



ALGORITHMIC-STATISTICAL MODEL UNDER A COMPUTATIONAL SCHEME FOR FORECASTING INSOLVENCY UNDER FINANCIAL RISK

Alexander Haro-Sarango*¹

¹Unidad de Posgrados en Ciencias Económicas, Universidad Nacional de Trujillo, Av. Juan Pablo II s/n – Ciudad Universitaria, Trujillo, Perú

¹<https://orcid.org/0000-0001-7398-2760>^{id}

Email: *aharos@unitru.edu.pe

ARTICLE INFO

Article History

Received: January 8, 2026
Reviewed: February 10, 2026
Accepted: March 22, 2026
Published: April 30, 2026

Keywords:

Data;
Supervised Learning;
Financial System;
Financial Risk;
Insolvency.

ABSTRACT

This research focuses on the analysis of insolvency risk in credit unions in Ecuador using a predictive approach based on automatic learning. Based on international and regional precedents on the determinants of financial solvency, the study aims to develop a model capable of classifying credit unions as solvent or not solvent, according to a return on equity (ROE) threshold of 5 %. The methodology adopted was quantitative, explanatory and predictive, using the Random Forest algorithm on a structured database with coded financial variables. The slope variable, called ROE - LOGIC, classifies as “not solvent” (1) those observations with ROE lower than 5 %, and as “solvent” (0) the rest. The model was trained with 80 % of the data and validated with the remaining 20 %. The results show excellent performance, with precision, recall and F1-score metrics above 0.88 in the test set, and an AUC of 0.95, indicating a high discriminative power. The most influential variables were the net interest margin on promised assets, operating expenses relative to net interest margin and the proportion of performing assets. These were used to construct an interpretive formula that estimates the probability of insolvency.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Financial insolvency among resource-intermediating institutions—banks, mutuals, cooperatives, and other financial entities—constitutes one of the most critical risks to economic stability and public confidence. In contexts of high volatility, narrow margins, and increasing regulatory pressure, a gradual deterioration of solvency may not be evident in traditional financial statements, yet it can manifest in subtle patterns of inefficiency, over-indebtedness, or risk concentration. Hence the importance of having forecasting schemes that not only describe the current situation, but also anticipate stress scenarios and enable intervention before losses become irreversible. Historically, insolvency risk assessment has relied on predominantly statistical approaches: financial ratio analysis, construction of solvency indices, linear or logistic regression models, and scoring systems that synthesize risk into a single indicator.

These methods have been foundational and remain an obligatory point of reference; however, their capacity to capture non-linear relationships, high-order interactions, and threshold effects is limited. In practice, the dynamics of insolvency respond to complex combinations of profitability, operating efficiency, asset quality, capital structure, and environmental conditions that are difficult to represent under strict assumptions of linearity or independence. In parallel, recent advances in machine learning have made available algorithms capable of modeling highly complex data structures.

Methods such as random forests, gradient boosting, and deep neural networks have demonstrated remarkable predictive power in classification tasks and risk forecasting. Nonetheless, their adoption in regulated financial contexts has been tempered by two key challenges: model interpretability—that is, the ability to explain why an entity is classified as solvent or non-solvent—and the need to guarantee computational schemes that are reproducible, auditable, and consistent with traditional statistical logic. Against this backdrop, it is pertinent to move towards hybrid “algorithmic–statistical” models that combine the learning power of machine-learning algorithms with interpretable, formalized structures drawn from applied statistics.

From this perspective, the present study proposes an *algorithmic-statistical model under a computational scheme for forecasting insolvency under financial risk*, which integrates three core components: (i) a set of carefully selected and transformed financial indicators, (ii) an algorithmic classification model based on supervised learning techniques, and (iii) the derivation of an interpretable statistical representation that approximates the probability of insolvency from the most influential variables. The model is developed on a database constructed from official financial statements, processed under a computational scheme in Python that includes data cleaning, refinement, normalization of financial ratios, and coding of the solvency target variable. Insolvency is operationalized through a performance threshold on return-on-equity indicators, such that the “non-solvent” class represents risk levels that, although not yet implying formal bankruptcy, do anticipate a trajectory of financial deterioration. On this basis, a classification algorithm is trained to learn non-linear patterns across dimensions of capital strength, operating efficiency, balance-sheet productivity, financial intermediation, and credit risk. The central contribution of the study therefore goes beyond merely “applying an algorithm” to the insolvency problem; it lies in formalizing a complete algorithmic–statistical scheme. First, it is shown that the model achieves robust predictive performance according to criteria such as accuracy, sensitivity, F1-score, and AUC, positioning it as a reliable tool for the early detection of vulnerable entities. Second, variable-importance measures and other internal outputs of the algorithm are exploited to construct an interpretive equation that acts as a proxy for insolvency risk.

While this equation does not replace the model in terms of accuracy, it translates the logic of the “black box” into a statistical structure that is understandable for managers, auditors, and regulators. The relevance of this approach lies in its ability to articulate three needs that often appear disconnected in financial risk management: the need for predictive accuracy, the need for interpretability to support decision-making, and the need for a replicable computational scheme that can be integrated into continuous monitoring systems. From the perspective of financial institutions, this translates into the ability to identify which combinations of intermediation margins, operating expenses, asset unproductiveness, and portfolio quality push the organization toward states close to insolvency. From the regulatory perspective, the model offers an objective and scalable instrument to prioritize supervision, design early-warning alerts, and strengthen system-wide stability. In sum, the article seeks to demonstrate that it is possible to move from descriptive or purely statistical approaches to an integrated algorithmic–statistical model capable of anticipating insolvency under financial risk with high levels of performance and, at the same time, with an explanatory structure that supports managerial and regulatory action. The following sections detail the methodological design of the implemented computational scheme, present the results obtained, and discuss their implications for the theory and practice of insolvency risk management in the financial system.

II. LITERATURE REVIEW

Corporate insolvency is one of the most visible and disruptive manifestations of financial risk, because it concentrates the cumulative effects of illiquidity, excessive leverage, declining profitability and fragile capital structures in a single event with severe consequences for creditors, investors, regulators and the real economy. Classical research in accounting and finance laid the foundations for quantitative insolvency prediction by linking firm failure to accounting ratios. Beaver was the first to show that specific financial ratios, such as cash flow to total debt, exhibit strong discriminatory power between failed and non-failed firms [1]. Building on this, Altman proposed the well-known Z-score model, a multivariate discriminant analysis that combines profitability, leverage, liquidity and activity ratios into a composite index of bankruptcy risk [2]. Ohlson later introduced a logistic regression specification that directly estimates the probability of bankruptcy as a function of financial ratios, relaxing some restrictive assumptions of discriminant analysis and aligning failure modelling with modern probability-based risk management [3]. Beyond these seminal contributions, subsequent work extended the statistical toolbox available for insolvency prediction. Härdle et al. combined variable selection and oversampling techniques with smooth support vector machines (SVMs) to improve default-risk prediction by explicitly addressing class imbalance and high dimensionality [4].

