Advances in Artificial Intelligence and Machine Learning; Research (Ahead of Print) Received 20-06-2026; Accepted 01-11-2026; Published online 08-11-2026

Predicting Financial Habits and Saving Propensity Using Machine Learning and Agile Methodologies

Ulises Roman-Concha nromanc@unmsm.edu.pe
Professors at the Faculty of Systems Engineering and Computer Science, UNMSM, Lima-Peru

Jhonathan Centti jcenti@unmsm.edu.pe
Graduated from the Professional School of Software Engineering, UNMSM, Lima-Perú

José Piedra jpiedrai@unmsm.edu.pe
Professors at the Faculty of Systems Engineering and Computer Science, UNMSM, Lima-Perú

Ronald Melgarejo rmelgarejos@unmsm.edu.pe
Professors at the Faculty of Systems Engineering and Computer Science, UNMSM, Lima-Perú

Carlos Chávez-Herrera cchavezh@unmsm.edu.pe
Professors at the Faculty of Systems Engineering and Computer Science, UNMSM, Lima-Perú

Luzmila Pró lproc@unmsm.edu.pe
Professors at the Faculty of Systems Engineering and Computer Science, UNMSM, Lima-Perú

Pablo Romero promeron@unmsm.edu.pe
Professors at the Faculty of Systems Engineering and Computer Science, UNMSM, Lima-Perú

Deisi Pomajulca dpomajulca@une.edu.pe
Professors at the Faculty of Sciences at the National University of Education, UNE, Lima-Perú

María Elena Ruiz mruizr@unmsm.edu.pe
Professors at the Faculty of Systems Engineering and Computer Science, UNMSM, Lima-Perú

Augusto Cortez acortezv@unmsm.edu.pe
Professors at the Faculty of Systems Engineering and Computer Science, UNMSM, Lima- Perú

Corresponding Author:

Copyright © 2026 Ulises Roman-Concha, et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Citation: Ulises Roman-Concha et al. Predicting Financial Habits and Saving Propensity Using Machine Learning and Agile Methodologies. **Advances in Artificial Intelligence and Machine Learning**. 2026. [Ahead of Print]. https://dx.doi.org/10.54364/AAIML.2026.66356

Abstract

Many economically active individuals struggle to manage their personal finances due to a lack of personalized technological tools that enable them to anticipate financial behavior. This article addresses this issue by presenting a machine learning model to predict financial habits and saving propensity. A hybrid methodology was used that integrates CRISP-DM (Cross-Industry Standard Process for Data Mining) to structure the data science process and a framework for agile software development (SCRUM) for agile project management. Heuristic rules were defined to create two variables: "Financial Habits Level" (multi-class, 1–5) and "Savings Propensity" (binary), based on sociodemographic and financial data. Seven supervised classification algorithms were compared, and XGBoost was selected as the best algorithm, as it uses advanced techniques such as SMOTE (Synthetic Minority Over-Sampling Technique) and a machine learning hyperparameter optimization technique to find good combinations of a model's hyperparameters by randomly testing different configurations and evaluating them via cross-validation (RandomizedSearchCV), and then rigorously evaluated using an independent test set. The "Financial Habits Level" model achieved an accuracy of 80%, while the "Savings Propensity" model achieved an accuracy of 94.5% and a ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) of 0.9920, demonstrating high predictive power.

Keywords: Financial habits, Propensity, Machine learning, prediction, CRISP-DM, Scrum, XGBoost.

1. INTRODUCTION

Various studies have shown that combining structured methodologies, such as CRISP-DM and SCRUM, yields consistent results in the development of predictive and management systems [1, 2]. They report successful experiences in applying CRISP-DM for disease classification, idea mining, etc., and emphasise the importance of a rigors process in data preparation and modelling. Conversely, [3] suggests a modification to CRISP-DM that aims to guarantee high quality in the analytical process, further validating the flexibility and applicability of this approach in data science.

Machine learning models are being applied in the financial sector [4, 5] to develop more accurate predictions of customer behavior and price forecasting. Their reported outcomes indicate trat predictive systems have the potential to substantially improve the decision-making processes of financial institutions by providing more accurate estimates on customer behavior and requirements.

Today, individuals from diverse socioeconomic backgrounds face difficulties in efficiently managing their personal finances. Users' ability to understand and predict their own financial habits is hampered by variations in financial literacy, behavioral biases, and the absence of accessible, personalized digital tools. Dedicated IT tools for analysis, visualization and predictive modelling are often short-sighted, simply collecting financial data while overlooking socioeconomic diversity, thus limiting their ability to guide real-world decisions that are static and limited across a range of environments. The strength of an intelligent platform and machine learning models is in the balance between the data rigour (quality), interface usability and the unprecedented hybridization of methodologies (CRISP-DM and SCRUM). This integration between data science and data engineering is necessary, otherwise the predictive and educational capabilities are stunted.

The present proposal develops an advanced predictive model (XGBoost, Random Forest and ANN) adapted to the socioeconomic context of the city of Lima. The value is in the balance of predicting financial habits, with the use of advanced analytical capabilities to make a direct human impact through education and financial self-management for users that are using financial services on a continuous basis.

2. RELATED WORK

The scientific literature has increasingly addressed the application of machine learning methods to predict financial behavior in recent years. Initial studies have shown that machine learning models are capable of identifying patterns in spending decisions, debt accumulation, and consumer preferences through economic and behavioral characteristics. Thus, in [4, 6], and [7], neural networks have been used to analyze cognitive processes related to financial decisions, while more traditional applications, such as the rejection inference methods used in credit scoring, remain relevant due to their statistical robustness [3]. Studies employing deep learning approaches to financial data and external economic data achieved significantly higher accuracy but with lower interoperability [5, 8].

