`, `pet_segment_flag`, `chose_declaration`, `classification_valid_date`, `own_property_flag`, `printed_signed_doc`, `producer_dap_flag`, `activity_segment`, `structure_size`, `producer_type`, `fiscal_regime`, `inss_withholding_type`, `producer_classification`, `registration_date`.

IV. **Imbalance**: application of undersampling to balance the classes without generating synthetic instances, preserving the distribution of the minority class and mitigating bias. Figure 2 summarizes the preprocessing stage.

https://doi.org/10.3390/a19030180

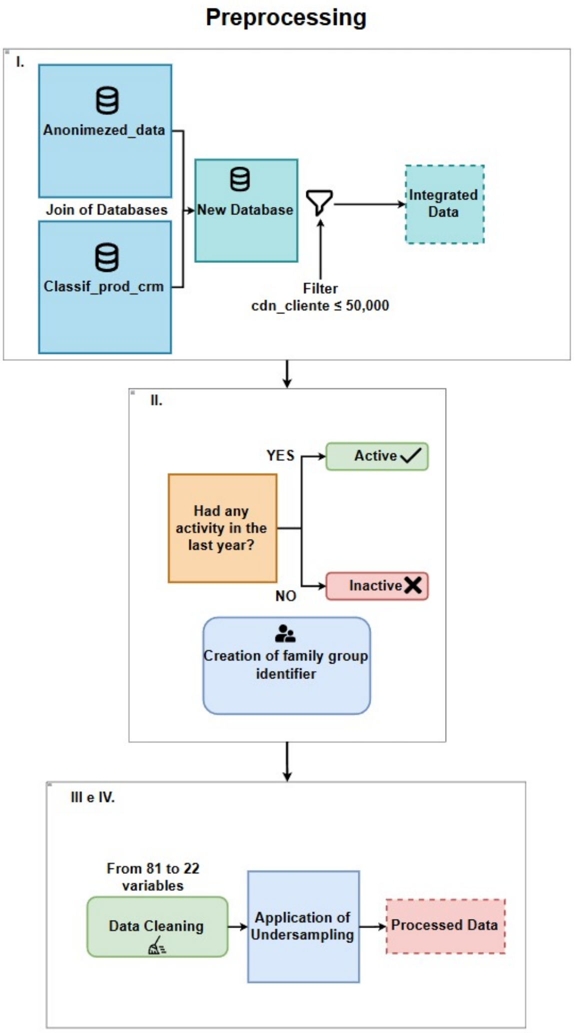


Figure 2. Summary of the preprocessing stage.

3.3. Transformation

An indicator of propensity to churn was defined based on one year of inactivity, or two crop seasons. This operational definition follows the cooperative's governance and monitoring practice, in which member engagement is naturally assessed on a season-based basis rather than on fixed monthly cycles. Accordingly, two consecutive crop seasons without recorded participation (approximately one annual cycle in the studied context) were adopted as a pragmatic proxy for sustained disengagement, consistent with how inactivity is managed in the CRM. We note that this criterion is organization- and calendar-dependent and should be calibrated when applied to cooperatives operating under different crop schedules or seasonal regimes. Synthetic attribute creation and one-hot encoding of categorical variables were performed as required for processing by machine learning algorithms [21,35–38]. The variables `structure_size`, `producer_type`, `fiscal_regime`, `inss_withholding_type`, `producer_classification`, and `registration_date` were binarized. The product of this stage is a unified, clean, and coherent dataset for training and evaluation [37], as summarized in Figure 3.

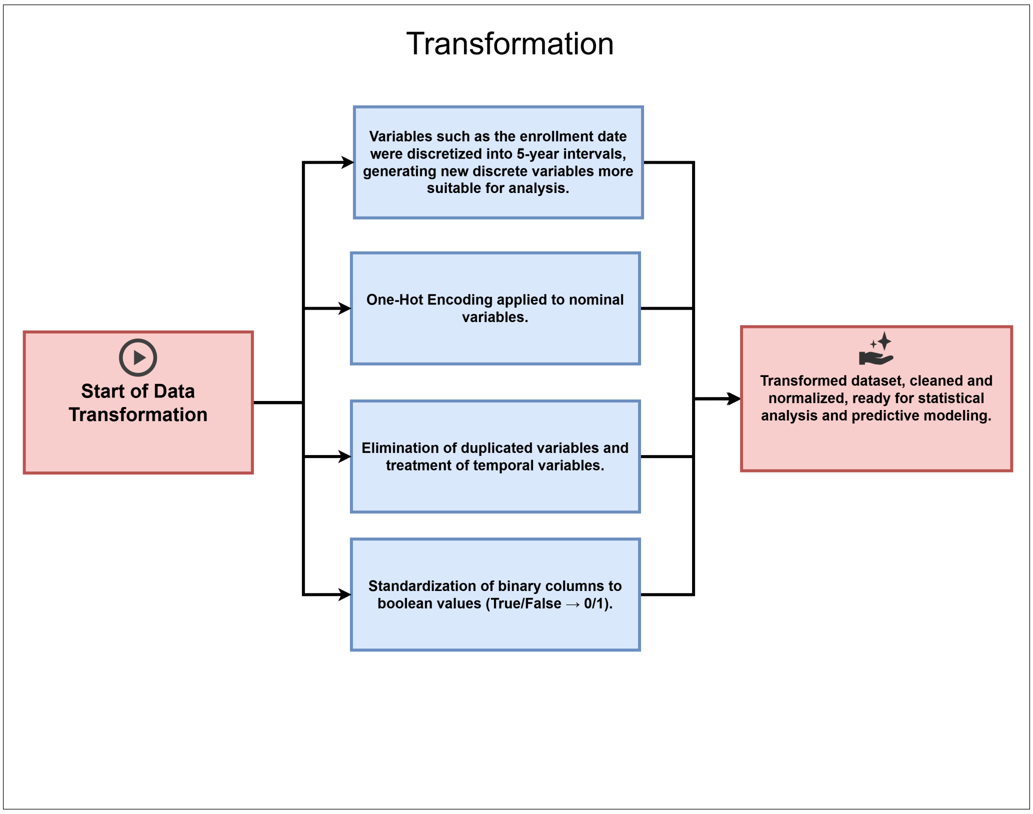


Figure 3. Summary of the transformation stage.

