COLLEGE ED MATHEMATIQUES
DOCTORAL STIC
PAYS DE LA LOIRE

université angers

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

THESE DE DOCTORAT

DE
L'UNIVERSITÉ D'ANGERS
SOUS LE SCEAU DE
LA COMUE ANGERS – LE MANS
ECOLE DOCTORALE N° 641
Mathématiques et Sciences et Technologies du numérique, de l'Information et de la Communication
Spécialité : « *Computer science* »

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA - DOCTORADO EN INDUSTRIA INTELIGENTE

Par
«Alex PAZ BECERRA »

« Machine Learning and Metaheuristic Feature Selection for Predicting Income Declaration Behavior in Student Loan Systems»

Thèse présentée et soutenue à « Valparaíso », le « 09 avril 2026 »
Unité de recherche : « Laboratoire d'Études et de Recherche en Informatique d'Angers, LERIA»

Rapporteurs avant soutenance :

Anarosa FRANCO BANDAO Professeur, Universidad de Sao Paulo, Brasil.
José Antonio GARCIA Professeur, Pontificia Universidad Católica de Valparaíso, Chile.

Composition du Jury :

Président :	Álvaro PEÑA	Professeur, Pontificia Universidad Católica de Valparaíso, Chile.
Examinateurs :	Dreidy VASQUEZ	Professeur, Pontificia Universidad Católica de Valparaíso, Chile.
	Ricardo SOTO	Professeur, Pontificia Universidad Católica de Valparaíso, Chile.
Dir. de thèse :	Broderick CRAWFORD, Professeur, Pontificia Universidad Católica de Valparaíso, Chile.	
Co-dir. de thèse :	Eric MONFROY, Professeur, Université de Angers, France.	

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO

FACULTAD DE INGENIERÍA

DOCTORADO EN INDUSTRIA INTELIGENTE

MACHINE LEARNING AND METAHEURISTIC FEATURE SELECTION FOR PREDICTING INCOME DECLARATION BEHAVIOR IN STUDENT LOAN SYSTEMS.

ALEX PATRICIO PAZ BECERRA

Profesor a Guía: Broderick Crawford Labrin

Profesor Co-tutor: Eric Monfroy

TESIS PARA OPTAR AL GRADO DE DOCTOR EN INDUSTRIA INTELIGENTE

Valparaíso, Enero 2026

Contents

1 Introduction **1**
- 1.1 Background and Motivation . . . 1
- 1.2 Problem Statement . . . 2
- 1.3 Solution Proposal . . . 3
- 1.4 Contributions . . . 4
- 1.5 Thesis Organization . . . 5

2 State of the Art **6**
- 2.1 Introduction . . . 6
- 2.2 Conceptual Framework . . . 6
  - 2.2.1 Income-Contingent Student Loan Systems as Institutional Frameworks 6
  - 2.2.2 Declarative Compliance as an Observable Behavioral Process . . . 8
  - 2.2.3 Predictive Modeling in Institutional Decision-Support Systems . . . 9
  - 2.2.4 Supervised Machine Learning for Classification Tasks . . . 10
  - 2.2.5 Evaluation Metrics for Compliance-Oriented Predictive Models . . . 11
  - 2.2.6 Feature Selection in High-Dimensional Institutional Data . . . 12
  - 2.2.7 Metaheuristic Optimization as a Search Paradigm . . . 13
  - 2.2.8 Interpretability and Stability as Design Requirements . . . 14
- 2.3 Literature Review . . . 15
  - 2.3.1 Machine Learning and Individual Credit Risk . . . 16
  - 2.3.2 Deep Learning Models for Credit Risk and Compliance Prediction . . . 19
  - 2.3.3 Feature Selection Techniques in Predictive Modeling . . . 20
  - 2.3.4 Metaheuristics Applied to Feature Selection and Optimization . . . 21
  - 2.3.5 Interpretability and Robustness in Optimized Predictive Models . . . 23
- 2.4 Existing Methods, Tools, and Technologies . . . 24
  - 2.4.1 Dominant Modeling Paradigms . . . 26
  - 2.4.2 Data Characteristics and Experimental Practices . . . 27
  - 2.4.3 Role of Feature Selection and Optimization . . . 30
  - 2.4.4 Interpretability and Robustness Considerations . . . 31
  - 2.4.5 Evaluation Metrics and Validation Practices . . . 32
- 2.5 Research Gaps and Open Challenges . . . 37
- 2.6 Research Justification and Contribution Opportunity . . . 39
- 2.7 Chapter Conclusions . . . 40

3 Thesis Research Proposal **42**
- 3.1 Operational Problem Definition . . . 42
- 3.2 Research Objectives, Questions, and Hypotheses . . . 43
  - 3.2.1 Main Objective . . . 43

3.2.2 Specific Objectives . . . 43
3.2.3 Research Questions . . . 44
3.2.4 Hypotheses . . . 44
3.3 Research Design . . . 45
3.4 Research Methodology Stages . . . 46
3.5 Modeling and Evaluation Strategy . . . 49
3.5.1 Evaluation Strategy . . . 49
3.5.2 Implementation Details and Reproducibility . . . 49

4 Results **51**
4.1 Data Pipeline Design and Baseline Predictive Modeling (Stage I) . . . 52
4.1.1 Stage I Overview and Baseline Objective . . . 52
4.1.2 Dataset Construction and Baseline Representation . . . 52
4.1.3 Baseline Modeling Setup . . . 53
4.1.4 Evaluation Metrics . . . 53
4.1.5 Baseline Predictive Performance . . . 53
4.1.6 Error Distribution and Class Imbalance Effects . . . 54
4.1.7 Limitations of the Baseline . . . 55
4.1.8 Summary of Stage I Results . . . 55
4.2 Interpretable Modeling under Pre-Declarative Constraints (Stage II) . . . 56
4.2.1 Stage II Overview . . . 56
4.2.2 Problem Definition and Predictive Scope . . . 57
4.2.3 Data Preparation and Experimental Protocol . . . 57
4.2.4 Modeling Framework . . . 61
4.2.5 Evaluation Metrics and Validation . . . 63
4.2.6 Experimental Results . . . 64
4.2.7 Model Interpretability Analysis . . . 70
4.2.8 Comparative Positioning with Subsequent Stages . . . 78
4.2.9 Summary of Stage II Results . . . 79
4.3 Metaheuristic Optimization for Predictive Modeling of Declaration Abandonment (Stage III) . . . 80
4.3.1 Stage III Overview . . . 80
4.3.2 Problem Definition and Experimental Scope . . . 81
4.3.3 Data Preparation and Experimental Protocol . . . 81
4.3.4 Optimization Framework . . . 97
4.3.5 Experimental Results . . . 105
4.3.6 Comparative Analysis with Previous Stages . . . 115
4.3.7 Summary of Stage III Results . . . 116
4.4 Bioinspired Deep Learning for Declaration Discontinuity Prediction (Stage IV) 117
4.4.1 Stage IV Overview . . . 117
4.4.2 Problem Definition and Predictive Scope . . . 118
4.4.3 Data Preparation and Experimental Protocol . . . 118
4.4.4 Modeling Framework . . . 122

4.4.5 Evaluation Metrics and Statistical Validation . . . 125
4.4.6 Experimental Results . . . 126
4.4.7 Model Interpretability Analysis . . . 132
4.4.8 Comparative Analysis with Previous Stages . . . 133
4.4.9 Summary of Stage IV Results . . . 134
4.4.10 Empirical Summary . . . 135

5 Discussion **136**
5.1 Cross-Stage Interpretation of Results . . . 136
5.1.1 Predictive Feasibility under Information Constraints . . . 137
5.1.2 Impact of Feature Space Enrichment and Optimization . . . 137
5.1.3 Trade-offs between Performance, Stability, and Interpretability . . . 138
5.1.4 Comparative Behavior across Modeling Paradigms . . . 139
5.2 Discussion by Research Question . . . 140
5.2.1 Discussion of RQ1 . . . 140
5.2.2 Discussion of RQ2 . . . 141
5.2.3 Discussion of RQ3 . . . 142
5.2.4 Discussion of RQ4 . . . 142
5.2.5 Discussion of RQ5 . . . 143
5.3 Comparison with Prior Studies . . . 143
5.3.1 Comparison with Machine Learning-Based Credit Risk Studies . . . 143
5.3.2 Comparison with Feature Selection and Optimization Approaches . . . 144
5.3.3 Methodological Contributions beyond Existing Benchmarks . . . 146
5.4 Implications . . . 147
5.4.1 Theoretical Implications . . . 147
5.4.2 Methodological Implications . . . 148
5.4.3 Practical and Institutional Implications . . . 149
5.5 Limitations . . . 150

6 Conclusion and future work **151**
6.1 Global Synthesis of the Research . . . 151
6.2 Summary of Contributions . . . 152
6.2.1 Practical and Institutional Contributions . . . 152
6.2.2 Scientific Contributions . . . 153
6.2.3 Methodological Contributions . . . 153
6.3 Scope and Limitations of the Study . . . 154
6.4 Future Research Directions . . . 155
6.4.1 Cross-Institutional and Contextual Validation . . . 155
6.4.2 Temporal and Longitudinal Extensions . . . 155
6.4.3 Adaptive and Hybrid Optimization Strategies . . . 156
6.4.4 Governance, Fairness, and Policy-Oriented Modeling . . . 156
6.5 Final Remarks . . . 156

7 References **158**

List of Figures

2.1 Reported Number of Classifier used in a Paper. 17
2.2 Frequency of Machine Learning Classifiers in Reviewed Documents. 18
2.3 Usage of Benchmark vs. Real-World Application Datasets. 27
2.4 Sources of Benchmark Datasets. 28
2.5 Distribution of Reported Metrics by Category. 33
2.6 Comparison of previously identified metrics and newly identified metrics by Category. 34
2.7 Trend of Previously Defined Metrics Distribution in Documents. 35
2.8 Trend of Newly Identified Metrics Distribution in Documents. 36

3.1 Simplified entity–relationship model supporting FSCU data integration. 47

4.1 Data processing pipeline for FSCU dataset construction and baseline analysis. . 53
4.2 Confusion matrices for selected baseline ensemble models 56
4.3 Confusion Matrices (Part I) 68
4.4 Confusion Matrices (Part II) 69
4.5 Average permutation feature importance across all trained models. 71
4.6 Linear models feature importance (tuned configurations). 73
4.7 Tree-based models feature importance (tuned configurations). 74
4.8 Decision Tree snapshots at different depths; Depth = 4. 75
4.9 Decision Tree snapshots at different depths; Depth = 5. 75
4.10 Decision Tree snapshots at different depths; Depth = 11. 76
4.11 SHAP values for LightGBM. 78
4.12 Missing values across dataset features 86
4.13 Distribution of selected numerical features (Part I) 87
4.14 Distribution of selected numerical features (Part II) 88
4.15 Distribution of selected numerical features (Part III 89
4.16 Distribution of selected numerical features (Part IV) 90
4.17 Distribution of selected categorical features (Part I) 91
4.18 Distribution of selected categorical features (Part II) 92
4.19 Distribution of selected categorical features (Part III) 93
4.20 Distribution of selected categorical features (Part IV) 94
4.21 Pearson correlation heatmap among numerical features 95
4.22 Distribution of enforceability dates 96
4.23 Decomposition of enforceability dates into day, month, and year components . 96
4.24 Confusion Matrices KNN 109
4.25 Confusion Matrices LightGBM 110
4.26 Confusion Matrices Random Forest 111
4.27 Heatmap Features Selected by GWO 112

4.28 Heatmap Features Selected by PSO 112
4.29 Heatmap Features Selected by WOA 112
4.30 Heatmap Features Used by KNN 113
4.31 Heatmap Features Used by Random Forest 113
4.32 Heatmap Features Used by LightGBM 113
4.33 Mean ROC Curve of the bioinspired DNN model 128
4.34 Mean ROC Curve of the RF baseline model 129
4.35 Boxplot AUC 130
4.36 Boxplot Matthews Correlation 130
4.37 Feature Importance (Integrated Gradients) 133

List of Tables

2.1 Public datasets (benchmarks). 28
2.2 Real-world applications datasets. 29

4.1 Baseline predictive performance of classification models on the FSCU dataset . 54
4.2 Initial pool of pre-declarative academic and administrative features (Stage II) . 59
4.3 Final pre-declarative feature set used for interpretable modeling (Stage II) . . . 60
4.4 K-Nearest Neighbors Results 64
4.5 Gaussian Naive Bayes Results 65
4.6 Logistic Regression Results 65
4.7 Linear Support Vector Machine Results 65
4.8 Decision Tree Results 65
4.9 Random Forest Results 66
4.10 Light Gradient Boosting Machine Results 66
4.11 Faculty dummy feature values used to interpret faculty-related predictors. . . . 72
4.12 Example decision path from the optimized Decision Tree (depth = 4) 77
4.13 Final feature set: variable names, data types, and ranges 83
4.14 Feature set: detailed descriptions of each variable 85
4.15 Features with missing values 86
4.16 Recall minority class 106
4.17 F-Score minority class 106
4.18 Precision minority class 106
4.19 Recall majority class 107
4.20 F-Score majority class 107
4.21 Precision majority class 107
4.22 Evaluation Metrics with 95% Confidence Intervals 127
4.23 Results of the paired Student's t-test between the bioinspired DNN and RF baseline models 128
4.24 Confusion matrix of the bioinspired DNN model 131
4.25 Confusion matrix of the RF baseline model 131

Abstract

Income-contingent student loan systems rely critically on borrowers' compliance with income declaration requirements to ensure financial sustainability, administrative effectiveness, and equitable repayment. Anticipating declaration non-compliance, however, remains challenging due to strict informational constraints, institutional complexity, and the evolving academic and financial trajectories of borrowers. This doctoral research addresses this challenge through the design, validation, and application of a stage-based predictive modeling framework aimed at anticipating income declaration compliance and declaration abandonment using real-world institutional data from the University Credit Solidarity Fund (*Fondo Solidario de Crédito Universitario*, FSCU) system.

The research adopts a progressive methodological design structured into five interconnected stages. Initial stages focus on the construction of a reproducible institutional data pipeline and the evaluation of interpretable machine learning models for predicting first-time income declaration compliance using exclusively pre-declarative academic and administrative information. These stages assess the feasibility, robustness, and limits of early-stage prediction under strict temporal and causal constraints. Subsequent stages address the prediction of declaration abandonment among previously compliant borrowers, incorporating enriched behavioral and administrative representations and metaheuristic-based feature selection to manage high-dimensional feature spaces while preserving interpretability. Advanced bioinspired modeling approaches are then explored to capture complex non-linear patterns associated with declaration discontinuity. A final integrative stage enables cross-stage comparison, supporting a unified interpretation of predictive behavior across compliance scenarios.

Empirical results demonstrate that income declaration behaviors can be anticipated with statistically meaningful and stable performance using institutional data alone, enabling early risk identification without reliance on post-declarative information. Metaheuristic-driven feature selection improves predictive robustness and parsimony while maintaining explanatory coherence, and a reduced subset of academic, financial, and administrative variables consistently explains compliance and non-compliance patterns across stages. Overall, the proposed framework constitutes a structured and extensible methodological contribution to compliance risk prediction in income-contingent loan systems, offering both scientific insights and practical relevance for institutional decision-making.

Keywords: Income-contingent student loans, Income declaration compliance, Predictive modeling, Machine learning, Optimization, Metaheuristics, Feature Selection, Institutional data, Model interpretability.

1 Introduction

1.1 Background and Motivation

In Chile, public instruments such as scholarships, tuition-free policies, and student loans have played a central role in expanding access to higher education. Among these instruments, the University Credit Solidarity Fund (*Fondo Solidario de Crédito Universitario*, FSCU) stands out as a university-based, income-contingent loan created by Law 19.287 and Decree No. 225 of 1994 [1, 2]. Managed by the Ministry of Education (*Ministerio de Educación de Chile*, MINEDUC), the program supports students belonging to the lowest 80% of the income distribution, covering between 50% and 100% of tuition costs depending on socioeconomic background [3, 4]. Its concessional design, characterized by a 2% real annual interest rate and income-contingent repayment capped at 5% of annual income, reflects an effort to balance equity in access with financial responsibility [5, 6].

Despite these design features, repayment and procedural compliance remain persistent challenges. Each year, the Council of Rectors of Chilean Universities (*Consejo de Rectoras y Rectores de las Universidades Chilenas*, CRUCH) publishes arrears lists that are transferred to the General Treasury of the Republic (*Tesorería General de la República*, TGR) for enforcement [7, 8]. Similar difficulties are observed in other state-backed programs. The State Guaranteed Credit for Higher Education (*Crédito con Aval del Estado*, CAE), for example, reports default rates approaching 48%, particularly among students who do not complete their studies or belong to low-income groups [9, 10]. These figures illustrate how repayment difficulties are closely linked to labor informality, delayed graduation, and heterogeneous post-graduation outcomes.

At the Pontificia Universidad Católica de Valparaíso (PUCV), a substantial proportion of undergraduate students finance their studies through the FSCU. This creates an ongoing institutional need to monitor repayment behavior, anticipate risk, and support borrowers beyond the ex-post publication of defaulting lists [7]. The university already manages extensive administrative records, including academic trajectories, income declarations, arrears, and collection data, which constitute a valuable but underutilized basis for predictive analysis [7, 8].

This study is motivated by the opportunity to leverage such institutional information more effectively. By applying data analysis and machine learning techniques to administrative records, the research seeks to identify early signals of declarative non-compliance and to support more proactive and evidence-based decision-making. Beyond its local relevance, this case reflects a broader international trend toward the application of data-driven credit scoring and predictive analytics in higher education financing, illustrating how machine learning can contribute to more equitable and sustainable student loan management across diverse contexts.

1.2 Problem Statement

The FSCU is a Chilean state loan granted to students enrolled in universities affiliated with the CRUCH [11]. The program covers part or all of the annual tuition fees of eligible students, who must belong to the lowest 80% of household income distribution and meet minimum academic admission requirements. Coverage reaches 100% of the reference tuition fee for students in the most vulnerable 60% and ranges between 50% and 100% for those between 60% and 80%.

The loan is denominated in Monthly Tax Units (*Unidad Tributaria Mensual*,UTM) and carries a real annual interest rate of 2%. Repayment begins two years after graduation and is structured as annual installments equivalent to 5% of the borrower's income from the previous year. The maximum repayment period is 12 years for debts below 200 UTM and 15 years for debts equal to or exceeding 200 UTM, after which any remaining balance is forgiven.

The FSCU is primarily regulated by Law No. 19.287 [12], enacted in 1994, and by Law No. 20.572 [13], which introduced substantial changes to the collection and compliance mechanisms in 2012. Prior to this reform, borrowers could alternate between income declaration and fixed-payment schemes, and non-declaration did not necessarily imply the loss of benefits in subsequent years. Additionally, defaulters could be reported to commercial credit registries, affecting their access to financial services.

The 2012 amendment established the annual income declaration as a mandatory procedural requirement to retain the benefits of income-contingent repayment. Borrowers are required to submit their declaration by the last business day of May each year. Failure to comply results in the loss of variable repayment conditions, exemptions for periods without income, debt forgiveness after the statutory term, and discounts for early full repayment.

From an operational perspective, the collection process differs significantly depending on whether the borrower submits an income declaration. For compliant borrowers, universities calculate an annual variable installment that expires in May of the following year. For non-declarants, the total outstanding debt is redistributed into fixed annual installments with maturities in December, over periods ranging from six to fifteen years depending on the amount owed [14].

While the reformed framework aims to incentivize compliance and improve institutional financial planning, it also introduces a critical source of uncertainty for universities, which must anticipate declarative behavior under strict temporal and informational constraints. In practice, institutions face the challenge of identifying, before declaration outcomes are observed, which borrowers are likely to comply, which are at risk of first-time non-declaration, and which may discontinue compliance after having previously declared.

Understanding and predicting income declaration compliance and abandonment therefore emerges as a central operational and analytical problem for the FSCU, requiring data-driven approaches capable of operating across different informational regimes while preserving causal validity, interpretability, and institutional plausibility.

> *The central problem addressed in this thesis arises from the institutional inability to anticipate income declaration behavior within the FSCU system before its consequences materialize. Universities must manage repayment conditions, cash flow projections, and compliance enforcement under strict legal deadlines, yet declara-*

tion outcomes are only observable ex post. This delayed observability limits the effectiveness of reactive administrative strategies and exposes the system to financial uncertainty and operational inefficiencies. The problem, therefore, is not merely to assess compliance after it occurs, but to predict, in advance and under realistic informational constraints, which borrowers are likely to comply, which are at risk of first-time non-declaration, and which may abandon declaration after prior compliance. Addressing this problem requires predictive approaches capable of operating across different stages of the declaration lifecycle, using heterogeneous institutional data, while preserving temporal validity, interpretability, and decision-support relevance.

1.3 Solution Proposal

This doctoral research adopts an applied, quantitative, and predictive methodological approach to analyze income declaration compliance within the FSCU system. The central purpose of the study is to design, validate, and apply advanced computational intelligence models capable of anticipating declarative non-compliance behaviors under real-world institutional conditions, thereby supporting more proactive and evidence-based decision-making in the administration of income-contingent student loan systems.

The research is grounded in the use of real institutional data and addresses the problem from an engineering and information systems perspective. It combines data engineering processes, supervised machine learning techniques, and optimization-oriented methods to model compliance behavior across different stages of the income declaration lifecycle. Rather than relying on synthetic benchmarks or simulated environments, the study emphasizes methodological rigor, temporal coherence, and operational realism.

Methodologically, the work is structured as a sequence of interrelated research stages, each corresponding to a specific analytical objective and increasing level of data availability and modeling complexity. These stages range from the construction and characterization of institutional datasets and baseline predictive analyses, to interpretable modeling of first-time income declaration compliance, and finally to the application of metaheuristic-based feature selection strategies for predicting declaration abandonment among previously compliant borrowers.

The general objective of this research is to assess to what extent predictive modeling and optimization techniques can effectively capture and explain declarative compliance behavior using institutional data, while balancing predictive performance, interpretability, and operational applicability. A comprehensive description of the research design, data sources, methodological stages, modeling strategies, evaluation protocols, and implementation details is provided in Chapter 3.

Rather than pursuing a single optimal predictive solution, this doctoral research adopts a stage-based methodological perspective that explicitly examines how predictive feasibility, robustness, interpretability, and expressive capacity evolve as informational constraints are progressively relaxed. This design enables the identification of empirical performance boundaries and methodological trade-offs across distinct compliance scenarios, providing a structured

foundation for both scientific analysis and institutional decision support.

By integrating predictive modeling, optimization, and interpretability within a unified and progressive framework, the research seeks to contribute not only empirical results, but also methodological guidance for the deployment of data-driven decision-support systems in income-contingent credit administration.

1.4 Contributions

The main contributions of this doctoral research can be summarized as follows:

- The conceptualization and formalization of declarative compliance behavior as a predictive modeling problem within income-contingent student loan systems, extending traditional credit risk research beyond the exclusive focus on financial default and explicitly addressing procedural non-compliance in income-contingent repayment schemes.
- The design and validation of an end-to-end data processing and modeling pipeline for institutional loan data, enabling the construction of analyzable datasets from heterogeneous administrative records under real-world operational and temporal constraints.
- The empirical assessment of baseline and interpretable machine learning models for predicting first-time income declaration compliance using strictly pre-declarative academic and administrative information, providing a controlled reference scenario grounded in temporal precedence and institutional causality.
- The development and evaluation of a metaheuristic-based feature selection framework for modeling declaration abandonment among previously compliant borrowers, demonstrating the role of optimization techniques in managing high-dimensional feature spaces derived from declaration processes while balancing predictive performance, robustness, and interpretability.
- An integrated experimental analysis contrasting predictive scenarios across different stages of the declaration lifecycle, highlighting trade-offs between data availability, model complexity, interpretability, and predictive effectiveness within institutional decision-support systems.
- A context-aware contribution to the management of publicly administered income-contingent student loan systems, providing empirical evidence to support early intervention strategies aimed at improving compliance monitoring and contributing to the long-term financial sustainability of public credit schemes.
- The generation of scientific outputs derived from different stages of the research, including peer-reviewed publications already published in indexed journals [15–17], as well as additional manuscripts currently under peer review, which validate the methodological soundness and relevance of the proposed approaches within the broader fields of computational intelligence, predictive modeling, and institutional analytics.

In addition to the methodological and conceptual contributions outlined above, this doctoral research has generated multiple scientific outputs that provide external validation of its relevance and originality. Different stages of the research have resulted in peer-reviewed publications and conference contributions addressing complementary aspects of the proposed framework.

Specifically, the predictive modeling of first-time income declaration compliance has been submitted to a Q1 journal indexed in the Web of Science, where it has passed editorial screening and is currently under first-round peer review. The metaheuristic-based feature selection framework for declaration abandonment prediction has been submitted to a Q1 Web of Science journal, while the bioinspired deep learning models for declaration discontinuity prediction have received editorial acceptance and are currently under peer review in the same venue. In addition, a systematic literature review supporting the theoretical foundations of this thesis has been published in a Q1 Web of Science journal.

Earlier methodological results related to baseline predictive modeling were also presented at an international engineering conference and accepted for publication in its peer-reviewed proceedings. Collectively, these scientific outputs demonstrate that the contributions of this thesis extend beyond a single experimental study and have been recognized as relevant advances within the fields of computational intelligence, predictive modeling, and institutional decision-support systems.

1.5 Thesis Organization

The remainder of this thesis is organized as follows. Chapter 2 presents the theoretical foundations and state of the art relevant to this research, covering income-contingent student loan systems, compliance behavior, predictive modeling, feature selection techniques, and metaheuristic optimization. This chapter synthesizes prior work and identifies the research gaps that motivate the proposed methodological framework.

Chapter 3 describes the general research methodology adopted in this study. It details the research design, data sources, dataset construction processes, modeling strategies, evaluation protocols, and implementation considerations that ensure reproducibility and methodological rigor across all experimental stages.

Chapter 4 reports the empirical results of the research through a sequence of methodological stages. This chapter presents the construction and characterization of the FSCU dataset, the baseline predictive analysis, the interpretable modeling of first-time income declaration compliance, and the metaheuristic-optimized prediction of declaration abandonment among previously compliant borrowers.

Chapter 5 provides an integrated discussion of the results, examining their methodological, theoretical, and practical implications. This chapter contrasts findings across predictive scenarios, relates the results to prior studies, and analyzes the limitations and scope of applicability of the proposed approaches in institutional contexts.

Finally, Chapter 6 concludes the thesis by summarizing the main findings, revisiting the research questions and objectives, outlining the scientific and practical contributions of the work, and proposing directions for future research.

2 State of the Art

2.1 Introduction

This chapter establishes the theoretical and analytical foundations that support the research. It is structured to clearly differentiate between conceptual definition, systematic examination of prior studies, and critical synthesis of existing approaches, thereby ensuring conceptual clarity and methodological coherence throughout the thesis.

The chapter begins with the *Conceptual Foundations* section, which defines the key theoretical constructs, analytical concepts, and design principles underlying the study. This section introduces and formalizes the notions of income-contingent student loan systems, declarative compliance, predictive modeling, feature selection, metaheuristic optimization, and institutional design requirements such as interpretability and stability. Its primary role is to provide a stable conceptual reference framework that is subsequently used, but not redefined, in the methodological and empirical chapters.

The *Literature Review* section then examines how these concepts have been addressed in prior research. This review is explicitly grounded in a systematic literature review conducted as part of this doctoral work and synthesizes peer-reviewed studies according to thematic dimensions aligned with the conceptual framework. Rather than offering a descriptive survey of individual papers, this section organizes the literature around dominant methodological approaches, modeling strategies, and optimization practices relevant to individual credit risk and compliance-related prediction.

Building on this foundation, the *State of the Art* section provides a higher-level synthesis of prevailing research patterns and dominant paradigms identified in the literature. Its purpose is not to introduce new studies, but to abstract and consolidate common assumptions, methodological trends, and recurring design choices that characterize the current state of research in the field.

Finally, the chapter advances to a *Critical Analysis* and an *Identified Research Gap*, where limitations and unresolved issues emerging from the state of the art are articulated, and concludes with a *Theoretical Positioning* that situates the present thesis within the existing body of knowledge. Together, these sections establish a coherent theoretical trajectory from foundational concepts to research positioning, ensuring that subsequent methodological and empirical decisions are firmly grounded in both theory and prior evidence.

2.2 Conceptual Framework

2.2.1 Income-Contingent Student Loan Systems as Institutional Frameworks

Income-contingent student loan systems constitute a specific class of financing arrangements in which repayment obligations are explicitly linked to the borrower's post-graduation income rather than to fixed repayment schedules. Unlike mortgage-style student loans, where repayment amounts and timelines are predefined and largely independent of post-study labor outcomes,

income-contingent schemes adjust repayment requirements dynamically according to observed income levels. This design aims to reduce repayment pressure during periods of low earnings and to mitigate the risk of financial distress among graduates, while redistributing part of the repayment risk from borrowers to the lending institution or the state [18].

From an institutional perspective, income-contingent systems fundamentally alter the nature of credit risk. Risk is no longer concentrated exclusively on short-term payment failure, but is instead distributed across a broader temporal horizon and mediated by administrative and procedural mechanisms. Repayment assessment depends not only on income realization but also on the borrower's continued interaction with the institutional framework, particularly through the periodic submission of income declarations. These declarative processes function as mandatory gateways that enable the system to determine repayment obligations and enforce income-based rules [19, 20].