Likewise, various studies have included the analysis of data from social media, financial news, and transaction records. In [7], it is explored that financial transactions reflect users' economic trends, while [9] uses models to extract relevant information from the web and create predictions based on textual patterns. On the other hand, it is known that machine learning can generate, through its algorithms, factors associated with saving, planning, propensity, and financial risk, as can be seen in the works of [10] and [11]. However, many of these studies are conducted on global variables, which limits their generalizability to local contexts.

A systematic review of the literature on tools and methodologies for integration and application in the field of data mining and software engineering is important. In [12], an explicit integration between CRISP-DM and SCRUM is proposed, enabling the articulation of iterative cycles of modeling, evaluation, and deployment. The papers [1] and [2] discuss the difficulties in managing machine learning projects and emphasise the need for continuous iterations to adapt models to new data used for testing. These methodological proposals are of great importance for predictive systems, although until now there are few practical cases in the financial domain.

Regarding the architectures and technologies studied, layered architectures [7, 13], API-based services to connect models with user interfaces [14, 15], and microservices to allow predictive components to scale independently [5, 16] are preferred. Implementations generally use Python for modeling, web frameworks such as Django or Flask, SQL/NoSQL databases, and MLOps tools that ensure automation and traceability [6, 13, 17]. Additionally, there has been growing interest in security and privacy mechanisms [18], which are particularly relevant for handling sensitive financial data.

Significant limitations to the deployment of financial predictive models in real-world contexts include the lack of reliable local data [19, 20], regulatory constraints regarding the use of user-sensitive personal data [11, 21], and gaps in digital and financial literacy [10]. Technological and infrastructure issues have also been extensively documented in [22, 23], while other studies warn

of the lack of sustainability of the developed models due to budgetary constraints for creation and maintenance [3, 14].

Conclusion: The literature reviewed provides evidence that machine learning can substantially aid in the understanding and prediction of financial behaviors. There is consensus that in order to have an impact, these solutions require methodological and contextual adaptation. Unlike previous studies, this research integrates the most recent financial behavior theories, merges CRISP-DM and SCRUM to establish an instructional design for model development, and contextualizes the framework within the Peruvian setting of Metropolitan Lima, fulfilling the methodological, infrastructural, and technological requirements for framework adoption.

3. METHODOLOGY

In this research, a robust machine learning model was created based on solid practices from CRISP-DM and SCRUM agile methodologies, as described in [24], to predict the residents of Metropolitan Lima's financial habits and saving tendencies. The model offers empirical personalized feedback and aids users in effectively and easily analyzing behavioral financial patterns in order to assist users in the optimization process (FIGURE 1).



Figure 1: CRISP-DM phases within the SCRUM framework [25]

The model estimates saving propensity and financial habits in the city of Lima using transactional and sociodemographic data, identifying patterns to deliver automated and personalized assessments. It integrates pre-processing, feature engineering and the XGBoost algorithm (tuned with SMOTE and Grid/RandomizedSearchCV),m within a CRISP-DM/SCRUM methodological framework that ensures a robust and flexible development process.

3.1 The CRISP-DM Methodology for Model Development

3.1.1 Business understanding

Two predictive tasks in this phase:

1. Develop a model using multiclass classification that assigns financial habits levels to users on an ordinal scale. The levels would range from '1 = Very Poor' to '5 = Very Good.'. This will help segment user behavior.
2. Binary classification: develop a model to predict the user's saving propensity (0 = low propensity, 1 = high propensity). This will facilitate the identification of specific profiles for making recommendations or carrying out interventions.

The measure of success will be the predictive ability of the model. Later phases will use Accuracy and F1-Score as metrics to evaluate the success. A first plan is defined, which incorporates CRISP-DM phases within an agile development framework (SCRUM). Specific sprints will be dedicated to the CRISP-DM phases of Understanding, Preparation, Modelling and Evaluation, allowing for iterative and incremental development of the model.

3.1.2 Data understanding

The analysis process begins from the initial data collection, being a stage in which the need to obtain representative information of the population of Metropolitan Lima is identified. These data were obtained through a survey of key sociodemographic and financial characteristics, including age, gender, education level, income, expenses, savings, debt, and use of financial products.

This is followed by a detailed description of the dataset where the types of variables available are considered and their behaviour analysed using frequency distributions, graphs such as histograms and boxplots as well as basic statistics that enable an understanding of general trends and levels of variability.

In the exploratory phase, correlation matrices for numerical data and contingency tables for categorical data are used to investigate preliminary relationships between variables. The aim is to identify relevant predictors of financial habits or the saving propensity.

The dataset was collected using a questionnaire. To evaluate the models, 20% of the data (500 records) was set aside; these records were not used for training or cross-validation and retained a stratified distribution of classes. This procedure allowed for an unbiased evaluation of the models

3.1.3 Data preparation

This is a key stage where raw data are transformed to be used by machine learning algorithms.