Marassi and Pediroda framed insolvency modelling as a decision problem and proposed a maximum-expected-utility approach, emphasizing that the choice of prediction rule should reflect the asymmetric costs associated with misclassification of solvent versus insolvent firms [5]. Lai et al. developed a support-vector-based metamodel for business risk identification, showing that kernel methods can capture complex, non-linear relations among financial indicators that are beyond the reach of linear discriminant analysis or logistic regression [6]. Collectively, these contributions illustrate a gradual transition from low-dimensional, linear models to more flexible, algorithmically sophisticated approaches within a statistically grounded framework. The rapid development of machine learning (ML) and artificial intelligence (AI) has transformed insolvency and bankruptcy prediction into a broad, data-driven field. Barboza, Kimura and Altman compared several ML models against classical statistical benchmarks and documented substantial gains in out-of-sample performance, especially for ensemble and non-linear classifiers [7]. Sector-specific studies confirm these advantages in distinct institutional settings. According to [8], evaluated artificial neural networks and decision trees for bankruptcy prediction in the agricultural sector, showing that AI models can exploit subtle patterns in sector-specific financial indicators.

In [9], [10], conducted comparative analyses of ML algorithms—including SVMs, random forests, gradient boosting and neural networks—for bankruptcy prediction in different national contexts, such as Pakistani listed companies, emphasizing the role of feature engineering, hyperparameter tuning and robust validation in achieving reliable predictions. In [11],[12], proposed an intelligent, ML-empowered framework for bankruptcy prediction, while Stasko et al. designed a country-specific bankruptcy prediction model for Latvian firms, illustrating how ML pipelines can be adapted to small open economies and local accounting regimes. In parallel, a complementary stream of research has focused on systematic mapping of AI applications in insolvency-related domains and sectoral early-warning systems. Jayawardana et al. conducted a systematic review and empirical feature derivation for AI-based insolvency prediction in the construction industry, identifying key financial and project-level indicators that drive failure in this sector [13]. Reviewed AI applications for proactive dispute management and risk mitigation in construction, highlighting how contractual conflicts, delayed payments and project disruptions can amplify insolvency risk and how early-warning models can support intervention [14].

Mirindi and Mukala offered an in-depth pedagogical exploration of loan default classification using ML techniques, which—although aimed at teaching—illustrates the full pipeline from data preparation to model interpretation in credit-risk classification [15]. Goel et al. discussed the enhancement of financial-management efficiency through advanced predictive analytics and data-driven decision-making strategies, thus situating insolvency and credit-risk models within broader financial-management architectures [16]. The literature has also expanded from firm-level bankruptcy to credit-risk-based early-warning systems that operate at the level of loans, borrowers and portfolios. Hamada et al. proposed an early prediction and detection framework for retail and corporate credit risks using ML algorithms, demonstrating that non-linear classifiers can identify high-risk exposures earlier and more accurately than traditional scoring systems [17]. Teles et al. compared SVM and random-forest algorithms for credit-operation classification, highlighting trade-offs between predictive performance and computational cost [18]. Orlova introduced methodologies and models for individual creditworthiness assessment using digital-footprint data and ML methods, showing that alternative, non-traditional data can substantially improve risk segmentation [19].

In a subsequent work, Orlova developed a data-driven design for credit-risk management based on digital-footprint intelligence, integrating such models into organizational risk processes [20]. In [21],[22] used massive corpora of news articles to diagnose corporate insolvency for credit-management purposes, while combined ML with sentiment analysis to project bank-insolvency risk, embedding textual and market sentiment information into prudential risk models. According to [23] examined how green-credit policy affects commercial banks' credit-risk profiles using ML models fitted to panel data for Chinese banks, thereby illustrating how policy-driven changes in portfolio composition can non-trivially reshape insolvency risk. In [24] compared empirical methods for reproducing global manufacturing companies' credit ratings, showing that ML can approximate rating-agency assessments and provide proxies for insolvency and default risk when external ratings are unavailable. Beyond specific models, recent contributions also promote a broader data-science and risk-analytics perspective.

Lai and Xing provide a comprehensive treatment of data science and risk analytics in finance and insurance, formalizing the statistical and computational principles underlying modern risk models and emphasizing their integration into risk-management processes [25]. Gupta et al. explore how forensic accounting can be revitalized through data analytics to mitigate financial risk, reinforcing the role of algorithmic tools in fraud detection and early identification of irregularities that may precede insolvency [26]. At the level of crisis management and resolution, Loiacono and Rulli introduce the notion of “ResTech,” a suite of innovative technologies for bank-crisis resolution, arguing that advanced analytics and resolution-tech tools can support more efficient and less disruptive handling of distressed institutions [27]. Examine how CEO turnover affects firm performance and insolvency risk in a global sample, showing that leadership changes can materially alter risk trajectories and should therefore be considered in holistic risk-assessment frameworks [28]. In regulated contexts, propose an explainable-AI framework for credit-risk assessment and financial decision support that combines high-performing models with explanation tools understandable to regulators, managers and clients, addressing concerns about opacity and accountability in ML-based risk systems [29].

Within the specific segment of small and medium-sized enterprises (SMEs), insolvency prediction faces additional challenges related to limited disclosure, noisy accounting data and heightened sensitivity to macroeconomic shocks. Lee and Choi show that technological-feasibility indicators significantly improve SME-insolvency prediction beyond conventional financial ratios [30]. In [31] demonstrates that non-financial variables enhance forecasting accuracy for Croatian SMEs, especially at early-warning horizons when firms are in “sensitive” states rather than already failed. The propose a core rule-based decision tree that compresses a complex decision tree into an interpretable set of insolvency rules, so that the main causal patterns leading to SME failure can be communicated in a transparent and operational way [32]. These studies, together with digital-footprint and forensic-accounting contributions [19], [20], [26], indicate that the future of insolvency modelling is not purely numerical: it requires integrating heterogeneous data sources (accounting, non-financial, behavioural and textual) within architectures that remain interpretable to decision-makers and compliant with regulatory standards.