As a consequence, the operational effectiveness of income-contingent student loan systems relies heavily on borrowers' compliance with declarative requirements. Income declaration is not a peripheral administrative task, but a central mechanism through which the system translates labor market outcomes into repayment decisions. Failure to comply with declarative obligations disrupts this mechanism, regardless of the borrower's actual repayment capacity, and introduces uncertainty into institutional monitoring and enforcement processes [21, 22].

Importantly, non-compliance in this context does not necessarily coincide with immediate financial insolvency or inability to repay. Borrowers may fail to declare income despite being economically active, due to factors such as labor informality, limited administrative literacy, strategic avoidance, or gradual disengagement from institutional processes. This distinction challenges conventional interpretations of credit risk that equate adverse outcomes exclusively with financial default, and highlights the need to conceptualize declarative behavior as an analytically distinct phenomenon [23].

The literature synthesized in the systematic review indicates that most predictive studies in credit risk assessment implicitly assume that repayment-related outcomes capture the full risk process. Such approaches tend to overlook the procedural and behavioral stages that precede observable repayment events, particularly in income-contingent schemes where compliance acts as a prerequisite for repayment determination. As a result, models that focus solely on realized repayment outcomes provide an incomplete representation of system dynamics and limit the capacity for early institutional intervention [18, 24].

These characteristics position income-contingent student loan systems as institutionally mediated environments in which credit risk emerges from the interaction between economic conditions, borrower behavior, and administrative processes. Analytical frameworks applied in this context must therefore be capable of integrating behavioral, administrative, and socioeconomic information, and of distinguishing between different stages of borrower engagement with the system. This perspective provides the conceptual basis for treating declarative compliance as a primary predictive outcome, rather than as a secondary indicator of repayment performance, and motivates the modeling approach developed in this thesis.

2.2.2 Declarative Compliance as an Observable Behavioral Process

Declarative compliance refers to the borrower's adherence to formal administrative obligations required by income-contingent student loan systems, particularly the timely and accurate submission of income information to the responsible institution. In contrast to repayment outcomes, which reflect realized financial transfers, declarative compliance constitutes a procedural behavior that enables the system to operationalize income-based repayment rules. As such, it represents an observable and institutionally mediated process rather than a purely economic outcome [19, 20].

From a behavioral standpoint, declarative compliance can be conceptualized as a sequence of actions that unfolds over time, encompassing initial engagement with the declaration system, continued participation across repayment periods, and potential disengagement or abandonment. This process-oriented view departs from conventional binary representations of compliance and non-compliance by recognizing that borrowers may transition between compliant and non-compliant states without immediately exhibiting repayment failure or financial distress [21, 22]. Such transitions are particularly relevant in income-contingent schemes, where declaration behavior functions as a prerequisite for repayment assessment rather than as a direct indicator of solvency.

The literature reviewed indicates that most predictive studies in individual credit risk assessment implicitly treat compliance-related variables as secondary or auxiliary indicators, subordinated to terminal outcomes such as default or delinquency. This perspective overlooks the analytical value of declarative behavior as an early and actionable signal of institutional risk. In income-contingent systems, failure to declare income interrupts the flow of information required for repayment determination, thereby generating uncertainty independently of the borrower's actual economic capacity [18, 24].

Declarative non-compliance may arise from heterogeneous and context-dependent factors. Empirical studies report associations with labor market informality, administrative complexity, limited awareness of obligations, and strategic behavior aimed at delaying or avoiding institutional oversight. Importantly, these factors do not necessarily coincide with immediate repayment incapacity, which complicates their interpretation within traditional credit risk frameworks [23]. Treating declarative non-compliance as equivalent to financial default therefore risks conflating distinct behavioral mechanisms and obscuring opportunities for targeted intervention.

Modeling declarative compliance as a predictive outcome requires an explicit distinction between different stages of borrower behavior. In particular, first-time compliance and subsequent abandonment of declarative obligations constitute analytically distinct phenomena, with potentially different determinants and institutional implications. The systematic review underlying this thesis suggests that existing models rarely differentiate between these stages, instead aggregating them into undifferentiated risk labels [18, 21]. This aggregation limits the capacity of predictive systems to support early monitoring and adaptive administrative responses.

By framing declarative compliance as an observable behavioral process, this thesis adopts a dynamic and institution-centered perspective on risk. Compliance is treated neither as a static attribute nor as a terminal outcome, but as a measurable manifestation of ongoing engagement with the loan system. This conceptualization provides the foundation for predictive modeling approaches that leverage pre-declarative and historical institutional data to anticipate both initial

compliance and subsequent abandonment behaviors. It also establishes the basis for evaluating predictive models not only in terms of accuracy, but in relation to their capacity to support timely and interpretable decision-making within income-contingent student loan systems [20, 24].

2.2.3 Predictive Modeling in Institutional Decision-Support Systems

Predictive modeling in institutional contexts refers to the systematic use of historical and administrative data to anticipate future states or behaviors that are relevant for organizational decision-making. Unlike exploratory or descriptive analytics, predictive models are designed to generate forward-looking estimates that support anticipatory actions, resource allocation, and policy interventions. In financial and educational environments, these models increasingly function as core components of decision-support systems rather than as standalone analytical tools [19, 20].

In the domain of individual credit risk and educational-financial systems, predictive modeling has traditionally focused on estimating the likelihood of adverse terminal outcomes, such as default or delinquency. These approaches typically assume that risk materializes primarily through realized repayment failure and that predictive accuracy constitutes the principal criterion for model usefulness. While such assumptions may be appropriate in conventional credit markets, they are insufficient in income-contingent student loan systems, where repayment obligations are mediated by procedural and behavioral requirements [18, 23].

Within institutional decision-support systems, predictive models are expected to satisfy requirements that extend beyond numerical performance. In particular, they must operate under constraints related to interpretability, stability, and accountability. Predictions often inform administrative actions that affect individuals directly, such as monitoring, communication, or enforcement measures. Consequently, institutional stakeholders require models whose outputs can be examined, justified, and explained within formal governance structures [20, 24].

Predictive modeling in these settings therefore entails a shift from purely outcome-oriented estimation toward process-aware risk anticipation. Rather than modeling only end states, such as repayment default, decision-support systems benefit from predictions that capture intermediate behavioral stages that precede adverse outcomes. In income-contingent loan schemes, declarative compliance represents one such stage, as it conditions the system's capacity to evaluate repayment obligations and to engage borrowers proactively [21, 22].

The literature reviewed indicates that many existing predictive frameworks implicitly embed decision-support assumptions without explicitly articulating them. Models are often evaluated in isolation from the institutional processes they are intended to support, leading to misalignment between predictive objectives and operational needs. For example, high-performing models may rely on unstable predictors or opaque decision rules that limit their applicability in longitudinal or policy-driven contexts [18, 24].

In response to these limitations, predictive modeling within institutional decision-support systems increasingly emphasizes alignment between modeling choices and organizational objectives. This includes careful definition of target variables, selection of appropriate evaluation metrics, and consideration of how model outputs will be used in practice. In the context of income-contingent student loan systems, this perspective motivates the use of predictive models

that explicitly target declarative compliance behaviors and that are designed to support early monitoring, targeted intervention, and transparent decision-making [20, 23].

Accordingly, this thesis adopts a view of predictive modeling as an institutional artifact embedded within a broader decision-support framework. Models are not evaluated solely on their capacity to minimize prediction error, but on their ability to generate stable, interpretable, and actionable insights that align with the operational logic of income-contingent repayment schemes. This conceptualization informs subsequent methodological choices related to algorithm selection, feature selection, and optimization strategies, which are introduced in the following sections.

2.2.4 Supervised Machine Learning for Classification Tasks

Supervised machine learning provides a formal framework for learning predictive relationships between a set of observed attributes and a predefined target variable using labeled data. In classification tasks, the objective is to assign each observation to one of a finite number of outcome categories based on patterns inferred from historical examples. This paradigm has become central to predictive analytics in institutional contexts, where categorical decisions often underpin administrative actions and policy interventions [18, 23].

In individual-level risk assessment, classification problems are most commonly formulated as binary tasks. Examples include distinguishing between compliant and non-compliant borrowers, identifying individuals at low or high risk of adverse outcomes, or predicting the occurrence of specific behavioral events within a defined time horizon. Binary classification aligns naturally with institutional decision-making processes, which frequently rely on threshold-based rules to trigger monitoring, communication, or enforcement mechanisms [20, 24].

A wide range of supervised learning algorithms has been applied to classification problems in financial and educational domains. Linear models, such as logistic regression, remain widely used due to their interpretability, computational efficiency, and well-understood statistical properties. However, their capacity to capture complex relationships is limited when data exhibit nonlinearities or interaction effects among predictors. To address these limitations, non-linear methods, including decision trees, ensemble-based classifiers, and kernel-based approaches, have been increasingly adopted in applied research [23, 25].

Ensemble methods, in particular, have received considerable attention due to their ability to combine multiple weak learners into more robust predictive models. By aggregating diverse decision rules, ensembles can reduce variance and improve generalization performance in heterogeneous and high-dimensional datasets. This property is especially relevant in institutional settings, where predictors originate from multiple administrative and socioeconomic processes and may exhibit complex dependencies [18, 25].

Despite their predictive advantages, the deployment of supervised classification models in institutional environments introduces methodological and practical challenges. Administrative datasets frequently present class imbalance, noisy labels, and missing or inconsistently recorded variables. Moreover, observed outcomes may reflect procedural or behavioral conditions rather than purely economic states, complicating both model training and interpretation [20, 24]. These characteristics necessitate careful consideration of model selection, evaluation strategies, and validation protocols.

In the context of income-contingent student loan systems, supervised classification models are used not merely to estimate abstract probabilities, but to support concrete administrative decisions related to compliance monitoring and intervention. Consequently, algorithm selection cannot be guided solely by predictive accuracy. Models must also exhibit stable behavior across data partitions and time periods, and their decision mechanisms should be sufficiently transparent to support institutional accountability [19, 20].

Accordingly, this thesis adopts supervised machine learning as the core predictive paradigm for modeling declarative compliance behavior. Classification algorithms are selected and evaluated within a decision-support perspective that prioritizes interpretability, robustness, and operational relevance alongside predictive performance. This framing provides the conceptual basis for the evaluation metrics and feature selection strategies introduced in the following sections.

2.2.5 Evaluation Metrics for Compliance-Oriented Predictive Models

Evaluation metrics constitute a central component of predictive modeling, as they determine how model performance is quantified, interpreted, and compared. In classification tasks, metrics provide a formal link between model outputs and the practical objectives of the decision-support system in which the model is embedded. Consequently, metric selection is not a neutral technical choice, but a substantive methodological decision that shapes how predictive success is defined and operationalized [20, 24].

In binary classification settings, performance evaluation is commonly grounded in the confusion matrix, which summarizes the relationship between predicted and observed class labels. Metrics derived from this representation capture different aspects of model behavior, including the ability to correctly identify positive cases, avoid false alarms, and balance errors across classes. While overall accuracy is frequently reported due to its simplicity, it provides a limited and potentially misleading view of performance in contexts characterized by class imbalance or asymmetric decision costs [20].

Precision and recall are particularly relevant in compliance-oriented predictive modeling. Precision quantifies the proportion of predicted non-compliant cases that are correctly identified, reflecting the reliability of alerts or interventions triggered by the model. Recall measures the proportion of actual non-compliant cases that are successfully detected, capturing the model's capacity to identify relevant events of interest. In institutional settings, these metrics correspond to different operational priorities, such as minimizing unnecessary administrative actions or maximizing early detection of disengagement [19, 24].

The F1-score combines precision and recall into a single harmonic measure, providing a balanced summary of model performance when both types of error are considered relevant. This metric is especially useful in imbalanced classification problems, where high accuracy may be achieved by trivially predicting the majority class while failing to capture meaningful minority-class behavior. As a result, F1-based evaluation has become common in studies addressing individual-level risk and behavioral prediction [18, 23].

Beyond point estimates, evaluation metrics must be interpreted in light of the institutional context in which predictions are applied. In income-contingent student loan systems, predictive

models are used to support monitoring and intervention strategies rather than to automate irreversible decisions. Accordingly, metrics should reflect the model's capacity to provide timely and actionable signals, rather than solely its ability to minimize aggregate classification error. This perspective underscores the importance of metrics that emphasize minority-class detection and balanced performance [21, 22].

The literature reviewed indicates that many predictive studies report multiple evaluation metrics without explicitly linking them to the underlying decision-support objectives. In some cases, metric selection appears driven by convention rather than by institutional relevance, limiting the interpretability and practical value of reported results [24]. This limitation is particularly salient in optimized predictive frameworks, where improvements in one metric may coincide with degradation in others.

In this thesis, evaluation metrics are treated as integral components of the modeling framework rather than as post hoc reporting tools. Metrics are selected to align with the objectives of predicting declarative compliance behavior and supporting institutional decision-making within an income-contingent loan system. This alignment ensures that subsequent empirical results can be interpreted consistently with the conceptual definitions established in this chapter and facilitates transparent comparison across modeling stages and experimental configurations.

2.2.6 Feature Selection in High-Dimensional Institutional Data

Institutional datasets used in educational and financial contexts are typically characterized by high dimensionality, heterogeneous variable sources, and complex dependency structures. Variables are often collected for administrative or operational purposes rather than for predictive modeling, resulting in feature spaces that include redundant, weakly informative, or noisy attributes. In such settings, the direct application of supervised learning algorithms without prior feature selection may lead to increased model complexity, reduced generalization capacity, and limited interpretability [19, 26].

Feature selection addresses these challenges by identifying a subset of original variables that are most relevant to the predictive task, while preserving the semantic meaning of institutional attributes. Unlike dimensionality reduction techniques that transform the feature space into latent representations, feature selection retains the original variables, facilitating traceability and explanation. This property is particularly important in public-sector and policy-driven applications, where models must support transparent decision-making and allow stakeholders to interpret the role of specific academic, socioeconomic, or administrative factors [20, 27].

From a methodological perspective, feature selection can be formulated as a combinatorial optimization problem. Given a set of candidate variables, the objective is to identify the subset that optimizes a predefined criterion, typically related to predictive performance, while controlling model complexity. The search space grows exponentially with the number of features, rendering exhaustive exploration infeasible in high-dimensional scenarios commonly encountered in institutional data [28, 29].

The literature synthesized in the systematic review indicates that feature selection approaches are commonly categorized into filter, wrapper, and embedded methods. Filter methods evaluate variables independently of the predictive model using statistical or information-theoretic

criteria, offering computational efficiency but limited capacity to capture interaction effects. Embedded methods integrate feature selection into model training through mechanisms such as regularization or tree-based splitting criteria. Wrapper methods, in contrast, evaluate candidate feature subsets directly through model performance, enabling the identification of feature interactions at the cost of increased computational complexity [28–30].

In individual credit risk and compliance-related predictive modeling, wrapper-based feature selection has attracted growing attention due to its flexibility in handling nonlinear relationships and heterogeneous predictors. Studies reviewed suggest that performance improvements often depend on combinations of academic, behavioral, and contextual variables rather than on isolated features. However, the effectiveness of wrapper methods is critically dependent on the strategy used to explore the feature space, particularly in high-dimensional institutional datasets [31–33].

Beyond predictive performance, feature selection has direct implications for model stability and interpretability. Feature subsets that vary substantially across data partitions, temporal samples, or optimization runs may lead to inconsistent model behavior and undermine institutional trust. Despite this, many existing studies prioritize accuracy gains without systematically assessing the robustness or temporal consistency of selected features, especially in administrative and policy-oriented applications [23, 24].

In the context of income-contingent student loan systems, these considerations are particularly salient. Predictive models are expected to inform longitudinal monitoring and intervention strategies, requiring feature relevance to remain stable over time and across cohorts. Feature selection strategies that balance predictive performance with robustness and interpretability are therefore essential to ensure that models remain operationally meaningful and institutionally acceptable. This perspective motivates the integration of controlled and stability-aware feature selection mechanisms within the modeling framework developed in this thesis.

2.2.7 Metaheuristic Optimization as a Search Paradigm

Metaheuristic optimization refers to a class of general-purpose search strategies designed to address optimization problems for which exact or deterministic methods are computationally impractical. These problems are typically characterized by high-dimensional, nonlinear, multimodal, or discrete search spaces, where objective functions may be expensive to evaluate or lack analytical structure. In such settings, metaheuristics provide flexible and scalable mechanisms for exploring large solution spaces without imposing restrictive assumptions on problem formulation [28, 30].

A defining characteristic of metaheuristic methods is their reliance on stochastic search mechanisms rather than deterministic optimization rules. Through iterative exploration guided by randomization and heuristic operators, these methods aim to balance global exploration of the search space with local exploitation of promising regions. This balance is commonly regulated through algorithmic parameters or control strategies that influence population diversity, convergence dynamics, and search intensity over time [29, 32].

In predictive modeling, metaheuristic optimization has been widely applied to tasks that cannot be efficiently solved through analytical optimization. These include feature selection, hyperparameter tuning, and model structure configuration. The systematic review underlying this

thesis indicates that, in individual credit risk and institutional analytics, metaheuristics are most frequently employed as wrapper-based optimization mechanisms, where candidate solutions are evaluated directly through model performance metrics [31, 33].

Feature selection constitutes a particularly natural application domain for metaheuristic optimization. When the objective is to identify an optimal subset of variables from a large candidate pool, the number of possible feature combinations grows exponentially with dimensionality. Exhaustive search becomes infeasible even for moderately sized feature sets, and greedy or deterministic heuristics may fail to capture complex interaction effects. Metaheuristic approaches address this challenge by enabling guided exploration of the combinatorial search space while maintaining flexibility with respect to feature dependencies [28, 29].

Despite their flexibility and empirical effectiveness, the use of metaheuristic optimization introduces additional methodological considerations. Due to their stochastic nature, optimization outcomes may vary across runs, and performance is often sensitive to initialization strategies, parameter settings, and stopping criteria. The literature reviewed shows that these aspects are frequently underreported or insufficiently controlled, particularly in applied and institution-oriented studies, limiting reproducibility and interpretability [23, 31].

In public-sector and policy-driven applications, such as income-contingent student loan systems, these limitations acquire particular relevance. Predictive models are expected to support stable and accountable decision-making over extended time horizons. Optimization strategies that yield unstable feature subsets or highly variable model behavior may undermine institutional trust and reduce operational usability. Consequently, metaheuristic optimization should be employed as a means to enhance feature relevance, robustness, and interpretability rather than as an end in itself [20, 24].

Within this thesis, metaheuristic optimization is positioned as a supporting search paradigm integrated into the predictive modeling framework. Its role is to guide feature selection in high-dimensional institutional data under explicit constraints related to stability and interpretability. This positioning aligns optimization objectives with the requirements of income-contingent loan systems and provides the conceptual foundation for the methodological choices and experimental designs presented in subsequent chapters.

2.2.8 Interpretability and Stability as Design Requirements

Interpretability and stability have emerged as fundamental design requirements in the application of predictive models to institutional and policy-oriented contexts. As machine learning systems increasingly inform administrative actions, regulatory decisions, and resource allocation, the capacity to understand, justify, and trust model outputs becomes as important as predictive accuracy. In such settings, predictive models function as decision-support artifacts embedded within governance structures, rather than as autonomous optimization tools [19, 26].

Interpretability refers to the extent to which the relationship between input variables and model outputs can be understood and communicated to human stakeholders. In educational and financial institutions, interpretability is closely linked to accountability, auditability, and the ability to translate predictive insights into actionable policies. Models that rely on opaque decision mechanisms may achieve high predictive performance, but their limited transparency

constrains their usefulness in environments where explanations are required for institutional legitimacy and compliance [20, 27].

Stability, in turn, captures the consistency of model behavior across different data samples, time periods, and experimental conditions. In predictive modeling, instability may manifest as substantial variation in performance metrics, decision boundaries, or selected feature subsets when models are retrained on slightly different data. In institutional contexts, such variability undermines confidence in model outputs and complicates longitudinal deployment, particularly when predictive systems are used to monitor behavior over extended periods [23, 24].

The integration of optimization techniques, especially metaheuristic-based feature selection, intensifies the relevance of interpretability and stability considerations. While optimization can enhance predictive performance and feature compactness, it may also introduce additional sources of variability due to stochastic search dynamics and sensitivity to algorithmic parameters. The systematic review underlying this thesis indicates that many studies report performance improvements without systematically evaluating the robustness or interpretability of optimized models [28, 29].

In income-contingent student loan systems, these challenges are particularly salient. Predictive models are expected to support early monitoring, targeted intervention, and policy evaluation, often under conditions of legal and social scrutiny. Models that exhibit unstable behavior or that rely on non-transparent decision rules risk generating inconsistent signals and eroding institutional trust, regardless of their predictive accuracy [18, 22].

Treating interpretability and stability as design requirements rather than as post hoc evaluation criteria implies that they must be explicitly considered throughout the modeling process. This includes choices related to algorithm selection, feature selection strategies, optimization objectives, and evaluation protocols. By embedding these requirements into the conceptual framework, predictive modeling can be aligned more closely with the operational and governance needs of income-contingent loan systems [20, 24].

Accordingly, this thesis adopts an approach in which interpretability and stability are integral components of model design. Predictive models are evaluated not only in terms of their capacity to anticipate declarative compliance behavior, but also with respect to their consistency, transparency, and suitability for institutional decision support. This conceptual commitment provides the foundation for the methodological strategies and empirical analyses developed in the subsequent chapters.

2.3 Literature Review

This section reviews the body of academic literature related to predictive modeling, feature selection, and optimization techniques applied to individual credit risk and compliance-related problems. The review is grounded in a systematic literature review conducted as part of this doctoral research and previously reported in [15], which identified, filtered, and analyzed peer-reviewed studies addressing machine learning-based risk prediction in financial and institutional contexts.

Rather than presenting an unsystematic narrative of prior work, this section synthesizes

the literature according to thematic dimensions that are directly aligned with the conceptual framework established in the previous section. The reviewed studies are organized around four main axes: predictive modeling approaches for individual credit risk, feature selection techniques in predictive modeling, the application of metaheuristic optimization methods, and considerations related to interpretability and robustness in optimized predictive systems.

The systematic review underlying this section enables a structured and comprehensive examination of how predictive models have been designed, optimized, and evaluated in related domains. It also provides a consistent basis for identifying methodological patterns, recurring design choices, and dominant analytical assumptions across studies, without relying on isolated or anecdotal evidence. Detailed methodological aspects of the review process, including data sources, inclusion criteria, and screening procedures, are reported elsewhere in this thesis and are not repeated here.

By grounding the literature review in a formal systematic process, this section ensures traceability between prior research and the methodological decisions adopted in the subsequent chapters. The reviewed literature thus serves not only as contextual background, but as an empirical foundation from which the research design, modeling strategies, and analytical focus of this thesis logically emerge.

2.3.1 Machine Learning and Individual Credit Risk

The methodologies employed in individual credit risk assessment have expanded substantially with the incorporation of machine learning techniques, which enable the modeling of complex and non-linear relationships between borrower attributes and adverse financial outcomes. The systematic literature review conducted as part of this doctoral research evidences a broad and sustained adoption of machine learning approaches across diverse financial and institutional contexts, reflecting their increasing relevance for predictive modeling tasks involving individual-level credit behavior.

Within the reviewed studies, machine learning methods are applied primarily as supervised classification tools aimed at estimating default risk, payment behavior, or related adverse outcomes. Following the taxonomy proposed by [34], the literature can be grouped into supervised learning (SL), unsupervised learning (UL), and reinforcement learning (RL) approaches. However, the reviewed body of work is heavily concentrated on supervised learning, which aligns with the availability of labeled institutional datasets and the predominance of binary decision problems in credit risk assessment.

Supervised learning methods rely on labeled data to train predictive models that estimate the probability of default or related adverse outcomes. The reviewed studies report extensive use of traditional and well-established classifiers, including Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), and Gradient Boosting Machines (GBM). These algorithms are favored due to their compatibility with tabular institutional data and their capacity to support binary classification tasks central to credit risk modeling.

Unsupervised learning techniques, such as clustering and anomaly detection, appear less frequently in the reviewed literature. When employed, methods such as K-means or Principal Component Analysis (PCA) are typically used for exploratory analysis, data preprocessing, or the

identification of atypical borrower profiles rather than as primary predictive tools. Reinforcement learning approaches are only marginally represented and are generally explored in experimental or conceptual contexts rather than in operational credit risk systems.

Machine Learning Algorithms Used in Individual Credit Risk

The systematic review indicates that most studies adopt a limited number of classifiers within their experimental designs. Figure 2.1 illustrates the distribution of the number of machine learning classifiers employed across the reviewed documents. The majority of studies rely on a single classifier [35–43], while a smaller number of studies explore multiple classifiers within the same experimental framework [44–57].

This tendency suggests that many studies prioritize depth of optimization or methodological novelty over broad algorithmic benchmarking. At the same time, the observed variation in the number of classifiers reflects heterogeneous experimental strategies and differing levels of methodological complexity across studies.

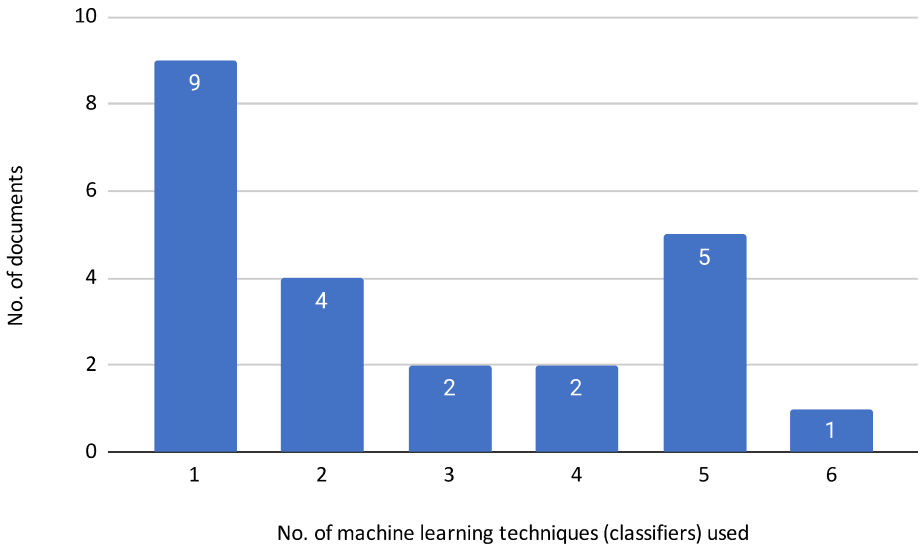


Figure 2.1: Reported Number of Classifier used in a Paper.

Across the reviewed literature, a wide variety of machine learning algorithms have been applied to credit risk assessment, often in combination with feature selection or optimization techniques. These include ensemble-based models [49, 51, 54], rule-based and hybrid approaches [35, 41], Bayesian models [40], and deep learning architectures [36, 42, 55]. Several studies also integrate optimization techniques such as Particle Swarm Optimization, Genetic Algorithms, Bayesian optimization, or multi-objective evolutionary algorithms to enhance predictive performance or interpretability [39, 48, 52, 53].

An exception to the dominant classification-oriented focus is presented by [37], who address the problem of missing data through imputation using Generative Adversarial Networks. This

approach highlights the relevance of data quality and preprocessing as complementary challenges in credit risk modeling.

Figure 2.2 summarizes the frequency of machine learning classifiers reported across the reviewed studies. Random Forest emerges as the most frequently used algorithm, followed by Logistic Regression, Support Vector Machines, and Neural Networks. The prevalence of these models reflects a balance between predictive capacity, robustness, and practical applicability in institutional environments.

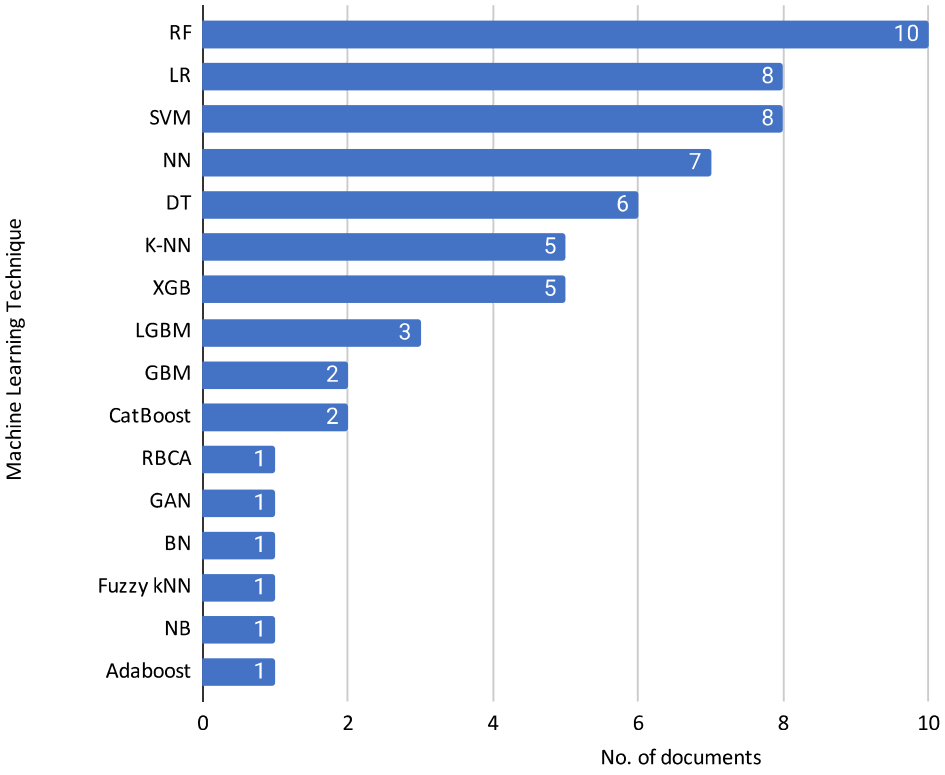


Figure 2.2: Frequency of Machine Learning Classifiers in Reviewed Documents.