- Data selection: The final variables to include in the modelling are selected based on the results of the exploration and knowledge of the personal finance field. Variables that are redundant or have little predictive power can be dropped.
- Data cleaning: The defined strategies are applied to handle missing data (e.g., by imputation with mean or median, or by removal of records) and outliers (e.g., by winsorization, removal, or logarithmic transformation). Consistent formats (e.g. dates and currencies).
- Data construction (feature engineering):
  - Transformation of existing variables: Categorical variables are converted to numerical form. If the variable is nominal, e.g. Gender or Education Level, we use one-hot encoding (i.e. dummy variables). Ordinal variables, such as "Card Usage Frequency," are mapped to whole numbers (from 1 to 5). Absolute monetary variables (e.g., "monthly savings" or "monthly debt") are converted to income-relative variables (percentages) so that the information is normalised and comparison across individuals with different income levels is possible.
  - Creation of target variables (application of heuristic rules): an important step to turn business knowledge into measurable variables.
    - Financial Habits Level (Score_H ? ordinal 1–5): A heuristic formula is defined that weights positive factors (such as a high percentage of savings) and negative factors (such as a high percentage of debt and a high frequency of credit use). The result (SCORE_H) is transformed into a balanced ordinal variable by dividing the distribution into quintiles (each level represents approximately 20% of the sample).
    - Saving propensity (Score_A ? binary 0-1): another heuristic formula is defined that positively combines the percentage savings rate and budgeting discipline and negatively combines the debt level. SCORE_A, or the raw score, is transformed into a balanced binary variable using the median as the cutoff point (approx. 50% in class 0 and 50% in class 1).
- The final dataset contains only the transformed predictor variables and the two created target variables ("Financial Habits Level" and "Savings propensity"). Dummy, ordinal, and percentage variables are included in the final dataset. Intermediate variables for calculation purposes only (e.g., SCORE_H and SCORE_A) are excluded in the final dataset. The final dataset is then exported to an appropriate format (CSV) to prepare for the next phase in the project (modeling).

3.1.4 Modeling

In this phase, predictive models are selected, constructed, and subjected to an initial evaluation.

- Selection of modeling techniques: A family of supervised classification algorithms is selected. For this specific problem, a set of models is chosen to enable comparative evaluation of their performance:
  - *Linear:* Logistic Regression (noted for its interpretability).

  - *Tree-based models:* Decision Tree (interpretable), Random Forest (robust), Gradient Boosting, XGBoost (high performance).
  - *Others approaches: K-Nearest Neighbors (KNN), Support Vector Machine (SVM)*.

- Test Design Generation: A rigorous protocol is defined to ensure unbiased evaluation of the models:
  - Train-Test Split: The prepared dataset divided into a training set (80%) and a test set (20%). The split must be stratified with respect to the target variable to preserve class proportions in both subsets. A fixed random seed (random_state) is used to guarantee the reproducibility of the results.
  - Cross-Validation: A 5-fold stratified cross-validation procedure is applied to the training set to obtain a more robust estimate of model performance and to support hyperparameter tuning. Stratification is also required cross-validation to maintain the distribution of the target variable across folds.

- Model Construction:
  - Baseline Training: Train all selected algorithms on the training set with their default hyperparameters.
  - Dealing with Imbalance: The target variables were intended to maintain some degree of equilibrium. If significant residual inequality is identified, particularly prior to the conclusive optimization, methods such as SMOTE (Synthetic Minority Over-sampling Technique) on the training data are implemented.
  - Hyperparameter Optimization: For the remaining selected algorithms (such as XGBoost), utilize GridSearchCV (complete enumeration of the hyperparameters) or RandomizedSearchCV (random search, more efficient while exploring large spaces) to define the optimal hyperparameter (such as learning_rate, n_estimators, max_depth) configuration that improves the target evaluation metric with cross-validation (Accuracy, F1-score).

- Preliminary Evaluation: Model's performance metrics must be evaluated on an ongoing basis. Metrics for this problem include Accuracy, Precision, Recall, F1-score, and Classification Behavior as derived from a confusion matrix. For probabilistic models, the model's discriminative ability was evaluated using the AUC-ROC.

3.1.5 Evaluation

This phase focuses assesses the final model(s) in terms of how well they business objectives.

- Evaluation of Results: Each target variable's final optimized model is assessed using the test set, which includes data the model is unfamiliar with. Relevant metrics are calculated:
  - For Variable 1 (Multiclass): Overall Accuracy and the evaluation metrics Precision, Recall, and F1 per class, along with Macro and Weighted averages, and Confusion Matrix are calculated.

  - For Variable 2 (Binary): Accuracy, Precision, Recall, F1-Score, ROC-AUC, and Confusion Matrix are computed.

- Process Review: Compare outcomes produced on the test set with the success criteria outlined in Business Understanding. Here, we assess if the performance is satisfactory for the job in the application (e.g. whether 80% accuracy is sufficient to issue financial recommendations). We analyse the previous phases to assess the model's strengths and weaknesses (e.g. in which scenarios do the predictions perform well/poorly?). Which variables do you think are the most important?
- Determining Next Steps:
  - Deploy: If the model satisfies the set criteria.
  - Iterate: If the model is in need of enhancements (for example, return to Data Preparation to generate additional features, experiment with more advanced and intricate model architectures, gather supplementary data).
  - Abandon: If the model perpetually cannot meet even the minimally acceptable performance level.

3.2 Data Labeling and Heuristic Rules

This section elucidates the core conceptual frameworks that guide the Data Preparation phase of CRISP-DM, particularly Data Construction (Feature Engineering). It identifies the input variables and the heuristic rules established for the construction of the preliminary target variables.

3.2.1 Analysis variable

Using domain knowledge and literature, we define an initial set of predictor variables (features) relevant to modeling the propensity to save and financial behaviors. Based on the external literature and expert knowledge of the factors that determine the propensity to save and financial behaviors, we group the variables into:

- Demographics:
  - Age: A relevant factor in the financial life cycle (for example, young people may have different patterns than older adults). Conceptual range: 18–65+ years.
  - Gender: May influence financial behaviors due to social or economic factors. Categories: Male, Female, Other.
  - Education Level: Often correlates with income level and financial literacy. Categories: From completed primary school to completed university, including intermediate levels.
- Financial:
  - Monthly Income: Something that plays a large role in how much a person can actually start to save or spend is the potential range (from 800 to 8,000 + soles) that accounts for the different socioeconomic levels in Metro Lima.

  - Other Raw Financial Variables: To create and change target variables, semi-quantitative variables like approximate monthly savings, approximate monthly debt, approximate monthly leisure spending, number of credit cards owned, budgets expenses, frequency of credit card use, and pays only the minimum amount of the credit card will be included in the study.

