Churn Detection in Agro-Industrial Cooperatives Using Machine Learning: A Practical Case Study for Strategic Decision-Making

Sergio Akio Tanaka
Universidade Tecnológica Federal do Paraná (UTFPR)
Londrina, Brazil
sergio.tanaka@unifil.br

Sergio Kenji Sawasaki Tanaka
Centro Universitário Filadélfia (UniFil)
Londrina, Brazil
serginho.tanaka@edu.unifil.br

Danilo Sipoli Sanches
Universidade Tecnológica Federal do Paraná (UTFPR)
Londrina, Brazil
danilosanches@utfpr.edu.br

Hugo Valadares Siqueira
Universidade Tecnológica Federal do Paraná (UTFPR)
Londrina, Brazil
hugosiqueira@utfpr.edu.br

***Abstract*—This article proposes a predictive model for classifying churn behavior among cooperative members using Decision Tree algorithms. The solution integrates Machine Learning (ML) and data engineering to predict inactivity behaviors, utilizing three datasets: registration information, financial history, and production data of the members. The predictive model achieved 98.36**

***Index Terms*—Artificial Intelligence, Machine Learning, CRM, Cooperatives, Churn detection, Decision Tree, Prediction**

I. Introduction

The advancement of AI has revolutionized customer relationship management (CRM), enabling automation, personalization, and predictive analytics. As Yocupicio-Zazueta *et al.* [1] noted, ML techniques like logistic regression enhance CRM systems by identifying patterns and optimizing resource allocation.

AI comprises subfields such as Machine Learning (ML), Deep Learning (DL), and Data Science (DS). ML supports predictive modeling based on historical data, widely applied in churn analysis. DL, via deep artificial neural networks, identifies complex patterns in large datasets. DS combines statistical methods, AI techniques, and engineering practices to generate strategic insights [2].

In this study, the Decision Tree algorithm was chosen due to its interpretability and effectiveness [3]. Although DL was not implemented, it is identified as a future improvement area [4]. DS supported the project by structuring the analytical pipeline, applying models, and visualizing results [5].

Agro-industrial cooperatives, key drivers of sustainable development, face challenges with member disengagement. Churn—understood as the withdrawal or inactivity of members—affects service quality and financial health [6]. This study applies ML to predict churn using integrated datasets and automated preprocessing via `scikit-learn` [7].

This study was supported by UniFil and UTFPR.

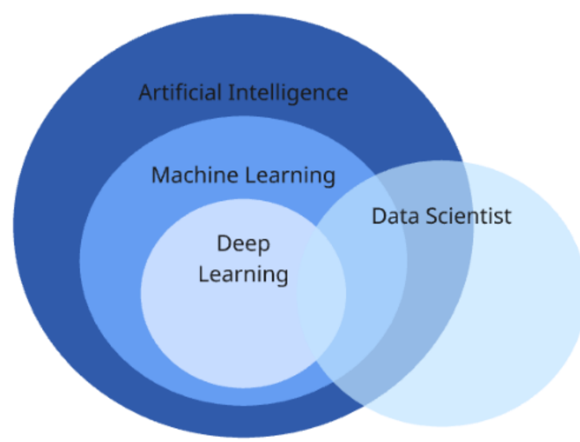

Fig. 1. Conceptual AI Hierarchy: AI, ML, DL, and DS.

II. Methodology

The methodology adopted in this study was structured based on a hybrid approach that integrated data engineering practices, machine learning techniques, and the development of an interactive interface aimed at end-users. The methodological process involved everything from the collection and integration of datasets to the delivery of a functional decision-support system.

Initially, three essential anonymized datasets were gathered and integrated to construct the predictive model: (i) member registration data, containing information such as age and size of the producer; (ii) financial records, including the history of banking transactions and operations carried out with the cooperative; and (iii) agricultural production history, consisting of records of harvest seasons, types of crops, production volumes, and yields over the years. Although heterogeneous, these information sources were consolidated through key relationships and later subjected to an initial exploratory analysis to evaluate variable patterns and distributions.