Credit Risk Assessment in Practice

Beyond methodological considerations, the reviewed literature applies machine learning-based credit risk assessment across diverse operational contexts. In the banking sector, studies focus on improving predictive accuracy, regulatory compliance, interpretability, and operational efficiency [36, 44, 48, 50, 53–56]. These works frequently emphasize alignment with regulatory frameworks and the integration of predictive models into existing decision-making processes.

In peer-to-peer lending platforms, machine learning techniques are primarily employed to classify borrowers and manage platform-specific risk dynamics, often under conditions of limited historical data and higher uncertainty [45, 47]. These studies highlight the adaptability of machine learning models to alternative lending environments.

Finally, a smaller subset of studies addresses credit risk assessment in educational and student loan contexts. Research in this area focuses on predicting default or payment behavior within student loan systems, reflecting growing interest in applying predictive analytics to education-related financial instruments [43, 51]. These works provide a direct conceptual bridge to the focus of the present thesis.

Taken together, the reviewed studies reveal substantial methodological diversity in the application of machine learning to individual credit risk assessment, both in terms of algorithmic choices and experimental design strategies. Rather than converging on a single modeling approach, the literature reflects heterogeneous practices shaped by data availability, institutional context, and specific research objectives.

These observations provide an empirical basis for identifying dominant modeling paradigms, recurrent experimental practices, and commonly accepted design choices within the field. A structured synthesis of these patterns, together with their implications for robustness, interpretability, and operational deployment, is developed in the subsequent State of the Art section.

2.3.2 Deep Learning Models for Credit Risk and Compliance Prediction

Recent advances in credit risk prediction have increasingly incorporated deep learning models as an extension of traditional machine learning approaches, motivated by their ability to capture complex nonlinear relationships and hierarchical interactions among heterogeneous variables. In student credit and educational finance contexts, deep neural networks have been explored alongside tree-based models and penalized regressions, showing potential improvements over classical statistical techniques, particularly in scenarios characterized by high dimensionality and complex variable interactions [15, 58].

The literature highlights that biologically inspired deep learning architectures, such as deep neural networks, are especially suited to modeling nonlinear dependencies among demographic, academic, and economic variables, which are common in institutional student loan datasets [15]. These models have demonstrated competitive predictive performance in comparison to conventional machine learning methods, although their effectiveness remains strongly dependent on data preprocessing quality, class imbalance handling, and experimental design [59–63].

Several studies have addressed challenges related to missing values and data heterogeneity when applying deep learning models, emphasizing the relevance of advanced imputation strategies such as model-based methods and ensemble-driven approaches, which outperform simplistic imputation schemes in nonlinear settings [64–67]. However, many empirical evaluations rely on limited experimental protocols, often based on single data partitions, which constrains the generalizability and robustness of reported results [15].

Class imbalance constitutes an additional challenge in student credit prediction, particularly when rare but operationally critical events are modeled. Although oversampling techniques such as SMOTE have been widely adopted in conjunction with deep learning models, prior work has noted potential risks of overfitting and instability, especially when highly expressive architectures are combined with synthetic data generation [68–70]. These concerns are accentuated in institutional contexts where robustness and reproducibility are essential for decision support.

More recently, deep learning models have been combined with hyperparameter optimization

strategies to enhance predictive performance [71–73]. Nevertheless, the literature reveals a lack of controlled comparative studies that evaluate deep neural networks against optimized classical models under identical data representations, multiple random partitions, and reproducible experimental protocols [15, 70]. Moreover, interpretability remains a persistent limitation, particularly in sensitive financial applications such as publicly administered student loan systems.

Within this context, deep learning models are best understood as highly expressive predictive tools whose contribution must be assessed through rigorous experimental design and systematic comparison with interpretable and optimized classical approaches. This theoretical positioning directly motivates the inclusion of bioinspired deep learning models as a comparative and upper-bound modeling paradigm in the later stages of this research.

2.3.3 Feature Selection Techniques in Predictive Modeling

Feature selection constitutes a central step in the development of effective machine learning models, particularly in high-dimensional settings where redundant, irrelevant, or noisy variables may degrade predictive performance and hinder interpretability. In the context of credit risk assessment, feature selection is commonly employed to enhance model generalization, reduce computational complexity, and improve transparency by identifying the most informative borrower attributes. Following the extensive literature reviewed in this thesis [74–85], feature selection methods are typically classified into three broad categories: filter methods, wrapper methods, and embedded methods.

Filter methods assess feature relevance based on statistical relationships between individual variables and the target outcome, independently of any specific predictive model. These approaches are computationally efficient and are frequently used as preliminary screening mechanisms. Common techniques reported in the reviewed studies include mutual information, which measures statistical dependence between features and the target variable; chi-squared tests, which evaluate associations between categorical predictors and class labels; and correlation-based measures, which examine linear relationships and are often used to detect multicollinearity among numerical variables.

Several studies included in the systematic review employ filter-based strategies as part of their modeling pipelines. For example, [45] apply the CfsSubsetEval evaluation method combined with a BestFirst search strategy, while [57] rely on Pearson correlation analysis to eliminate redundant predictors. Similarly, [38] conduct univariate analysis using KS and AUC metrics, followed by correlation-based filtering to remove highly correlated features. Additional applications include the removal of collinear variables [54] and the use of information value and correlation analysis [56].

Wrapper methods evaluate subsets of features by directly assessing their impact on the performance of a specific machine learning algorithm. These approaches are capable of capturing feature interactions but typically incur higher computational costs due to repeated model training. Within the reviewed literature, wrapper-based feature selection is frequently combined with metaheuristic optimization techniques. Notable examples include the use of the Firefly Algorithm for feature selection in conjunction with multiple classifiers [53], the integration of Particle Swarm Optimization (PSO) with decision tree learning [39], and the application of the Boruta

algorithm for identifying relevant predictors [43].

Embedded methods perform feature selection as part of the model training process itself, leveraging algorithm-specific mechanisms to identify relevant variables. These methods are commonly associated with regularization techniques or tree-based models. In the reviewed studies, embedded feature selection is implemented through logistic regression and neural networks [44], Recursive Feature Elimination (RFE) combined with Random Forests [48], and the Bolasso approach, which integrates bootstrap-based LASSO for stable feature selection [50]. Additional embedded strategies include the use of LASSO, Ridge, Elastic Net, feature importance measures, and chi-square statistics [41].

Hybrid feature selection approaches, which combine elements of multiple categories, are less common but illustrate the methodological diversity present in the literature. An example is provided by [52], who integrate a Kernel Partial Least Squares (KPLS)-based filter method with a Quantum Particle Swarm Optimization (QPSO)-based wrapper strategy to address challenges associated with high-dimensional and limited-sample credit risk datasets.

Not all reviewed studies explicitly incorporate feature selection mechanisms. For instance, [37] focus primarily on data imputation using generative adversarial networks rather than on feature relevance assessment. This variation reflects differing modeling priorities and highlights the heterogeneous role that feature selection plays across credit risk prediction studies.

Overall, the reviewed literature demonstrates that feature selection techniques are widely employed to support predictive performance and interpretability in credit risk assessment. The diversity of methods reported underscores the absence of a single dominant approach and illustrates how feature selection strategies are adapted to specific data characteristics, modeling objectives, and institutional constraints.

2.3.4 Metaheuristics Applied to Feature Selection and Optimization

Optimization techniques, and metaheuristics in particular, are general-purpose search algorithms that can be adapted to a wide range of optimization problems. They are commonly characterized by their capacity to obtain high-quality solutions within practical computational times by balancing exploration and exploitation of the search space [86]. In the context of machine learning, metaheuristics are frequently employed to support tasks such as feature selection, hyperparameter configuration, and hybrid model design, where the underlying search space is typically large, discrete, and not amenable to exhaustive exploration.

Across the studies included in the systematic literature review, metaheuristics are primarily used as auxiliary mechanisms to improve model performance, enhance interpretability, or address challenges related to high dimensionality and limited sample sizes. Commonly reported approaches include Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Simulated Annealing (SA), although the reviewed literature also includes a variety of more recent or specialized metaheuristics. The following methods were observed in the reviewed articles:

- **Biogeography Based Optimization (BBO):** Biogeography Based Optimization (BBO) is inspired by biogeography, which studies the distribution of species across habitats over time [87]. In BBO, each candidate solution is modeled as a habitat with a habitat suitability index

(HSI) representing its fitness. Information exchange between high- and low-HSI habitats is interpreted as migration, enabling exploration of the search space through iterative feature sharing and adaptation.

- **Adaptive Marine Predators Algorithm (AMPA):** The Adaptive Marine Predators Algorithm (AMPA) mimics adaptive foraging behavior in marine predator-prey interactions [88]. It adjusts pursuit and search strategies to balance exploration and exploitation across dynamic landscapes, supporting optimization in complex problem spaces.
- **Variable Color Harmony Algorithm (VCHA):** The Variable Color Harmony Algorithm (VCHA) is inspired by improvisation principles associated with harmony processes [89]. It employs dynamic parameter adjustment to regulate exploration and exploitation, enabling effective optimization of complex functions.
- **Quantum Particle Swarm Optimization (QPSO):** Quantum Particle Swarm Optimization (QPSO) extends Particle Swarm Optimization by incorporating principles associated with quantum behavior [90]. Particle positions are updated based on probabilistic rules rather than deterministic dynamics, aiming to enhance global search ability and reduce premature convergence.
- **Firefly Algorithm (FFA):** The Firefly Algorithm (FFA) is inspired by the flashing behavior of fireflies [91]. Candidate solutions are modeled as fireflies whose attractiveness is proportional to brightness (fitness), with attractiveness decreasing with distance. This mechanism induces movement toward promising regions of the search space while preserving exploration.
- **Particle Swarm Optimization (PSO):** Particle Swarm Optimization (PSO) is a population-based method inspired by social behavior observed in flocking and schooling [92]. Each particle updates its position by combining its own experience with information shared by neighbors, enabling iterative refinement of candidate solutions through a balance of exploration and exploitation.
- **Non-dominated Sorting Genetic Algorithm II (NSGA-II):** NSGA-II is an evolutionary algorithm designed for multi-objective optimization [93]. It employs fast non-dominated sorting to identify Pareto fronts and uses a crowding-distance mechanism to maintain diversity, supporting the joint optimization of conflicting objectives such as predictive performance and interpretability.
- **Genetic Algorithm (GA):** Genetic Algorithms (GA) are inspired by natural selection and genetics [94]. Candidate solutions evolve across generations through selection, crossover, and mutation operators, enabling broad search and reducing susceptibility to local optima in complex optimization spaces.

Several reviewed studies integrate these metaheuristics into credit risk assessment pipelines as mechanisms for rule discovery, feature selection, or parameter optimization. The LGBBO-RuleMiner proposed by [35] is a rule-based classification technique designed to predict credit

risk using a Biogeography Based Optimization (BBO) strategy. In this approach, BBO guides the discovery of an optimized rule set from datasets containing both categorical and continuous attributes through iterative migration and mutation operators.

[36] introduced the ABSMPNN model to identify customer credit quality in the banking sector by integrating Binarized Spiking Neural Networks (BSNN) with the Adaptive Marine Predators Algorithm (AMPA). In addition, the Variable Color Harmony Algorithm (VCHA) is incorporated to support feature selection within the overall modeling framework.

In [52], a hybrid approach integrates Wasserstein Generative Adversarial Networks (WGAN) for data augmentation with a Kernel Partial Least Squares (KPLS)-based filter stage and a Quantum Particle Swarm Optimization (QPSO)-based wrapper stage for feature selection. This pipeline addresses data scarcity through synthetic sample generation, ranks feature relevance, and optimizes the feature subset to improve predictive model performance.

[53] employed the Firefly Algorithm (FFA) for feature selection in conjunction with multiple classifiers, including KNN, Fuzzy KNN, Random Forest, Decision Tree, and SVM. The study frames FFA as the search mechanism for identifying informative feature subsets, and it also reports the use of SMOTE to address class imbalance.

The model proposed by [39] combines Particle Swarm Optimization (PSO) with Structure Decision Tree Learning (SDTL) for credit risk assessment. In this framework, PSO is used to support feature selection, while SDTL provides the classification component intended to enhance predictive reliability.

In [41], interpretability is addressed through a two-stage rule extraction approach from tree ensemble models. Local rule extraction is performed first, and the full rule set is subsequently optimized using NSGA-II, enabling a multi-objective treatment that balances predictive performance with interpretability.

Finally, [42] combined a Backpropagation (BP) neural network with a mutation genetic algorithm to improve credit risk assessment in commercial banking. In this approach, the mutation genetic algorithm is used as an optimization mechanism for network parameters, with the stated aim of improving convergence behavior and mitigating local optimum issues.

Beyond methodological diversity, the reviewed literature occasionally reports computational aspects of metaheuristic-based pipelines. In particular, [36, 52] explicitly report computational time as part of their empirical evaluation, reflecting an emphasis on efficiency considerations in addition to predictive outcomes. This aspect is relevant in institutional contexts where computational cost and scalability may influence the feasibility of adopting optimized predictive models.

2.3.5 Interpretability and Robustness in Optimized Predictive Models

As predictive models for credit risk assessment increasingly incorporate feature selection and metaheuristic optimization techniques, concerns related to interpretability and robustness emerge as recurrent, though unevenly addressed, themes in the literature. This is particularly relevant in institutional and financial contexts, where predictive outputs are often expected to support decision-making processes subject to regulatory, ethical, and operational constraints.

Within the reviewed studies, interpretability is approached through a variety of strategies,

ranging from explicit design choices to indirect methodological effects. Some works explicitly incorporate interpretability as a modeling objective. For instance, [41] propose a two-stage framework in which rules are extracted from tree ensemble models and subsequently optimized using NSGA-II, with interpretability treated as an explicit optimization criterion alongside predictive performance. Similarly, [46] introduce an interpretable selective learning framework that dynamically balances the use of logistic regression and neural networks, prioritizing simpler models when adequate predictive performance can be achieved.

In other studies, interpretability is addressed implicitly through model or feature selection choices. The use of linear models, decision trees, rule-based classifiers, and sparsity-inducing feature selection techniques contributes to improved transparency without being framed as a primary objective. Approaches based on LASSO, recursive feature elimination, tree-based feature importance, or rule discovery mechanisms [35, 48, 50] illustrate how interpretability can emerge as a by-product of methodological decisions aimed at dimensionality reduction or model simplification.

Robustness and stability are less consistently defined across the reviewed literature. In most cases, robustness is implicitly associated with standard validation practices, such as cross-validation, train-test splits, or performance consistency across datasets. Several studies report average performance metrics over multiple experimental runs or folds, thereby providing an indirect indication of model stability [42, 53]. However, explicit analyses of stability with respect to feature selection variability, optimization stochasticity, or temporal data shifts are comparatively rare.

Only a limited subset of the reviewed works explicitly report computational aspects or performance consistency as part of their evaluation. For example, [36] and [52] include computational time as a reported outcome, reflecting an awareness of efficiency and scalability considerations in optimized modeling pipelines. Nevertheless, systematic treatments of robustness that distinguish between predictive accuracy, stability across runs, and resilience to data perturbations are not uniformly present across studies.

Overall, the literature reflects a heterogeneous treatment of interpretability and robustness in optimized predictive models for credit risk assessment. While several studies incorporate mechanisms that enhance transparency or report validation-based indicators of stability, these aspects are addressed through diverse and often implicit strategies, rather than through standardized or unified evaluation frameworks. This diversity highlights the multifaceted role that interpretability and robustness play within optimized predictive modeling approaches in financial and institutional contexts.

2.4 Existing Methods, Tools, and Technologies

The current state of the art in predictive modeling for individual credit risk assessment reflects a growing convergence between machine learning techniques, feature selection strategies, and metaheuristic optimization methods. This convergence has been driven by the increasing availability of high-dimensional institutional data, the need for improved predictive accuracy, and the demand for models that can operate under practical constraints related to interpretability,

scalability, and regulatory compliance.

Recent studies predominantly frame credit risk assessment as a supervised classification problem, relying on labeled historical data to predict default or non-compliance events. Across the reviewed literature, ensemble-based machine learning models, particularly tree-based approaches such as Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost, are among the most frequently adopted techniques [41, 48, 50–57]. These models are often favored due to their strong predictive performance and their ability to handle heterogeneous and nonlinear feature spaces.

Alongside ensemble methods, classical algorithms such as logistic regression, support vector machines, decision trees, and k-nearest neighbors continue to be widely used, either as baseline models or as components within hybrid frameworks [44–47]. Neural network-based approaches, including deep learning architectures and spiking neural networks, have also gained visibility in recent years, particularly in studies aiming to capture complex behavioral patterns or temporal dependencies in credit-related data [36, 42, 55].

Feature selection has emerged as a central component in many state-of-the-art pipelines, motivated by the high dimensionality and multicollinearity commonly present in credit datasets. Filter-based methods remain prevalent as preliminary steps, while wrapper and embedded approaches are increasingly integrated into model construction [39, 45, 48, 53]. In several studies, feature selection is tightly coupled with optimization procedures, enabling joint exploration of feature subsets and model performance criteria [35, 41, 52].

Metaheuristic optimization techniques play a prominent role in the current state of the art, particularly in wrapper-based feature selection and hybrid model design. Algorithms such as Genetic Algorithms, Particle Swarm Optimization, Firefly Algorithm, Biogeography Based Optimization, Quantum Particle Swarm Optimization, and NSGA-II are employed to navigate complex search spaces that arise from combinatorial feature selection and multi-objective optimization problems [35, 39, 41, 52, 53]. These methods are commonly used to balance predictive accuracy against secondary objectives such as model simplicity or interpretability.

Hybrid modeling frameworks represent an important trend within the state of the art. Several studies combine machine learning models with metaheuristic optimization and auxiliary techniques such as data augmentation, rule extraction, or ensemble integration [36, 52, 57]. These frameworks reflect an emphasis on adaptability and performance enhancement in challenging data environments, including imbalanced datasets and limited sample sizes.

Despite methodological diversity, evaluation practices across the literature show a degree of convergence. Most studies rely on standard performance metrics such as accuracy, AUC, precision, recall, and F1-score, typically reported through cross-validation or hold-out validation schemes. A smaller subset of works additionally reports computational time or efficiency-related indicators, particularly in studies involving metaheuristic optimization [36, 52]. However, comprehensive assessments that jointly consider predictive performance, stability, interpretability, and operational feasibility remain relatively uncommon.

Overall, the state of the art in credit risk predictive modeling is characterized by increasingly complex and hybridized approaches that integrate machine learning, feature selection, and optimization techniques. These developments reflect ongoing efforts to improve predictive effectiveness while addressing practical challenges associated with real-world institutional data

and decision-support requirements.

2.4.1 Dominant Modeling Paradigms

The state of the art in individual credit risk assessment is characterized by a clear predominance of supervised machine learning paradigms, where the modeling task is framed as a classification problem using labeled historical data. Across the studies included in the systematic literature review, predictive modeling is primarily oriented toward estimating the probability of default or non-compliance at the individual level, relying on institutional, financial, and behavioral variables.

A defining feature of current modeling practices is the widespread adoption of ensemble-based learning methods. Tree-based ensembles, including Random Forest, Gradient Boosting variants, XGBoost, LightGBM, and CatBoost, emerge as the most frequently employed classifiers in the reviewed literature. This prevalence reflects their capacity to handle heterogeneous feature spaces, nonlinear relationships, and complex interactions commonly present in credit-related datasets [41, 48, 50–57]. The empirical basis for this observation is documented in the literature review (Section 2.3.1).

In parallel with ensemble approaches, classical machine learning models such as logistic regression, support vector machines, decision trees, and k-nearest neighbors continue to play a significant role within dominant modeling paradigms. These models are frequently employed either as baseline classifiers, as components of comparative experimental setups, or as interpretable alternatives within hybrid frameworks [44–47]. Their persistent use reflects both their interpretability advantages and their relevance in regulatory or operational contexts.

Neural network-based approaches represent a complementary paradigm within the state of the art. While less prevalent than ensemble methods, neural networks, including deep learning architectures and spiking neural networks, are increasingly explored to capture complex behavioral patterns and nonlinear dependencies in credit risk data [36, 42, 55]. These approaches are typically combined with feature selection or optimization mechanisms to mitigate overfitting and control model complexity.

Another notable characteristic of dominant modeling paradigms is the limited number of classifiers employed within individual studies. The reviewed literature shows a clear tendency toward focused modeling pipelines, where methodological effort is concentrated on optimizing one or two selected algorithms rather than conducting exhaustive algorithmic comparisons [35–43].

Overall, the dominant modeling paradigms reflected in the literature combine a strong preference for supervised ensemble learning with selective use of classical and neural approaches. This configuration illustrates a state of the art oriented toward predictive performance and operational applicability, and it provides the conceptual foundation for analyzing the role of feature selection, optimization strategies, and robustness considerations in subsequent sections.

2.4.2 Data Characteristics and Experimental Practices

The state of the art in individual credit risk prediction is strongly shaped by the nature of the datasets used and by recurring experimental practices that condition model development, evaluation, and reported performance. Across the studies included in the systematic literature review, datasets are drawn from two broad sources: public benchmark datasets, commonly used to enable methodological replication, and proprietary or institutional datasets, used to reflect operational settings and domain-specific constraints.

Figure 2.3 summarizes the distribution of dataset usage in the reviewed studies, distinguishing benchmark datasets, real-world application datasets, and studies that employ both. This figure provides a compact view of how the literature balances reproducibility-oriented experimentation with institutionally grounded modeling.

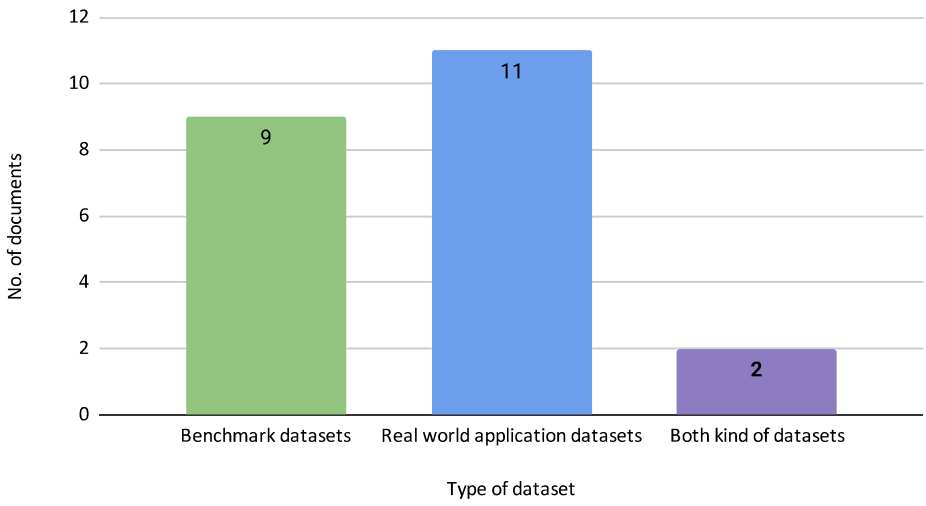


Figure 2.3: Usage of Benchmark vs. Real-World Application Datasets.

Benchmark datasets and their role in methodological consolidation.

Public benchmark datasets remain central to the empirical validation of credit scoring and default prediction methods. The reviewed literature shows frequent reliance on repositories such as UCI and Kaggle, which provide standardized datasets that facilitate controlled experimentation and cross-study comparison. Figure 2.4 summarizes the sources of benchmark datasets reported in the review, and Table 2.1 enumerates representative benchmark datasets, including widely used UCI Statlog variants and large-scale Kaggle datasets used for default prediction and credit scoring [95–105].

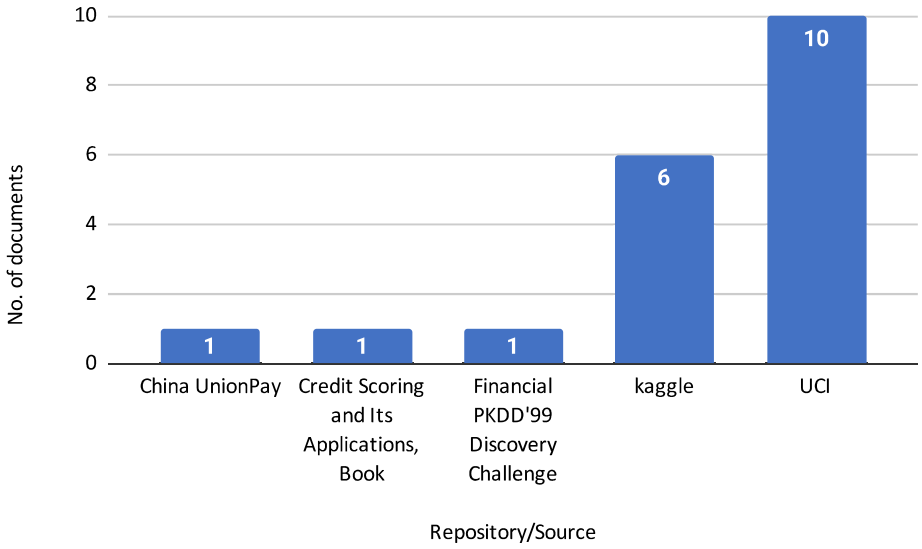


Figure 2.4: Sources of Benchmark Datasets.

Dataset Name	Source	Intances	Features	Labels
Statlog (Australian Credit Approval)	UCI	690	14	2
Statlog (German Credit Data)	UCI	1000	20	2
Default of Credit Card Clients	UCI	30,000	23	2
South German Credit Dataset	UCI	1000	21	2
Credit risk dataset	Kaggle	239	11	2
Kaggle's Bank Loan Status dataset	Kaggle	12,535	19	2
Loan default prediction dataset	Kaggle	105,471	769	2
Give me some credit dataset	Kaggle	120,969	10	2
Credit card econometrics	Kaggle	1320	12	2
Kaggle Home Credit Default Risk	Kaggle	307,511	122	2
Czech Financial Dataset	Financial PKDD'99 Discovery Challenge	682	55	2
Thomas dataset	Credit Scoring and Its Applications, Book	1225	14	2
China UnionPay credit dataset	China UnionPay	11,017	199	2

Table 2.1: Public datasets (benchmarks).

Real-world and proprietary datasets as drivers of practical constraints.

A substantial portion of the reviewed literature relies on real-world application datasets sourced from financial institutions, credit registries, and operational lending platforms. These datasets exhibit broader heterogeneity in size, feature dimensionality, and label structure, and

they are often aligned with specific institutional objectives. Table 2.2 summarizes representative real-world datasets, including data from Lending Club platforms, commercial banks, credit registries, and student loan contexts [37–45, 50, 51, 55–57].

Dataset Name	Source	Intances	Features	Labels
Lending Club loan data	Lending Club	n.a.	n.a.	n.a.
Lending club dataset	Lending Club	42,538	143	2
Credit risk assessment data	Anonymous local bank in China	10,744	10	n.a.
LendingClub (LC)	Lending Club	500,000	120	n.a.
Commercial Bank Credit Records	Anonymous commercial bank in China	27,520	27	5
Business credit score dataset	Data from a Latin American bank	20,835	585	2
Personal credit score	Data from a Latin American bank	76,209	936	2
General Data Protection Regulation (GDPR)	Commercial banks and savings institutions	1,000,000	n.a.	n.a.
Advanced Analytics of Credit Registry Dataset	undetermined	n.a.	n.a.	n.a.
WIND Dataset	Commercial bank's personal credit database	n.a.	n.a.	n.a.
Croatian Bank Credit Risk Dataset 2009–2013	Large Croatian Bank	870,710	109	2
Croatian Bank Credit Risk Dataset 2004–2018	Large Croatian Bank	782,875	108	2
North Macedonia Credit Registry Data	Central Bank of the Republic of North Macedonia	1,000,000,000	52	5
Tunisian Bank Loan Data	Data from several Tunisian banks	n.a.	9	2
Lending Club (LC) dataset 2017–2018	Lending Club	477,131	16	2
Bank credit risk data	Commercial banks	360	n.a.	n.a.
National Student Loans Dataset	University in Beijing, China	18,000	20	2
Student Payment Behavior Dataset	Private university in Peru	8,495	13	2

Table 2.2: Real-world applications datasets.

Variable families commonly represented in credit risk datasets.

Despite dataset heterogeneity, the reviewed studies show recurring families of predictors used in individual credit risk modeling. These typically include demographic attributes, financial capacity indicators, loan contract characteristics, credit history and payment behavior variables, and employment-related features. A smaller subset of studies includes education-related variables, particularly in student loan contexts [43, 51]. This pattern suggests that, across institutional settings, credit risk modeling is commonly grounded in variables that capture ability-to-pay proxies (income, debt burden), behavioral history (past due events, repayment patterns), and contractual exposure (loan amount, term).

Experimental practices and data challenges.

The reviewed literature reports recurring data challenges that influence experimental practice, including missing values and class imbalance. Missing data is explicitly addressed in [37], where Multiple Generative Adversarial Imputation Networks (MGAIN) are proposed as an imputation strategy building on Generative Adversarial Imputation Networks [106]. In addition, class imbalance is frequently mitigated through resampling techniques such as SMOTE, as reported in studies combining feature selection and optimized classifiers [53, 57]. These practices illustrate how data characteristics often motivate the inclusion of preprocessing and data engineering steps as integral components of credit risk predictive pipelines.