Taken together, the literature reveals a clear evolution: from linear, low-dimensional statistical models [1-6] to high-capacity AI/ML models for bankruptcy, credit risk and insolvency [7-12], [17-24], and then to system-level, governance-aware perspectives that incorporate data science, forensic accounting, crisis-resolution technologies and explainable AI [13-16], [25-29]. Yet three gaps remain particularly salient for the present study. First, many high-performing AI models are implemented as ad hoc pipelines whose preprocessing, feature-engineering, resampling and hyperparameter-tuning steps are not formalized as a reproducible computational scheme, which limits their auditability and transferability. Second, interpretability is often treated as an optional, post-hoc add-on—via feature-importance measures or local explanation tools—rather than a structural requirement embedded in the model design [29],[32]. Third, relatively few contributions explicitly frame their proposals as algorithmic–statistical hybrids, where the algorithmic (ML) component is systematically coupled with a statistically coherent layer that delivers calibrated insolvency probabilities, confidence measures and analytically tractable representations of how financial indicators interact under different risk regimes.

In this context, the present work positions itself as a response to these gaps by proposing an algorithmic–statistical model under a computational scheme for forecasting insolvency under financial risk. Building on robust evidence that ML methods improve predictive accuracy in bankruptcy and credit-risk tasks [7-12], [17-24], and on the need for interpretable, regulator-compatible frameworks [13], [16], [25-29], [30-33], the model integrates: (i) a carefully engineered set of financial indicators linked to profitability, leverage, liquidity, asset productivity and capital structure; (ii) an ML-based classification core trained under a transparent and reproducible computational pipeline (data cleaning and transformation, treatment of class imbalance, and rigorous validation); and (iii) a statistical representation that approximates insolvency probabilities using the most influential variables, thereby bridging the internal “black-box” logic of the algorithm with the explanatory requirements of managers, auditors and supervisors. This hybrid perspective aligns with the broader trajectory of the literature while making an explicit, operationally oriented contribution to the design of integrated, interpretable tools for forecasting insolvency under financial risk.

III. MATERIALS AND METHODS

The study adopts a quantitative, explanatory, and predictive design aimed at constructing an algorithmic–statistical model capable of estimating the financial solvency of cooperatives using a set of standardized financial ratios. Formally, we work with a historical dataset structured as.

$$\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N,$$

Where each observation $i \in \{1, \dots, N\}$ corresponds to a cooperative in a given period. The feature vector.

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^\top \in \mathbb{R}^p$$

Represents the $p = 12$ selected financial variables (ratios and proportions), while.

$$\square \square \in \{0,1\}$$

Is a binary variable that encodes the solvency state (0 = solvent, 1 = non-solvent). The research is framed within data science applied to financial risk management, using supervised machine learning algorithms to model the conditional probability

$$\mathbb{P}(Y = 1 \mid \mathbf{X} = \mathbf{x}),$$

Interpreted as the probability that a cooperative with financial profile \mathbf{x} is in a non-solvent state under financial risk. The central objective is to construct a decision function

$$\hat{f}: \mathbb{R}^p \rightarrow \{0,1\},$$

Capable of approximating the true mapping

$$f^*(\mathbf{x}) = y$$

With high predictive performance and, at the same time, with sufficient structure to be interpreted and used for decision-making. The dependent variable is defined from the Return on Equity (ROE). For each observation i , ROE_i is expressed as an annual proportion. Based on the researcher’s criterion, the logical variable ROE–LOGIC is constructed using the indicator function

$$y_i = \mathbb{I}(ROE_i < \tau), \quad \tau = 0.05,$$

where $\mathbb{I}(\cdot)$ denotes the indicator function. Explicitly,

$$y_i = \begin{cases} 1, & \text{if } ROE_i < 0.05, \\ 0, & \text{if } ROE_i \geq 0.05. \end{cases}$$

Thus, the solvency state is converted into a binary classification problem, compatible with tree-based models, logistic regression, Random Forest and other supervised algorithms. Regarding the independent variables, we work with $p = 12$ financial indicators grouped into five conceptual dimensions: (i) capital strength, (ii) operating efficiency, (iii) asset productivity and unproductivity, (iv) loan portfolio management, and (v) financial intermediation. Denoting by $j = 1, \dots, 12$ each indicator, the observed value x_{ij} corresponds to a precomputed ratio from balance sheet and income statement figures. The twelve predictors used in the model are defined as follows (for a generic observation i):

Net non-productive assets / total assets (var_3):

$$\text{var}_3(i) = x_{i,3} = \frac{\text{Net non-productive assets}_i}{\text{Total assets}_i}.$$

Productive assets / total assets (var_4):

$$\text{var}_4(i) = x_{i,4} = \frac{\text{Productive assets}_i}{\text{Total assets}_i}.$$

Productive assets / interest-bearing liabilities (var_5):

$$\text{var}_5(i) = x_{i,5} = \frac{\text{Productive assets}_i}{\text{Interest-bearing liabilities}_i}.$$

Uncovered non-performing loans / (equity + earnings) (var_8):

$$\text{var}_8(i) = x_{i,8} = \frac{\text{Uncovered non-performing loans}_i}{\text{Equity}_i + \text{Earnings}_i}.$$

Coverage of the problem loan portfolio (var_18):

$$\text{var}_{18}(i) = x_{i,18} = \frac{\text{Allowance for loan losses}_i}{\text{Problem loan portfolio}_i}.$$

Improductivity index (var_19, Fi_i):

$$\text{var}_{19}(i) = Fi_i = 1 + \frac{\text{Non-productive assets}_i}{\text{Total assets}_i}.$$

Adjusted capital strength (var_20, Fk_i):

$$\text{var}_{20}(i) = Fk_i = \frac{\text{Equity}_i + \text{Earnings}_i - \text{Extraordinary income}_i}{\text{Total assets}_i}.$$

Operating expenses / financial margin (var_22):

$$\text{var}_{22}(i) = x_{i,22} = \frac{\text{Operating expenses}_i}{\text{Financial margin}_i}.$$

Estimated operating expenses / average total assets (var_23):

$$\text{var}_{23}(i) = x_{i,23} = \frac{\text{Estimated operating expenses}_i}{\text{Average total assets}_i}.$$

Estimated personnel expenses / average total assets (var_24):

$$\text{var}_{24}(i) = x_{i,24} = \frac{\text{Estimated personnel expenses}_i}{\text{Average total assets}_i}.$$