3.4. Data Mining

The KDD data mining stage was conducted in accordance with the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, which provides an iterative and business-oriented framework for structuring predictive analytics projects. Within this framework, the present study operationalizes the Data Preparation, Modeling, and Evaluation phases, while maintaining continuous feedback with the Business and Data Understanding stages, as illustrated in Figure 4.

The data mining stage sought to uncover relevant patterns in the dataset by mapping transaction-level records to predictive models [37]. Supervised learning was employed [36] to predict churn, with the target variable defined from transaction history and the absence of activity across consecutive seasonal periods. Predictors comprised financial and registry attributes, along with derived features, totaling 32 variables.

Consistent with the Modeling phase of CRISP-DM, multiple classification algorithms were explored and compared in parallel, allowing both manual and automated strategies to coexist within the same methodological protocol. RF, XGBoost—part of the decision-tree family and widely used for churn prediction [8]—and a Support Vector Machine designed for classification tasks, hereafter referred to as Support Vector Classifier (SVC), alongside an ensemble of these models, were evaluated. In addition, an AutoML approach was incorporated as a complementary modeling strategy, enabling automated model selection, hyperparameter optimization, and internal ensembling under identical data constraints.

Training and validation relied on stratified subsets to mitigate sampling bias [39,40], reflecting the CRISP-DM emphasis on representativeness during model assessment.

Figure 4 summarizes the modeling pipeline. Following the CRISP-DM logic, Data Preparation and Modeling are treated as tightly coupled and iterative processes, allowing preprocessing decisions to be revisited based on intermediate modeling results. After algorithm selection (AutoML, RF, XGBoost, SVM) and data preprocessing (splitting and optimization), base learners are trained and combined via a voting ensemble. Finally, both the ensemble and the individual models are assessed using cross-validation, corresponding to the Evaluation phase of CRISP-DM, prior to potential deployment decisions.

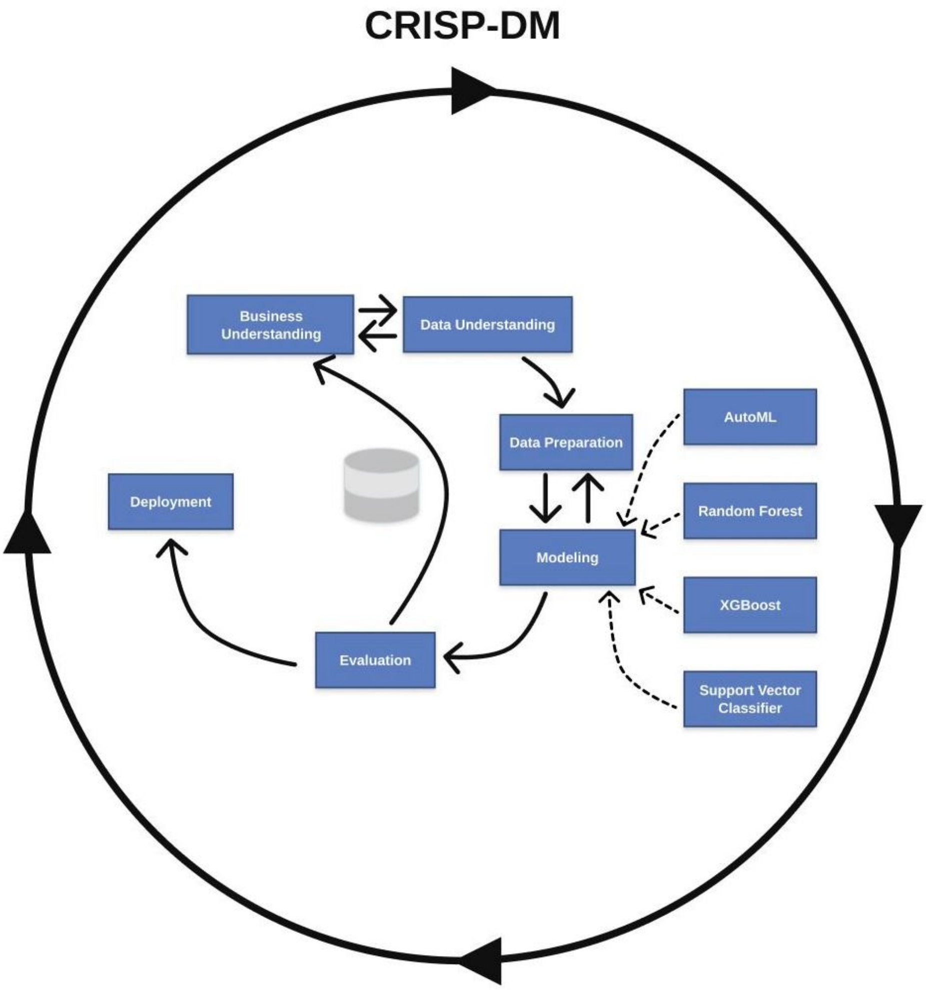


Figure 4. CRISP-DM-oriented data mining pipeline adopted in this study. The arrows indicate the sequential flow of phases and the iterative feedback loops between stages.