Taken together, the dataset landscape and experimental practices described in the reviewed literature indicate that state-of-the-art credit risk modeling is shaped not only by algorithm choice, but also by data availability, feature dimensionality, label structure, and preprocessing requirements. These aspects provide the empirical context for understanding why contemporary approaches increasingly integrate feature selection and optimization mechanisms within end-to-end predictive modeling workflows.

2.4.3 Role of Feature Selection and Optimization

In the current state of the art, feature selection and optimization are not treated as auxiliary steps, but rather as structural components of end-to-end predictive modeling pipelines for individual credit risk assessment. This prominence is largely explained by the dimensionality and heterogeneity of credit-related datasets, where redundant variables, collinearity, and noisy proxies can reduce generalization and complicate the operational use of predictive systems. Accordingly, many reviewed studies incorporate feature selection as a deliberate design choice to support predictive effectiveness and to manage the complexity of institutional data environments [39, 41, 45, 48, 52, 53].

A dominant pattern in the reviewed literature is the use of feature selection as a mechanism to improve model performance while implicitly constraining model complexity. Filter-based methods are frequently employed as initial screening steps, particularly through correlation analysis, univariate testing, and subset evaluation strategies [38, 45, 54, 56, 57]. These approaches are attractive for their computational efficiency and their suitability for high-dimensional settings.

However, the state of the art increasingly reflects a shift toward wrapper and embedded strategies, which are more tightly coupled to predictive modeling and can capture feature interactions and model-specific relevance patterns [41, 48, 50].

Wrapper-based feature selection is strongly associated with metaheuristic optimization in the reviewed studies, reflecting the combinatorial structure of subset selection and the practical infeasibility of exhaustive search in large feature spaces. Metaheuristics such as Firefly Algorithm and Particle Swarm Optimization appear as search mechanisms for identifying informative feature subsets in pipelines that combine feature selection with supervised classifiers [39, 53]. Hybrid strategies also emerge, where filter stages reduce dimensionality prior to metaheuristic-driven wrapper optimization, as illustrated by the combination of KPLS-based filtering with QPSO-based wrapper selection [52]. This pattern indicates that state-of-the-art pipelines often operationalize feature selection as a multi-stage process, balancing computational tractability with modeling precision.

Embedded feature selection techniques constitute another prominent trend, particularly through regularization-based approaches and model-intrinsic relevance mechanisms. Studies employing LASSO-type methods, stability-enhancing variants such as Bolasso, or tree-based importance measures illustrate how feature relevance is sometimes integrated directly into model training [41, 44, 50]. In these cases, feature selection is not an external optimization problem but a property of the modeling framework itself. This trend aligns with the broader dominance of ensemble learning in credit risk assessment, where tree-based models naturally induce structured variable selection through split-based criteria.

Beyond feature selection, optimization is increasingly used to refine predictive pipelines through parameter configuration and multi-objective model design. Several reviewed works combine machine learning models with metaheuristic optimization not only to select features, but also to optimize model parameters, decision rules, or hybrid design elements [35, 36, 41, 42]. Multi-objective optimization becomes particularly salient when interpretability is considered alongside predictive performance, as in approaches that optimize rule sets extracted from ensemble models using NSGA-II [41]. Similarly, hybrid models integrating neural architectures with metaheuristic optimizers illustrate how optimization can be embedded into model training or configuration to improve empirical outcomes [36].

Taken together, the reviewed literature indicates that feature selection and optimization play a central operational role within contemporary credit risk modeling, supporting the construction of predictive pipelines that manage high dimensionality and data complexity. Current state-of-the-art approaches frequently implement feature selection as either a staged filtering and optimization process or as an embedded model property, while metaheuristic optimization is leveraged to navigate combinatorial search spaces and to refine predictive configurations under practical constraints [39, 41, 52, 53].

2.4.4 Interpretability and Robustness Considerations

Within the current state of the art, interpretability and robustness are increasingly recognized as relevant considerations in the design of optimized predictive models for individual credit risk assessment, although they are addressed in heterogeneous and often non-uniform ways. This

variability reflects differences in application contexts, regulatory expectations, and modeling priorities across studies, as well as the tension between predictive performance and operational transparency.

Interpretability is most explicitly incorporated in studies that employ rule-based models, linear classifiers, or hybrid frameworks designed to balance model complexity with explanatory capacity. Approaches based on rule extraction from ensemble models exemplify this trend, where interpretability is operationalized through human-readable decision rules optimized alongside predictive accuracy [41]. Similarly, selective learning frameworks that alternate between simpler models, such as logistic regression, and more complex architectures, such as neural networks, represent attempts to retain interpretability when high-capacity models are not strictly necessary [46]. These approaches position interpretability as a design objective rather than as a post hoc analysis step.

In many other studies, interpretability is addressed implicitly through methodological choices rather than explicit evaluation criteria. The use of sparse feature selection techniques, regularization-based embedded methods, and tree-based models contributes to reducing model complexity and enhancing transparency without formal interpretability metrics [35, 48, 50]. In these cases, interpretability emerges as a by-product of dimensionality reduction or model structure, rather than as a primary modeling goal.

Robustness considerations in the reviewed literature are most commonly reflected through standard validation practices, including cross-validation, repeated experimental runs, and performance averaging across folds or datasets. These practices provide indirect evidence of model stability with respect to data partitioning and sampling variability [42, 53, 57]. However, robustness is rarely formalized as a distinct evaluation dimension, and explicit analyses of stability with respect to feature selection variability, stochastic optimization behavior, or temporal data shifts are comparatively uncommon.

A limited subset of studies reports additional indicators related to computational efficiency or scalability, particularly in contexts where metaheuristic optimization is employed. For example, some works explicitly report computational time as part of their experimental evaluation, acknowledging the operational relevance of efficiency in optimized predictive pipelines [36, 52]. Nevertheless, such reporting remains inconsistent across the literature and is typically secondary to predictive performance metrics.

Overall, the state of the art reflects a fragmented treatment of interpretability and robustness in optimized credit risk models. While certain studies integrate explicit mechanisms to enhance transparency or implicitly promote robustness through validation practices, there is no unified or standardized approach to evaluating these aspects across predictive pipelines. Instead, interpretability and robustness are addressed through diverse strategies that are closely tied to model choice, feature selection design, and optimization objectives, reinforcing their contextual and application-dependent nature in contemporary credit risk assessment research.

2.4.5 Evaluation Metrics and Validation Practices

A defining characteristic of the current state of the art is that empirical contributions are commonly positioned and justified through the lens of *evaluation*, yet the set of metrics actually reported

across studies remains uneven and strongly concentrated on a narrow subset of classifier-centric indicators. To structure this space with explicit bibliographic support, this thesis adopted the taxonomy proposed by [107], which organizes evaluation metrics into four groups: *performance metrics*, *metaheuristics metrics*, *feature metrics*, and *statistical test metrics*. This taxonomy was used as an analytic scaffold to map the metric choices observed in the systematic literature review and, crucially, to identify additional metrics that expanded the original classification.

Figure 2.5 summarizes the distribution of reported metrics by category across the reviewed documents. Performance metrics dominate the evaluative landscape, whereas metaheuristics metrics are reported in a smaller subset of studies, and feature metrics and statistical test metrics are comparatively infrequent. This pattern indicates that, even in studies that incorporate optimization or feature selection components, evaluation practices remain largely anchored in predictive performance, with limited emphasis on documenting optimization behavior, feature set properties, or the statistical reliability of observed differences.

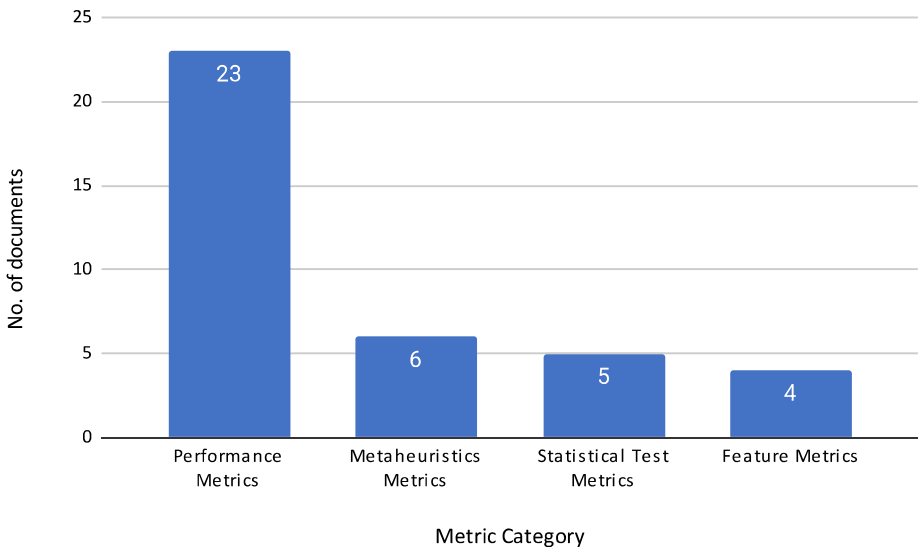


Figure 2.5: Distribution of Reported Metrics by Category.

Beyond the relative predominance of categories, the review also revealed an expansion in metric usage that is not fully captured by previously established taxonomies. Figure 2.6 distinguishes between metrics previously identified within the taxonomy of [107] and those newly observed in the reviewed literature. The expansion is especially visible in the performance metrics category, but it also appears in the remaining categories, suggesting that evaluative practice is diversifying without converging toward a stable and shared reporting standard.

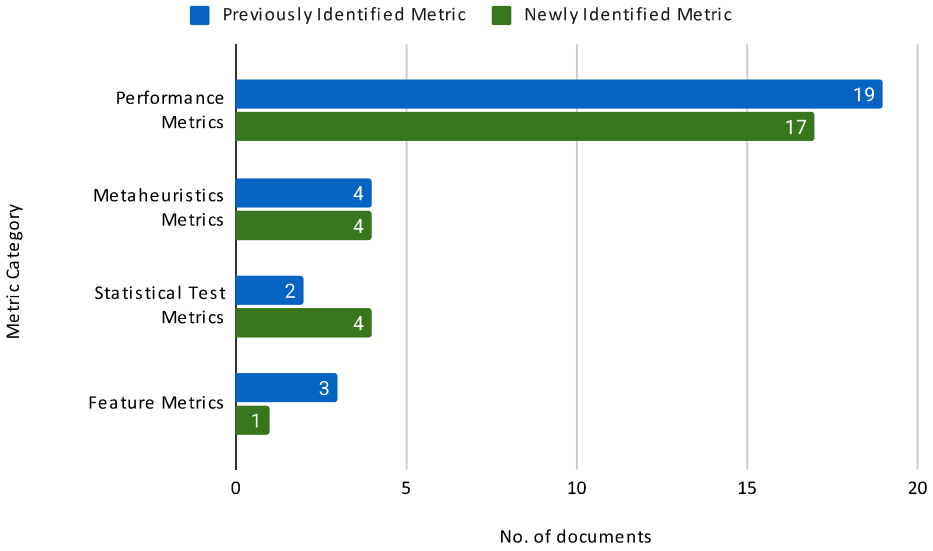


Figure 2.6: Comparison of previously identified metrics and newly identified metrics by Category.

Dominant performance metrics and their implications

The previously identified metrics exhibit a strongly skewed frequency distribution. Figure 2.7 shows that Accuracy is the most frequently reported metric, followed by Recall (TPR), Precision, and F1-score, with substantially lower frequency for other indicators such as computational time, specificity, and feature-related measures. This pattern reflects an evaluation culture primarily oriented toward aggregate predictive correctness, despite the well-known sensitivity of accuracy to class imbalance and asymmetric costs in risk-related decision tasks.

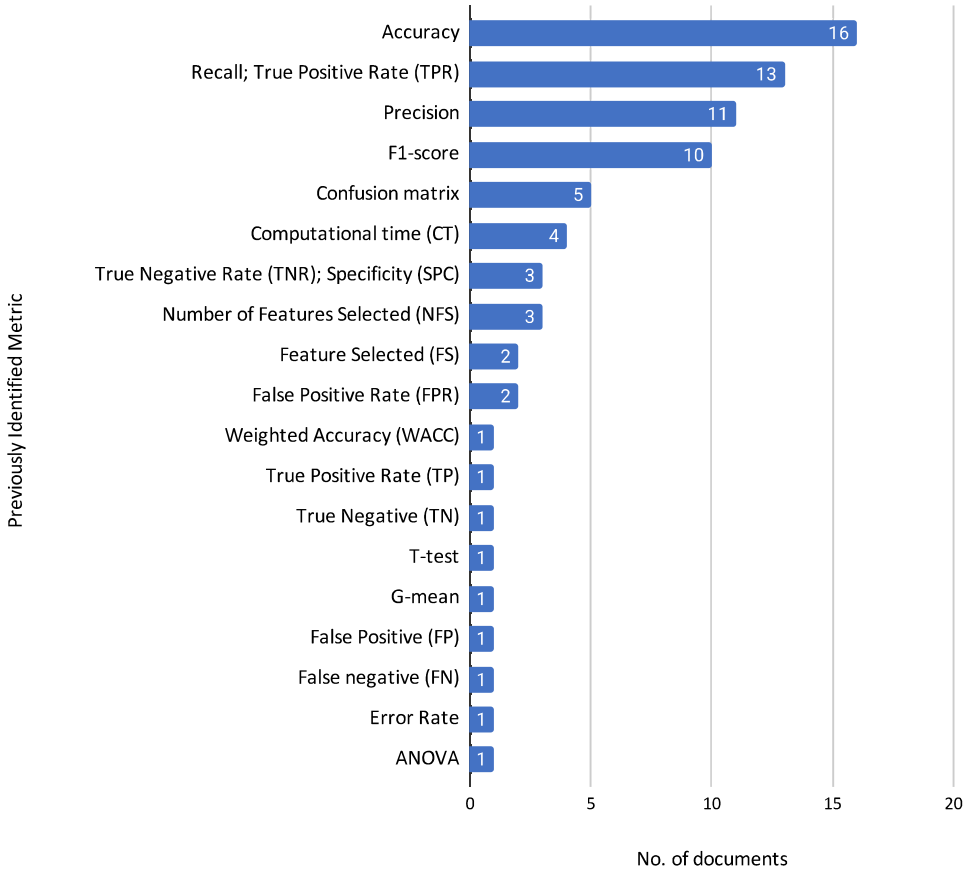


Figure 2.7: Trend of Previously Defined Metrics Distribution in Documents.

Accordingly, a large portion of the reviewed studies rely on standard confusion-matrix-derived measures, including Accuracy [35, 36, 39, 40, 42, 44, 45, 47–51, 53, 54, 56, 57], Precision [35, 36, 40, 41, 47–49, 51, 53, 56, 57], Recall/TPR [35, 36, 40, 41, 46–49, 51–53, 56, 57], and F1-score [35, 36, 40, 47–49, 51, 53, 56, 57]. The confusion matrix itself is explicitly reported in multiple works [36, 44, 46, 47, 57]. In some cases, variants adapted to imbalance are also employed, such as Weighted Accuracy (WACC) in [52].

The newly identified performance metrics further indicate an increased interest in ranking-oriented and probability-sensitive evaluation. Figure 2.8 shows that AUC-ROC is the most frequently reported among newly identified metrics, appearing across a substantial fraction of studies [35, 36, 38, 41, 43, 44, 46, 47, 49–52, 54, 55]. ROC curves are also reported in several studies [36, 44, 46, 51, 56], and complementary metrics appear in specific cases, including Balanced Accuracy [41], AUC-PRC [35, 43], and the KS statistic [38]. Error-oriented metrics such as MAE and RMSE are reported in [35, 37], while Log Loss is used in [43] and the Brier score is reported in [55]. These results suggest that, although confusion-matrix metrics remain dominant, the literature is gradually incorporating more discriminative and probabilistic

evaluation tools.

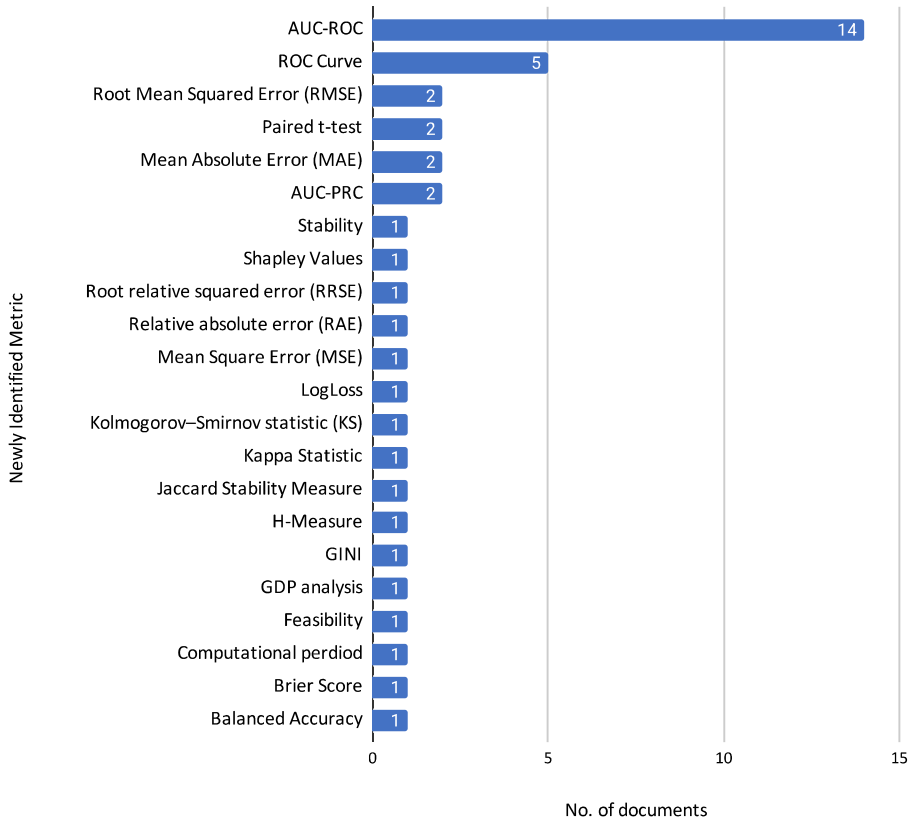


Figure 2.8: Trend of Newly Identified Metrics Distribution in Documents.

Metaheuristics, feature, and statistical-testing metrics: limited but revealing adoption

While many studies incorporate feature selection or optimization procedures, explicit reporting of *metaheuristics metrics* is comparatively rare. The most frequently reported indicator is computational time, used to describe efficiency or operational viability [36, 42, 48, 50, 52]. Additional metaheuristic-related indicators include feasibility [39] and computational period [36]. Notably, stability-related measures are reported in only a small number of works, including a stability metric based on pairwise similarity in [48] and the Jaccard stability measure in [50]. This limited adoption indicates that optimization behavior is often treated as an implementation detail rather than as a first-class object of evaluation, even when metaheuristics materially shape the final predictive pipeline.

Feature metrics are also sparsely reported, most commonly as the number of selected features (NFS) and the explicit identity of selected features (FS) [50, 52, 53]. In contrast, model-agnostic feature attribution measures such as Shapley values are observed only in isolated

cases [38]. This observation is consistent with the broader pattern that evaluation concentrates on predictive performance, with limited systematic reporting on the properties of the selected feature subsets.

Finally, *statistical test metrics* are among the least frequently reported. The literature includes examples of t-tests [37], ANOVA [36], paired t-tests [36, 52], Welch's t-test [48], and the H-measure [55]. The relative rarity of these practices suggests that, in much of the reviewed work, claims of improvement are primarily communicated through point estimates of predictive metrics rather than through systematic significance testing and robustness-oriented inference. In the context of institutional decision-support settings, where reproducibility and reliability are central requirements, this gap in reporting practices becomes theoretically and practically consequential.

Synthesis for compliance-oriented, institutional decision support

Overall, the state of the art reflects a metric culture dominated by performance indicators, increasingly complemented by discriminative measures such as AUC-ROC, but with comparatively limited emphasis on optimization-specific evaluation, feature-set properties, and inferential statistical testing. Within the scope of this thesis, this synthesis directly motivates the explicit treatment of interpretability and stability as design requirements (Section 2.4.4) and establishes the need to operationalize evaluation beyond predictive accuracy by incorporating feature-oriented and robustness-oriented metrics grounded in [107]. This positioning is essential to ensure that subsequent empirical chapters can reference a clear, theory-backed evaluative framework when justifying model selection, feature selection behavior, and the practical readiness of predictive decision-support artifacts.

2.5 Research Gaps and Open Challenges

The systematic literature review indicated that contemporary research on individual credit risk assessment has largely converged toward supervised learning pipelines applied to tabular institutional data and evaluated primarily through predictive performance metrics [34]. Within this dominant paradigm, ensemble models and tree-based learners have been recurrently adopted as the core modeling backbone, frequently combined with feature selection or optimization components to improve discrimination capacity under heterogeneous and partially correlated feature spaces [49, 51, 52, 54]. While this configuration reflects a mature and practically oriented body of work, the evidence synthesized in the review revealed a set of methodological regularities that constrain the transferability of reported results to institutional decision-support contexts.

A first limitation concerns the narrow experimental breadth that characterizes a substantial portion of the reviewed studies. Many contributions relied on a limited number of classifiers per experimental design, prioritizing the refinement of a preferred modeling pipeline over systematic comparisons across families of learners [35, 36, 38–40, 43]. As a result, reported performance improvements were often attributable to intertwined design choices involving preprocessing, sampling, feature selection, and model configuration, which were not consistently disentangled

through controlled ablation or comparative analysis. This practice reduced the interpretive value of reported gains and limited the extent to which conclusions could be generalized beyond specific experimental setups.

A second recurrent issue relates to dataset dependence and the incomplete reporting of data diagnostics. Empirical evidence was divided between benchmark datasets and real-world application datasets, with a non-trivial share of studies continuing to rely on public repositories such as UCI and Kaggle for evaluation [96, 97, 99, 100, 102]. Although benchmark usage supports reproducibility, it also increases the risk of over-fitting methodological claims to a small set of repeatedly reused datasets. More importantly for institutional deployment, many studies described datasets at a high level while omitting systematic analyses of data quality, preprocessing decisions, and diagnostic checks that would clarify the reliability and governance of the learning task [35, 39, 41, 43, 45, 46, 49, 52, 53, 57]. Only a limited subset documented comprehensive end-to-end data preparation and variable analysis pipelines [56], leaving open questions regarding reproducibility, auditability, and institutional trust.

Feature selection emerged as a recurring methodological component across filter, wrapper, embedded, and hybrid strategies [39, 50, 52, 53, 107]. However, the review showed that feature selection outcomes were rarely evaluated under a consistent and stability-oriented framework. Several studies reported improvements in predictive accuracy without explicitly quantifying feature-level consequences such as subset size, robustness across resamples, or temporal consistency of selected variables, even though stability-aware measures were occasionally introduced [48, 50]. This heterogeneity limited the extent to which selected feature subsets could be interpreted as reliable institutional signals rather than as dataset-specific artifacts.

Optimization techniques and metaheuristics were frequently justified as general-purpose search procedures capable of balancing exploration and exploitation [86]. In practice, a diverse set of metaheuristics and evolutionary schemes was applied to feature selection, rule extraction, and parameter optimization tasks [87–94]. Nevertheless, only a limited number of contributions documented computational efficiency in a manner that supported operational interpretation, despite computational time being a recognized evaluation dimension [36, 42, 48, 50, 52]. This lack of efficiency reporting is particularly consequential in institutional contexts where model retraining cycles, decision latency, and resource constraints directly affect deployment feasibility.

Interpretability and robustness were acknowledged across the literature, with some contributions proposing compact rule sets or selective combinations of simpler and more complex models [41, 46]. However, interpretability was not consistently treated as a primary design requirement. Robustness-related practices, including class imbalance handling through resampling techniques such as SMOTE, were commonly implemented but not always validated through class-disaggregated or stability-aware analyses [53, 57]. Consequently, claims regarding improved minority-class behavior or enhanced robustness often remained difficult to generalize without additional supporting evidence.

Finally, evaluation practices in the reviewed literature remained heavily oriented toward conventional classification metrics, including accuracy, precision, recall, F1-score, and AUC-ROC, with comparatively sparse use of metaheuristic-oriented, feature-oriented, or statistical test-based measures [107]. Even when additional metrics such as stability indicators, computational period, or statistical comparisons were reported, their adoption was fragmented and inconsistent

across studies [36, 48, 50]. This pattern revealed an unresolved gap concerning evaluation protocols capable of jointly addressing predictive effectiveness, stability, interpretability, and efficiency under institutional decision-support constraints.

Taken together, the evidence synthesized in the systematic review revealed a misalignment between what is most frequently modeled and evaluated in the credit risk literature and what is required for institutional decision-support in income-contingent student loan systems, where compliance is expressed as a recurring administrative behavior rather than as a single financial default event. Although domain proximity exists, most studies operationalized risk within banking and lending contexts [40, 44, 55, 56], with only a limited subset directly addressing educational lending or student-related repayment behavior [43, 51]. This misalignment delineates a clear research gap concerning the design of predictive modeling frameworks that integrate supervised learning and optimization while explicitly accounting for stability, interpretability, efficiency, and temporal validity under real-world institutional constraints.

2.6 Research Justification and Contribution Opportunity

In response to the gaps evidenced by the systematic review, this thesis was theoretically positioned at the intersection of (i) supervised predictive modeling for institutional records, (ii) metaheuristic optimization as a search paradigm for selecting informative representations and improving model configurations, and (iii) decision-support design requirements centered on interpretability, stability, and operational feasibility.

At the modeling level, the thesis adopted supervised learning as the primary analytical paradigm, consistent with the dominant approach in the reviewed literature on individual credit risk [34]. Rather than framing risk solely as default, the thesis treated compliance as an observable behavioral process expressed through administrative actions and their discontinuities within an income-contingent student loan system. This framing remained compatible with classification-based pipelines commonly used in credit risk assessment while aligning the prediction target with institutional compliance management needs [43, 51].

At the optimization level, the thesis grounded the use of metaheuristics in their role as general-purpose procedures that searched complex spaces through an exploration–exploitation balance [86]. The review provided evidence that metaheuristics were already used to support feature selection, rule extraction, and parameter optimization in credit-related prediction tasks [87, 90–94]. However, given the observed heterogeneity in reporting and efficiency evidence, the thesis positioned optimization not as an auxiliary enhancement but as a controlled design mechanism whose outputs were evaluated against explicit institutional requirements, including computational time where appropriate [36, 52].

At the evaluation level, the thesis aligned its assessment logic with a metric taxonomy capable of distinguishing predictive performance, optimization behavior, feature outcomes, and statistical validation, as formalized by the metric organization proposed in [107]. This positioning ensured that performance claims were not treated as sufficient in isolation, but were interpreted alongside stability- and feature-related outcomes when such evidence was available in the reviewed literature [48, 50]. Where interpretability was relevant, the thesis drew on the

existence of interpretability-oriented strategies in prior work, including rule-based extraction and selective modeling frameworks [41, 46], and treated interpretability and robustness as design requirements rather than post hoc considerations.

Overall, the theoretical stance of the thesis positioned declarative compliance prediction as a specialized institutional risk modeling problem requiring the integration of supervised learning and metaheuristic search under explicit decision-support constraints. These commitments are operationalized through a stage-based methodological design that translates the identified gaps into concrete modeling, optimization, and evaluation decisions, as detailed in the following chapter.

2.7 Chapter Conclusions

This chapter established the theoretical, conceptual, and analytical foundations that support the development of the research. Through a structured progression from conceptual definitions to critical synthesis, the chapter articulated how predictive modeling, feature selection, and metaheuristic optimization have been addressed in the literature on individual credit risk and compliance-related problems, with particular attention to institutional decision-support contexts.

The conceptual framework formalized income-contingent student loan systems as institutionally mediated environments in which risk emerges not only from financial outcomes, but also from administrative and behavioral processes. Within this framework, declarative compliance was defined as an observable and dynamic behavioral process that conditions the operational functioning of income-based repayment schemes. This perspective justified the treatment of compliance as a primary predictive target and motivated the adoption of supervised machine learning models oriented toward early and actionable institutional interventions.

The systematic literature review provided a structured synthesis of prior research on machine learning, feature selection, and optimization techniques applied to individual credit risk assessment. The reviewed studies revealed a strong predominance of supervised classification paradigms, widespread use of ensemble-based models, and increasing integration of feature selection mechanisms to address high-dimensional institutional data. Metaheuristic optimization emerged as a common strategy for navigating combinatorial search spaces, particularly in wrapper-based feature selection and hybrid modeling pipelines. At the same time, the review highlighted substantial heterogeneity in experimental practices, evaluation metrics, and reporting standards.

Building on this evidence, the state-of-the-art analysis identified dominant modeling paradigms, recurring data characteristics, and prevalent methodological choices. While contemporary approaches have achieved notable improvements in predictive performance, the synthesis showed that interpretability, robustness, and stability are often treated implicitly or inconsistently. Evaluation practices remain heavily oriented toward performance metrics, with limited and fragmented use of feature-oriented, optimization-related, and statistical validation measures that are critical for institutional deployment.