Estimated intermediation margin / average total assets (var_26):

$$\text{var}_{26}(i) = x_{i,26} = \frac{\text{Estimated intermediation margin}_i}{\text{Average total assets}_i}.$$

Total portfolio delinquency (var_34):

$$\text{var}_{34}(i) = x_{i,34} = \frac{\text{Past-due loan portfolio}_i}{\text{Total loan portfolio}_i}.$$

Collecting these predictors in a single feature vector, we have

$$\mathbf{x}_i = (\text{var}_3(i), \text{var}_4(i), \text{var}_5(i), \text{var}_8(i), \text{var}_{18}(i), \text{var}_{19}(i), \text{var}_{20}(i), \text{var}_{22}(i), \text{var}_{23}(i), \text{var}_{24}(i), \text{var}_{26}(i), \text{var}_{34}(i))^T \in \mathbb{R}^{12}.$$

Equivalently, in terms of the generic indices x_{ij} we can write

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{i12})^T, \quad p = 12.$$

If an additional homogenization step were needed, each variable j could be standardized as

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j},$$

where μ_j and σ_j denote the mean and standard deviation of variable j , respectively. In this study, the model is trained directly on x_i in their original financial scales, preserving the interpretability of each ratio.

The resulting design matrix is

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \dots & x_{Np} \end{bmatrix},$$

with $x_i \in \mathbb{R}^{12}$ for $i = 1, \dots, N$ and $p = 12$.

An initial data cleaning step is performed, removing observations with missing values in any of the selected variables. The resulting dataset is denoted by \mathcal{D}^* , with size N^* . A stratified partition is then applied into training and test sets, preserving the empirical class distribution:

$$\mathcal{D}^* = \mathcal{D}_{\text{train}} \cup \mathcal{D}_{\text{test}}, \quad |\mathcal{D}_{\text{train}}| = 0.8 N^*, \quad |\mathcal{D}_{\text{test}}| = 0.2 N^*,$$

so that

$$\mathbb{P}(Y = 1 | \mathcal{D}_{\text{train}}) \approx \mathbb{P}(Y = 1 | \mathcal{D}_{\text{test}}).$$

For predictive modeling, a Random Forest classifier is employed. Mathematically, the model is defined as a collection of B decision trees:

$$\mathcal{F} = \{h_b(\mathbf{x}; \theta_b)\}_{b=1}^B,$$

]

where h_b is the b -th tree and θ_b collects the randomness introduced by bootstrap sampling and random feature selection at each split. In this study, we use

$$B = 100$$

and a fixed global random seed to ensure reproducibility. Each tree produces a class prediction $h_b(\mathbf{x}) \in \{0,1\}$. The estimated probability of non-solvency is

$$\hat{p}(y = 1 | \mathbf{x}) = \frac{1}{B} \sum_{b=1}^B h_b(\mathbf{x}).$$

The final decision rule of the forest is expressed as

$$\hat{f}(\mathbf{x}) = \begin{cases} 1, & \text{if } \hat{p}(y = 1 | \mathbf{x}) \geq \tau_c, \\ 0, & \text{if } \hat{p}(y = 1 | \mathbf{x}) < \tau_c, \end{cases}$$

Where the classification threshold is set to

$$\tau_c = 0.5.$$

Training the forest amounts to approximately minimizing the empirical 0–1 loss over the training set:

$$\min_{\mathcal{F}} \frac{1}{|\mathcal{D}_{\text{train}}|} \sum_{(\mathbf{x}_i, y_i) \in \mathcal{D}_{\text{train}}} \mathbb{I}(\hat{f}(\mathbf{x}_i) \neq y_i),$$

Subject to the structural constraints of the Random Forest (maximum depth, minimum samples per leaf, number of candidate variables per split, etc.). In practice, each tree is built on a bootstrap sample of $\mathcal{D}_{\text{train}}$, and at each node the split is chosen to maximize impurity reduction (for example, Gini decrease). Once the model is trained, the relative importance of each variable is estimated. For each predictor j , the aggregated Gini importance is

$$I_j = \sum_{b=1}^B \sum_{t \in T_b(j)} \Delta G_{b,t},$$

Where $T_b(j)$ is the set of nodes in tree b where variable j is used for splitting, and $\Delta G_{b,t}$ is the Gini impurity reduction at node t . Importances are normalized to obtain relative weights:

$$w_j = \frac{I_j}{\sum_{k=1}^p I_k}, \quad j = 1, \dots, p,$$

with

$$\sum_{j=1}^p w_j = 1, \quad w_j \geq 0.$$

Based on these weights, an algorithmic–statistical insolvency index is defined as a linear combination of the financial variables:

$$S(\mathbf{x}) = \sum_{j=1}^p w_j x_j.$$

The value $S(\mathbf{x})$ acts as a scalar projection of the cooperative’s financial profile onto the importance structure learned by the model. Although the Random Forest remains the main classification tool, $S(\mathbf{x})$ provides an interpretable and parsimonious representation of non-solvency risk: higher values of $S(\mathbf{x})$ are associated with higher insolvency exposure given the relative weight of variables such as total delinquency, asset unproductiveness, or operating cost burden.

If an explicit probability approximation based on this index is desired, a secondary logistic function can be fitted:

$$\hat{\pi}(\mathbf{x}) = \frac{1}{1 + \exp(-\alpha - \beta S(\mathbf{x}))},$$

Where α and β are estimated by maximum likelihood in a simple logistic regression of y on $S(\mathbf{x})$. This additional layer does not replace the Random Forest in terms of accuracy, but yields a closed-form analytical expression for $\hat{\pi}(\mathbf{x})$, useful for communication and sensitivity analysis. Finally, model performance is evaluated on $\mathcal{D}_{\text{test}}$ using standard binary classification metrics. Let TP, TN, FP and FN denote true positives, true negatives, false positives and false negatives, respectively. The metrics are defined as

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

$$Sensitivity = \frac{TP}{TP + FN},$$

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

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$$

In addition, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) can be used to evaluate the overall discriminative capacity of the model. Together with the interpretative index $S(\mathbf{x})$, this configuration defines a complete algorithmic–statistical scheme for forecasting insolvency under financial risk in cooperatives.

IV. RESULTS AND DISCUSSIONS

It begins with the development of quality measures for the Random Forest model:

Table 1: Performance metrics.

Precision	Recall	F1-Score	AUC	MCC (Matthews)	Cohen’s Kappa	Accuracy	Set
1.000	1.000	1.000	1.000	1.000	1.000	1.000	Training
0.883	0.925	0.904	0.946	0.780	0.779	0.891	Test

Note: The model was implemented in Python and evaluated using a 20% training split.