The data preprocessing stage was conducted with the aim of ensuring quality and consistency in the unified dataset. Missing and inconsistent values were handled using techniques such as Min-Max Scaling, which adjusts variables to a common range, and Label Encoding, used to convert categorical attributes into numerical representations. A particular

III. System Architecture

The proposed system adopts a distributed microservices architecture, leveraging containerization to ensure scalability, modularity, and maintainability. This design is composed of three main layers: Front-End, Back-End, and Data/AI Services.

A. Front-End Layer

The Front-End is developed using the React library within the Next.js framework, and is deployed inside a Node.js container. This configuration enables both server-side rendering and static site generation, improving performance and user experience. The effectiveness of Next.js in building modern, high-performance web applications has been well-documented in the literature [8].

B. Back-End Layer

The Back-End is implemented using the NestJS framework, also deployed in a Node.js container. NestJS provides a modular and extensible structure and supports microservices natively, making it well-suited for building scalable enterprise applications [9].

Data persistence is handled by PostgreSQL due to its proven reliability, support for complex queries, and strong consistency guarantees. Redis is used as an in-memory caching system to accelerate data retrieval, reduce response times, and alleviate database load [10].

To support asynchronous operations and decoupled communication between services, the system integrates RabbitMQ, a robust message broker widely used in distributed systems [10].

C. Data and AI Services

A dedicated AI module, `Docker_IA`, was implemented in Python 3.12 and executes the `cooperadaAlgorithm`, a machine learning model responsible for predicting cooperative member churn. This module receives inference requests from the Back-End via RabbitMQ, processes them, and returns the results. The adoption of Python-based AI modules in microservice environments is supported by recent studies, highlighting their flexibility and integration capabilities [11].

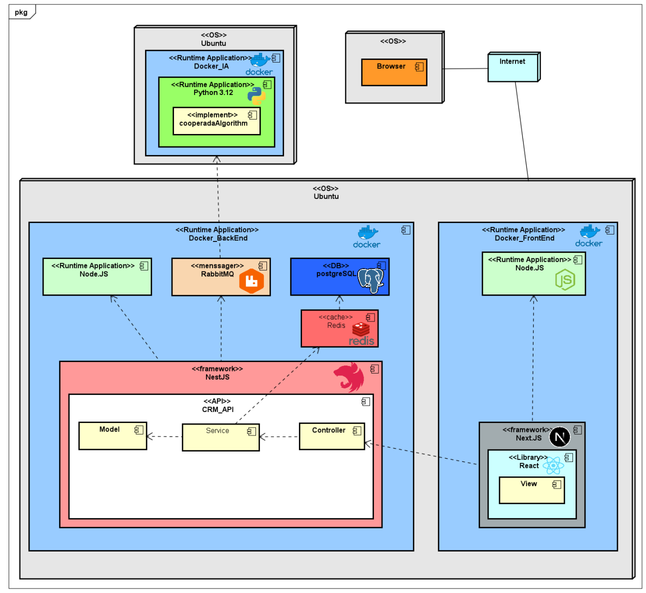

Fig. 2. Deployment diagram of the proposed architecture.

D. Use Case Representation

Figure 3 illustrates the main actors and interactions within the system through a UML use case diagram. The two primary user roles are `SA_integrada_vendor` and `SA_integrada_admin`. The former can access dashboards, sales data, member information, and churn analytics, while the latter extends those privileges with administrative capabilities, including the ability to update AI services.

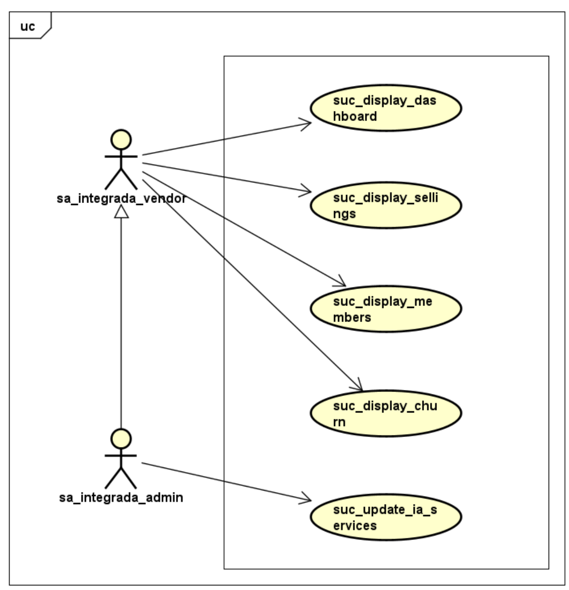

Fig. 3. Use case diagram showing system-user interactions.