The critical analysis and research gap identification consolidated these observations by exposing a misalignment between prevailing modeling practices and the requirements of institutional decision-support systems in income-contingent student loan contexts. In particular,

the literature rarely addresses compliance as a staged behavioral process, seldom evaluates feature selection outcomes as stable institutional signals, and inconsistently reports robustness and efficiency under operational constraints. These limitations delineate a clear opportunity for methodological contributions that integrate predictive accuracy with interpretability, stability, and practical feasibility.

Overall, the chapter positioned the thesis at the intersection of supervised predictive modeling, metaheuristic optimization, and decision-support design requirements. By grounding methodological choices in the empirical regularities and gaps identified in the literature, the chapter established the conceptual and analytical rationale that guides the stage-based methodological design, experimentation, and interpretation of results developed in the remainder of the thesis.

3 Thesis Research Proposal

Chapter Overview

This chapter presents the methodological framework that supports the design, validation, and application of predictive models developed in this doctoral research. The methodology was conceived to ensure coherence across research questions, hypotheses, modeling decisions, and evaluation strategies, while maintaining transparency, rigor, and reproducibility.

The chapter defines the research design and analytical scope, describes the institutional data sources and dataset construction processes, and details the modeling framework integrating machine learning and metaheuristic optimization. Evaluation protocols and implementation principles are also specified to guarantee methodological consistency across all empirical proposals presented in Chapter 4.

3.1 Operational Problem Definition

From an operational perspective, this research addresses the problem of anticipating income declaration compliance behavior within the FSCU system under real-world institutional constraints. Unlike traditional formulations of credit risk, the focus is not limited to financial default outcomes, but rather to procedural compliance events that precede or condition subsequent financial processes.

Two predictive phenomena are defined as the core operational targets of the research. The first corresponds to first-time income declaration compliance, defined as the binary outcome associated with whether a borrower submits the required income declaration when first obligated to do so. The second corresponds to declaration abandonment, defined as the interruption of income declaration behavior among borrowers who had previously complied with declaration requirements. These phenomena are temporally ordered and reflect distinct stages of interaction between borrowers and the FSCU system.

The operationalization of these targets is subject to strict temporal and informational constraints. For first-time declaration prediction, only academic and administrative information available prior to any declaration event is considered. For declaration abandonment prediction, additional information generated through the declaration process itself becomes available, resulting in higher-dimensional and more heterogeneous feature spaces. This distinction motivates the adoption of differentiated modeling strategies across research stages.

At an institutional level, the problem is characterized by delayed observability and limited opportunities for early intervention. Administrative actions are typically reactive, relying on ex-post detection of non-compliance. The predictive tasks addressed in this research are therefore framed to support anticipatory decision-making, enabling earlier identification of borrowers at risk of non-compliance based on available institutional data.

From a methodological standpoint, the problem is constrained by the use of real-world

administrative data, which entails challenges related to class imbalance, missing information, and heterogeneous data representations. These constraints directly inform the choice of evaluation metrics, modeling paradigms, and optimization strategies adopted throughout the research.

This operational problem definition establishes the foundation for the stage-based methodological design presented in the following sections. The progressive incorporation of data sources and modeling techniques across research stages reflects the evolving informational context of the FSCU system and supports a systematic analysis of how predictive capacity, interpretability, and robustness change as methodological complexity increases.

3.2 Research Objectives, Questions, and Hypotheses

In this research, objectives, research questions, and hypotheses are defined as operational elements of the methodological design rather than as abstract statements of intent. Given the stage-based structure adopted, these elements are formulated to reflect how the research problem is addressed through successive methodological stages, each characterized by different data availability conditions and modeling strategies.

3.2.1 Main Objective

The main objective of this doctoral research is to design, validate, and apply predictive models based on machine learning and metaheuristic optimization to anticipate income declaration compliance and declaration abandonment behaviors in income-contingent student loan systems, using real-world institutional data from the FSCU.

This objective is pursued through a progressive methodological framework in which predictive capacity, interpretability, and robustness are analyzed as data representations and modeling complexity evolve across research stages.

3.2.2 Specific Objectives

The specific objectives of the research are defined in alignment with the methodological stages described in Section 3.4:

- To design and implement a reproducible data processing pipeline and construct a structured dataset derived from FSCU administrative records (Stage I).
- To develop and evaluate interpretable machine learning models for predicting first-time income declaration compliance using strictly pre-declarative academic and administrative information (Stage II).
- To model and predict the abandonment of income declaration among previously compliant borrowers using enriched institutional data representations (Stage III).
- To apply metaheuristic-based feature selection techniques to the declaration abandonment problem and assess their impact on predictive performance, stability, and interpretability (Stage III).

- To explore advanced bioinspired deep learning models for capturing complex patterns associated with income declaration discontinuities (Stage IV).
- To identify and analyze the most influential academic, financial, and administrative variables associated with declaration compliance and non-compliance behaviors across different stages of the declaration process (Stages II–IV).

3.2.3 Research Questions

The research questions are formulated to reflect the operational scope of each methodological stage:

- RQ1: To what extent can machine learning models predict first-time income declaration compliance using institutional data available prior to any declaration event? (Stage II)
- RQ2: How effectively can predictive models anticipate the abandonment of income declaration among borrowers who have previously complied with declaration requirements, under real-world institutional data conditions? (Stage III)
- RQ3: How does metaheuristic-based feature selection influence predictive performance, model stability, and interpretability in the prediction of declaration abandonment? (Stage III)
- RQ4: To what extent do bioinspired deep learning models improve predictive performance in the prediction of income declaration abandonment when compared to optimized classical models under identical data representations and evaluation protocols? (Stage IV)
- RQ5: Which academic, financial, and administrative variables are consistently associated with declaration compliance and non-compliance behaviors across different stages of the income declaration process? (Stages II–IV)

3.2.4 Hypotheses

Based on the operational problem definition and the stage-based methodological design, the following hypotheses are evaluated:

- H1: Machine learning models trained exclusively on pre-declarative institutional data can predict first-time income declaration compliance with statistically significant performance (Stage II).
- H2: Declaration abandonment among previously compliant borrowers can be anticipated using historical academic, administrative, and declaration-related information (Stage III).
- H3: Metaheuristic-based feature selection improves predictive performance and robustness in the prediction of declaration abandonment by reducing feature dimensionality while preserving interpretability (Stage III).

- H4: A reduced and interpretable subset of academic, financial, and administrative variables captures the dominant explanatory structure underlying declaration compliance and non-compliance behaviors across different stages of the income declaration process (Stages II–IV).
- H5: Bioinspired deep learning models provide statistically significant predictive performance improvements over optimized classical models in the prediction of income declaration abandonment, when linearity and interpretability constraints are relaxed under identical data representations and evaluation protocols (Stage IV).

These hypotheses are not tested in isolation, but within the context of the stage-based methodological framework, allowing their evaluation under progressively enriched data representations and modeling strategies.

3.3 Research Design

This doctoral research follows an applied, quantitative, and predictive research design, oriented toward the development and validation of data-driven models under real-world institutional conditions. The study is grounded in the analysis of observational administrative data and does not involve experimental manipulation of subjects or controlled interventions.

The research design is implemented through a stage-based methodological framework, in which successive stages address increasing levels of data richness and modeling complexity. This approach enables a structured analysis of how predictive performance, interpretability, and robustness evolve as additional information sources and methodological components are incorporated. Rather than adopting a single static experimental setup, the design supports incremental methodological refinement aligned with the operational context of the FSCU system.

From a temporal perspective, the design explicitly accounts for causal precedence between available information and predicted outcomes. Predictive models are trained and evaluated using only data that would have been available to the institution at the time decisions could realistically be made. This constraint is central to the validity of the predictive tasks addressed and differentiates the present research from retrospective or ex-post analytical studies.

The study employs supervised machine learning techniques as the primary modeling paradigm, complemented by metaheuristic optimization methods in stages characterized by high-dimensional feature spaces. Model evaluation is conducted under a comparative framework, using consistent data partitions, performance metrics suitable for imbalanced classification, and repeated experimental runs to assess stability and robustness.

Overall, the research design integrates data engineering, predictive modeling, and optimization within a coherent methodological structure that reflects both the technological nature of the doctoral work and the operational requirements of institutional decision-making in income-contingent student loan systems.

3.4 Research Methodology Stages

The methodological implementation of this research is organized into a sequence of research stages, each addressing a specific aspect of the operational problem and contributing incrementally to the fulfillment of the research objectives and the evaluation of the proposed hypotheses. This stage-based structure reflects the technological nature of the doctoral work and supports a systematic progression from data infrastructure development to advanced predictive modeling.

Rather than treating the research as a single experimental setup, the stages are designed to isolate the effects of data availability, modeling choices, and optimization strategies on predictive performance, interpretability, and robustness. Each stage builds upon the outputs of the preceding one, ensuring methodological continuity and enabling cross-stage comparative analysis.

Problem Definition and Institutional Framing (Stage 0)

The purpose of this stage is to formalize the research problem within its institutional context and define the predictive phenomena under investigation. This stage establishes the operational definitions of first-time income declaration compliance and declaration abandonment, as well as the temporal and informational constraints governing their prediction.

By grounding the research in real-world institutional processes, this stage provides the conceptual basis for the subsequent methodological stages and informs the formulation of research questions and hypotheses.

Data Pipeline Design and Dataset Construction (Stage I)

Stage I addresses the design and implementation of a reproducible data processing pipeline for transforming heterogeneous FSCU administrative records into structured datasets suitable for predictive modeling. The central contribution of this stage is methodological, as it establishes data consistency, traceability, and temporal validity under real-world institutional constraints.

Although limited exploratory analysis is performed, this stage does not pursue empirical optimization. Its outputs provide the necessary data infrastructure that enables the predictive modeling, feature analysis, and optimization tasks developed in subsequent stages.

Figure 3.1 presents a simplified entity–relationship representation of the FSCU information system, illustrating the core institutional data structures integrated during this stage to support dataset construction.

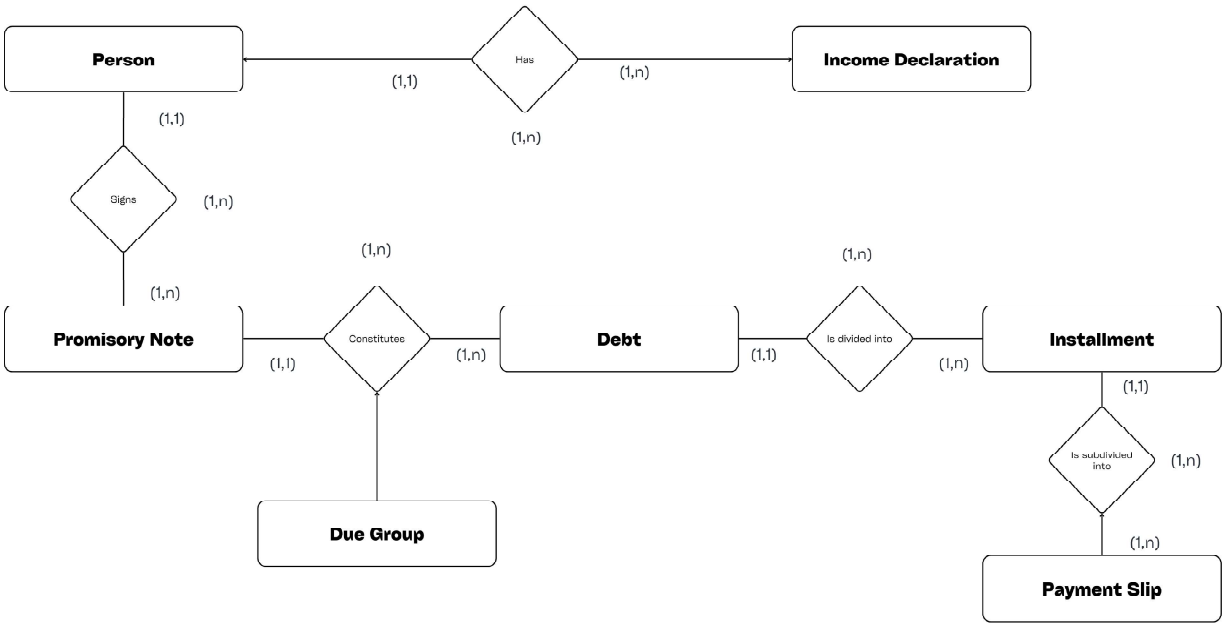


Figure 3.1: Simplified entity–relationship model supporting FSCU data integration.

Interpretable Predictive Modeling under Pre-Declarative Constraints (Stage II)

The objective of this stage is to evaluate the feasibility of predicting first-time income declaration compliance using exclusively pre-declarative academic and administrative information. This stage directly addresses the research questions related to early-stage prediction and the identification of institutionally meaningful predictors under informational constraints.

The methodological focus is placed on interpretable machine learning models calibrated through standard hyperparameter tuning via Grid Search. This tuning process is treated as conventional model configuration rather than as an optimization objective in itself, and no metaheuristic-based optimization or feature selection mechanisms are introduced at this stage. To account for class imbalance in declaration outcomes, selected experiments incorporate controlled imbalance-handling strategies (e.g., under-sampling or synthetic over-sampling), evaluated as preprocessing variants rather than optimization targets.

By combining tuned yet interpretable models with systematic stability and interpretability analyses, this stage establishes a controlled predictive baseline that isolates the contribution of antecedent information. The results delineate the attainable limits of predictive performance and explanation quality when operating without declaration-generated data, providing a reference point for subsequent stages that introduce higher model complexity or optimization-driven feature selection.

Metaheuristic-Optimized Feature Selection for Declaration Abandonment (Stage III)

This stage addresses the prediction of declaration abandonment among previously compliant borrowers under conditions of increased data richness and feature dimensionality. Building upon the modeling framework established in Stage II, metaheuristic optimization techniques are introduced to manage high-dimensional feature spaces and to identify informative and parsimonious subsets of variables.

The methodological contribution of this stage lies in evaluating how optimization-driven feature selection affects predictive performance and robustness, while explicitly characterizing the trade-offs introduced in terms of model interpretability. This stage provides empirical evidence relevant to the hypotheses concerning the benefits and limitations of metaheuristic-based feature selection in complex institutional datasets.

Bioinspired Deep Learning Models for Discontinuity Prediction (Stage IV)

The purpose of this stage is to explore advanced non-linear predictive models based on bioinspired deep neural network architectures applied to income declaration discontinuity prediction. By relaxing linearity and interpretability constraints, this stage examines whether increased model expressiveness yields additional predictive gains beyond those attainable through optimization-driven feature selection.

This stage contributes to the comparative assessment of modeling paradigms by contrasting deep learning approaches with the optimized classical models analyzed in previous stages. The results support the analysis of trade-offs between model complexity, predictive performance, robustness, and practical applicability within the FSCU context.

Integrated Cross-Stage Analysis (Stage V)

The final stage synthesizes the methodological and empirical insights obtained across all previous stages through a structured cross-stage comparison. This analysis integrates results related to predictive performance, feature space characteristics, data requirements, model complexity, robustness, and interpretability, allowing the cumulative effects of successive methodological decisions to be systematically examined.

Rather than introducing new modeling experiments, this stage consolidates evidence across stages to identify consistent patterns, trade-offs, and performance boundaries associated with different predictive paradigms and information constraints. The integrated analysis supports a coherent interpretation of how progressively enriched representations, optimization strategies, and model classes influence predictive behavior within the FSCU context.

This stage provides the methodological foundation for the discussion of results and the articulation of the scientific and practical contributions of the thesis, directly informing the evaluation of the research questions and hypotheses.

3.5 Modeling and Evaluation Strategy

Across all research stages, a consistent modeling and evaluation framework was adopted to ensure methodological coherence and comparability of results. Rather than defining independent experimental setups for each stage, common principles were applied to guide model development, training, and assessment under different data availability conditions.

Supervised machine learning constitutes the core modeling paradigm throughout the research. Classification models were trained to predict declarative compliance outcomes using structured representations derived from institutional data. Depending on the research stage, models varied in complexity and interpretability, ranging from transparent classifiers to advanced non-linear architectures. These variations were introduced deliberately to assess trade-offs between predictive performance, interpretability, and robustness.

Evaluation procedures were designed to reflect the operational characteristics of the problem. Given the presence of class imbalance and asymmetric misclassification costs, performance metrics suitable for imbalanced classification were employed consistently across stages. Data partitioning strategies and repeated experimental runs were applied to mitigate variance and support stability analysis.

Metaheuristic optimization techniques were incorporated selectively in stages characterized by high-dimensional feature spaces. When applied, these techniques were integrated within a wrapper-based framework, ensuring that feature selection was guided by predictive performance while preserving compatibility with the underlying classifiers.

Throughout all stages, strict temporal constraints were enforced to prevent information leakage. Models were trained and evaluated using only data that would have been available at the time of prediction in a real institutional setting. This principle ensures that reported performance reflects realistic predictive scenarios rather than retrospective reconstruction.

Together, these modeling and evaluation principles provide a unified methodological foundation that supports cross-stage comparison and underpins the interpretation of results presented in the subsequent chapter.

3.5.1 Evaluation Strategy

Across all research stages, model performance is assessed appropriate classification metrics, including accuracy, precision, recall, F1-score, area under the ROC curve, and Matthews Correlation Coefficient, depending on the class balance and decision context. Cross-validation and hold-out validation strategies are applied to ensure robustness and to mitigate overfitting. Comparative evaluations are conducted against baseline models to quantify the contribution of feature selection and optimization techniques.

3.5.2 Implementation Details and Reproducibility

All experiments were implemented using open-source tools and libraries. The computational environment, software versions, random seed configurations, and experimental parameters were explicitly controlled to ensure reproducibility. The modular structure of the implementation

allows the proposed strategy to be extended to additional datasets or compliance scenarios.

4 Results

Chapter Overview

This chapter presents the empirical results of the doctoral research following the stage-based methodological framework defined in Chapter 3. The results are organized according to the progressive research stages through which the predictive modeling problem was addressed, reflecting increasing levels of data availability, modeling complexity, and methodological intervention.

Not all methodological stages yield empirical results in the strict sense. Stage 0, dedicated to problem definition and institutional framing, establishes the conceptual and operational foundations of the study and therefore does not generate empirical outcomes reported in this chapter. Similarly, Stage V corresponds to an integrative cross-stage synthesis and is addressed exclusively in the discussion chapter, where results are interpreted comparatively and situated within the broader research context.

Empirical results are reported for Stages I through IV. Stage I presents descriptive and baseline results associated with the construction of the FSCU dataset and the validation of predictive feasibility under constrained informational conditions. These results provide the empirical foundation of the study by characterizing data structure, quality, and baseline model behavior, and by supporting subsequent stage-based experimentation.

Stage II reports results obtained from interpretable machine learning models applied to the prediction of first-time income declaration compliance using strictly pre-declarative academic and administrative information. This stage directly addresses the early-stage prediction problem defined in the methodological framework and evaluates predictive feasibility, stability, and interpretability under causal and transparency constraints, providing empirical evidence relevant to Research Question RQ1 and Hypothesis H1.

Stage III presents the results of predictive modeling for declaration abandonment among previously compliant borrowers, incorporating enriched data representations derived from declaration related information. This stage evaluates both the feasibility of abandonment prediction under real-world institutional conditions and the impact of metaheuristic-based feature selection on predictive performance, robustness, and feature compactness. The reported results therefore address Research Questions RQ2 and RQ3, and provide empirical support for Hypotheses H2 and H3.

Finally, Stage IV reports complementary results obtained from bioinspired deep learning models designed to capture complex non-linear patterns associated with income declaration discontinuities. These results are presented within a comparative framework, assessing whether increased model expressiveness yields predictive improvements over optimized classical approaches under identical data representations and evaluation protocols. This stage directly addresses Research Question RQ4 and Hypothesis H5, while contributing additional evidence to the broader analysis of modeling trade-offs.

Throughout the chapter, results are presented descriptively and comparatively, without redefining methodological choices or introducing new modeling assumptions. Each stage reports

observable empirical outcomes aligned with the research questions and hypotheses defined in Chapter 3, while interpretative synthesis, theoretical implications, and cross-stage integration are deliberately deferred to Chapter 5.

4.1 Data Pipeline Design and Baseline Predictive Modeling (Stage I)

4.1.1 Stage I Overview and Baseline Objective

Stage I constitutes the first empirical stage of the research and serves a primarily methodological and exploratory role within the stage-based framework. Its objective is to establish the empirical feasibility of predictive modeling for income declaration compliance within the FSCU system using real-world institutional data, under realistic informational and operational constraints. At this stage, the emphasis is not placed on predictive performance optimization, but on validating whether declarative behavior can be meaningfully learned from structured representations derived from heterogeneous administrative records.

The predictive task considered in this stage corresponds to first-time income declaration compliance. Only academic, financial, and administrative information available prior to the first declaration deadline is included, resulting in a constrained feature space that reflects the information realistically available to the institution at the earliest decision point. No feature selection, optimization, or advanced modeling strategies are introduced.

The empirical objective of Stage I is therefore to characterize baseline predictive behavior, assess data suitability, and identify initial modeling limitations, establishing a neutral technical reference that supports and motivates the more focused and interpretable predictive analyses conducted in subsequent stages.

4.1.2 Dataset Construction and Baseline Representation

The empirical results reported in this stage are based on a structured dataset constructed from institutional records associated with the administration of the FSCU at PUCV. The dataset integrates heterogeneous administrative information into a borrower-level analytical representation suitable for supervised learning.

Each observation represents a unique borrower at the time of their first income declaration obligation. To ensure temporal validity, only information observable prior to the declaration deadline was included, thereby preventing information leakage. The resulting feature space includes numerical, categorical, and temporally derived variables describing borrower characteristics, loan conditions, and academic trajectory.

Figure 4.1 summarizes the data processing workflow that transforms raw institutional records into an analyzable dataset and supports baseline predictive evaluation. This pipeline constitutes a key empirical outcome of Stage I, as it establishes traceability between raw FSCU data and observed model behavior.

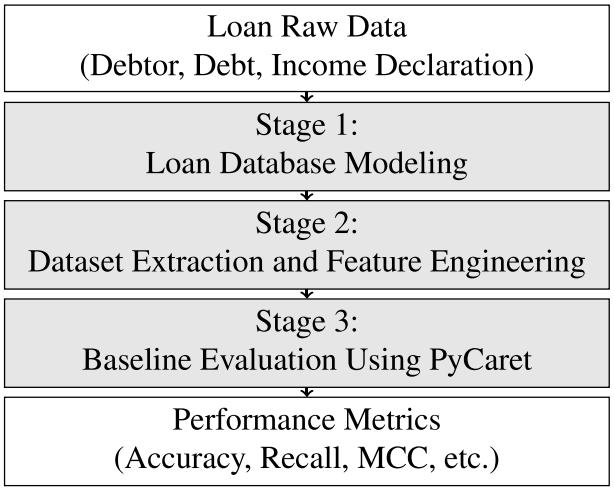


Figure 4.1: Data processing pipeline for FSCU dataset construction and baseline analysis.

4.1.3 Baseline Modeling Setup

Baseline predictive modeling was conducted using a diverse set of supervised classification algorithms representing linear, tree-based, ensemble, and distance-based paradigms. All models were trained using default hyperparameter configurations. No hyperparameter tuning, resampling, cost-sensitive learning, or feature selection mechanisms were applied at this stage in order to isolate intrinsic model behavior and preserve comparability.

A unified experimental framework and fixed validation protocol were employed to ensure reproducibility and consistency across models. Under this configuration, observed differences in predictive performance can be attributed to model structure rather than procedural variability.

4.1.4 Evaluation Metrics

Model performance was evaluated using accuracy, precision, recall, F1-score, area under the ROC curve (AUC), Cohen's Kappa, and the Matthews Correlation Coefficient (MCC). Given the imbalanced distribution of compliant and non-compliant borrowers, MCC is treated as a primary robustness indicator, as it provides a balanced assessment that accounts for all elements of the confusion matrix. AUC is reported as a threshold-independent measure of discriminative capacity and is not interpreted as a direct measure of classification accuracy. These metrics are reported exclusively for exploratory and baseline characterization purposes and are not intended to support fine-grained model selection or explanatory analysis at this stage.

4.1.5 Baseline Predictive Performance

Table 4.1 reports the predictive performance of baseline classification models evaluated on the FSCU dataset under strictly pre-declarative informational conditions. All results correspond to default model configurations and reflect intrinsic model behavior in the absence of optimization

or imbalance mitigation strategies. Accordingly, the reported results should be interpreted as indicative of baseline learnability rather than as optimized predictive performance.

Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
Extreme Gradient Boosting	0.9626	0.9860	0.9888	0.9677	0.9781	0.8484	0.8509
Light Gradient Boosting Machine	0.9606	0.9843	0.9906	0.9638	0.9770	0.8379	0.8421
Gradient Boosting Classifier	0.9280	0.9635	0.9798	0.9380	0.9584	0.6918	0.7014
Decision Tree Classifier	0.9275	0.8612	0.9569	0.9575	0.9572	0.7212	0.7214
Random Forest Classifier	0.9135	0.9567	0.9809	0.9219	0.9505	0.6104	0.6296
Logistic Regression	0.9053	0.9422	0.9699	0.9223	0.9455	0.5866	0.5985
Extra Trees Classifier	0.8958	0.9383	0.9758	0.9080	0.9407	0.5169	0.5409
Ada Boost Classifier	0.8865	0.9288	0.9545	0.9152	0.9344	0.5154	0.5222
Linear Discriminant Analysis	0.8732	0.8993	0.9413	0.9117	0.9263	0.4731	0.4765
Ridge Classifier	0.8658	0.8999	0.9841	0.8734	0.9255	0.2754	0.3376
Dummy Classifier	0.8466	0.5000	1.0000	0.8466	0.9169	0.0000	0.0000
K Neighbors Classifier	0.8389	0.7817	0.9409	0.8776	0.9081	0.2584	0.2696
Naive Bayes	0.7714	0.8143	0.7802	0.9397	0.8525	0.3634	0.3968
SVM - Linear Kernel	0.7471	0.8588	0.8145	0.8999	0.8110	0.1455	0.2210
Quadratic Discriminant Analysis	0.3816	0.6767	0.3043	0.9165	0.4090	0.0537	0.1015

Table 4.1: Baseline predictive performance of classification models on the FSCU dataset

As shown in Table 4.1, accuracy values are generally high across most models. However, performance differences become more pronounced when assessed using imbalance-aware metrics. Ensemble-based methods, particularly Extreme Gradient Boosting and Light Gradient Boosting, achieve the highest MCC values, indicating superior balance between correct classification of compliant and non-compliant cases.

In contrast, models such as Logistic Regression and distance-based classifiers achieve competitive accuracy but lower MCC values, reflecting reduced robustness under class imbalance. The Dummy Classifier illustrates this limitation explicitly, achieving high recall for the majority class while yielding null MCC and Kappa values, thereby confirming that accuracy alone is insufficient to characterize predictive performance in this context.

4.1.6 Error Distribution and Class Imbalance Effects

The impact of class imbalance on baseline model behavior is illustrated in Figure 4.2, which presents confusion matrices for representative ensemble classifiers. These matrices provide a granular view of classification outcomes that complements aggregate performance metrics.

As shown in Figure 4.2, baseline models correctly classify the majority of compliant borrowers, resulting in a high number of true positives. Nevertheless, false negatives remain non-negligible, indicating limitations in the consistent identification of non-compliant borrowers under constrained informational conditions.

Comparative inspection of the confusion matrices reveals differences in error trade-offs across models that are not fully captured by accuracy values alone. These differences directly

influence imbalance-aware metrics such as MCC and F1-score reported in Table 4.1, highlighting the relevance of examining instance-level error distributions in institutional datasets.

4.1.7 Limitations of the Baseline

Despite demonstrating that first-time income declaration compliance constitutes a learnable predictive problem, Stage I is subject to several structural limitations inherent to its baseline design. First, all models are evaluated under default hyperparameter configurations, without systematic calibration or controlled exploration of the parameter space. As a result, reported performance reflects intrinsic algorithm behavior rather than optimized predictive capacity.

Second, no class imbalance mitigation strategies are applied at this stage. Given the asymmetric distribution of compliant and non-compliant borrowers, this constraint directly affects error profiles and limits the interpretability of accuracy-based comparisons. Although imbalance-aware metrics such as MCC partially address this issue, the baseline design does not attempt to correct or control class distribution effects.

Third, interpretability is not examined explicitly in Stage I. While some models inherently support transparent decision structures, no systematic feature attribution, rule extraction, or explanation analysis is conducted. Consequently, the stage does not support institutionally actionable explanations regarding the drivers of compliance behavior.

Finally, the feature space is treated as fixed and uncurated beyond basic preprocessing. No semantic refinement, feature reduction, or representation analysis is performed, limiting insight into which pre-declarative variables are most informative. These limitations delimit the analytical scope of Stage I and motivate the structured, interpretable, and controlled modeling strategy introduced in Stage II.

4.1.8 Summary of Stage I Results

The results of Stage I demonstrate that income declaration compliance within the FSCU system constitutes an empirically learnable predictive problem when modeled using real institutional data, even under strictly pre-declarative and informationally constrained conditions. As evidenced by the performance metrics reported in Table 4.1, baseline machine learning models achieve discrimination capacity beyond trivial or random classification, confirming the feasibility of early-stage predictive modeling.

At the same time, the confusion matrices presented in Figure 4.2 reveal systematic error asymmetries associated with class imbalance and limited feature availability. These patterns indicate that overall accuracy alone does not adequately characterize predictive behavior in this context and underscore the need for imbalance-aware evaluation metrics to properly assess model performance under institutional data conditions.