The hyperparameters in Tables 3–5 were determined via *grid search* with stratified resampling under a unified validation protocol across all models, ensuring fair comparisons and reducing overfitting risk. This systematic tuning strategy aligns with CRISP-DM best practices by enforcing methodological consistency across candidate models. These results pertain exclusively to models tuned via *grid search* and do not include any configurations derived from AutoML, owing to its automated model-discovery and stacking procedures.

Table 3. Optimal hyperparameters for XGBoost.

Hyperparameter	Value
`learning_rate`	0.01
`n_estimators`	1000
`max_depth`	4
`subsample`	0.7
`colsample_bytree`	0.7
`gamma`	0.2

Source: authors' elaboration (2025).

Table 4. Optimal hyperparameters for Random Forest.

Hyperparameter	Value
`n_estimators`	500
`max_depth`	10
`min_samples_split`	5
`min_samples_leaf`	2
`max_features`	`log2`
`class_weight`	`balanced`

Source: authors' elaboration (2025).

Table 5. Optimal hyperparameters for linear-kernel SVM.

Hyperparameter	Value
`C`	0.1
`gamma`	`scale`
`kernel`	`linear`
`class_weight`	`balanced`

Source: authors' elaboration (2025).

3.5. Evaluation and Interpretation

Evaluation aimed to measure performance and generalization capacity [41] through stratified splitting [42] and cross-validation [35,36], which are critical practices in imbalanced scenarios [8,39,42]. The metrics considered were accuracy [32,37], precision, recall, and F1 score [32,37,43,44], in addition to ROC curves and AUC [5,43,45]. For interpretation, SHAP [14] was applied with two purposes. The first was to verify the coherence of risk determinants with domain knowledge. The second was to support managerial communication within the CRM. Sensitivities to collinearity and to the choice of background were acknowledged, and stability checks were conducted.

4. Results

The data mining stage followed the prescribed methodology, applying k-fold cross-validation together with grid search to optimize hyperparameters. After preprocessing and transformation, we evaluated three supervised algorithms: RF, SVC, and XGBoost—alongside an AutoML pipeline (AutoGluon).

Table 6 summarizes ROC-AUC per fold for the four approaches. AutoML achieved the highest mean ROC-AUC (0.8844) with a standard deviation of 0.0155 across ten folds; XGBoost followed with a mean of 0.8690; RF reached 0.8660; and SVC 0.8611. The lowest fold-specific performance typically occurred in fold 4 for XGBoost (0.8528), RF (0.8432), and SVC (0.8375), whereas AutoML reached its minimum in fold 10 (0.8658) and its maximum in fold 9 (0.9098). The relatively small dispersion (standard deviations between 0.012 and 0.015) indicates a stable cross-validation protocol. Given that ROC-AUC is threshold-independent, this metric is well aligned with the pipeline's selection criterion and robust to prevalence shifts induced by undersampling.

Beyond fold-wise results, we report class-wise and global metrics for the same algorithms, adding an ensemble for completeness, as shown in Tables 7 and 8.

Table 6. ROC-AUC results per fold (K-Fold).

Fold	XGBoost	RF	SVC	AutoML
1	0.8809	0.8826	0.8739	0.8910
2	0.8587	0.8557	0.8536	0.9046
3	0.8752	0.8755	0.8686	0.8866
4	0.8528	0.8432	0.8375	0.8734
5	0.8829	0.8786	0.8678	0.8806
6	0.8854	0.8849	0.8725	0.8673
7	0.8749	0.8688	0.8643	0.8710
8	0.8561	0.8544	0.8562	0.8942
9	0.8696	0.8682	0.8697	0.9098
10	0.8539	0.8486	0.8473	0.8658
Mean	0.8690	0.8660	0.8611	0.8844
Standard Deviation	0.0120	0.0140	0.0114	0.0155

Table 7. Metricsby class (Class 0 and Class 1).

Model	Class 0			Class 1		
	Precision	Recall	F1	Precision	Recall	F1
RF	0.83	0.68	0.75	0.73	0.86	0.79
SVC	0.81	0.73	0.77	0.75	0.83	0.79
XGBoost	0.80	0.74	0.77	0.76	0.82	0.79
AutoML	0.81	0.77	0.79	0.78	0.82	0.80
Ensemble	0.81	0.78	0.79	0.79	0.81	0.80

Table 8. Global metrics.