3.2.2 Labeling criteria

Heuristic rules are established for monthly income thresholds to derive target variables (Financial Habits Level and Saving Propensity). The thresholds tend to align with general recommendations on finance and credit risk assessment, but are modified to the specific context.

- Savings Level (% of Income): The capacity for saving is partitioned into five ordered categories.
  - Low: 0% - 8%
  - Medium-Low: >8% - 12%
  - Medium: >12% - 20%
  - High: >20% - 30%
  - Very High: >30%
- Debt Level (% of Income): The five gradational levels also classify debt burden.
  - Low: 0% - 10%
  - Medium-Low: >10% - 15%
  - Medium: >15% - 30%
  - High: >30% - 40%
  - Very High: >40%
- Other Criteria (Transformed Qualitative Variables): Key behavioral variables which have been transformed to ordinal or binary scales are included:
  - Estimated Expenses: Frequency (such as 'Never' to 'Very Frequent') is transformed into a numerical scale (1 to 5). Higher frequency indicates better spending habits.
  - Credit Card Use: Frequency is also transformed into a numerical scale (1 to 5). Very high frequency may indicate credit dependency or lack of liquidity (negative habit).
  - Minimum Credit Card Payment: The frequency of paying only the minimum is transformed into a numerical scale (1 to 5). Frequently paying only the minimum is a negative indicator.
  - Level of Leisure Spending: This is conceptually categorized (as Low to Very High) as an indicator of control over discretionary spending.

These transformed and categorized variables are the direct inputs for the heuristic formulas (SCORE_H and SCORE_A) defined in section 4.2.3 to create the final target variables ("Financial Habits Level" and "Saving Propensity").

3.3 Machine Learning Model Construction Process

The labelling of the data and the creation of the target variables are based on an initial logical basis of heuristic rules. The final prediction is proposed to be made by the central component of a supervised machine learning model for various conceptual reasons. The purpose of the model development is to overcome the limitation of the heuristics used in the first labelling phase, because these rules are simplifications and do not always sufficiently reflect the actual complexity of the financial behaviour, but are useful as a starting point. Thus the model tries to learn deeper and more subtle patterns directly from the data which gives a better understanding of the users' financial habits. For this reason, a supervised learning approach is applied, where the model learns the relation between the predictor variables (obtained by transforming the demographic and economic characteristics) and the target categories generated from heuristics. Such labels are a starting point, but the model should not simply replicate the classification process, but should also generalise it, exploiting its capacity to detect complex and non-linear interactions that cannot be captured by fixed rules.

3.4 Evaluation Method

3.4.1 Test dataset

The final evaluation was performed on the held-out test set (20% of the entire data) that was not used in any of the aforementioned training or cross-validation. The test set contains 500 records and maintains the stratified class distribution of the first split. The test set provides an evaluation of the models' generalization performance that is free of bias.

3.4.2 Evaluation metrics

To assess how well the classification models performed, standard evaluation metrics were applied. Each metric was conceptually defined and computed using the predictions generated from the test set:

1. Accuracy: The overall share of correctly instances relative to the total number of observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. Precision: What proportion of correct predictions were positive? Significant in reducing false positive predictions. (Calculated per class).

$$Precision = \frac{TP}{TP + FP}$$

3. Recall: What proportion of actual positive cases were correctly identified? Important for reducing false negatives. (Calculated per class).

$$Recall = \frac{TP}{TP + FN}$$

4. F1-Score: The harmonic mean of Precision and Recall. It is most helpful in cases of unbalanced classes.. (Calculated per class and as averages).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

5. Macro and Weighted Averages (for F1, Precision, Recall):
   - Macro: Simple average of the metric computed for each class (disregards number of instances in a class).
   - Weighted: Average of the metric with samples in each class counted as a class size (larger classes contribute more to the average).
6. Confusion Matrix: A matrix providing a description of performance with respect to each class. It outlines True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).
7. ROC-AUC (Area Under the ROC Curve - For Binary Model 2 Only): Assesses the overall performance of the model in differentiating the positive and negative classes. A score nearing 1 suggests outstanding discrimination.

4. RESULTS AND DISCUSSION

This section presents the quantitative results of the evaluations of the machine learning models predicting "Financial Habits Level" (Model 1, multi-class) as shown in TABLE 1, and "Saving Propensity" (Model 2, binary). As part of the described experiment, we assessed the performance and accuracy of the selected models (XGBoost optimized with SMOTE and RandomizedSearchCV in both cases) on an independent test data set.

4.1 Results

Model 1 Results: Financial Habits Level (Multi-class). The following table shows the performance metrics of the optimized final model (XGBoost + SMOTE + RandomizedSearchCV) for predicting "Financial Habits Level" (scale 1–5) on the test set (Figure 2).

Table 1: Performance Metrics - Model 1 (Financial Habits Level)

Class	Precision	Recall	F1-Score	Support
1	0.96	0.93	0.94	55
2	0.72	0.78	0.75	40
3	0.70	0.75	0.72	51
4	0.61	0.61	0.61	23
5	0.96	0.84	0.90	31
Accuracy			**0.80**	**200**
Macro Avg	**0.79**	**0.78**	**0.78**	**200**
Weighted Avg	**0.81**	**0.80**	**0.80**	**200**