Source: Authors, (2026).

In Table 1, the reported model quality indicators allow assessment of performance on both the training and test sets. The training set refers to the data used to build and fit the model, whereas the test set is used to verify how well the model generalizes to unseen data. In this case, the model achieves perfect performance on the training data (all metrics equal to 1.000), which may indicate overfitting. Regarding the test set—the most relevant for judging the model’s practical utility—the results remain strong. Precision (0.883) indicates that approximately 88.3% of the positive predictions produced by the model were correct. Sensitivity (recall = 0.925) shows that the model correctly identified 92.5% of the true positive cases. The F1-score (0.904) reflects the balance between precision and recall; its high value reinforces that the model maintains a solid trade-off between these two dimensions. The AUC (0.946) further indicates an excellent capacity to discriminate between classes, as values close to 1 are typically associated with outstanding performance in binary classification. Moreover, the Matthews correlation coefficient (MCC) reaches 0.780 on the test set. This metric is particularly informative under class imbalance because it incorporates all cells of the confusion matrix. The obtained value suggests a strong association between predicted and true labels. Likewise, Cohen’s Kappa (0.779) reflects very good agreement beyond chance. Overall accuracy is 0.891, indicating that the model correctly classifies nearly 89% of the test observations. In summary, although the model may be somewhat overfitted to the training data, its test performance suggests that it is robust, accurate, and able to generalize effectively. Nonetheless, additional evaluations—such as cross-validation or testing on external datasets—would be advisable to confirm the stability of these metrics.

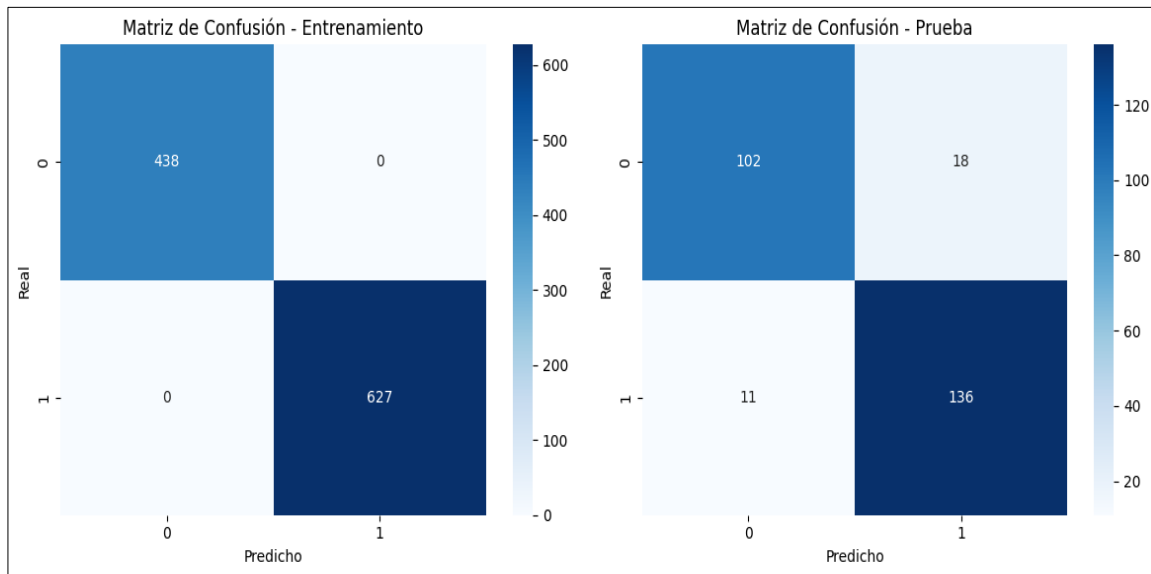


Figure 1: Confusion matrix.

Source: Authors, (2026).

Considering Figure 1, the model correctly classified all 438 cases from class 0 (negative) and all 627 cases from class 1 (positive) in the training set, without any errors. That is, there were no false positives or false negatives, which results in precision, recall, and F1-score equal to 1.000. While such outcomes are theoretically ideal, in practice they often suggest overfitting, as the model may have memorized the training data rather than learned generalizable patterns. In contrast, the test-set confusion matrix provides a more realistic and informative view of generalization performance. Of the 120 true class-0 cases, the model correctly classified 102 (true negatives) and misclassified 18 as class 1 (false positives). For class 1, out of 147 true cases, the model correctly predicted 136 (true positives) and misclassified 11 as class 0 (false negatives). This indicates that, while the model remains strong, it incurs some errors that were not present during training—an expected and natural outcome under proper model evaluation.

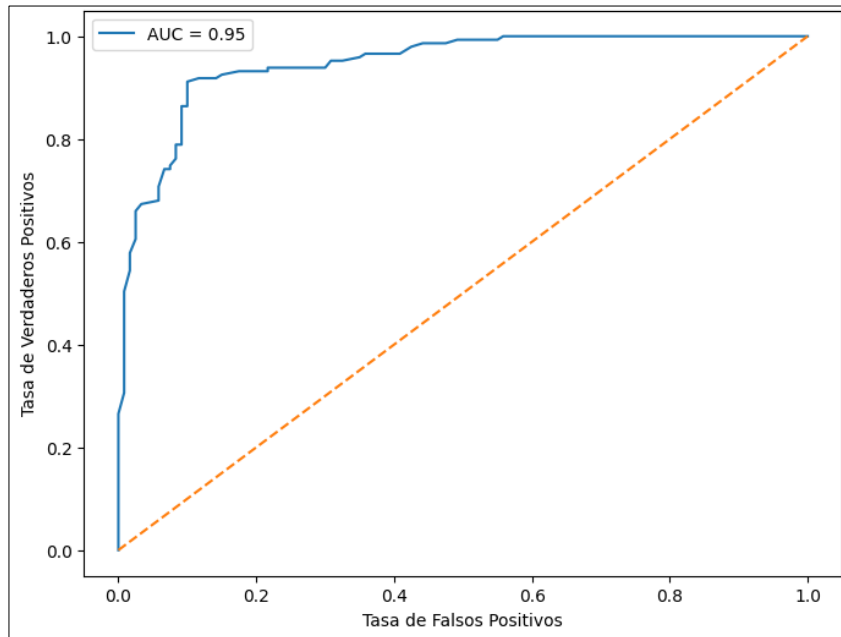


Figure 2: ROC curve.
Source: Authors, (2026).