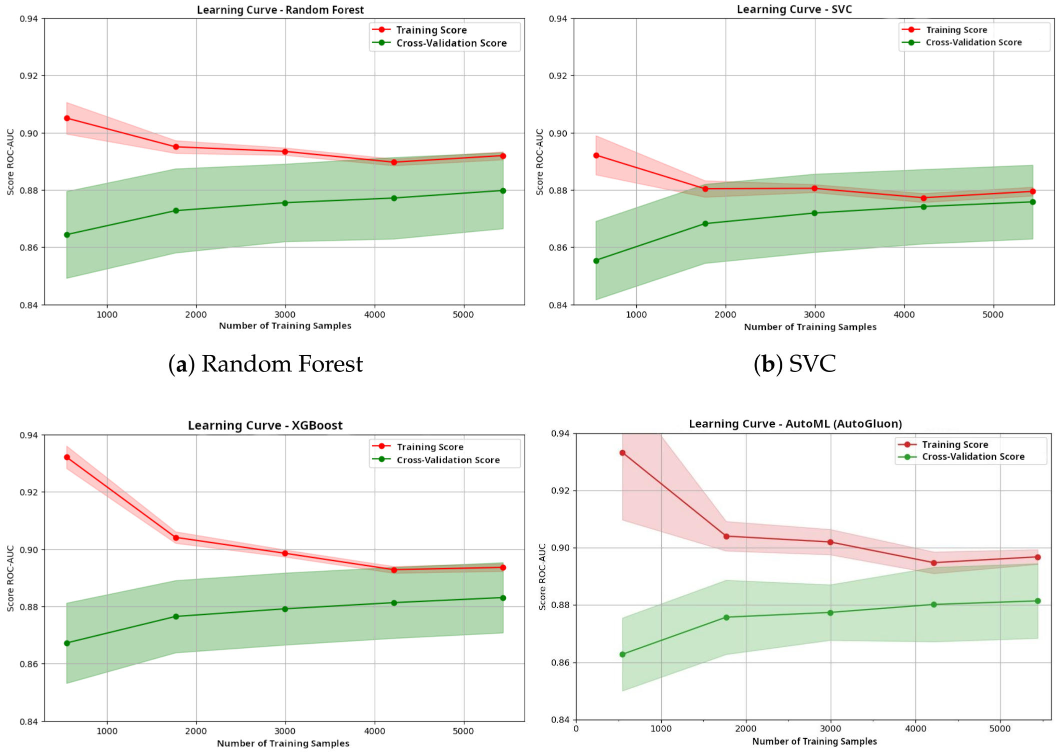
Model	Accuracy (Global)	ROC-AUC (Global)
RF	0.77	0.8675
SVC	0.78	0.8633
XGBoost	0.78	0.872
AutoML	0.80	0.8876
Ensemble	0.80	0.8845

Aggregating by macro-averages, RF, SVC, and XGBoost yield macro-F1 around 0.77–0.78 with closely matched macro-precision and macro-recall, indicating balanced classifiers. AutoML attains a macro-F1 of 0.795 with macro-precision of 0.795 and macro-recall of 0.795, marginally above manual models; the Ensemble reaches a macro-F1 of 0.795 with macro-precision of 0.800 and macro-recall of 0.795, reflecting a small precision gain without sacrificing recall. Global metrics confirm this ordering: accuracy is 0.77–0.78 for manual models and 0.80 for AutoML and the Ensemble; global ROC-AUC ranks AutoML first (0.8876), followed by Ensemble (0.8845) and XGBoost (0.8720).

Reported confidence intervals and summary statistics are consistent with the cross-validated picture. Accuracies for RF, SVC, and XGBoost cluster tightly (e.g., intervals RF 0.8535–0.9058; SVC 0.8508–0.9009; XGBoost 0.8591–0.9071), while ROC-AUC values are 0.8836 (RF), 0.8755 (SVC), 0.8866 (XGBoost), and 0.8845 (Ensemble), corroborating the hierarchy seen in the folds.

4.1. Learning Curves

The learning curves in Figure 5a–d show balanced bias–variance trade-offs, with no clear signs of underfitting or overfitting.



(**a**) Random Forest (**b**) SVC

(**c**) XGBoost (**d**) AutoML

Figure 5. Learning curves of the evaluated models. (**a**) Learning curve of the Random Forest model. (**b**) Learning curve of the SVC model. (**c**) Learning curve of the XGBoost model. (**d**) Learning curve of the AutoML framework.

4.1.1. AutoML

To enable a fair comparison, we contrasted the manually tuned models with AutoGluon models using the common metric ROC-AUC. In AutoML, the reported *val_score* corresponds to an out-of-fold validation ROC-AUC (bagging with saved folds), conceptually aligned with our cross-validation estimates, as shown in Table 9.

Table 9. Comparison of ROC-AUC: manual models vs. AutoML.

Model	Origin	ROC-AUC
RF	Manual	0.8836
SVC	Manual	0.8755
XGBoost	Manual	0.8866
NeuralNetTorch_r14_BAG_L1	AutoML	0.8845
NeuralNetFastAI_r100_BAG_L1	AutoML	0.8844
XGBoost_r31_BAG_L1	AutoML	0.8844
NeuralNetTorch_r79_BAG_L1	AutoML	0.8843

These top AutoML models are numerically comparable to the best manual models, with a slight edge for manual XGBoost (0.8866) over the best AutoML entries (0.8843–0.8845), indicating that the automated search effectively explored the solution space, as detailed in Table 10.

Table 10. Top 10 models on the leaderboard.

#	Model	ROC-AUC (val)
1	WeightedEnsemble_L2	0.8864
2	NeuralNetTorch_r22_BAG_L1	0.8835
3	NeuralNetTorch_r197_BAG_L1	0.8831
4	NeuralNetTorch_r143_BAG_L1	0.8831
5	XGBoost_r31_BAG_L1	0.8828
6	NeuralNetFastAI_r143_BAG_L1	0.8828
7	CatBoost_r50_BAG_L1	0.8828
8	CatBoost_r5_BAG_L1	0.8828
9	CatBoost_r49_BAG_L1	0.8827
10	NeuralNetTorch_BAG_L1	0.8826

4.1.2. SHAP Analysis

SHAP analysis for XGBoost (Figure 6) highlights both the magnitude and direction of contributions among the top ten predictors. For instance, higher values of *chose_declaration* and *uses_member_portal* shift contributions toward the positive side, indicating lower churn propensity. Conversely, *structure_MINI* shows associations consistent with elevated churn risk, suggesting producer profiles with smaller productive structures may be more susceptible. Among the ten most relevant variables, seven act protectively against churn, while three positively indicate churn.