IV. Results

The exploratory data analysis revealed consistent behavioral patterns associated with member churn. The application of the

Decision Tree algorithm resulted in an accuracy of 98.36%, with strong dependence on the variable `dt_ult_movto`, corresponding to the date of the member's last transaction.

As shown in Figure 4, the high relevance of this feature suggests a strong temporal correlation with inactivity. However, when this variable was removed, other significant attributes emerged, such as `log_ativo_pa`—indicating financial activity in the previous year—revealing new latent behavioral patterns. This approach aligns with the findings of Brown *et al.* [12], who emphasize the value of experiments involving the exclusion of dominant variables to avoid biases in predictive models.

With the removal of both `dt_ult_movto` and `log_ativo_pa`, other features gained prominence in the analysis. In this scenario, the variable `cod_transp`—corresponding to the transporter code—stood out, reinforcing the hypothesis that logistical aspects, combined with production history, are relevant indicators of member churn. This observation is consistent with recent literature, which highlights the impact of multiple seasonal and productive characteristics on user behavior [13], [14].

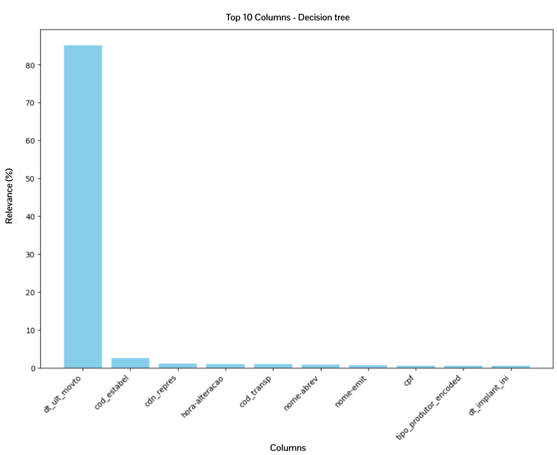

Fig. 4. Feature Importance in the Decision Tree Model

Figure 5 presents the visual interface of the application developed as part of this study, aimed at offering an interactive and accessible tool for the integrated analysis of cooperative member data. The dashboard was designed with a focus on usability, visual clarity, and decision-making support, especially for non-technical users such as cooperative managers and analysts.

The dashboard is structured into multiple components, each representing a critical dimension for monitoring member behavior:

- **Top Indicators (analytical cards)**: KPIs such as accumulated earnings, available balance, and number of transactions with percentage comparisons.
- **Overview Bar Chart**: Displays seasonal earnings trends from 2021 to 2025.
- **New Member Gauge**: Shows new member enrollment as a percentage.

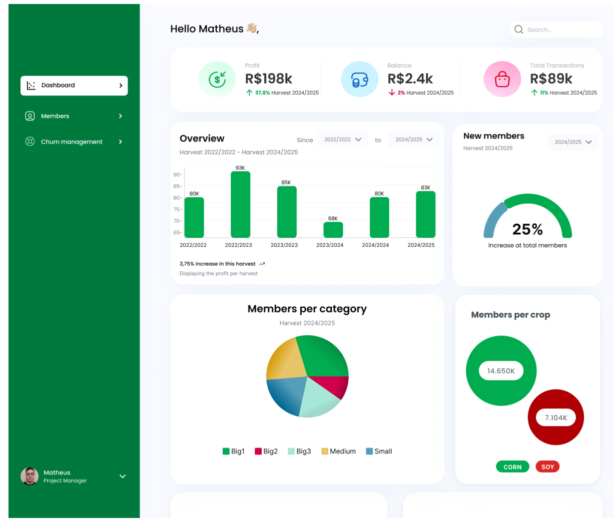

Fig. 5. Developed System Dashboard Interface

- **Members by Category Pie Chart**: Segments producers by size (Small to Large).
- **Crop Type Distribution**: Displays member count per crop type.