Figure 2 exhibits a markedly convex shape and lies well above the orange diagonal, which represents a random classifier with no discriminative ability. That diagonal corresponds to an AUC (Area Under the Curve) of 0.5, commonly interpreted as the minimum baseline for a model that performs no better than chance. By contrast, the evaluated model achieves an AUC of 0.95, which is excellent. This value indicates a very high ability to discriminate between classes, correctly separating positive and negative cases in most situations. An AUC of 0.95 also implies that, when randomly selecting one positive and one negative instance, the model assigns a higher score to the positive instance 95% of the time.

Table 2: Variable relevance.

Variable	Importance
var_26	0.222
var_22	0.197
var_4	0.096
var_18	0.081
var_20	0.078
var_5	0.058
var_19	0.054
var_3	0.051
var_23	0.048
var_24	0.046
var_34	0.045
var_8	0.024

Note: The model was implemented in Python and the results were sorted in descending order.

Source: Authors, (2026).

The most influential variable is **var_26**, corresponding to the intermediation margin estimated over average assets, with an importance of 22.2%. This indicator is central because it captures the core of financial intermediation: the spread between earnings from lending and costs of funding. A low margin implies that basic revenues barely cover (or fail to cover) financial costs, which can drive weak performance. In this context, lower values of this variable are associated with higher insolvency risk and therefore a higher probability of being classified as “1”. The second most important variable is **var_22**, representing the ratio of operating expenses to financial margin. With an importance of 19.7%, this variable captures the efficiency with which the cooperative manages its financial income. Higher values indicate that administrative costs absorb a substantial portion of the available margin, leaving little or no surplus. This not only undermines immediate profitability but also constrains capitalization, investment capacity, and risk absorption.

Consequently, elevated values shift predictions toward the non-solvent class. The third variable in importance is **var_4**, measuring productive assets as a share of total assets (9.6%). A low ratio suggests that a sizable portion of assets is not generating income. Balance-sheet unproductivity is a warning sign because it reflects inefficient resource allocation; combined with weak operating profitability, it amplifies insolvency risk. Next, **var_18**—coverage of the problem loan portfolio (8.1%)—indicates that measuring portfolio quality alone is insufficient; the cooperative’s preparedness to absorb potential losses is also critical. Insufficient coverage for delinquent or high-risk loans increases financial exposure, weakens earnings, and raises the likelihood of being classified as non-solvent. **var_20**, defined as the Fk index $(\text{equity} + \text{results} - \text{extraordinary income}) / \text{total assets}$, carries a weight of 7.8%.

This metric provides an adjusted view of structural solvency by excluding non-recurring income. Its relevance confirms that underlying capital strength—beyond exceptional events—substantially influences the model’s risk assessment. Other relevant variables include **var_5** (productive assets over interest-bearing liabilities, 5.8%) and **var_19** (FI index: 1+1+1+ unproductive assets / total assets, 5.4%). Both relate to the balance between income-generating resources and financial burdens. Unfavorable values are interpreted by the model as early warning signals. Variables **var_23** (estimated operating expenses over average assets, 4.8%) and **var_24** (personnel expenses over average assets, 4.6%) further reinforce the model’s sensitivity to cost efficiency. When a substantial share of resources is tied to fixed costs without a proportional return, the cooperative may struggle to sustain profitability. Finally, **var_34** (total portfolio delinquency, 4.5%) and **var_8** (uncovered unproductive portfolio over equity plus results, 2.4%) act as more direct proxies for credit risk. Although they remain relevant, their smaller weights relative to efficiency and structural variables suggest that the model places greater emphasis on the ability to generate, manage, and protect the financial margin than on pure default risk alone. Based on these results, the following equation is proposed as a forecasting model for insolvency among Segment 1 savings and credit cooperatives:

$$\begin{aligned} \text{Prob Insolv} = & 0.2223 * \text{var}_{26} + 0.1968 * \text{var}_{22} + 0.0957 * \text{var}_{4} + 0.0807 * \text{var}_{18} + 0.0775 * \text{var}_{20} \\ & + 0.0583 * \text{var}_{5} + 0.0536 * \text{var}_{19} + 0.0515 * \text{var}_{3} + 0.0481 * \text{var}_{23} + 0.0460 * \text{var}_{24} \\ & + 0.0453 * \text{var}_{34} + 0.0240 * \text{var}_{8} \end{aligned}$$

The resulting formula does not produce a true ROE value; instead, it provides an approximation of insolvency risk as a function of the most influential variables. When evaluating a cooperative with this equation, higher estimated values imply greater proximity to class “1” (non-solvent). Thus, values closer to 1 indicate higher insolvency risk, whereas values closer to 0 indicate greater solvency. The evidence obtained in this study reinforces—and operationalizes within an algorithmic–statistical scheme—the long-standing view that insolvency risk is ultimately rooted in the joint behavior of profitability, operating efficiency, asset productivity, and capital strength. Classical insolvency research established that accounting ratios can discriminate between failed and non-failed entities (Beaver; Altman; Ohlson), thereby motivating ratio-based early-warning models. In our cooperative setting, this tradition is echoed in the dominance of variables tied to intermediation profitability and efficiency—particularly the intermediation margin over average assets (**var_26**) and operating expenses relative to the financial margin (**var_22**)—which carry the largest importance weights in the Random Forest. This alignment suggests that, even under modern ML implementations, the “economic content” emphasized by ratio-based frameworks remains structurally central: cooperatives approach non-solvency when the margin that sustains intermediation is compressed and/or absorbed by operating costs. From a predictive standpoint, our results are consistent with the broader empirical finding that non-linear and ensemble ML models tend to outperform purely linear statistical benchmarks, particularly in out-of-sample classification.

Barboza, Kimura, and Altman document systematic performance gains from ML—especially ensemble methods—in bankruptcy prediction tasks, and our test-set performance (AUC = 0.946; accuracy = 0.891; F1 = 0.904) provides convergent evidence in the context of Ecuadorian Segment 1 cooperatives. At the same time, the perfect training scores (all metrics = 1.000) raise the standard interpretive caution highlighted across comparative ML studies (e.g., the need for robust validation and sensitivity checks): our own results explicitly acknowledge that such perfection may signal overfitting, motivating additional evaluation strategies (e.g., cross-validation or external testing) to verify stability. A central substantive contribution of the model lies in what it learns as the main discriminators of insolvency status: the pattern indicates that deterioration in the business model’s margin mechanics and cost discipline is detected as a primary route toward non-solvency, even before (or beyond) pure delinquency metrics. Specifically, **var_26** (intermediation margin over average assets, 22.2%) and **var_22** (operating expenses/financial margin, 19.7%) dominate, while total portfolio delinquency (**var_34**) has comparatively lower weight (4.5%). This ranking is informative in light of credit-risk ML literature that frames early-warning detection as the identification of non-linear combinations of financial pressures that precede visible default events (e.g., ML-based early detection of credit risks and classification of credit operations).