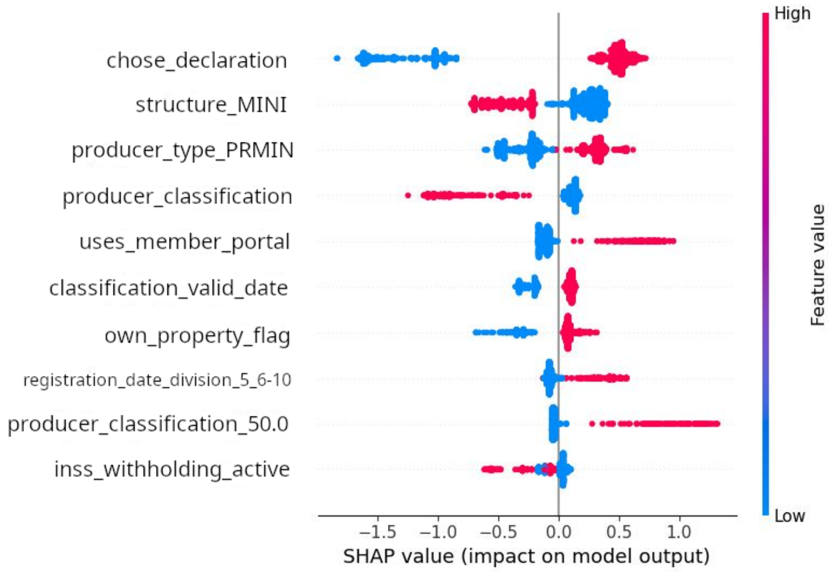


Figure 6. SHAP summary plot illustrating the top ten most influential features for the XGBoost churn model.

5. Discussion

From a methodological perspective, the primary contribution of this study lies in the formalization and empirical validation of an end-to-end churn modeling protocol tailored to cooperative CRM environments. Rather than proposing a novel learning algorithm, the work demonstrates how established tabular models, when embedded in a leakage-aware and governance-aligned pipeline, can achieve stable and reproducible performance under real-world constraints such as seasonality, severe class imbalance, and data confidentiality.

Placing these results in context, the stability across folds and the consistent superiority of AutoML in mean ROC-AUC illustrate that automated pipelines can match or slightly surpass carefully tuned manual baselines, while also offering reproducibility and operational simplicity. The relatively modest performance gap between AutoML and manually tuned models is consistent with prior evidence on structured tabular data, where gains are often incremental once strong baselines are established. In this context, AutoML's contribution should be interpreted less as algorithmic superiority and more as improved reproducibility, reduced human bias in model selection, and robustness across retraining cycles. The Ensemble's small yet coherent gain in precision at stable recall aligns with expected variance reduction from combining heterogeneous learners. Because ROC-AUC underpinned model selection, global ROC-AUC serves as the anchor for comparing classifiers, whereas threshold-dependent metrics (precision, recall, F1) are best used to calibrate decision thresholds for class-specific operational goals.

Beyond the internal comparison of models, these findings are consistent with trends reported in churn prediction studies conducted in mature domains such as telecommunications, banking, and other subscription-based services. In these settings, the literature commonly reports diminishing marginal gains in global discrimination metrics once strong tabular baselines are established, with improvements concentrating on ranking stability, robustness, and operational consistency rather than large increases in ROC-AUC. The results observed in this study suggest that similar dynamics apply to agro-industrial cooperative CRM data, despite its distinct characteristics, including seasonal activity cycles, heterogeneous member profiles, and governance constraints. This alignment reinforces the interpretation that the contribution of the proposed protocol lies not in outperforming existing algorithms, but in empirically validating their transferability and reliability in a domain where systematic evidence remains scarce.

The results further highlight an inherent trade-off between predictive performance, computational cost, and interpretability. Business-wise, elevated discriminative performance can translate into more reliable churn risk estimates and more assertive retention strategies. That said, computational cost matters: AutoML and XGBoost typically demand more training time and resources, which may affect feasibility at scale. Model choice should, therefore, balance marginal performance gains against computational efficiency, retraining frequency, and service-level constraints in operational CRM deployments.

The SHAP results provide actionable interpretability. Variables reflecting land structure and banking information were especially influential for churn prediction. Understanding which behavioral and structural dimensions most strongly impact churn risk enables targeted interventions, from incentives for recurrent operations to personalized outreach to inactive producers. These outputs were integrated into a CRM risk dashboard where each member receives a churn propensity score and is ranked in descending order of risk, enabling segmented retention actions such as targeted campaigns, technical visits, credit offers, or loyalty benefits. Seasonal re-scoring supports monitoring of inactivity indicators and evaluation of intervention impact over time, fostering a data-driven learning cycle. Although instantiated in a single agro-industrial cooperative, the proposed protocol generalizes to CRM settings characterized by recurrent interactions, seasonal activity cycles, and governance constraints. The methodological steps—leakage detection, imbalance treatment, unified validation, and explainability—are domain-agnostic and transferable to other cooperative or membership-based organizations.

From an operational standpoint, the engineering architecture was designed to support the analytical protocol rather than to introduce software engineering novelty. From an engineering perspective, the pipeline was implemented in Python 3.13, while the CRM application was built with Next.js and NestJS to expose insights generated by the AI models. Access is controlled via Cloudflare for pre-authorized emails, and both APIs provide visual documentation via Swagger. The project adhered to strong principles of experimental reproducibility and transparency: fixed random seeds, execution logs, and full traceability from preprocessing to inference. All artifacts—preprocessing scripts, data dictionary, and changelogs—were versioned in a secure repository, preserving the historical record of pipeline and dataset changes. Containerized environments ensured consistency across development, validation, and deployment stages, encapsulating dependencies and configuration to eliminate discrepancies between local and production settings.