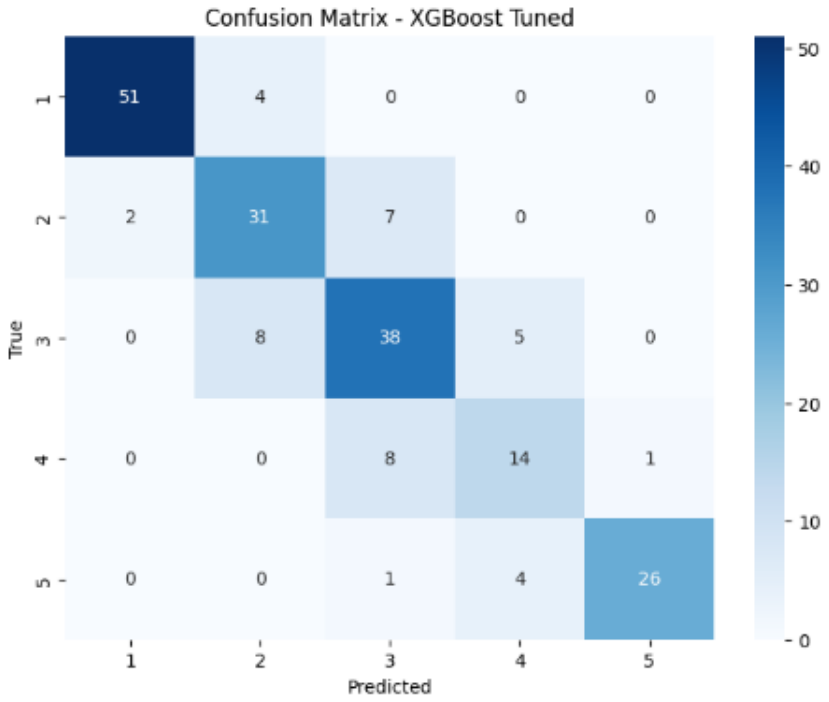


Figure 2: Confusion Matrix - Model 1 (Financial Habits Level)
Note. Prepared by the authors based on the results of Sprint 3.

The model, based on a tuned XGBoost, shows a high performance in the multiclass classification of financial habits. From the confusion matrix, it can be seen that most of the observations fall on the main diagonal, indicating correct classifications across the five evaluated categories. Class 1 is well done by the model (51 correctly classified). Class 2 and 3 are also steady, showing only some weakness in differentiating similar financial profiles (which is expected) (Figure 3).

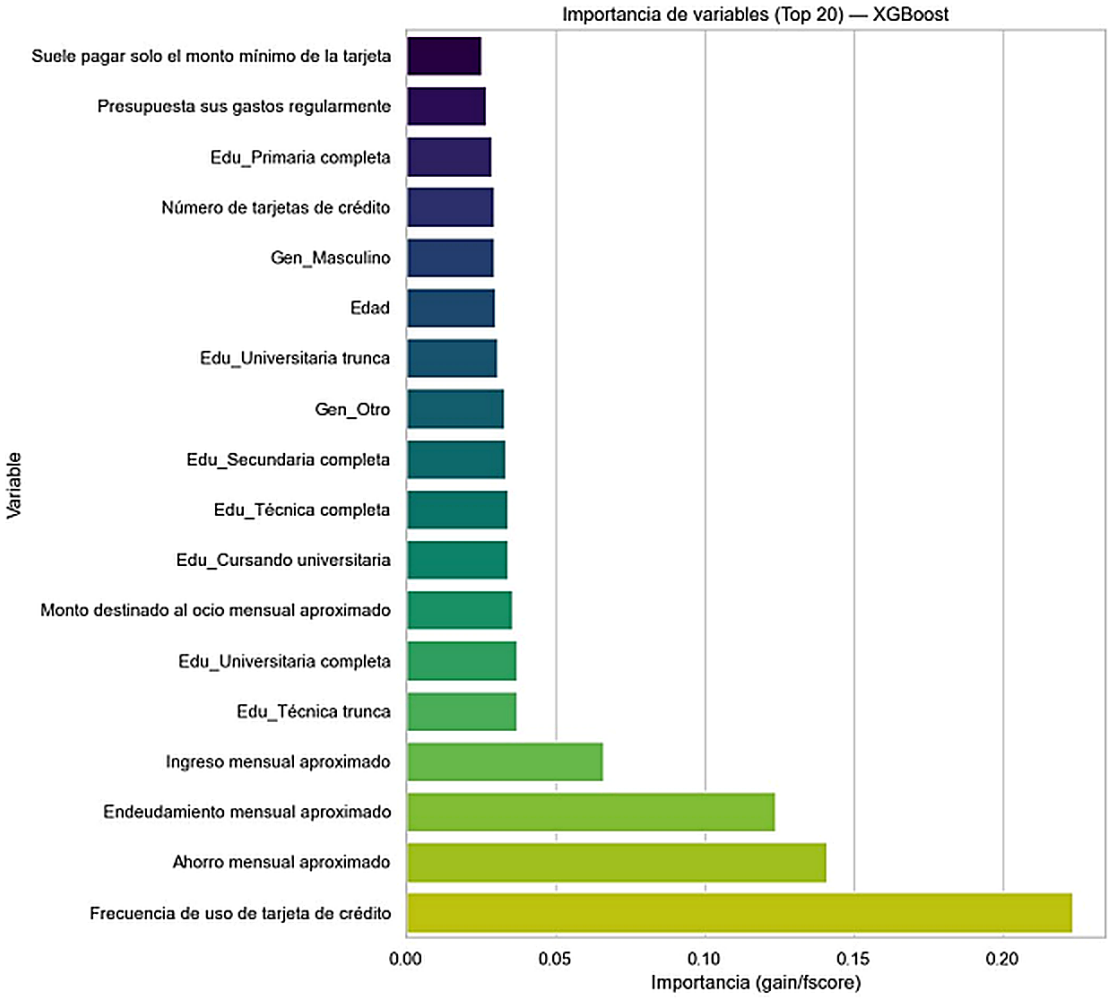


Figure 3: Importance of variables in Model 1

The most important predictors of the model are the frequency of credit card use and monthly estimates of savings, debt and income These results are consistent with studies on the financial behaviour of companies in this sector . These variables have the highest gain values which means that they have the highest contribution to the error reduction in the model training . Similarly, the sociodemographic variables such as educational level, age and gender have a moderate effect, which indicates that although they give some context, variables directly related to financial behaviour are more important in the predictive power of the model (FIGURE 4).