In practical terms, the model implies that margin compression and cost absorption may operate as leading signals of vulnerability, whereas delinquency—though relevant—may act more as a lagging manifestation once structural pressures have already accumulated. The confusion-matrix results further support a realistic interpretation of model behavior under generalization. While the training matrix shows zero errors, the test matrix exhibits non-trivial false positives and false negatives, which is expected when moving from in-sample fit to out-of-sample performance. This matters in light of decision-oriented perspectives such as Marassi and Pediroda’s framing of insolvency prediction under asymmetric misclassification costs: in supervisory and managerial contexts, the relative tolerance for false alarms (false positives) versus missed distress signals (false negatives) should guide thresholding and deployment. Accordingly, although the ROC/AUC indicates excellent discrimination (≈ 0.95), the operational use of the classifier should be paired with explicit cost preferences (e.g., favoring sensitivity if the priority is early intervention). Importantly, the study directly addresses a core constraint repeatedly emphasized in regulated financial ML: interpretability. Recent work argues that high-performing credit-risk models must be explainable to regulators and decision-makers, rather than remaining opaque black boxes (e.g., explainable AI frameworks; rule-based explanatory trees).

In that direction, our approach leverages Random Forest importance weights to construct an interpretable insolvency index and a tractable probability approximation layer, explicitly designed to translate algorithmic outputs into an auditable statistical form. This is not merely a presentation choice: it operationalizes the “algorithmic–statistical hybrid” positioning stated as a gap in prior work (interpretability embedded structurally, not treated as an optional add-on). Finally, the results should be read alongside clear scope boundaries that emerge from the literature. Multiple streams show that insolvency and credit-risk prediction can improve when heterogeneous sources are incorporated—digital footprints, textual/news information, sentiment signals, and non-financial/technological feasibility indicators—especially for earlier warning horizons. Our model is deliberately financial-ratio-based, which strengthens transparency and alignment with traditional supervisory data, but it also implies that future extensions could integrate these complementary data modalities to test whether (and how) predictive structure changes under broader information sets. In other words, the present findings provide a strong financial “core,” while the literature indicates a natural research frontier: coupling this core with alternative and non-financial signals within the same reproducible, regulator-compatible computational scheme.

V. CONCLUSIONS

The results of this study indicate that insolvency risk in Segment 1 savings and credit cooperatives in Ecuador can be anticipated with high precision and reliability by using machine-learning techniques, specifically the Random Forest algorithm. The developed model achieved strong predictive performance on the test set, including an AUC of 0.946 and an accuracy of 89.1%, supporting its practical relevance as a preventive tool for financial risk management.

The analysis of the most influential predictors shows that insolvency risk is not driven solely by classical delinquency or credit exposure indicators; rather, it is primarily shaped by operating efficiency, capital structure, and balance-sheet productivity. Variables such as the intermediation margin over average assets, operating expenses relative to the financial margin, and the share of productive assets explain a substantial portion of insolvency probability. This implies that weak internal resource management and low operating profitability are more decisive determinants than the mere level of delinquency.

Moreover, the evidence confirms that financial solvency is not an isolated phenomenon, but the outcome of a set of sustainable, structured, and efficient financial practices. In this sense, the proposed model can support both cooperative management teams and regulatory authorities by providing an objective tool to monitor and classify institutions at higher risk of capital deterioration.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Alexander Haro-Sarango.

Methodology: Alexander Haro-Sarango..

Investigation: Alexander Haro-Sarango.

Discussion of results: Alexander Haro-Sarango.

Writing – Original Draft: Alexander Haro-Sarango.

Writing – Review and Editing: Alexander Haro-Sarango.

Resources: Alexander Haro-Sarango.

Supervision: Alexander Haro-Sarango.

Approval of the final text: Alexander Haro-Sarango.