Threats to validity persist. Reliance on historical CRM data can introduce collection or entry biases; inconsistencies, gaps, or errors may degrade predictive quality. Because models were trained on a specific data profile, distributional shifts could erode performance. In this context, while the unified stratified grid-search procedure and the observed cross-fold consistency provide an internally comparable baseline, the manuscript does not separately quantify the marginal contribution of feature groups, robustness under controlled noise perturbations, or performance sensitivity across the hyperparameter space. Such analyses would complement the current evidence by clarifying failure modes and robustness margins under operational variability.

A key limitation is that the empirical evaluation is based on a single organization's dataset and relies on i.i.d. cross-validation, which—while appropriate for establishing a consistent internal baseline on real-world data—does not fully characterize generalization across time, seasons, or branches. Accordingly, strengthening external validity through temporal validation and branch/filial-based holdout (and, when feasible, multi-organization evaluation) is a necessary next step for broader deployment claims.

Continuous monitoring is advisable, with a designated owner maintaining data consistency, tracking statistics, and updating the pipeline and models across seasons using existing training and prediction routes. Information leakage analysis revealed that temporal variables such as "last crop movement" could unduly influence predictions; removing them caused a modest reduction in recall and F1, reinforcing the importance of careful preprocessing and domain-aware feature curation to secure generalizable results. These threats reinforce the need for protocol-level safeguards—such as leakage analysis, season-aware retraining, and continuous monitoring—to preserve external validity over time.

Taken together, the empirical evidence indicates that robust churn prediction in cooperative CRM systems depends less on algorithmic novelty and more on disciplined methodology, validation rigor, and governance-aware deployment. Finally, the empirical evidence supports the study's working hypotheses: the supervised models (RF, SVC, XGBoost, Ensemble) achieved stable, satisfactory discrimination (e.g., ROC-AUC around or above 0.88 with test accuracies between 0.77 and 0.80), demonstrating precise and generalizable churn risk estimation; tree-based models showed resilience to imbalance after undersampling, with minority-class recall between 0.82 and 0.86; AutoML delivered slightly superior averages and higher reproducibility across training cycles; and SHAP successfully surfaced business-coherent determinants such as *chose_declaration* and *uses_member_portal*, enabling the operationalization of insights in the risk dashboard.

6. Conclusions

This study designed, implemented, and evaluated a CRM module integrated with machine learning to predict member churn in an agro-industrial cooperative. Supervised algorithms—RF, XGBoost, and SVC—combined with an AutoML framework produced reliable churn propensity estimates once class imbalance was addressed via undersampling. Performance improved markedly after balancing, especially in the minority class recall, underscoring the centrality of data treatment for classifier generalization. Information leakage checks and the removal of problematic temporal variables modestly reduced recall and F1 but increased confidence in the external validity of the models. SHAP-based interpretability clarified feature contributions and reinforced business alignment.

https://doi.org/10.3390/a19030180

Strategically, embedding predictive models in the CRM environment supports smarter resource allocation, targeted retention initiatives, and stronger member relationship strategies. By illuminating behavioral patterns associated with churn, the cooperative can prioritize higher-impact loyalty actions, reduce acquisition costs, and improve engagement outcomes—turning data-driven insight into operational advantage.

Future work will broaden data sources. The present churn prediction relied on the *anonymized_data* dataset, complemented by business-rule features. The cooperative also maintains the *classif_prod_crm* base with detailed information on productive structure and delivered crops. Integrating this second base into the existing methodology is the main next step, enriching the feature set with quantitative agricultural indicators such as cultivated area and revenue, and potentially yielding further gains in discriminative performance and actionable insight.

Author Contributions: Conceptualization, S.A.T. and J.V.d.C.A.; methodology, S.A.T. and J.V.d.C.A.; software, S.A.T. and J.V.d.C.A.; validation, S.A.T., J.V.d.C.A. and A.B.B.; formal analysis, S.A.T. and J.V.d.C.A.; investigation, S.A.T. and J.V.d.C.A.; resources, A.B.B. and A.C.; data curation, J.V.d.C.A.; writing—original draft preparation, S.A.T.; writing—review and editing, J.V.d.C.A., A.B.B. and A.C.; visualization, S.A.T. and J.V.d.C.A.; supervision, D.S.S. and H.V.S.; project administration, D.S.S. and H.V.S.; funding acquisition, D.S.S. and H.V.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study analyzed secondary CRM records provided by the data owner under its data-governance framework, with records de-identified prior to transfer and no direct interaction with individuals.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset metadata are publicly available on Zenodo (DOI: 10.5281/zenodo.18473557); however, access to the files is restricted. File access is granted only upon reasonable request to the corresponding author, evaluated in coordination with the cooperative (data owner), and subject to compliance with the cooperative's data-governance agreement and data-use conditions.

Acknowledgments: The authors thank UniFil—Centro Universitário Filadélfia (Londrina), the Cooperativa Integrada, the UniFil Artificial Intelligence Research Group, and UTFPR—Federal University of Technology—Paraná for their institutional and technical support.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
API	Application Programming Interface
AUC	Area Under the Curve
AUC-PR	Area Under the Precision–Recall Curve
AUC-ROC (ROC-AUC)	Area Under the Receiver Operating Characteristic Curve
AutoML	Automated Machine Learning
BPMN	Business Process Model and Notation
CatBoost	Categorical Boosting
CNN	Convolutional Neural Network
CRM	Customer Relationship Management
CV	Cross-Validation
DAP/AF	Family Farming Registry (Brazil)