True Class \ Predicted Class	F	T
F	1033	204
T	71	6760

(a) XGBoost

True Class \ Predicted Class	F	T
F	989	248
T	63	6768

(b) LGB

True Class \ Predicted Class	F	T
F	808	429
T	136	6695

(c) GB

Figure 4.2: Confusion matrices for selected baseline ensemble models

Beyond predictive outcomes, the successful construction of a structured, temporally valid, and reproducible dataset constitutes a central empirical contribution of this stage. The data processing pipeline summarized in Figure 4.1 validates the integration of heterogeneous administrative records into a coherent analytical framework suitable for subsequent modeling stages.

Taken together, the results of Stage I establish a neutral and methodologically constrained empirical baseline that confirms the feasibility of early-stage prediction under strictly pre-declarative conditions. At the same time, the observed limitations related to class imbalance, model calibration, and the absence of interpretability analysis motivate the transition to a more structured and explanatory modeling setting. These findings directly inform the design of Stage II, which builds upon the same informational regime while introducing interpretable models, controlled preprocessing strategies, and robustness-oriented evaluation protocols.

4.2 Interpretable Modeling under Pre-Declarative Constraints (Stage II)

4.2.1 Stage II Overview

Stage II addresses the predictive modeling of first-time income declaration compliance under a strictly pre-declarative information regime, directly corresponding to the early-stage prediction problem defined in the methodological framework. This stage extends the baseline established previously by relying exclusively on academic and administrative records that temporally precede any declaration event, thereby preserving causal precedence and preventing information leakage. The analytical scope is explicitly restricted to antecedent information available prior to the first declaration deadline, ensuring institutional validity and operational feasibility.

The primary objective of Stage II is to empirically characterize the predictive performance, stability, and interpretability attainable when machine learning models are trained exclusively on pre-declarative data. In contrast to subsequent stages, this stage deliberately excludes feature selection procedures, metaheuristic optimization, and post-declarative variables. Model development relies on transparent and interpretable classifiers calibrated through standard hyperparameter

tuning, allowing observed effects to be attributed to data content and modeling assumptions rather than to algorithmic search or dimensionality reduction mechanisms. The results of this stage provide direct empirical evidence relevant to Research Question RQ1 and Hypothesis H1.

Methodologically, Stage II serves as a controlled and causally grounded reference point within the stage-based design of the thesis. It establishes an empirically validated benchmark for interpretable predictive modeling under realistic institutional constraints, against which subsequent stages that introduce increased data richness, optimization mechanisms, and adaptive modeling strategies can be rigorously compared. In this sense, Stage II delineates the capabilities and limitations of early-stage predictive modeling in terms of both classification performance under class imbalance and the stability and coherence of explanatory patterns derived from institutional data.

4.2.2 Problem Definition and Predictive Scope

Stage II focuses on the prediction of first-time income declaration compliance among borrowers subject to an initial declaration obligation within the institutional credit system. The predictive task is formally defined as a binary classification problem, where the target variable indicates whether a borrower submitted their first income declaration by the required deadline. The analysis is restricted to a homogeneous cohort facing this initial declarative event, excluding borrowers with prior declaration history.

The predictive scope of this stage is constrained by a strictly pre-declarative information regime. Only academic, administrative, demographic, and financial attributes that temporally precede any declaration activity are considered admissible inputs. Variables generated as a consequence of declaration behavior, repayment activity, or post-deadline processes are explicitly excluded to preserve causal ordering and institutional plausibility.

Under these constraints, Stage II aims to assess whether meaningful predictive signal exists prior to any observed declaration behavior, and to what extent such signal can be captured by interpretable machine learning models. This scope deliberately avoids considerations related to feature optimization, metaheuristic search, or post-declarative enrichment, which are deferred to subsequent stages. As such, the problem definition in Stage II isolates the earliest feasible decision point for anticipatory risk assessment based solely on antecedent institutional data.

4.2.3 Data Preparation and Experimental Protocol

Dataset Description and Cohort Definition

The dataset analyzed in Stage II is derived from institutional administrative records associated with the university credit system and encompasses borrowers subject to a first-time income declaration requirement. The analytical cohort is defined by selecting individuals who reached their initial declaration deadline within the observation window and for whom complete pre-declarative academic and administrative information was available prior to that deadline.

Only borrowers facing their first declarative obligation are included, thereby excluding cases with any prior declaration history. This cohort definition ensures conceptual homogeneity of

the predictive task and avoids conflating first-time compliance behavior with patterns influenced by prior exposure to the declaration process. The resulting population represents a well-defined institutional scenario in which borrowers transition from a purely academic status to an initial compliance obligation.

All records are temporally aligned to guarantee that input variables strictly precede the declaration deadline. Academic trajectory variables reflect enrollment history accumulated before enforceability, while financial attributes correspond to loan conditions established prior to any declaration activity. This design preserves causal precedence and ensures that observed predictive patterns are attributable to antecedent institutional information rather than to behavioral feedback effects. The dataset construction and inclusion criteria follow the general preprocessing principles described in the methodological chapter, with Stage II operating as the first fully constrained empirical evaluation under these rules.

Feature Space Characterization

The pre-declarative feature space considered in Stage II integrates demographic attributes, loan characteristics, and structured representations of academic trajectories derived from institutional records. This enriched yet strictly antecedent information regime reflects the breadth of data available prior to the first income declaration deadline, while preserving temporal and causal validity.

Table 4.2 summarizes the initial pool of academic and administrative variables prior to preprocessing. This pool includes demographic indicators (e.g., sex, nationality, marital status), financial exposure measures related to the loan contract, and variables capturing enrollment history, institutional affiliation, and academic continuity. At this stage, the feature set represents the full scope of potentially relevant pre-declarative information, without regard to redundancy, scale, or representation constraints.

Name	Data type	Feature type	Detail
estado_civil	Boolean	Categorical	1 and 2
nacionalidad	Boolean	Categorical	1 and 2
sexo	Boolean	Categorical	M and F
fecha_nacimiento	Date	Date	01/01/1900 to 26/03/1991
edad	Integer	Numerical	21 to 119
edad_dias	Integer	Numerical	7860 to 43646
deud_monto	Float	Numerical	0.571 to 1285.762
deud_fecha_exigibilidad	Date	Date	01/01/1994 to 01/01/2023
deud_t_deuda	Integer	Categorical	1 value
tiene_declaracion	Boolean	Target	0 and 1
monto_total_pagare	Float	Numerical	1.32 to 922.22
conteo_pagare	Integer	Numerical	1 to 29
anio_ult_matr	Integer	Categorical	13 values
e_ult_matr	Integer	Categorical	1 value
cod_carr_ult_matr	Integer	Categorical	80 values
carr_t_carrera	Integer	Categorical	1 value
cod_inst_ult_matr	Integer	Categorical	1 value
conteo_matr	Integer	Numerical	1 to 31
facultad	String	Categorical	9 values
escuela	String	Categorical	34 values
stem	Boolean	Categorical	0 and 1

Table 4.2: Initial pool of pre-declarative academic and administrative features (Stage II)

Following representation and preprocessing decisions defined in the methodological framework, a reduced and semantically coherent feature set was obtained. Table 4.3 reports the final set of variables used for model training in Stage II. This reduction reflects semantic curation and representation decisions aimed at preserving interpretability and institutional meaning, rather than any form of metaheuristic-based feature selection or optimization. Redundant, non-informative, or structurally dependent attributes were removed, yielding a compact feature space constrained to strictly antecedent information.

Name	Data type	Feature type	Detail
estado_civil	Boolean	Categorical	1 and 2
nacionalidad	Boolean	Categorical	Filtered (foreign category removed)
sexo	Boolean	Categorical	M and F
deud_monto	Float	Numerical	standardized
dia_exigibilidad	Integer	Numerical	1 to 31
anio_exigibilidad	Integer	Numerical	multiple years
anio_ult_matr	Integer	Categorical	filtered values
conteo_matr	Integer	Numerical	1 to 31
facultad	String	Categorical	9 values
stem	Boolean	Categorical	0 and 1
tiene_declaracion	Boolean	Target	0 and 1

Table 4.3: Final pre-declarative feature set used for interpretable modeling (Stage II)

The resulting representation balances informational richness with transparency, enabling interpretable modeling while maintaining fidelity to institutional data-generating processes. This curated feature space constitutes the complete conceptual input domain for Stage II and serves as a stable reference point for subsequent stages that introduce higher dimensionality and optimization mechanisms.

Missing Value Handling and Transformation

Missing value treatment and feature transformations in Stage II follow the preprocessing procedures defined in the methodological framework and are applied uniformly across all experimental configurations. These operations are restricted to standard data preparation steps and do not involve feature selection or optimization-driven transformations.

Categorical variables were encoded into machine-readable representations consistent with their institutional semantics, while numerical variables were transformed to ensure comparability across scales when required by the modeling techniques. In particular, continuous financial and academic variables were standardized prior to model training to support stable parameter estimation in linear and distance-based classifiers. Date-type attributes were decomposed into interpretable numerical components (e.g., year or day indicators) to preserve temporal information within the strictly pre-declarative regime.

Missing values were handled using predefined imputation rules specified in the methodology chapter, ensuring that imputation relied exclusively on information available prior to the declaration deadline. These procedures were applied identically to training and test partitions within each experimental run to prevent information leakage. As a result, the transformed dataset provides a complete and consistent representation of pre-declarative institutional data, suitable for interpretable modeling under the constraints imposed in Stage II.

Class Imbalance Handling

The first-time income declaration compliance outcome exhibits a moderate class imbalance, reflecting the institutional prevalence of compliant borrowers relative to non-compliant cases. To account for this imbalance without altering the underlying predictive objective, Stage II incorporates controlled class imbalance handling strategies as part of the experimental design.

Specifically, multiple resampling variants were evaluated, including random under-sampling (RUS) of the majority class and synthetic over-sampling techniques such as SMOTE and ADASYN. These methods were applied exclusively to the training data within each experimental run and were treated as alternative preprocessing configurations rather than as optimization objectives. No resampling was performed on the held-out test set.

The inclusion of imbalance-handling variants results in a set of predefined experimental configurations that allow systematic assessment of model sensitivity to class distribution shifts. Importantly, these strategies do not modify the feature space, introduce new variables, or involve metaheuristic search procedures. Their role in Stage II is limited to mitigating potential bias toward the majority class and to enabling fair comparison of interpretable models under different but controlled data balance conditions.

Stratified Partitioning and Robustness Design

To ensure robustness and comparability of results under class imbalance, Stage II adopts a stratified data partitioning strategy. The dataset is divided into training and test subsets using class-stratified sampling, preserving the original class distribution of first-time declaration outcomes in the held-out test set. This design supports unbiased evaluation of generalization performance under realistic institutional conditions.

All preprocessing operations, including encoding, transformation, imputation, and class imbalance handling, are confined to the training data within each experimental run. Model evaluation is performed exclusively on the corresponding test partition, which remains unseen during training and preprocessing. This separation prevents information leakage and ensures that reported results reflect genuine predictive behavior.

Robustness is further supported through repeated experimental runs with controlled random seeds, as specified in the methodological chapter. Performance metrics are aggregated across runs to reduce sensitivity to incidental partitioning effects and sampling variability. Within Stage II, this robustness design provides a stable empirical foundation for interpreting performance and interpretability outcomes, while avoiding methodological complexity beyond standard validation practices.

4.2.4 Modeling Framework

The modeling framework adopted in Stage II is designed to balance predictive capability with transparency under strictly pre-declarative constraints. The emphasis is placed on interpretable machine learning models whose decision mechanisms can be directly examined or reliably explained using established interpretability techniques. This choice aligns with the institutional

objective of producing verifiable and auditable predictions at an early decision point, prior to any observed declaration behavior.

All models in this stage are trained using the fixed pre-declarative feature space defined after preprocessing and representation decisions. No feature selection procedures, dimensionality reduction techniques, or metaheuristic optimization mechanisms are introduced. As a result, observed differences in predictive behavior can be attributed to model structure, calibration, and controlled preprocessing choices rather than to algorithmic search over the feature space.

Hyperparameter adjustment, when applied, is treated as standard model calibration within predefined parameter grids and does not alter the interpretability-oriented nature of the framework. The modeling strategy in Stage II therefore establishes a transparent empirical reference, enabling subsequent stages to evaluate the marginal contribution of increased model complexity, optimization strategies, and adaptive mechanisms relative to a well-defined interpretable baseline.

Interpretable Model Families

Stage II evaluates a set of machine learning model families selected for their interpretability and suitability under institutional constraints. The chosen classifiers span linear, probabilistic, instance-based, and tree-based paradigms, enabling comparison of diverse decision mechanisms while preserving transparency of model behavior.

Linear models, including Logistic Regression and Linear Support Vector Machines, are included due to their direct coefficient-based interpretations and stable behavior under standardized feature spaces. Probabilistic modeling is represented by Gaussian Naive Bayes, which provides a simple and auditable generative perspective on class separation under conditional independence assumptions. Instance-based learning is explored through the K-Nearest Neighbors classifier, offering locally interpretable decision logic based on similarity in the pre-declarative feature space.

Tree-based models are incorporated to capture nonlinear relationships while retaining rule-based interpretability. This group includes single Decision Trees, which allow explicit inspection of decision paths, as well as ensemble extensions such as Random Forest and Light Gradient Boosting Machine. Although ensemble methods increase structural complexity, their inclusion remains consistent with the interpretability objective through the use of model-wise importance measures and post hoc explanation techniques.

Together, these model families provide a comprehensive and interpretable modeling spectrum for assessing first-time income declaration compliance using exclusively pre-declarative institutional data. Their joint evaluation supports analysis of how different inductive biases affect predictive behavior and explanatory coherence within a constrained and transparent modeling framework.

Hyperparameter Calibration Strategy

Hyperparameter calibration in Stage II is performed using standard Grid Search procedures within predefined parameter spaces, as specified in the methodological chapter. This process is applied selectively to assess the impact of conventional model calibration on predictive performance and

stability, while preserving the interpretability-oriented nature of the modeling framework.

The parameter grids are defined a priori for each model family and reflect commonly accepted tuning ranges. Grid Search is conducted exclusively on the training data within each experimental configuration, using cross-validated performance estimates where applicable. No adaptive, heuristic, or metaheuristic search strategies are employed at this stage, and the feature space remains fixed throughout calibration.

Importantly, hyperparameter tuning in Stage II is treated as a controlled refinement of baseline models rather than as an optimization objective in itself. Its role is limited to aligning model capacity with the statistical properties of the pre-declarative data, enabling fair comparison between uncalibrated and calibrated configurations. This distinction ensures that performance gains observed in Stage II can be attributed to standard calibration practices, providing a transparent benchmark for the more advanced optimization mechanisms introduced in subsequent stages.

Baseline Configuration Definition

Within Stage II, a clear distinction is established between baseline and variant model configurations to support structured comparison under pre-declarative constraints. The baseline configuration for each model family corresponds to the pipeline trained on the preprocessed feature space without class rebalancing and without hyperparameter calibration beyond default settings. This configuration represents the most direct application of each algorithm to the antecedent institutional data.

All additional configurations are defined as controlled variants of this baseline. These include pipelines incorporating class imbalance handling strategies (e.g., random under-sampling or synthetic over-sampling) and pipelines subjected to standard hyperparameter calibration via Grid Search. Each variant modifies a single methodological aspect relative to the baseline, allowing isolated assessment of its effect on predictive behavior.

For analytical consistency, best-performing configurations referenced in subsequent analyses, such as confusion matrix inspection or interpretability assessment, are selected from among these predefined variants based on validation performance criteria. This selection does not alter the experimental design or introduce post hoc optimization, but serves to illustrate representative model behavior under the constraints imposed in Stage II.

4.2.5 Evaluation Metrics and Validation

Model evaluation in Stage II is conducted using a set of complementary classification metrics selected to reflect performance under class imbalance and to support balanced interpretation of predictive behavior. Given the asymmetric class distribution of first-time declaration outcomes, the Matthews Correlation Coefficient (MCC) is adopted as the primary evaluation metric. MCC provides a single summary measure that accounts for all elements of the confusion matrix and remains informative even when class proportions are skewed.

Accuracy is reported as a secondary metric to provide an overall view of correct classification rates, but its interpretation is treated cautiously due to its sensitivity to majority-class

dominance. Precision, recall, and F_1-score are included as complementary indicators to characterize the trade-offs between false positives and false negatives, and to assess minority-class sensitivity.

Validation follows the stratified partitioning and robustness design described earlier. All metrics are computed on the held-out test set, which remains untouched during model training, preprocessing, and calibration. When multiple experimental runs are performed, performance measures are aggregated to reduce variability associated with random partitioning. This validation strategy ensures that reported metrics reflect generalization performance under realistic pre-declarative conditions and provide a consistent basis for comparison across model families and experimental configurations.

4.2.6 Experimental Results

Predictive Performance under Pre-Declarative Constraints

This subsection reports the predictive performance achieved by the evaluated models when trained exclusively on strictly pre-declarative institutional data. Results are summarized across all predefined experimental configurations, including baseline pipelines, imbalance-handling variants, and calibrated models, and are evaluated on a held-out test set following the validation protocol described earlier.

Tables 4.4–4.10 present the test-set performance for all model families considered in Stage II. Performance is reported using Accuracy, Precision, Recall, F_1-score, and Matthews Correlation Coefficient (MCC), enabling complementary assessment of overall correctness, class-wise behavior, and balanced discrimination under class imbalance.

Pipeline	Accuracy	Precision	Recall	F1-Score	MCC
KNN	0.746	0.795	0.872	0.832	0.323
Smote KNN	0.701	0.814	0.758	0.785	0.298
Adasyn KNN	0.701	0.817	0.753	0.784	0.303
RUS KNN	0.720	0.805	0.807	0.806	0.304
OPT KNN	0.765	0.785	0.928	0.850	0.344
OPT Smote KNN	0.746	0.814	0.839	0.826	0.354
OPT Adasyn KNN	0.740	0.823	0.814	0.819	0.361
OPT RUS KNN	0.763	0.804	0.887	0.843	0.366

Table 4.4: K-Nearest Neighbors Results

Pipeline	Accuracy	Precision	Recall	F1-Score	MCC
Naive Bayes	0.615	0.874	0.543	0.670	0.309
Smote Naive Bayes	0.569	0.892	0.458	0.605	0.293
Adasyn Naive Bayes	0.562	0.888	0.449	0.596	0.282
RUS Naive Bayes	0.599	0.875	0.518	0.651	0.296
OPT Naive Bayes	0.715	0.847	0.738	0.789	0.366
OPT Smote Naive Bayes	0.754	0.810	0.860	0.835	0.361
OPT Adasyn Naive Bayes	0.735	0.827	0.800	0.813	0.360
OPT RUS Naive Bayes	0.725	0.835	0.770	0.801	0.360

Table 4.5: Gaussian Naive Bayes Results

Pipeline	Accuracy	Precision	Recall	F1-Score	MCC
Logistic Regression	0.770	0.792	0.923	0.853	0.366
Smote Logistic Regression	0.729	0.847	0.760	0.802	0.383
Adasyn Logistic Regression	0.729	0.850	0.757	0.801	0.387
RUS Logistic Regression	0.753	0.821	0.841	0.831	0.376
OPT Logistic Regression	0.770	0.792	0.922	0.852	0.364
OPT Smote Logistic Regression	0.752	0.822	0.837	0.830	0.377
OPT Adasyn Logistic Regression	0.738	0.840	0.785	0.812	0.384
OPT RUS Logistic Regression	0.756	0.820	0.846	0.833	0.379

Table 4.6: Logistic Regression Results

Pipeline	Accuracy	Precision	Recall	F1-Score	MCC
Linear SVM	0.775	0.788	0.939	0.857	0.374
Smote Linear SVM	0.727	0.847	0.758	0.800	0.380
Adasyn Linear SVM	0.729	0.851	0.756	0.801	0.388
RUS Linear SVM	0.754	0.818	0.846	0.832	0.373
OPT Linear SVM	0.776	0.789	0.940	0.858	0.376
OPT Smote Linear SVM	0.729	0.840	0.770	0.804	0.373
OPT Adasyn Linear SVM	0.750	0.828	0.824	0.826	0.382
OPT RUS Linear SVM	0.726	0.846	0.758	0.800	0.377

Table 4.7: Linear Support Vector Machine Results

Pipeline	Accuracy	Precision	Recall	F1-Score	MCC
DecisionTree	0.716	0.804	0.801	0.802	0.297
Smote DecisionTree	0.702	0.818	0.755	0.785	0.306
Adasyn DecisionTree	0.695	0.811	0.751	0.780	0.286
RUS DecisionTree	0.688	0.808	0.743	0.774	0.273
OPT DecisionTree	0.764	0.791	0.913	0.848	0.351
OPT Smote DecisionTree	0.704	0.868	0.696	0.772	0.384
OPT Adasyn DecisionTree	0.735	0.841	0.780	0.809	0.381
OPT RUS DecisionTree	0.742	0.844	0.787	0.814	0.394

Table 4.8: Decision Tree Results

Pipeline	Accuracy	Precision	Recall	F1-Score	MCC
Random Forest	0.741	0.808	0.840	0.824	0.337
Smote Random Forest	0.733	0.818	0.808	0.813	0.343
Adasyn Random Forest	0.733	0.821	0.804	0.813	0.348
RUS Random Forest	0.720	0.816	0.788	0.802	0.323
OPT Random Forest	0.786	0.802	0.933	0.863	0.416
OPT Smote Random Forest	0.770	0.828	0.859	0.843	0.412
OPT Adasyn Random Forest	0.762	0.834	0.835	0.835	0.407
OPT RUS Random Forest	0.752	0.844	0.805	0.824	0.408

Table 4.9: Random Forest Results

Pipeline	Accuracy	Precision	Recall	F1-Score	MCC
LightGBM	0.786	0.807	0.922	0.861	0.418
Smote LightGBM	0.760	0.825	0.845	0.835	0.392
Adasyn LightGBM	0.765	0.832	0.843	0.838	0.409
RUS LightGBM	0.765	0.825	0.855	0.840	0.400
OPT LightGBM	0.786	0.803	0.931	0.862	0.415
OPT Smote LightGBM	0.756	0.836	0.823	0.829	0.403
OPT Adasyn LightGBM	0.758	0.836	0.826	0.831	0.405
OPT RUS LightGBM	0.750	0.838	0.809	0.823	0.397

Table 4.10: Light Gradient Boosting Machine Results

Across models, predictive performance remains stable and consistently above trivial baselines, indicating that meaningful predictive signal is present in pre-declarative academic and administrative information. Linear classifiers exhibit stable behavior with competitive Recall and F_1-scores, while tree-based ensemble models achieve modest improvements in balanced performance, as reflected by higher MCC values. Accuracy is interpreted cautiously throughout, given the underlying class distribution.

Overall, these results establish that reliable generalization can be achieved under strictly pre-declarative constraints, without recourse to feature selection, post-declarative variables, or metaheuristic optimization. The reported performance therefore represents the attainable predictive capacity of interpretable models operating at the earliest feasible decision point in the institutional process.

Sensitivity to Class Imbalance

This subsection analyzes the sensitivity of predictive performance to class imbalance handling strategies under strictly pre-declarative constraints. The objective is not to optimize class distribution, but to assess how different controlled resampling variants influence model behavior and evaluation metrics when the underlying outcome is imbalanced.

Across all model families, imbalance-handling configurations based on random undersampling and synthetic over-sampling (SMOTE and ADASYN) exhibit systematic effects on class-wise performance. In general, these variants modify the trade-off between precision and

recall relative to the baseline configurations, with recall tending to increase under over-sampling strategies, while precision may decrease moderately. These effects are reflected consistently across the metrics reported in Tables 4.4–4.10.

The impact of imbalance handling is more pronounced in models with higher sensitivity to class distribution, such as Naive Bayes and Decision Trees, where resampling leads to noticeable shifts in recall, F_1-score, and MCC. In contrast, ensemble methods display greater robustness to resampling choices, maintaining comparatively stable balanced performance across variants. Importantly, no resampling strategy uniformly dominates across all models and metrics, underscoring the role of imbalance handling as a controlled experimental factor rather than as an optimization objective.

These results indicate that class imbalance handling affects predictive behavior in systematic but model-dependent ways under pre-declarative conditions. By explicitly incorporating and isolating these effects in Stage II, the analysis provides a transparent reference for evaluating more advanced balancing and optimization mechanisms introduced in subsequent stages.

Variability and Stability across Runs

Variability and stability of predictive performance in Stage II are assessed through repeated experimental runs conducted under controlled random seeds, following the robustness design outlined earlier. This procedure aims to evaluate whether observed performance patterns are consistent across different data partitions and resampling realizations, rather than being driven by incidental sampling effects.

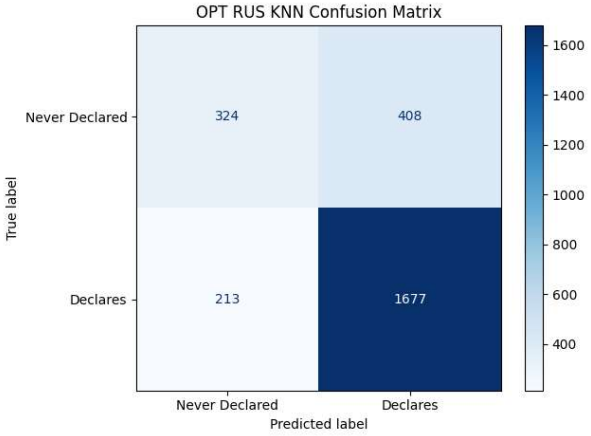
Across model families and experimental configurations, performance metrics exhibit limited dispersion between runs, indicating stable generalization behavior under strictly pre-declarative conditions. In particular, balanced metrics such as MCC and F_1-score remain within narrow ranges for a given model and configuration, suggesting that the extracted predictive signal is not highly sensitive to random variation in training data composition.

Models characterized by simpler decision structures, such as linear classifiers, show especially consistent behavior across runs, reflecting their reduced sensitivity to sampling fluctuations. Tree-based and ensemble models display slightly higher variability, as expected given their greater flexibility, but without evidence of unstable or degenerate solutions. Importantly, no configuration exhibits performance collapse or erratic behavior across repetitions.

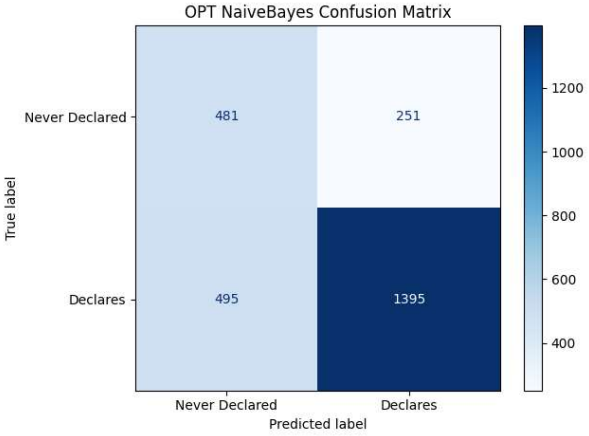
Overall, these observations support the robustness of the Stage II experimental design and indicate that the reported results reflect stable predictive patterns inherent to the pre-declarative institutional data, rather than artifacts of a particular data split or random initialization.

Class-wise Error Analysis

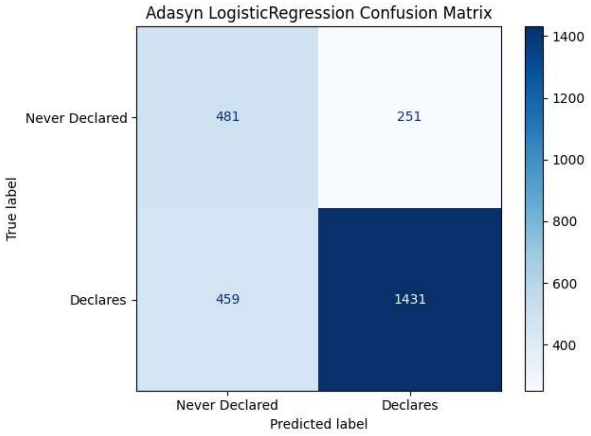
To complement aggregate performance metrics, a class-wise error analysis is conducted using confusion matrices corresponding to representative best-performing configurations for each model family. Figures 4.3 and 4.4 display these confusion matrices and provide a granular view of how models distinguish between borrowers who submitted their first income declaration and those who did not.