VII. REFERENCES

- [1] W. H. Beaver, "Financial ratios as predictors of failure," *J. Accounting Res.*, vol. 4, pp. 71–111, 1966.
- [2] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *J. Finance*, vol. 23, no. 4, pp. 589–609, 1968.
- [3] J. A. Ohlson, "Financial ratios and the probabilistic prediction of bankruptcy," *J. Accounting Res.*, vol. 18, no. 1, pp. 109–131, 1980.
- [4] W. Härdle, Y.-J. Lee, D. Schäfer, and Y.-R. Yeh, "Variable selection and oversampling in the use of smooth support vector machines for predicting the default risk of companies," *J. Forecasting*, vol. 28, no. 6, pp. 513–534, 2009. doi: 10.1002/for.1109.
- [5] D. Marassi and V. Pediroda, "Risk insolvency predictive model maximum expected utility," *Int. J. Business Performance Management*, vol. 10, nos. 2–3, pp. 174–190, 2008. doi: 10.1504/IJBPM.2008.016637.
- [6] K. K. Lai, L. Yu, W. Huang, and S. Wang, "A novel support vector machine metamodel for business risk identification," in *Lecture Notes in Computer Science*, vol. 4099, 2006, pp. 980–984. doi: 10.1007/11801603_118.
- [7] F. Barboza, H. Kimura, and E. I. Altman, "Machine learning models and bankruptcy prediction," *Expert Syst. Appl.*, vol. 83, pp. 405–417, 2017.
- [8] D. Gajdosikova and J. Michulek, "Artificial Intelligence Models for Bankruptcy Prediction in Agriculture: The Performance of Artificial Neural Networks and Decision Trees," *Agriculture (Switzerland)*, vol. 15, no. 10, 2025. doi: 10.3390/agriculture15101077.
- [9] G. K. Kamalam, S. N. Kumar, G. Subiga, and R. Yadhuvarshini, "Prediction of Bankruptcy Using Machine Learning Models," in *Lecture Notes in Networks and Systems*, vol. 1047, 2024, pp. 313–322. doi: 10.1007/978-3-031-64836-6_31.
- [10] D. Máté, H. Raza, and I. Ahmad, "Comparative Analysis of Machine Learning Models for Bankruptcy Prediction in the Context of Pakistani Companies," *Risks*, vol. 11, no. 10, 2023. doi: 10.3390/risks11100176.
- [11] N. E. Radwan, H. M. Alzoubi, N. Sahawneh, A. Fatima, A. Rehman, and S. Khan, "An Intelligent Approach for Predicting Bankruptcy Empowered with Machine Learning Technique," in *Proc. 2022 Int. Conf. on Cyber Resilience (ICCR)*, 2022. doi: 10.1109/ICCR56254.2022.9995890.
- [12] A. Stasko, I. Birzniece, and G. Kebers, "Development of Bankruptcy Prediction Model for Latvian Companies," *Complex Systems Informatics and Modeling Quarterly*, no. 27, pp. 15–28, 2021. doi: 10.7250/csimq.2021-27.02.
- [13] J. Jayawardana, P. Wijeratne, Z. Vrcelj, and M. Sandanayake, "Artificial Intelligence for Predicting Insolvency in the Construction Industry—A Systematic Review and Empirical Feature Derivation," *Buildings*, vol. 15, no. 17, 2025. doi: 10.3390/buildings15172988.
- [14] M. Francis, S. Perera, W. Zhou, and S. Nanayakkara, "Artificial Intelligence Applications for Proactive Dispute Management in the Construction Industry: A Systematic Literature Review," *J. Inf. Technol. Constr.*, 2025. doi: 10.36680/j.itcon.2025.029.
- [15] E. Mirindi and P. Mukala, "An In-Depth Pedagogical Exploration of Loan Default Classification with Machine Learning Techniques," in *Lecture Notes in Networks and Systems*, vol. 1539, 2026, pp. 83–101. doi: 10.1007/978-981-96-9248-4_7.
- [16] P. Goel, Umakanth, H. Thakker, A. Kumar Pathak, R. Joyce, and P. Mishra, "Enhancing Financial Management Efficiency through Advanced Predictive Analytics y estrategias de toma de decisiones basadas en datos," *Management (Montevideo)*, 2025. doi: 10.62486/agma2025162.
- [17] M. A. Hamada, K. Farag, and A. E. Abiche, "Early Prediction Detection of Retail and Corporate Credit Risks Using Machine Learning Algorithms," *Emerging Science Journal*, vol. 9, no. 2, 2025. doi: 10.28991/ESJ-2025-09-02-025.

- [18] G. Teles, J. J. P. C. Rodrigues, R. A. L. Rabêlo, and S. A. Kozlov, "Comparative study of support vector machines and random forests machine learning algorithms on credit operation," *Software—Practice and Experience*, 2021. doi: 10.1002/spe.2842.
- [19] E. V. Orlova, "Methodology and models for individuals' creditworthiness management using digital footprint data and machine learning methods," *Mathematics*, vol. 9, no. 15, 2021. doi: 10.3390/math9151820.
- [20] E. V. Orlova, "Data Driven Design to Credit Risk Management Using Digital Footprint Intelligence," in *Proc. 3rd Int. Conf. on Control Systems, Mathematical Modeling, Automation and Energy Efficiency (SUMMA)*, 2021. doi: 10.1109/SUMMA53307.2021.9632188.
- [21] H. Jin, J.-P. Hong, K.-H. Lee, and D.-W. Joo, "Diagnosis of Corporate Insolvency Using Massive News Articles for Credit Management," in *Proc. 2019 IEEE Int. Conf. on Big Data and Smart Computing (BigComp)*, 2019. doi: 10.1109/BIGCOMP.2019.8679267.
- [22] D. P. Pitta de Jesus and C. D. N. Besarria, "Machine learning and sentiment analysis: Projecting bank insolvency risk," *Research in Economics*, 2023. doi: 10.1016/j.rie.2023.03.001.
- [23] T. Feng, J. Xu, Z. Zhou, and Y. Luo, "How Green Credit Policy Affects Commercial Banks' Credit Risk: Machine-Learning-Based Modeling From 26 Listed Commercial Banks in China," *J. Cases Inf. Technol.*, 2023. doi: 10.4018/JCIT.333858.
- [24] S. Grishunin, S. Suloeva, A. Egorova, and E. Burova, "Comparison of Empirical Methods for the Reproduction of Global Manufacturing Companies' Credit Ratings," *Int. J. Technology*, vol. 11, no. 6, 2020. doi: 10.14716/ijtech.v11i6.4424.
- [25] T. L. Lai and H. Xing, *Data Science and Risk Analytics in Finance and Insurance*. Taylor and Francis, 2024. doi: 10.1201/9781315117041.
- [26] M. Gupta, P. K. Aggarwal, and R. Gupta, "Revitalizing the Forensic Accounting: An Exploratory Study on Mitigating the Financial Risk Using Data Analytics," *Int. J. Exp. Res. Rev.*, vol. 41 (special issue), pp. 227–238, 2024. doi: 10.52756/ijerr.2024.v41spl.019.
- [27] G. Loiacono and E. Rulli, "ResTech: innovative technologies for crisis resolution," *J. Banking Regulation*, 2022. doi: 10.1057/s41261-021-00154-4.
- [28] A. Salvi, A. Tron, and F. Colantoni, "The impact of CEO turnover on firm performance and insolvency risk—A global analysis," *Finance Res. Lett.*, vol. 62, 2024. doi: 10.1016/j.frl.2024.105093.
- [29] M. K. Nallakaruppan *et al.*, "Credit risk assessment and financial decision support using explainable artificial intelligence," *Risks*, vol. 12, no. 10, p. 164, 2024.
- [30] S. Lee and K. Choi, "Predicting the insolvency of SMEs using technological feasibility," *Sustainability*, vol. 12, no. 23, art. 9790, 2020.
- [31] M. T. Kuvék, "Improving the accuracy of firm failure forecasting using non-financial variables: The case of Croatian SMEs," *Eng. Proc.*, vol. 39, no. 1, p. 62, 2023.
- [32] S. Lee, K. Choi, and D. Yoo, "Building a core rule-based decision tree to explain the causes of insolvency in small and medium-sized enterprises more easily," *Humanit. Soc. Sci. Commun.*, vol. 10, art. 933, pp. 1–16, 2023.
- [33] Haro, A., Sandoval, G., Rodríguez, M., Armijo, V., Arana, I., Vasquez, W., Proaño, E., and Martínez, A., "Low Financial Risk of Default and Productive Use of Assets Through Hidden Markov Models," *Risks*, vol. 13, no. 12, Art. no. 230, 2025, doi: 10.3390/risks13120230.