GDP	Gross Domestic Product
HTTP	Hypertext Transfer Protocol
INSS	National Institute of Social Security (Brazil)
KDD	Knowledge Discovery in Databases
KNN	k-Nearest Neighbors
LightGBM	Light Gradient Boosting Machine
LIME	Local Interpretable Model-agnostic Explanations
LR	Logistic Regression
MILP	Mixed-Integer Linear Programming
NB	Naive Bayes
OS	Operating System
RF	Random Forest
R^2	Coefficient of Determination
ROC	Receiver Operating Characteristic
SHAP	SHapley Additive exPlanations
SPSS	Statistical Package for the Social Sciences
SVC	Support Vector Classifier
SVM	Support Vector Machine
UML	Unified Modeling Language
UTFPR	Universidade Tecnológica Federal do Paraná
UniFil	Centro Universitário Filadélfia
XAI	Explainable Artificial Intelligence
XGBoost	Extreme Gradient Boosting

References

1. CEPEA. PIB do Agronegócio Brasileiro. Available online: https://www.cepea.org.br/br/pib-do-agronegocio-brasileiro.aspx (accessed on 22 October 2025).
2. EMBRAPA. Soja: Dados Econômicos. Available online: https://www.embrapa.br/soja/cultivos/soja1/dados-economicos (accessed on 19 October 2025).
3. IBGE. Produção da Pecuária Municipal 2024—Aquicultura (Tilápia). Available online: https://www.ibge.gov.br/explica/producao-agropecuaria/tilapia/br (accessed on 19 October 2025).
4. Neves, M.d.C.R.; de Castro, L.S.; de Freitas, C.O. O impacto das cooperativas na produção agropecuária brasileira: Uma análise econométrica espacial. *Rev. Econ. Sociol. Rural* **2019**, *57*, 559–576. [CrossRef]
5. Chang, V.; Hall, K.; Xu, Q.A.; Amao, F.O.; Ganatra, M.A.; Benson, V. Prediction of customer churn behavior in the telecommunication industry using machine learning models. *Algorithms* **2024**, *17*, 231. [CrossRef]
6. Yocupicio-Zazueta, A.; Brau-Avila, A.; Cirett-Galán, F.; Valenzuela-Galván, M. Design and deployment of ML in CRM to identify leads. *Appl. Artif. Intell.* **2024**, *38*, 2376978. [CrossRef]
7. Sikri, A.; Jameel, R.; Idrees, S.M.; Kaur, H. Enhancing customer retention in telecom industry with machine learning driven churn prediction. *Sci. Rep.* **2024**, *14*, 13097. [CrossRef]
8. Vaudevan, M.; Narayanan, R.S.; Nakeeb, S.F.; Abhishek, A. Customer churn analysis using XGBoosted decision trees. *Indones. J. Electr. Eng. Comput. Sci.* **2022**, *25*, 488–495. [CrossRef]
9. Imani, M.; Joudaki, M.; Beikmohammadi, A.; Arabnia, H.R. Customer Churn Prediction: A Systematic Review of Recent Advances, Trends, and Challenges in Machine Learning and Deep Learning. *Mach. Learn. Knowl. Extr.* **2025**, *7*, 105. [CrossRef]
10. AbdelAziz, N.M.; Bekheet, M.; Salah, A.; El-Saber, N.; AbdelMoneim, W.T. A Comprehensive Evaluation of Machine Learning and Deep Learning Models for Churn Prediction. *Information* **2025**, *16*, 537. [CrossRef]
11. Alotaibi, M.Z.; Haq, M.A. Customer Churn Prediction for Telecommunication Companies using Machine Learning and Ensemble Methods. *Eng. Technol. Appl. Sci. Res.* **2024**, *14*, 14572–14578. [CrossRef]
12. H2O.ai. H2O AutoML: Automatic Machine Learning—Documentação. 2024. Available online: https://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html (accessed on 7 September 2025).
13. LeDell, E.; Poirier, S. H2O AutoML: Scalable Automatic Machine Learning. In Proceedings of the 7th ICML Workshop on Automated Machine Learning (AutoML), Vienna, Austria, 17–18 July 2020; pp. 1–9.
14. Lundberg, S.M.; Lee, S. A Unified Approach to Interpreting Model Predictions. *arXiv* **2017**, arXiv:1705.07874. [CrossRef]
15. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef]