Below are the performance metrics of the final optimized model (XGBoost + SMOTE + RandomizedSearchCV) for predicting “Savings Propensity” (classes 0 and 1) on the test set (TABLE 2).

Table 2: Performance Metrics - Model 2 (Saving Propensity)

Class	Precision	Recall	F1-Score	Support
0	0.95	0.94	0.94	99
1	0.94	0.95	0.95	101
Accuracy			**0.94**	200
Macro Avg	0.95	0.94	**0.94**	200
Weighted Avg	0.95	0.94	**0.94**	200
ROC-AUC Score			**0.9920**	

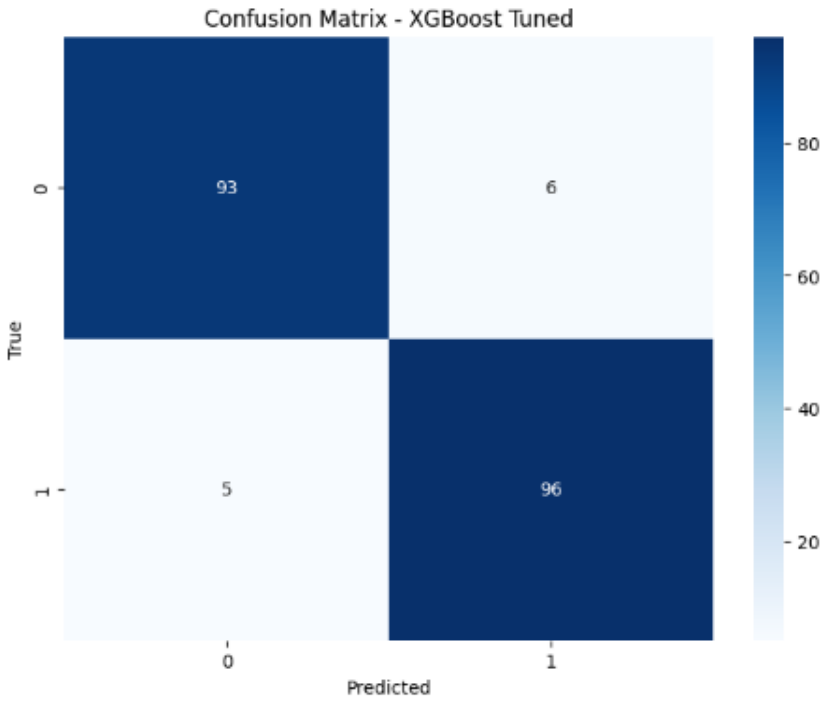


Figure 4: Confusion Matrix - Model 2 (Saving Propensity)
Source: Prepared by the authors based on the results of Sprint 4.

The proposed model is formulated as a binary classification problem and very satisfactory performance is obtained. The confusion matrix indicates good discriminative power with 96 true positives and 93 true negatives against a small number of errors (6 false positives and 5 false negatives) (FIGURE 5).

The evaluation metrics confirm the robustness of this approach with the overall accuracy of 94.5%, precision of 94.1% and sensitivity of 95.0%. The F1-score of 94.5% indicates a good trade-off between precision and recall. This implies that the model correctly detects the majority of positive cases with a few classification errors. The ROC curve is well above the diagonal reference line

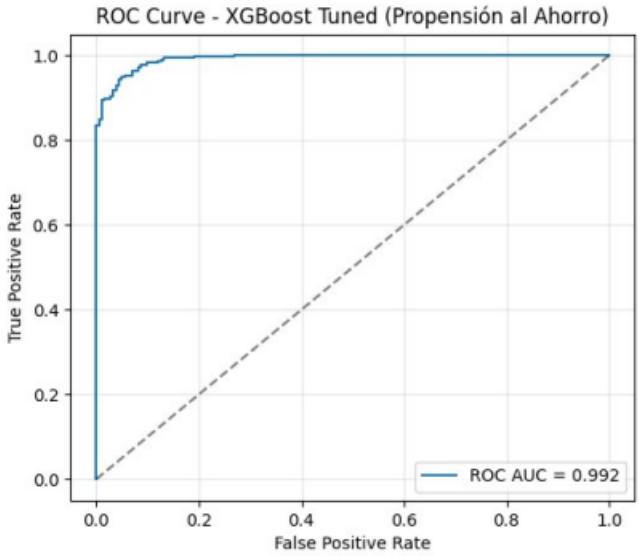


Figure 5: ROC Curve - Model 2*Source: Prepared by the authors based on the results of Sprint 4.*

(random classifier), demonstrating its strong discriminative ability. As shown in FIGURE 6, the area under the curve (AUC) is 0.992, which indicates a near-perfect classification ability in differentiating between individuals with high and low saving propensity.