The application supports cross-variable analyses (e.g., size vs. crop vs. transaction), enabling identification of churn risk profiles and supporting strategic decisions such as technical visits or customized offers. It was designed following Explainable AI (XAI) principles [15], [16].

Challenge involved the column related to harvest seasons, which contained two values per line; to resolve this ambiguity, an expanding strategy was applied, duplicating the rows and separating the values into distinct instances. After these transformations, the data were joined using inner join operations, forming a unified relational structure to proceed with modeling.

The predictive modeling was carried out using the Decision Tree algorithm, chosen for its robustness, ability to interpret heterogeneous variables, and ease of interpretation of results. The goal of the model was to classify members as either active or inactive, identifying the most relevant factors associated with churn. The decision to use this algorithm took into account its simplicity of implementation, clarity in decision rule structures, and satisfactory performance in tabular datasets with multiple attributes [17].

V. Conclusion and Future Work

This study demonstrated the potential of AI as a tool to support strategic decision-making in agro-industrial cooperatives, with a focus on member retention. The application of the Decision Tree model showed high effectiveness, achieving an accuracy of 98.36% and enabling the identification of key variables associated with churn. Among the most relevant attributes were the date of the last transaction (`dt_ult_movto`), recent financial activity, and harvest history—elements directly related to member engagement.

The architecture of the developed solution integrated a robust automated preprocessing pipeline, including Min-Max Scaling normalization and categorical encoding via Label Encoding. This approach follows widely recognized methodological best practices [18], [19], and ensures consistency and replicability of the model. Furthermore, the availability of the tool through an interactive interface favored adoption by non-technical analysts, promoting greater accessibility to data visualization—aligning with usability and explainable AI guidelines [15], [16].

Beyond providing predictive insights, the results reinforce the importance of data quality and model interpretability in building decision support systems. This premise is particularly critical in sensitive sectors such as agroindustry, where strategic decisions have a direct impact on operational sustainability [20].

However, the study identified a potential structural bias resulting from the duplication of records in the "harvest" column. This practice may have compromised the actual statistical distribution of the data. As a solution, the One-Hot Encoding technique will be adopted, as recommended by Smith *et al.* [21], aiming to represent temporal categories in a non-ordinal manner.

Additionally, complementary models such as Random Forest and XGBoost were explored, which offer significant performance gains in datasets with high cardinality [18]. These algorithms will later be evaluated in terms of interpretability and fairness, following guidelines proposed by Jo *et al.* [17].

As a future development path, the incorporation of Deep Learning (DL) models is being considered. These models demonstrate superior capacity for identifying complex and non-linear patterns, especially in contexts involving time series, crop imagery, or extensive histories.

To evolve the application, new areas for improvement have been identified: integration of external variables (e.g., climate, prices, policies) [22], recommendation systems for preventive actions [20], cross-regional validation, and more advanced dashboards with confidence metrics.

Through these initiatives, the study contributes to strengthening predictive solutions in the cooperative sector, promoting greater economic sustainability, member loyalty, and efficiency in data-driven decision-making.

REFERENCES

[1] A. Yocupicio-Zazueta and J. Ramirez, "Ai-powered crm: Predictive modeling and customer retention," *Journal of Data-Driven Marketing*, vol. 9, no. 1, pp. 22–39, 2024.

[2] N. Rane, M. Paramesha, S. Choudhary, and J. Rane, "Artificial intelligence, machine learning, and deep learning for advanced business strategies: a review," *Journal of PUII*, vol. 2, pp. 147–171, 2024.