16. Peng, K.; Peng, Y.; Li, W. Research on customer churn prediction and model interpretability analysis. *PLoS ONE* **2023**, *18*, e0289724. [CrossRef] [PubMed]
17. GhorbanTanhaei, H.; Boozary, P.; Sheykhan, S.; Rabiee, M.; Rahmani, F.; Hosseini, I. Predictive analytics in customer behavior: Anticipating trends and preferences. *Results Control Optim.* **2024**, *17*, 100462. [CrossRef]
18. Boozary, P.; Sheykhan, S.; GhorbanTanhaei, H.; Magazzino, C. Enhancing customer retention with machine learning: A comparative analysis of ensemble models for accurate churn prediction. *Int. J. Inf. Manag. Data Insights* **2025**, *5*, 100331. [CrossRef]
19. Cheng, H.; Ng'ombe, J.N.; Choi, Y.; Kalinda, T.H.; Zheng, S. Understanding the drivers of smallholder dairy cooperative participation in developing countries: Evidence from rural Zambia. *Agric. Syst.* **2025**, *224*, 104261. [CrossRef]
20. Zhongchen, G.; Jie, H.; Chen, C. Intelligent transformation of financial services of agricultural cooperatives based on edge computing and deep learning. *Soft Comput.* **2023** . [CrossRef]
21. Patil, S.; Mohammed, A.S. Proactive CRM: Predicting Customer Behavior and Churn Using Machine Learning Models. *SSRN Electron. J.* **2023**, *2*, 61–74. [CrossRef]
22. Özkurt, C. Transparency in Decision-Making: The Role of Explainable AI (XAI) in Customer Churn Analysis. *Inf. Technol. Econ. Bus.* **2025**, *2*, 1–11. [CrossRef]
23. Li, Y.; Yan, K. Prediction of bank credit customers churn based on machine learning and interpretability analysis. *Data Sci. Financ. Econ.* **2025**, *5*, 19–34. [CrossRef]
24. Bhuria, R.; Gupta, S.; Kaur, U.; Bharany, S.; Rehman, A.U.; Hussen, S.; Tejani, G.G.; Jangir, P. Ensemble-based customer churn prediction in banking: A voting classifier approach for improved client retention using demographic and behavioral data. *Discov. Sustain.* **2025**, *6*, 28. [CrossRef]
25. Abdellaoui Alaoui, E.A.; Elgamouz, N.; Sallah, A.; Filali, A.; Hessane, A.; Agoujil, S. Balancing Speed and Insight: Computational Efficiency of XAI Methods for Telecommunication Customer Churn Analysis. In Proceedings of the IEEE International Conference on Computing Sciences and Communication (ICCSC), Fez, Morocco, 19–20 June 2025; pp. 1–7. [CrossRef]
26. Duru, I.; Ak, N.; Dede, R. A Comparative Study of Feature Selection and Resampling Techniques for Customer Churn Prediction with Explainable AI in the Telecommunications Sector. In Proceedings of the IEEE Conference on Artificial Intelligence and Data Science Applications (ACDSA), Antalya, Turkiye, 7–9 August 2025; pp. 1–12. [CrossRef]
27. Asif, D.; Arif, M.S.; Mukheimer, A. A data-driven approach with explainable artificial intelligence for customer churn prediction in the telecommunications industry. *Results Eng.* **2025**, *26*, 104629. [CrossRef]
28. Oprea, S.V.; Bâra, A. Customer-Centric Decision-Making with XAI and Counterfactual Explanations for Churn Mitigation. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 129. [CrossRef]
29. Taher, S.S.; Ameen, S.Y.; Ahmed, J.A. Advanced fraud detection in blockchain transactions: An ensemble learning and explainable AI approach. *Eng. Technol. Appl. Sci. Res.* **2024**, *14*, 12822–12830. [CrossRef]
30. Ahmad, A.; Jafar, A.; Aljoumaa, K. Customer churn prediction in telecom using machine learning in big data platform. *J. Big Data* **2019**, *6*, 28. [CrossRef]
31. Mittal, K. Leveraging Artificial Intelligence to Enhance Customer Service in Cooperative Banks in India. Master's Thesis, Tampere University of Applied Sciences, Tampere, Finland, 2025.
32. Sabbeh, S.F. Machine-learning techniques for customer retention: A comparative study. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 273–281. [CrossRef]
33. Mouli, K.C.; Raghavendran, C.V.; Bharadwaj, V.Y.; Vybhavi, G.Y.; Sravani, C.; Vafaeva, K.M.; Hussein, L. An analysis on classification models for customer churn prediction. *Cogent Eng.* **2024**, *11*, 2378877. [CrossRef]
34. He, C.; Ding, C.H.Q. A novel classification algorithm for customer churn prediction based on hybrid Ensemble-Fusion model. *Sci. Rep.* **2024**, *14*, 20179. [CrossRef]
35. Witten, I.H.; Frank, E. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed.; Morgan Kaufmann Publishers/Elsevier: San Francisco, CA, USA, 2005; p. 525.
36. James, G.; Witten, D.; Hastie, T.; Tibshirani, R.; Taylor, J. *An Introduction to Statistical Learning: With Applications in Python*; Springer Texts in Statistics; Springer Nature AG: Cham, Switzerland, 2023. [CrossRef]
37. Aggarwal, C.C. *Data Mining: The Textbook*; Springer: New York, NY, USA, 2015; p. 746.
38. Larose, D.T. *Wiley Series on Methods and Applications in Data Mining: An Introduction to Data Mining*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014; p. 336.
39. Santos, L.L.; Bianchi, R.A.; Costa, A.H. FinBERT-PT-BR: Análise de sentimentos de textos em português do mercado financeiro. In *Proceedings of the Anais do 2nd Brazilian Workshop on Artificial Intelligence in Finance*; Sociedade Brasileira de Computação (SBC): Porto Alegre, Brazil, 2023; pp. 144–155.
40. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
41. Bishop, C.M. *Pattern Recognition and Machine Learning*; Information Science and Statistics; Springer Science+Business Media, LLC: New York, NY, USA, 2006.

42. Fontanari, T.; Fróes, T.; Recamonde-Mendoza, M. Cross-validation Strategies for Balanced and Imbalanced Datasets. In *Brazilian Conference on Intelligent Systems*; Springer International Publishing: Cham, Switzerland, 2022; pp. 626–640. [CrossRef]
43. Wang, S.; Šuster, S.; Baldwin, T.; Verspoor, K. Predicting publication of clinical trials using structured and unstructured data: Model development and validation study. *J. Med. Internet Res.* **2022**, *24*, e38859. [CrossRef]
44. Lalwani, P.; Mishra, M.K.; Chadha, J.S.; Sethi, P. Customer churn prediction system: A machine learning approach. *Computing* **2022**, *104*, 271–294. [CrossRef]
45. Teodorescu, V.; Obreja Brașoveanu, L. Assessing the Validity of k-Fold Cross-Validation for Model Selection: Evidence from Bankruptcy Prediction Using Random Forest and XGBoost. *Computation* **2025**, *13*, 127. [CrossRef]