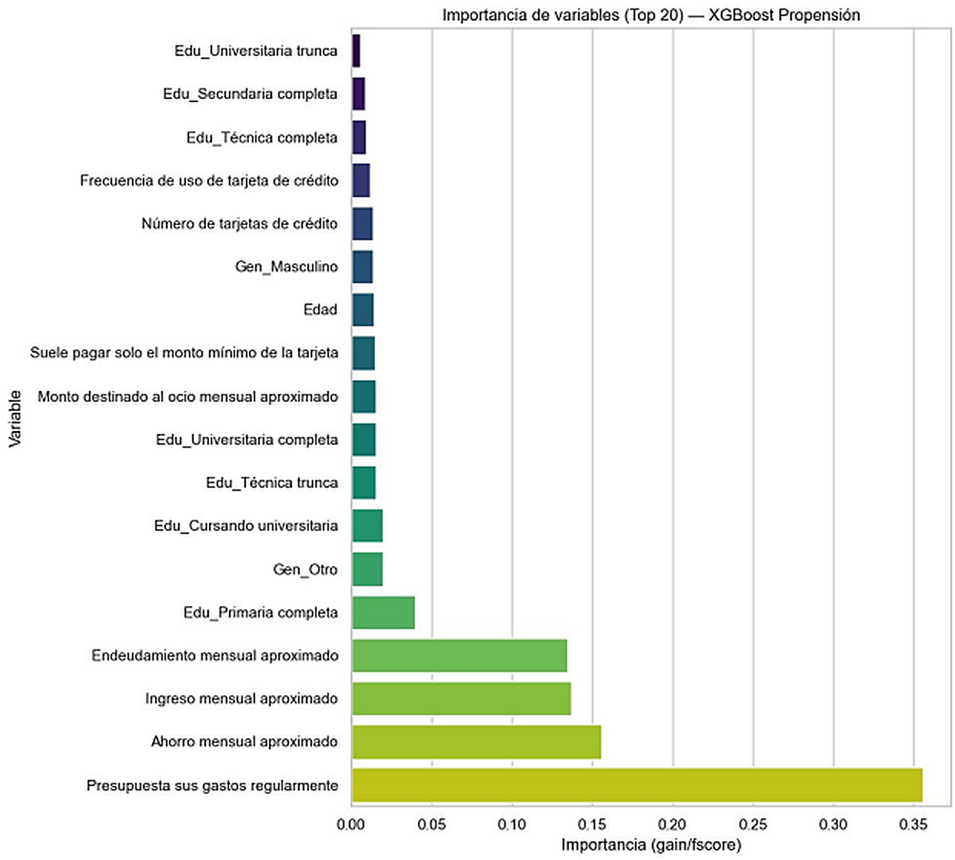


Figure 6: Importance of Variables - Model 2
Source: Prepared by the authors based on results of Sprint

The most important predictor of saving propensity is financial planning, particularly "budgeting expenses regularly", as measured by the "gain" metric in XGBoost, followed by the other quantitative economic variables (savings, income, and monthly debt); sociodemographics, on the other hand, contribute marginally to the performance of Model 2.

4.2 Discussion

The variables "financial habits level" (multiclass) and "saving propensity" (binary) were effective, indicating different patterns and reflecting the complexity of each prediction. Model 2 (saving propensity) had excellent and highly robust performance with 94.5% accuracy and 0.9920 ROC-AUC. This success is due to its lower structural complexity as it is a binary classification. Nevertheless, the accuracy of Model 1 (financial habits level) was 80%, a good result, but lower because the boundaries between the intermediate financial behaviours are blurred. The model performs best at the extremes, reaching an F1-score of 94% in the lower class, and 90% in the upper class, where

behavioural patterns are stable and clear. but it shows weakness in the middle ranges, especially in class 4 (F1 = 61%); this is due to the ambiguity and heterogeneity of the financial profiles of each customer.

On the other hand, the variables indicate that each financial behaviour is a response to psychological and behavioural mechanisms of the different users. This causes the predictors to be different between the two models, as the most influential variable was "Budget your expenses regularly."

This suggests that disciplined financial planning for savings is the strongest predictor, so it is evident that financial education is important as it recognises intention and organisation as consistent predictors of healthy financial behaviours. In the case of Model 1, frequent credit card use is the dominant predictor, showing that financial habits are linked to debt and liquidity management. Thus, it is clear that habits stem from daily credit, loan, and transaction practices, while saving depends on financial planning.

The choice of XGBoost as the primary model is consistent with [26], due to its effectiveness in financial forecasting and its ability to model complex and nonlinear relationships. Likewise, Random Forest demonstrates consistent performance, in line with [27], which highlights its robustness and stability when dealing with multiple predictor variables.

The results obtained are consistent with current trends in financial predictive systems; the use of the XGBoost algorithm is supported in [8] due to its effectiveness with heterogeneous data and nonlinear relationships. Regarding the performance of Random Forest as a robust alternative with multiple variables, it aligns with [6] and [17]. Consequently, the accuracies achieved (80% in multi-class prediction and 94.5% in binary) technically validate the approach proposed in this study. The hybrid CRISP-DM and SCRUM approach proved its viability for the continuous iteration of the study, consistent with [9] and [24], authors who advocate the use of agile cycles to strengthen the feedback loop between analysis, visualization, and predictive modeling of large volumes of data (Big Data). The experience of this study confirms its effectiveness by having facilitated the constant adjustment of the models' parameters and hyperparameters.

5. CONCLUSIONS

1. During the experimentation and optimization of multi-class and binary classification (SMOTE and RandomizedSearchCV), it was found that XGBoost delivered the best performance and robustness for both variables, achieving an accuracy of 80% for the multi-class model and 94.5% for the binary model.
2. The accuracy of the final predictive models was rigorously verified using an independent test set (20% of the data) and standard classification metrics. The model for "Financial Habits Level" achieved an accuracy of 80% and a weighted F1-score of 80%, while the model for "Savings Propensity" achieved an accuracy of 94.5%, a weighted F1-score of 95%, and an ROC-AUC of 0.9920, validating its high predictive capability.
3. The future development based on this article would be to implement the Explainable Artificial Intelligence (XAI) techniques such as SHAP or LIME to provide clear explanations of the model's predictions and thus increase its educational value. In addition, we propose to develop a personalised recommendation module for recommending practical actions to improve

users' financial habits. We will explore the use of deep learning architectures such as neural networks and LSTMs to find more complex relationships in large data volumes. Finally, a cloud deployment strategy based on MLOps will be studied to ensure the model scalability, security and continuous monitoring as its use increases.