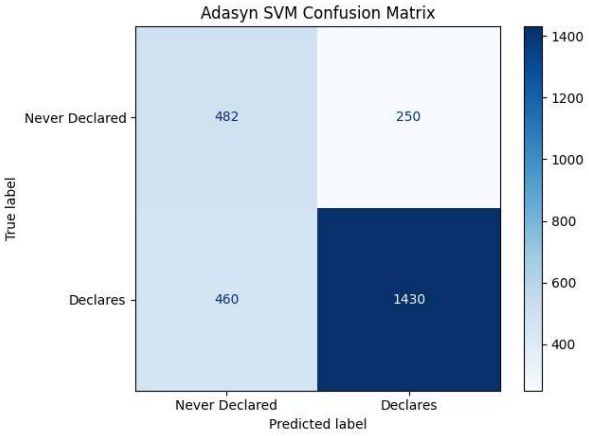
(a) Optimized HP KNN with RUS



(b) Optimized HP Naive Bayes

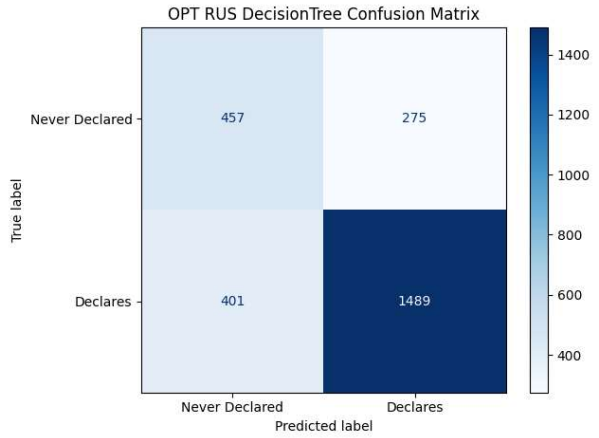


(c) Logistic Regresion with ADASYN

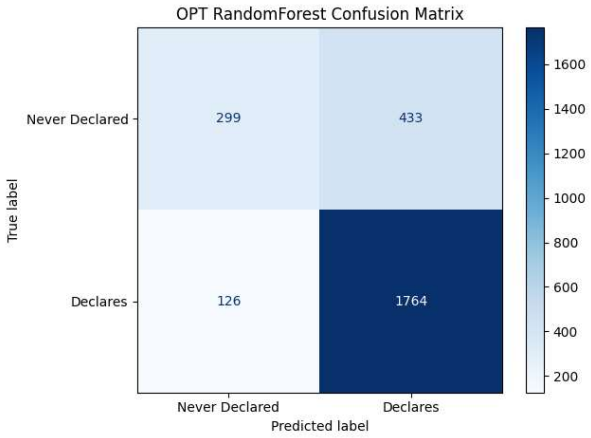


(d) Linear SVM with ADASYN

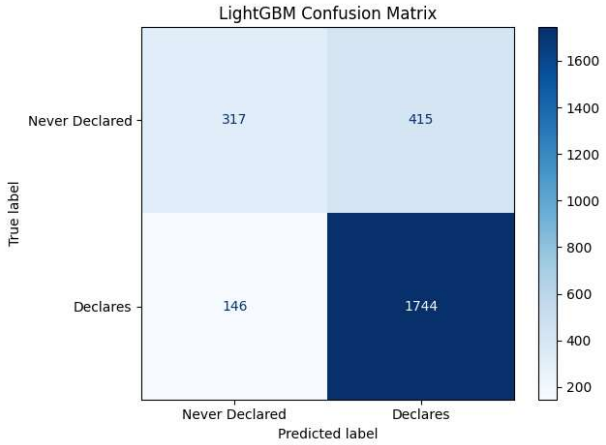
Figure 4.3: Confusion Matrices (Part I)



(a) Optimized HP DT with RUS



(b) Optimized HP Random Forest



(c) Base Light Gradient Boosting Machine

Figure 4.4: Confusion Matrices (Part II)

The confusion matrices reveal systematic differences in error profiles across model families. Several classifiers, including Naive Bayes, Logistic Regression, Linear SVM, and Decision Tree, exhibit a predominance of Type II errors, corresponding to false negatives where borrowers who eventually declared are predicted as non-declarants. This behavior is consistent with their recall values and reflects a more conservative classification tendency under class imbalance.

In contrast, instance-based and ensemble models such as KNN, Random Forest, and LightGBM show a greater inclination toward Type I errors, corresponding to false positives where borrowers are predicted to declare when they ultimately do not. While this tendency slightly reduces precision, it preserves higher recall and prevents systematic omission of true declarants. As a result, these models achieve more balanced error distributions, which is reflected in their comparatively higher MCC values.

From an analytical perspective, the observed asymmetry between Type I and Type II errors highlights an inherent trade-off in early-stage predictive modeling under uncertainty. Configurations biased toward Type I errors prioritize sensitivity to potential non-compliance at

the expense of additional false alarms, whereas models dominated by Type II errors achieve higher precision but risk missing true positive cases. Within Stage II, this analysis contextualizes metric-based comparisons and clarifies how different model families operationalize risk discrimination under strictly pre-declarative constraints.

4.2.7 Model Interpretability Analysis

This section analyzes the interpretability of the predictive models evaluated in Stage II, with the objective of identifying which pre-declarative institutional variables systematically drive model behavior and how these variables influence predictions across different modeling paradigms. Interpretability is treated as a primary analytical dimension, on equal footing with predictive performance, given the institutional requirement for transparency, auditability, and explanatory coherence at early decision stages.

The analysis integrates multiple complementary explanation perspectives to avoid reliance on a single interpretability mechanism. Specifically, global model-agnostic relevance is examined through permutation feature importance, model-specific contributions are analyzed using coefficients and tree-based importance measures, rule-based reasoning is illustrated through decision tree structures and representative paths, and case-level explanations are provided using SHAP attributions for an ensemble model. Figures and tables referenced in the following subsections (e.g., Figures 4.5, 4.6, 4.7, 4.8–4.10, and 4.11, as well as Table 4.11) support this multi-view interpretability assessment.

By examining convergence and consistency across these explanation paradigms, Stage II aims to establish whether interpretability conclusions are robust to model choice and explanation technique. This approach enables transparent identification of institutionally meaningful predictors under strictly pre-declarative constraints, while providing a stable explanatory baseline for comparison with subsequent stages that introduce higher model complexity and optimization mechanisms.

Permutation Feature Importance

Permutation Feature Importance (PFI) is used as a global, model-agnostic mechanism to assess the relative contribution of each pre-declarative variable to predictive performance across the evaluated models. This analysis evaluates the impact of randomly permuting individual features on model predictions, thereby quantifying their importance independently of any specific model structure.

Figure 4.5 reports the averaged permutation importance scores aggregated across all trained models in Stage II. The results show a consistent dominance of a small subset of predictors. In particular, *deud_monto* (total loan amount) and *conteo_matr* (total number of enrollments) emerge as the most influential variables, followed by *estado_civil* (marital status) and *anio_ exigibilidad* (year of enforceability). This ranking indicates that financial exposure, academic trajectory, and basic demographic structure concentrate the majority of the predictive signal under strictly pre-declarative conditions.

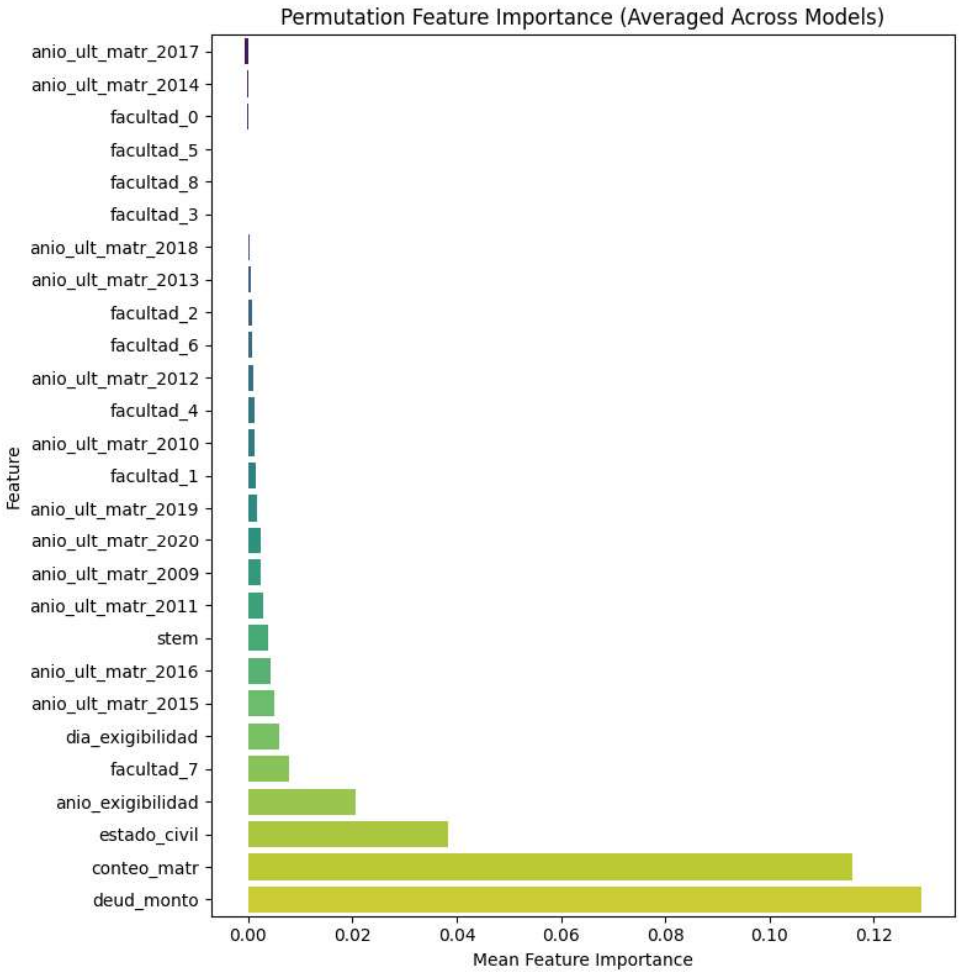


Figure 4.5: Average permutation feature importance across all trained models.

The remaining variables exhibit progressively lower importance values. Most faculty-related indicators contribute marginally to overall discrimination, with the notable exception of the dummy corresponding to the Faculty of Law, which consistently appears among the higher-ranked features. For clarity and institutional interpretability, Table 4.11 provides the mapping between faculty dummy codes and their corresponding institutional units.

Dummy Feature	Real Value
0	Ecclesiastical Faculty of Theology
1	Faculty of Sciences
2	Faculty of Philosophy and Education
3	Faculty of Economic and Administrative Sciences
4	Faculty of Engineering
5	Faculty of Marine and Geographical Sciences
6	Faculty of Agronomic and Food Sciences
7	Faculty of Law
8	Faculty of Architecture and Urbanism

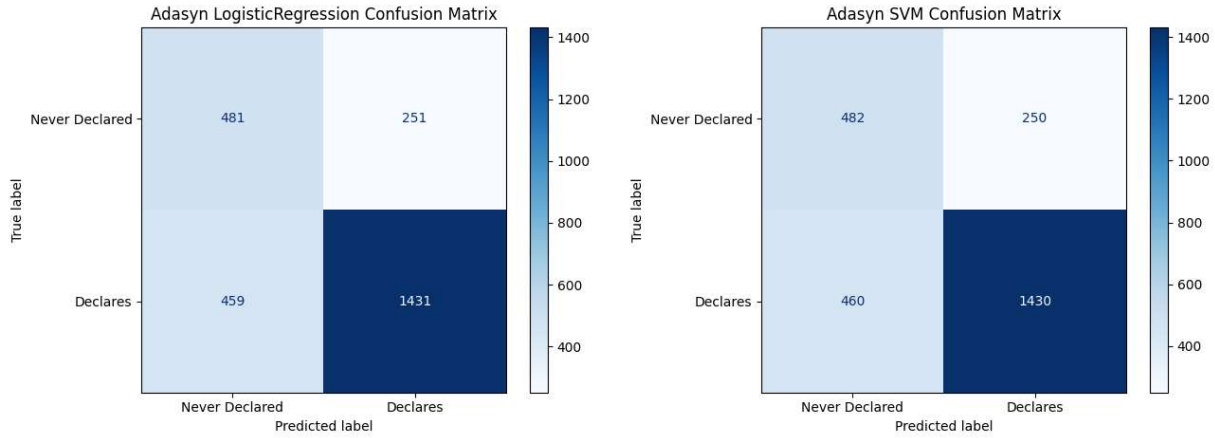
Table 4.11: Faculty dummy feature values used to interpret faculty-related predictors.

At the lower end of the importance distribution, slightly negative PFI values are observed for some variables. Given their small magnitude and the known sensitivity of permutation importance to sampling noise and feature collinearity, these values are interpreted as negligible rather than as evidence supporting variable exclusion. Overall, the PFI analysis provides a stable, model-agnostic view of feature relevance that supports subsequent, model-specific interpretability assessments.

Model-Specific Feature Contributions

To complement the global relevance analysis provided by permutation feature importance, model-specific feature contribution profiles are examined for representative tuned configurations of linear and tree-based models. This analysis focuses on how different model families internalize and weight pre-declarative variables within their respective decision mechanisms.

For linear models, coefficient-based importance profiles are reported for Logistic Regression and Linear Support Vector Machine configurations, as illustrated in Figure 4.6. Under the standardized feature representation, *estado_civil* exhibits a dominant marginal contribution, indicating a strong linear association with declaration compliance. Academic and institutional indicators, including selected *anio_ult_matr* categories, *stem*, and faculty-related dummies (notably *facultad_7*), provide secondary discriminatory information. The consistency of coefficient signs and relative magnitudes across linear models suggests stable linear effects within the pre-declarative feature space.

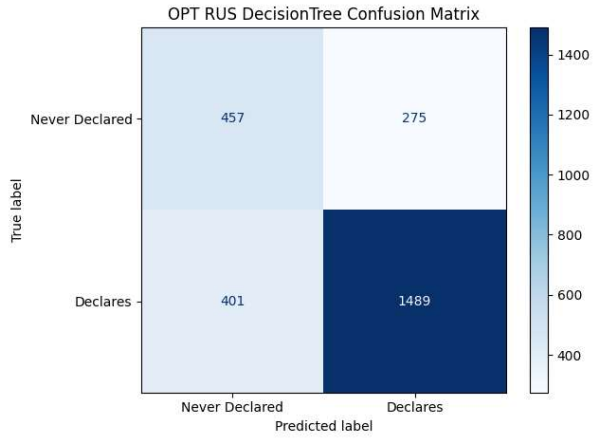


(a) Logistic Regression with ADASYN feature importance

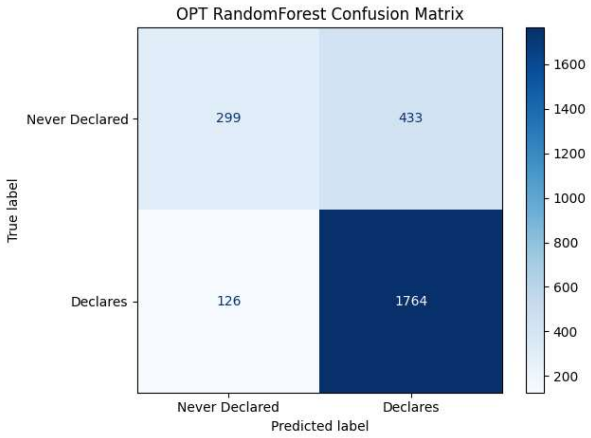
(b) Linear SVM with ADASYN feature importance

Figure 4.6: Linear models feature importance (tuned configurations).

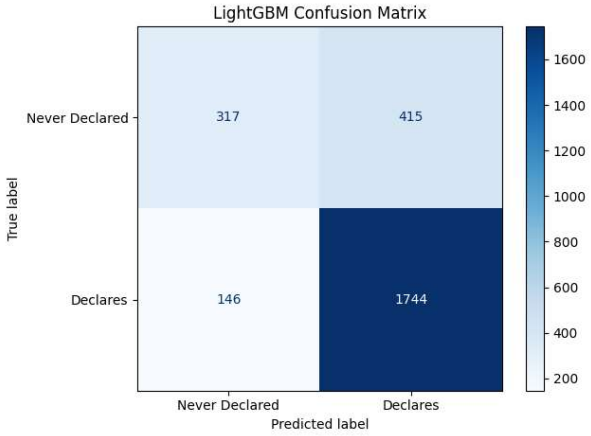
For tree-based models, feature importance profiles derived from Decision Tree, Random Forest, and LightGBM configurations are summarized in Figure 4.7. Across these models, importance rankings are highly consistent and closely aligned with the global PFI results. Decision Tree models prioritize *conteo_matr* and *deud_monto* as primary splitting variables, while ensemble aggregation in Random Forest and LightGBM increases the relative contribution of financial exposure (*deud_monto*) while preserving enrollment history (*conteo_matr*) among the strongest predictors. The recurrent presence of *anio_exigibilidad* among the top-ranked variables highlights the relevance of repayment timeline information within the pre-declarative regime.



(a) Decision Tree with random under-sampling feature importance



(b) Random Forest feature importance



(c) LightGBM feature importance

Figure 4.7: Tree-based models feature importance (tuned configurations).

Overall, the strong agreement between model-specific importance patterns and the global permutation analysis indicates that explanatory conclusions are robust across modeling paradigms. A compact and institutionally meaningful subset of predictors governs model behavior regardless of whether linear or nonlinear decision structures are employed, reinforcing the interpretability and coherence of Stage II results.

Decision Tree-Based Interpretations

Decision tree–based interpretations are used to illustrate how predictive structure can be translated into explicit and auditable decision rules under strictly pre-declarative constraints. This analysis focuses on representative Decision Tree configurations and complements the global and model-specific importance assessments by providing human-readable rule extraction.

Figures 4.8–4.10 display the same Decision Tree model visualized at increasing maximum

depths ($d = 4$, $d = 5$, and $d = 11$). Shallow representations emphasize a limited number of high-yield splits with broad population coverage, while deeper trees introduce finer partitions that increase structural complexity and reduce direct interpretability. Across all depths, early splits consistently involve the same dominant variables identified in previous analyses, notably *estado_civil*, *deud_monto*, *conteo_matr*, and *anio_exigibilidad*. This recurrence indicates that rule-based structures are driven by stable predictive signals inherent to the pre-declarative feature space.

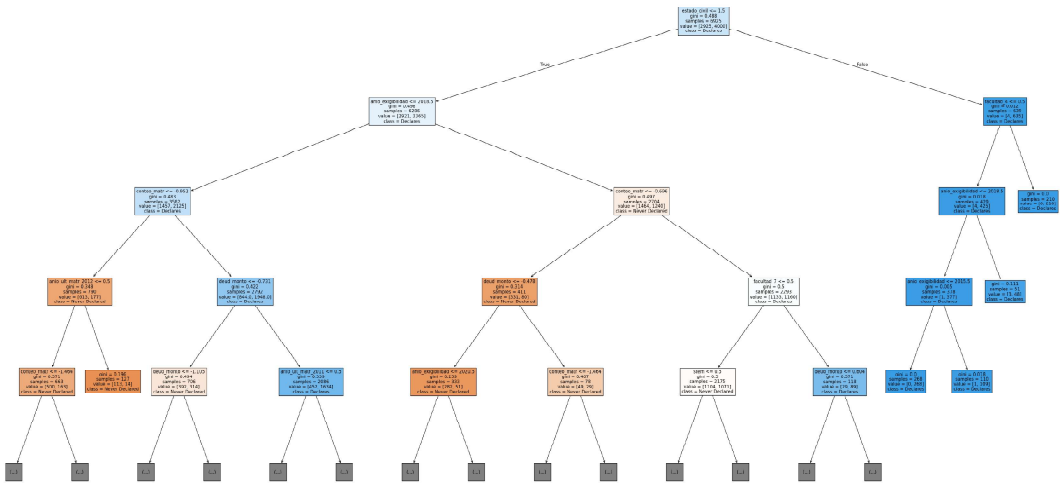


Figure 4.8: Decision Tree snapshots at different depths; Depth = 4.

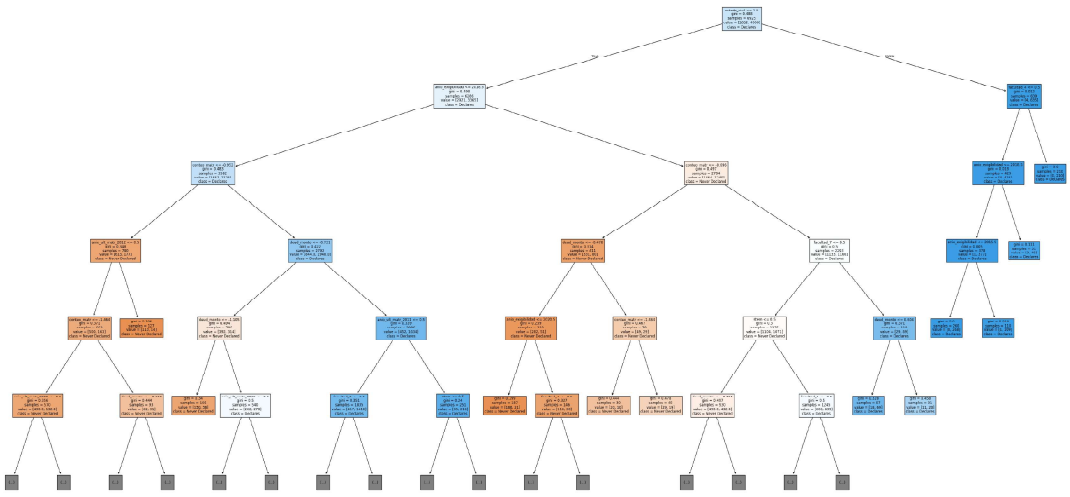


Figure 4.9: Decision Tree snapshots at different depths; Depth = 5.

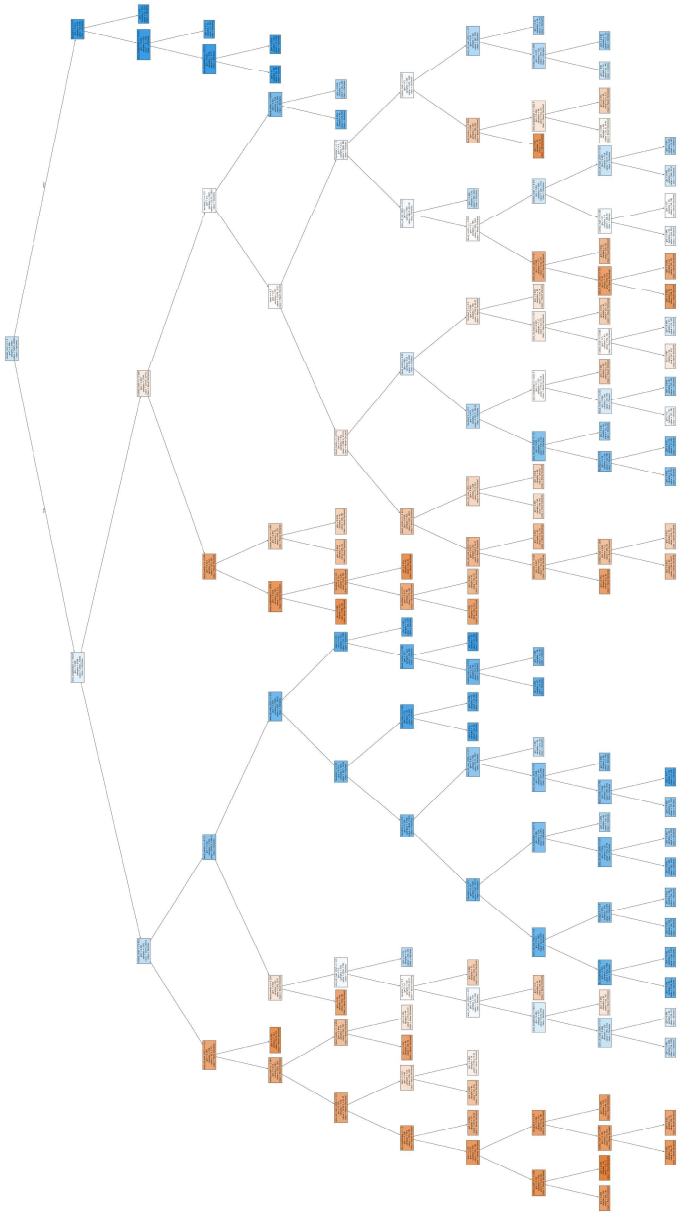


Figure 4.10: Decision Tree snapshots at different depths; Depth = 11.

To provide a concrete and auditable example of rule-based reasoning, Table 4.12 reports a representative decision path extracted from the depth-4 tree. The sequence of split conditions illustrates how demographic status, repayment timing, academic continuity, and financial exposure are jointly combined to yield a classification outcome. The reported terminal-node probability corresponds to the empirical class proportion observed at that node and is interpreted as an uncalibrated leaf frequency rather than as a probabilistic forecast.

Observed value	**Split condition**	**Feature meaning**	**Branch taken**
1	$estado_civil \leq 1.5$	Borrower is single or without dependents	**True (left branch)**
2019	$anio_exigibilidad \leq 2018.5$	Loan enforceability year (2019)	**False (right branch)**
-0.5	$conteo_matr \leq -0.696$	Total number of enrollments (standardized)	**True (left branch)**
-0.4	$deud_monto \leq -0.478$	Total debt amount (standardized)	**True (left branch)**
2019	$anio_exigibilidad \leq 2020.5$	Loan enforceability year (2021)	**End branch**
Predicted class: *Never Declared (estimated probability $\approx$ 0.85, 282 cases of no declaration over 333 total in this node)*			

Table 4.12: Example decision path from the optimized Decision Tree (depth = 4)

Overall, the decision tree–based interpretations demonstrate that interpretable rule structures derived from pre-declarative data align closely with the variable importance patterns observed across models. This alignment supports the transparency and verifiability of Stage II models and confirms that extracted decision logic reflects institutionally meaningful relationships rather than artifacts of model complexity.

SHAP-Based Attribution Analysis

To complement global and model-specific interpretability analyses with case-level explanations, SHAP (SHapley Additive exPlanations) values are examined for the Light Gradient Boosting Machine model. SHAP provides locally additive attributions that quantify the contribution of each feature to individual predictions, while remaining consistent with the overall model output.

Figure 4.11 presents the SHAP summary plot for the evaluated LightGBM configuration. The distribution of SHAP values confirms strong alignment with the global permutation importance and tree-based feature rankings. In particular, *deud_monto*, *conteo_matr*, *estado_civil*, and *anio_exigibilidad* exhibit the largest attribution magnitudes, indicating that these variables dominate both aggregate discrimination and individual prediction adjustments.

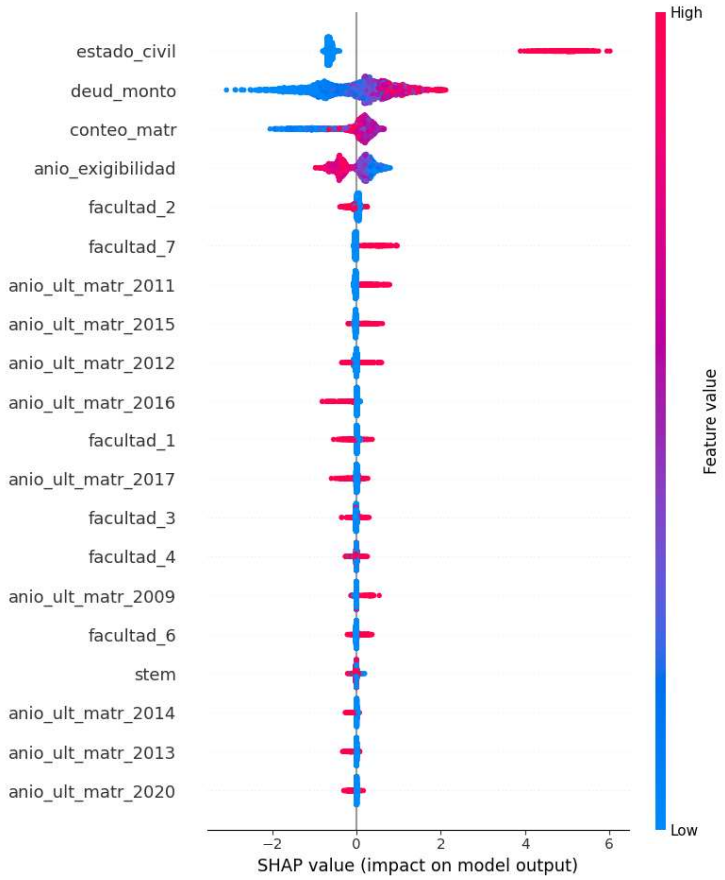


Figure 4.11: SHAP values for LightGBM.

The SHAP analysis further illustrates how variations in feature values influence the direction and magnitude of predictions at the individual level. Higher financial exposure and extended academic trajectories are associated with systematic shifts in predicted compliance behavior, while demographic and temporal indicators modulate predictions within narrower ranges. Importantly, no contradictory attribution patterns are observed relative to the global and model-wise analyses.

Overall, SHAP-based attribution reinforces the interpretability conclusions of Stage II by demonstrating that the same compact subset of pre-declarative variables governs predictive behavior across global, model-specific, and local explanation perspectives. This convergence supports the robustness and coherence of interpretability findings under strictly pre-declarative modeling constraints.

4.2.8 Comparative Positioning with Subsequent Stages

Stage II establishes a controlled and interpretable empirical reference for predictive modeling under strictly pre-declarative constraints, as defined in the methodological framework. By restricting inputs to academic and administrative information available prior to any declaration event, and by explicitly excluding feature selection, metaheuristic optimization, and post-declarative

variables, this stage isolates predictive behavior under conditions that preserve causal precedence, temporal validity, and institutional plausibility.

Rather than defining theoretical performance limits, Stage II characterizes the level of predictive capacity and explanatory coherence that can be empirically achieved at the earliest feasible decision point within the FSCU process under realistic informational constraints. Its results therefore provide an operationally grounded benchmark, reflecting what can be learned from antecedent institutional data without increasing model or data complexity.

Subsequent stages build explicitly upon this reference. In particular, Stage III relaxes several of the constraints imposed in Stage II by incorporating declaration-generated information and higher-dimensional feature spaces, together with metaheuristic-based feature selection mechanisms. Cross-stage comparisons therefore focus on assessing the marginal contribution of enriched data representations and optimization-driven search strategies relative to the causally grounded and interpretable baseline established in Stage II.

Within the stage-based methodological design of the thesis, Stage II functions as the interpretability and causal anchor of the empirical analysis. Its results enable rigorous evaluation of the benefits, trade-offs, and potential risks introduced by increased dimensionality, optimization, and model complexity in later stages, while maintaining a stable reference aligned with institutional decision-making requirements.