[3] M.-F. Balcan and D. Sharma, "Learning accurate and interpretable decision trees," in *Proceedings of the 39th International Conference on Machine Learning*, vol. 162, 2023, pp. 1523–1540. [Online]. Available: https://proceedings.mlr.press/v162/balcan22a.html

[4] M. Pichler, F. Hartig, F. Wesselingh, and H. Koch, "Machine learning and deep learning—a review for ecologists," *Methods in Ecology and Evolution*, vol. 14, no. 3, pp. 589–606, 2023. [Online]. Available: https://doi.org/10.1111/2041-210X.14061

[5] I. D. Dinov, *Data Science and Predictive Analytics: Biomedical and Health Applications Using R*, 2nd ed. Springer, 2023. [Online]. Available: https://doi.org/10.1007/978-3-031-17483-4

[6] A. Manzoor and Y. Zhou, "Challenges in member retention in agricultural cooperatives," *Journal of Cooperative Studies*, vol. 15, no. 2, pp. 45–59, 2024.

[7] F. Barrera, M. Segura, and C. Maroto, "Multiple criteria decision support system for intelligent customer segmentation," *Expert Systems with Applications*, 2024.

[8] M. Fariz and D. Anggraini, "Modern front end web architectures with react.js and next.js," *International Research Journal of Advanced Engineering and Science*, vol. 7, no. 1, pp. 162–167, 2022. [Online]. Available: https://irjaes.com/wp-content/uploads/2022/02/IRJAES-V7N1P162Y22.pdf

[9] V. Dolzhenko, "Development of microservices using nestjs: Architecture and practical examples," *Universum: Technical Sciences*, vol. 123, no. 6, 2024.

[10] A. Ćatović, N. Buzaija, and S. Lemeš, "Microservice development using rabbitmq message broker," *Science, Engineering and Technology*, vol. 2, no. 1, pp. 30–37, 2022.

[11] I. D. Mienye and T. G. Swart, "A comprehensive review of deep learning: Architectures, recent advances, and applications," *Information*, vol. 15, p. 755, 2024.

[12] J. Brown, K. Williams, and M. Harris, "Automated data cleaning and predictive analytics in agroindustrial cooperatives: a case study," *Computers and Electronics in Agriculture*, vol. 205, p. 105987, 2023.

[13] C. Liu, K. Wang, Y. Wang, and X. Yuan, "Learning deep multimanifold structure feature representation for quality prediction with an industrial application," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 9, pp. 5849–5858, 2022.

[14] Y. Chen, H. Tan, and R. Feng, "Bias reduction in machine learning models for crop prediction: a comprehensive review," *Expert Systems with Applications*, vol. 210, p. 18246, 2023.

[15] M. Graziani *et al.*, "A global taxonomy of interpretable ai: unifying the terminology for the technical and social sciences," *Artificial Intelligence Review*, vol. 56, no. 4, pp. 3473–3504, 2023. [Online]. Available: https://doi.org/10.1007/s10462-022-10256-8

[16] C. Li, Y. Fang, and Y. Chiang, "Can ai chatbots help retain customers? an integrative perspective using affordance theory and service-domain logic," *Technological Forecasting and Social Change*, vol. 197, p. 122921, 2023.

[17] N. Jo *et al.*, "Learning optimal fair decision trees: Trade-offs between interpretability, fairness, and accuracy," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 2023. [Online]. Available: https://dl.acm.org/doi/fullHtml/10.1145/3600211.3604664

[18] J. Pasaribu, N. Yudistira, and W. F. Mahmuddy, "Tabular data classification and regression: Xgboost or deep learning with retrieval-augmented generation," *IEEE Access*, vol. 12, pp. 191 719–191 732, 2024.

[19] Henderi, T. Wahyuningsih, and E. Rahwanto, "Comparison of min-max normalization and z-score normalization in the k-nearest neighbor algorithm to test the accuracy of types of breast cancer," *International Journal of Informatics and Information System*, vol. 3, no. 2, pp. 17–25, 2021. [Online]. Available: https://www.researchgate.net/publication/350556043

[20] R. Miller, S. Carter, and T. Davis, "Leveraging ai for member retention: an exploratory study in the agroindustrial sector," *Journal of Cooperative Studies*, vol. 14, no. 1, pp. 1–15, 2023.

[21] A. Smith and B. Thompson, "Encoding techniques in predictive agriculture," *Computers and Electronics in Agriculture*, vol. 202, p. 105934, 2023.

[22] I. D. Mienye and T. G. Swart, "A comprehensive review of deep learning: Architectures, recent advances, and applications," *Information*, vol. 15, p. 755, 2024.