References

[1] https://www.linkedin.com/pulse/crisp-dm-agile-scrum-methodology-data-science-project-abdollazadeh/

[2] https://atom-academy.mx/uniendo-scrum-y-crisp-dm/

[3] Ayele WY. Adapting Crisp-Dm for Idea Mining: A Data Mining Process for Generating Ideas Using a Textual Dataset. Int J Adv Comput Sci Appl. 2020;11:20-32.

[4] Berrospi GJC. Scrum Framework for Software Project Management in a Private Financial Institution, Lima 2023 [Bachelor's thesis]. César Vallejo University, 2023. Available URL: https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/128906/Caceres_BGJ-SD.pdf

[5] Calero M, Baldeon Calisto M, Bonilla S, Riofrio D. Application of machine learning algorithms for the prediction of payment by agreement in a debt collection company with the CRISP-DM methodology. In 3rd South American International Conference on Industrial Engineering and Operations Management. IEOM Society. 2022.

[6] Guamán Yunga RE. Development and Implementation of a Customer Classification Model Based on Probability of Retention [Master's Thesis]. National Polytechnic School, 2024. Available URL: https://bibdigital.epn.edu.ec/bitstream/15000/26141/1/CD

[7] Salazar EP. Scrum Methodology in the Management of Information Technology Projects at a Financial Institution, Lima 2024 [Master's Thesis]. César Vallejo University, 2024. Available URL: https://repositorio.ucv.edu.pe/bitstream/handle/20.500.12692/159302/Palomino_SE-SD.pdf?sequence=1isAllowed=y

[8] Cifuentes Ortega MF, Muñoz Samboni DA. BISSMES: An Agile Method for Implementing Business Intelligence in Software Development Companies [Master's thesis]. Technological Institute of Antioquia, University Institution. 2021. Available URL: https://dspace.tdea.edu.co/handle/tdea/1944

[9] Mejía Rodríguez CA, Rincon Pinzon MA, Palmera Quintero LM, Arevalo Vergel LM. Application of Machine Learning and the Crisp-Dm Methodology for the Accurate Classification of Severity in Dengue Cases. Rev Colomb Tecnol Avanz. 2024;1:78-85.

[10] Kunaifi A, Akbar AN. Financial Knowledge and Millennials' Investing Behaviors in Indonesia. Jurnal Sosial Humaniora. 2019;0:26.

[11] Idris NH, Wee MF, Ismail S, Abd Samad K. Financial Knowledge, Debt Literacy and Over-Indebtedness. J Soc Sci Res. 2018:22-29.

[12] Baijens J, Helms R, Iren D. Applying scrum in data science projects. 2020 IEEE 22nd conference on business informatics (CBI). IEEE. 2020;1:30-38.

[13] Olivera Cokan CA. Framework for the Development of Data Analytics Projects [Master's thesis]. Pontifical Catholic University of Peru, 2024. Available URL: https://tesis.pucp.edu.pe/server/api/core/bitstreams/bb1d219a-4612-43ac-b8bd-bd47e1c74438/content

[14] Rojas Pecho PY, Acevedo Mortola MA. Proposal for the design of a machine learning model to estimate costs in software projects [Master's thesis]. Peruvian University of Applied Sciences, 2025. Available URL: https://repositorioacademico.upc.edu.pe/entities/publication/e4610758-95a0-52e0-b069-02f686effb32

[15] Reyes-Riveros AJ, Castillo-Sarmiento JM, Santos-Fernández JP, Alcántara-Moreno OR, Sánchez-Ticona RJ. Web System for the Dynamization of Inventory Management and Marketing Strategies in Peruvian Supermarkets. Rev Cient Sist Inform. 2024;4:e673.

[16] https://medium.com/@jeremychrist23/xgboost-explained-the-go-to-algorithm-for-predictive-modeling-9e7a7c2036c3

[17] https://builtin.com/articles/crisp-dm-data-science

[18] https://agile-spain.org/

[19] Attanasi ED, Coburn TC. Random Forest. In: Daya Sagar BS, Cheng Q, McKinley J, Agterberg F, editors. Encyclopedia of Mathematical Geosciences. Encyclopedia of Earth Sciences Series. Springer. 2023:1182–1185.

[20] Kumar NA, Jangale L, Sathe V, Shelke A, Redij T. Study of Supervised Logistic Regression Algorithm. Alochana J. 2024;13:227-230.

[21] https://www.forbes.com/sites/forbesfinancecouncil/2020/04/14/dont-overlook-these-15-important-financial-health-habits/

[22] https://www.ibm.com/mx-es/topics/postgresql

[23] Kraut N, Transchel F. On the Application of Scrum in Data Science Projects. In 2022 7th International Conference on Big Data Analytics. IEEE. 2022:1-9.

[24] https://www.hiberus.com/crecemos-contigo/adaptacion-de-scrum-para-business-analytics/

[25] Amirian, E., Abdollahzadeh, A., and N. Sulaiman. "Synergizing Hybrid Agile-Scrum and CRISP-DM Approaches in Data Science Project Management." Paper presented at the SPE Canadian Energy Technology Conference and Exhibition, Calgary, Alberta, Canada, March 2024. doi: https://doi.org/10.2118/218114-MS

[26] Zhao T, Chen G, Suraphee S, Phoophiwfa T, Busababodhin P. A Hybrid TCN-XGboost Model for Agricultural Product Market Price Forecasting. PLOS One. 2025;20:e0322496.

[27] Heo ML. Development and Validation of Machine Learning Models to Predict Prediabetes Using Dietary Intake Data in Young Adults in Korea: A Cross-Sectional Study. J Korean Biol Nurs Sci. 2024;26:300-310.