4.2.9 Summary of Stage II Results

Stage II provides the first consolidated empirical evaluation of predictive modeling for first-time income declaration compliance under strictly pre-declarative conditions, directly addressing the early-stage prediction problem and the corresponding research question defined in the methodological framework. Unlike the exploratory analyses conducted in Stage I, which primarily validated predictive feasibility and data readiness, this stage focuses on a single, well-defined predictive phenomenon evaluated on a homogeneous cohort of borrowers facing their initial declaration obligation.

The results demonstrate that interpretable machine learning models, calibrated using standard hyperparameter tuning procedures and trained exclusively on antecedent academic and administrative data, are capable of extracting stable and meaningful predictive signal. Balanced performance metrics indicate reliable discrimination despite class imbalance and limited information availability, providing empirical support for Research Question RQ1 and Hypothesis H1. Importantly, these outcomes are achieved without the use of feature selection procedures, metaheuristic optimization, or declaration-generated variables, in full alignment with the methodological design of this stage.

From an interpretability perspective, the analyses converge consistently on a compact and institutionally meaningful subset of predictors associated with financial exposure, academic trajectory, and basic demographic structure. The stability of these variables across global, model-specific, rule-based, and local explanation techniques supports the robustness, transparency, and causal plausibility of the extracted insights.

Overall, Stage II establishes a transparent, causally coherent, and empirically grounded reference for early-stage predictive modeling within the FSCU system. This reference point provides

the baseline against which subsequent stages evaluate the incremental value of increased data richness, feature dimensionality, and optimization strategies, while preserving methodological rigor and interpretative clarity as required by the stage-based research design.

4.3 Metaheuristic Optimization for Predictive Modeling of Declaration Abandonment (Stage III)

4.3.1 Stage III Overview

Stage III focuses on the predictive modeling of income declaration abandonment among borrowers who have previously complied with declaration requirements, under conditions of increased data availability and feature dimensionality. This stage represents a methodological transition from the interpretable, pre-declarative modeling setting explored in Stage II to a post-declarative scenario in which richer administrative, academic, and declaration-derived information becomes available.

The central methodological challenge addressed in this stage arises from the high dimensionality and heterogeneity of the feature space associated with declaration history and institutional records. While the incorporation of post-declarative information increases the potential predictive signal, it also introduces risks related to overfitting, reduced model stability, and diminished interpretability. In this context, Stage III incorporates metaheuristic-based feature selection as a core methodological component, treating optimization not as an auxiliary enhancement but as a necessary design mechanism for managing complex and high-dimensional representations.

From a research perspective, this stage directly addresses Research Questions RQ2 and RQ3 by evaluating (i) the extent to which declaration abandonment can be anticipated using enriched institutional data, and (ii) how optimization-driven feature selection influences predictive performance, robustness, and interpretability. Accordingly, Hypotheses H2 and H3 are assessed within a controlled experimental framework designed to isolate the effects of feature selection from other sources of performance variation.

Methodologically, Stage III builds upon the modeling and evaluation protocols established in earlier stages, ensuring continuity in data preprocessing, temporal validation, and metric computation. At the same time, it extends the analytical scope by introducing explicit optimization processes aimed at identifying parsimonious and informative subsets of variables from high-dimensional feature spaces. This design enables a systematic examination of trade-offs between predictive accuracy, feature stability, and model transparency under operationally realistic conditions.

Overall, Stage III constitutes a pivotal step in the thesis by empirically grounding the role of metaheuristic optimization in institutional compliance prediction. The results obtained in this stage provide both standalone evidence on declaration abandonment modeling and a critical methodological reference point for subsequent comparisons with more expressive modeling paradigms explored in Stage IV.

4.3.2 Problem Definition and Experimental Scope

In Stage III, the predictive problem is defined as the anticipation of income declaration abandonment among borrowers who have previously complied with declaration requirements. Unlike the first-time compliance scenario addressed in Stage II, this stage focuses on a post-declarative context in which historical declaration behavior, administrative records, and enriched institutional information are available at prediction time. The target variable represents a binary outcome indicating whether a borrower who has declared income at least once subsequently discontinues the declaration process within the observation window defined by the institutional rules of the FSCU system.

From an experimental standpoint, this stage operates under relaxed informational constraints compared to earlier stages. The feature space incorporates academic, financial, administrative, and declaration-derived variables, resulting in a high-dimensional representation that reflects both structural borrower characteristics and dynamic declaration behavior. While this enriched representation increases the potential for predictive accuracy, it also amplifies challenges related to redundancy, noise, and instability, which motivate the explicit incorporation of feature selection mechanisms.

The experimental scope of Stage III is deliberately designed to isolate the contribution of metaheuristic-based feature selection within the predictive modeling pipeline. To this end, the prediction task, dataset partitions, temporal ordering, and evaluation metrics remain consistent with those defined in the general methodological framework. This controlled design ensures that observed differences in performance, robustness, or interpretability can be attributed to the introduction of optimization-driven feature selection rather than to changes in data handling or validation procedures.

Predictive models are evaluated under repeated stratified validation schemes in order to account for class imbalance and to assess performance stability across partitions. The assessment emphasizes not only standard discrimination metrics, but also robustness-related indicators and feature-level outcomes, reflecting the institutional requirement for reliable and interpretable decision-support tools. Feature selection results are therefore analyzed jointly with predictive performance, rather than treated as an auxiliary or purely technical outcome.

In summary, Stage III frames declaration abandonment prediction as a high-dimensional institutional risk modeling problem in which optimization plays a central role. The experimental scope is structured to provide empirical evidence on how metaheuristic-driven feature selection affects predictive behavior, model stability, and the identification of informative institutional signals, setting the foundation for both cross-stage comparison and subsequent exploration of more expressive modeling paradigms.

4.3.3 Data Preparation and Experimental Protocol

The data preparation process and experimental protocol adopted in Stage III are grounded in the institutional data pipeline previously established in Stages II. Nevertheless, this stage addresses a distinct predictive problem and relies on an expanded dataset that incorporates post-declarative information, requiring additional preprocessing and validation considerations.

Data Sources and Cohort Definition.

The dataset used in Stage III is constructed from the FSCU administrative records provided by the Finance Department of the PUCV. As described in the Materials and Methods section, the source database comprises multiple relational tables capturing borrower information, loan characteristics, academic records, and income declaration histories. The analysis is restricted to obligations enforceable from 2012 onward, following the legislative reform that standardized the annual income declaration process.

Unlike Stage II, which focuses on first-time declaration compliance, Stage III targets declaration abandonment among borrowers who have previously submitted at least one income declaration. The working dataset is therefore built at the borrower level using information available up to the time of the last submitted declaration. The binary target variable indicates whether a borrower continues declaring income (0) or abandons the declaration process (1).

Feature Construction and Dataset Composition.

The final feature set combines demographic, academic, financial, and declaration-derived variables. In addition to baseline institutional attributes, this stage incorporates variables summarizing historical income declaration behavior, such as counts, averages, temporal differences, and variability measures. A detailed overview of the resulting feature space, including variable names, data types, and value ranges, is provided in Table 4.13, while concise semantic descriptions are reported in Table 4.14.

Name	Data Type	Feature Type	Detail
estado_civil	Boolean	Categorical	1 and 2
nacionalidad	Boolean	Categorical	1 and 2
sexo	Boolean	Categorical	1 and 2
fecha_nacimiento	Date	Date	1900-01-01 to 1991-03-26
deud_t_deuda	Integer	Categorical	1 value
deud_e_deuda	Integer	Categorical	7 values
deud_monto	Float	Numerical	1.50 to 1285.76
deud_monto_cuota_fija	Float	Numerical	0.01 to 84.79
deud_fecha_exigibilidad	Date	Date	2012-01-01 to 2023-01-01
deud_numero_cuotas_fija	Integer	Numerical	0 to 15
paga_t_pagare	Integer	Categorical	1 value
paga_e_pagare	Integer	Categorical	1 value
monto_total_pagare	Float	Numerical	1.37 to 922.22
conteo_pagare	Integer	Numerical	1 to 29
num_declaraciones	Integer	Numerical	1 to 15
dejo_declarar	Boolean	Categorical	0 and 1
anio_ult_declaracion	Integer	Categorical	13 values
dein_c_afp	Integer	Categorical	12 values
dein_estado_civil	Boolean	Categorical	1 and 2

Table 4.13 continued from previous page

Name	Data Type	Feature Type	Detail
dein_hijo	Boolean	Categorical	0 and 1
dein_c_inst_conyuge_deud	Integer	Categorical	18 values
dein_c_afp_conyuge	Integer	Categorical	10 values
dein_total_deudor_2	Float	Numerical	0 to 4239.44
dein_total_conyuge_2	Float	Numerical	0 to 2487.19
deuda_conyuge	Boolean	Categorical	0 and 1
anio_primera_declaracion	Integer	Categorical	16 values
proporcion_ingreso_cuota	Float	Numerical	-1 to 242.16
prom_proporcion_ingreso_deuda	Float	Numerical	0 to 101.57
std_proporcion_ingreso_deuda	Float	Numerical	0 to 14.92
dif_ingresos_declarados	Float	Numerical	-580.04 to 2582.33
prom_deudor	Float	Numerical	0 to 4239.449
std_deudor	Float	Numerical	0 to 1130.33
prom_conyuge	Float	Numerical	0 to 2469.92
std_conyuge	Float	Numerical	0 to 1785.70
cantidad_cambios_estado_civil	Integer	Numerical	0 to 6
cuot_saldo	Float	Numerical	0 to 84.79
cuot_e_cuota	Integer	Categorical	5 values
avg_valor_cuota	Float	Numerical	0 to 80.36
conteo_cuota	Integer	Numerical	1 to 26
prom_cuota	Float	Numerical	0.28 to 80.19
std_cuota	Float	Numerical	0.12 to 33.09
bole_monto_deuda_prom	Float	Numerical	0.03 to 8.05
bole_count_prom	Float	Numerical	6.4 to 10
conteo_boletin	Integer	Numerical	10 to 366
matr_anio	Integer	Categorical	13 values
matr_c_carrera	Integer	Categorical	74 values
matr_e_matricula	Integer	Categorical	1 value
carr_t_carrera	Integer	Categorical	1 value
matr_c_institucion	Integer	Categorical	1 value
conteo_matr	Integer	Numerical	1 to 31
facultad	Integer	Categorical	9 values
escuela	Integer	Categorical	34 values

Table 4.13: Final feature set: variable names, data types, and ranges

Name	**Detail**
estado_civil	Last known marital status of the debtor. It can take the following values: 1 not married, 2 married

Table 4.14 continued from previous page

Name	Detail
nacionalidad	Whether the debtor is Chilean or foreign. 1 means Chilean, 2 mean foreign
sexo	Gender of the debtor. M means male and F means Female
fecha_nacimiento	Birth date of the debtor
deud_t_deuda	If it a loan shared with another institution
deud_e_deuda	Debt statement
deud_monto	Total loan amount
deud_monto_cuota_fija	Fixed installment amount
deud_fecha_exigibilidad	Date of enforceability of the loan
deud_numero_cuotas_fija	Number of fixed installments
paga_t_pagare	Type of promissory note. 1 means normal promissory note
paga_e_pagare	Promissory note statement. 1 means valid
monto_total_pagare	Total promissory note amount
conteo_pagare	Total promissory note count
num_declaraciones	Total income declarations counts
dejo_declarar	Whether the debtor stopped declaring their income or not. 0 means they keep declaring their income, 1 means they stopped
anio_ult_declaracion	Year of the last income declaration presented
dein_c_afp	Pension fund administrator code
dein_estado_civil	Marital status of the debtor in their last income declaration. 1 means not married, 2 means married
dein_hijo	Whether the debtor has children or not. 0 means they do not have children, 1 means that they have children
dein_c_inst_conyuge_deud	Institution code of the spouse of the debtor
dein_c_afp_conyuge	Pension fund administrator of the spouse of the debtor
dein_total_deudor_2	Total declared income of the debtor (in UTM)
dein_total_conyuge_2	Total declared income of the spouse of the debtor (in UTM)
deuda_conyuge	Whether the spouse of the debtor has a FSCU loan or not. 0 means they do not have a loan, 1 means they have.
anio_primera_declaracion	Year of the first income declaration of the debtor
proporcion_ingreso_cuota	Ratio of the average income of the debtor to the average installment amount
prom_proporcion_ingreso_deuda	Average ratio of the declared income of the debtor to the total amount of debt
std_proporcion_ingreso_deuda	Standard deviation of the ratio of the declared income of the debtor to the total amount of debt
dif_ingresos_declarados	Difference of declared income between the first and last income declaration of the debtor
prom_deudor	Average declared income of the debtor
std_deudor	Standard deviation of the declared income of the debtor
prom_conyuge	Average declared income of the spouse of the debtor

Table 4.14 continued from previous page

Name	Detail
std_conyuge	Standard deviation of the declared income of the spouse of the debtor
cantidad_cambios_estado_civ	Number of changes in the marital status of the debtor through their income declarations
cuot_saldo	Balance of the last installment of the loan
cuot_e_cuota	Status of the last installment of the loan
avg_valor_cuota	Average effective installment amount
conteo_cuota	Number of effective installments
prom_cuota	Average installment amount
std_cuota	Standard deviation of the installments amounts
bole_monto_deuda_prom	Average sub-installment amount
bole_count_prom	Average number of sub-installments per installment
conteo_boletin	Number of sub-installments
matr_anio	Year of the last college enrollment of the debtor
matr_c_carrera	Code of the degree program covered by the loan
matr_e_matricula	Status of the last college enrollment of the debtor. 1 means the enrollment has a valid status
carr_t_carrera	Type of degree program in the last college enrollment of the debtor. 1 means undergraduate program
matr_c_institucion	Institution code in the last college enrollment of the debtor
conteo_matr	Total amount of enrollments of the debtor within the degree program covered by the loan
facultad	Faculty of the degree program
escuela	School of the degree program
steam	Whether the degree program covered by the loan is a STEM one or not. 1 means the degree program is a STEM program, 0 means it is not

Table 4.14: Feature set: detailed descriptions of each variable

This expanded feature space increases both dimensionality and heterogeneity, motivating the subsequent use of metaheuristic-based feature selection to manage redundancy and improve predictive robustness.

Data Cleaning and Missing Value Treatment.

An initial assessment of data completeness is presented in Figure 4.12, which highlights the distribution of missing values across all extracted features. As summarized in Table 4.15, seven variables exhibit missing observations. Features with extensive and overlapping missingness (*sexo* and *fecha_nacimiento*) are removed from the dataset. Other missing values are treated according to their semantic origin: variables whose missingness arises from single-declaration

cases or structural zeros are imputed with zero, while features dominated by zero or near-zero values and limited variability are discarded.

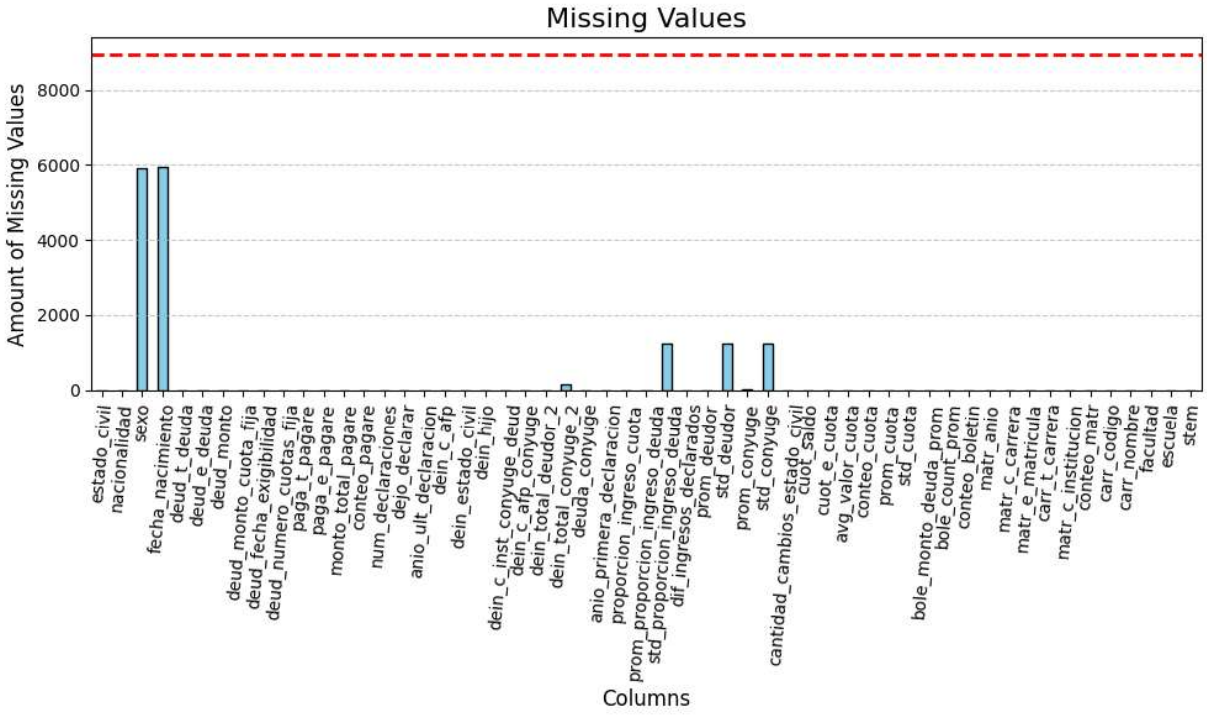


Figure 4.12: Missing values across dataset features

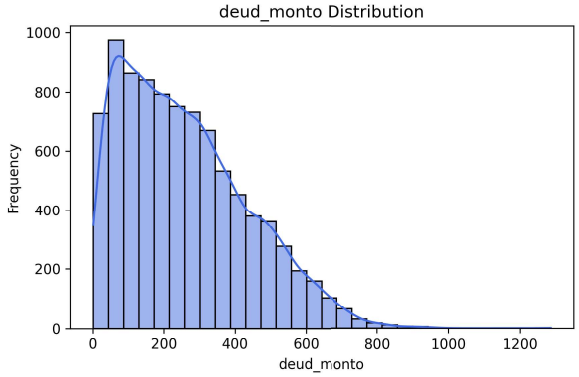
Column Name	Missing Values
sexo	5925
fecha_nacimiento	5949
dein_total_conyuge_2	144
std_proporcion_ingreso_deuda	1233
std_deudor	1233
prom_conyuge	11
std_conyuge	1237

Table 4.15: Features with missing values

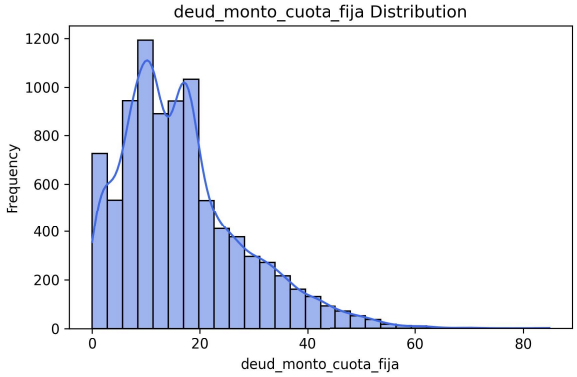
Categorical Consolidation and Redundancy Screening.

The distribution of numerical and categorical features is examined through a comprehensive set of histograms and count plots (Figures 4.13–4.16 and Figures 4.17–4.20). Based on these

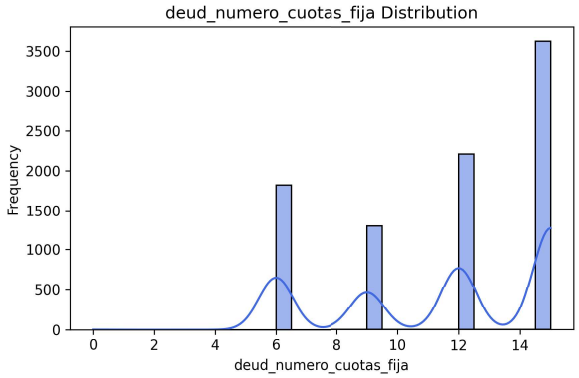
distributions, invariant and quasi-invariant categorical variables are removed, and rare categories are consolidated to reduce sparsity and cardinality.



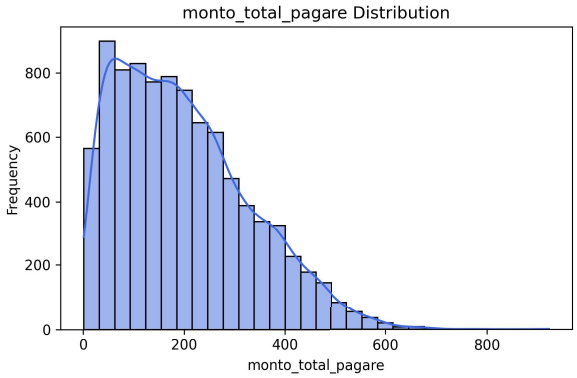
(a) deud_monto Distribution



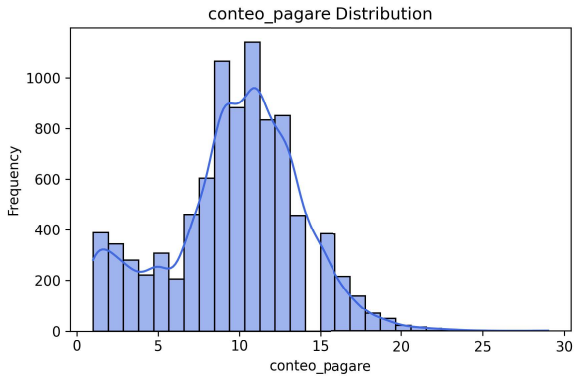
(b) deud_monto_cuota_fija Distribution



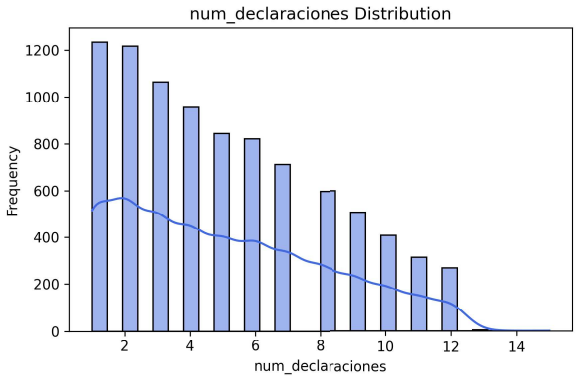
(c) deud_numero_cuotas_fijas Distribution



(d) monto_total_pagare Distribution

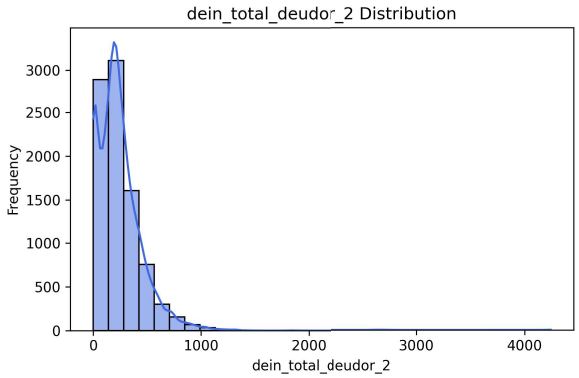


(e) conteo_pagare Distribution

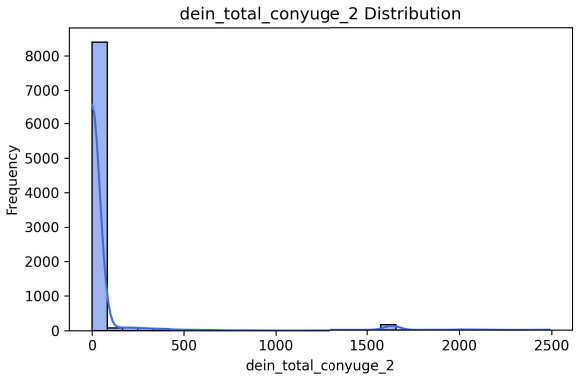


(f) num_declaraciones Distribution

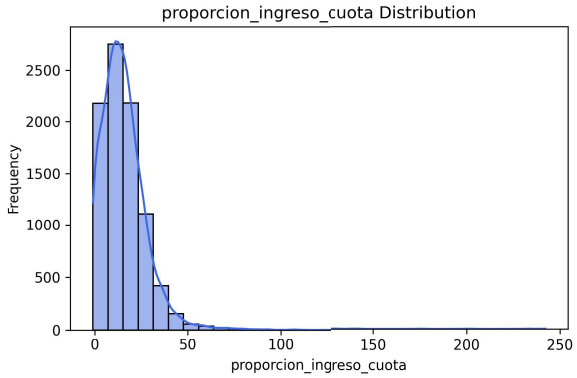
Figure 4.13: Distribution of selected numerical features (Part I)



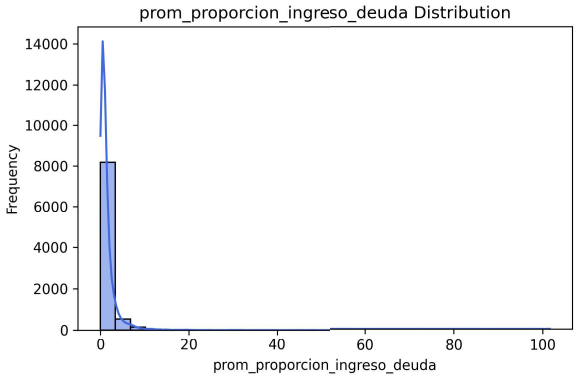
(a) dein_total_deudor_2 Distribution



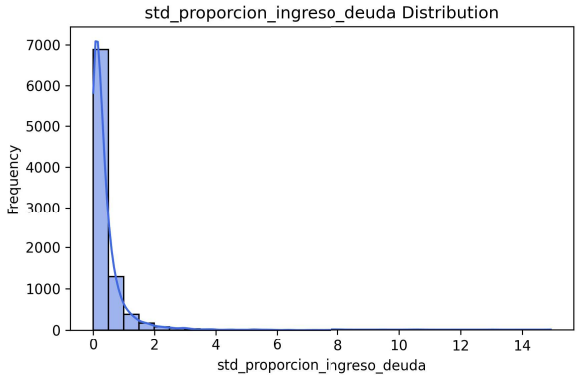
(b) dein_total_conyuge_2 Distribution



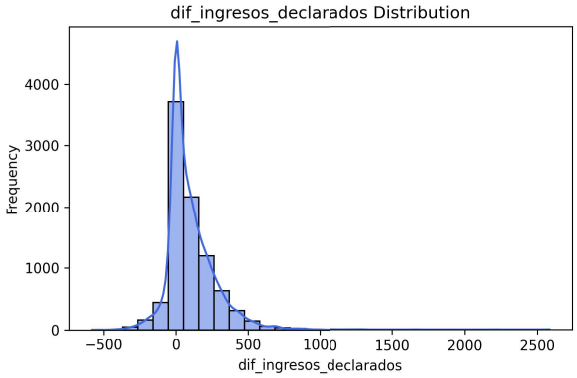
(c) proporcion_ingreso_cuota Distribution



(d) prom_proporcion_ingreso_deuda Distribution

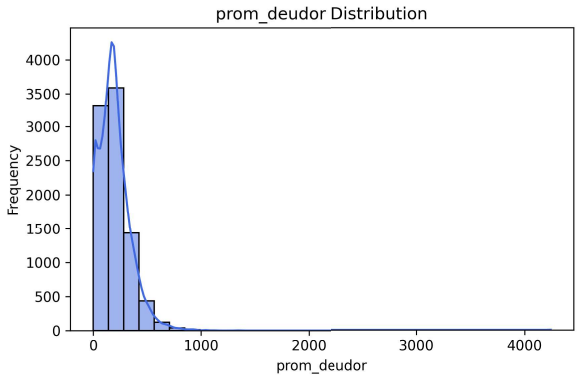


(e) std_proporcion_ingreso_deuda Distribution

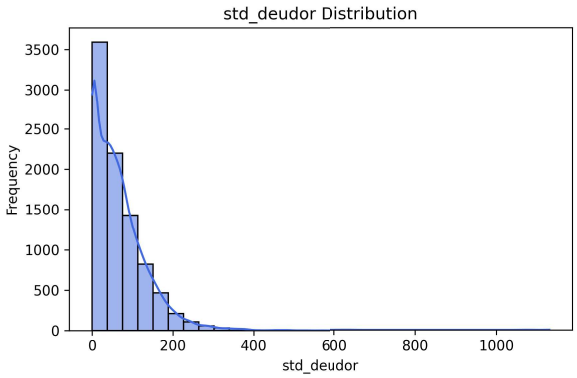


(f) dif_ingresos_declarados

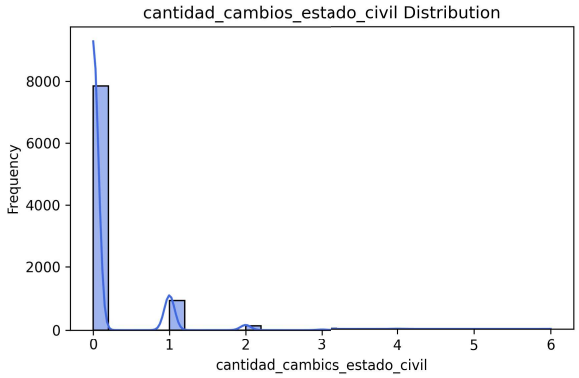
Figure 4.14: Distribution of selected numerical features (Part II)



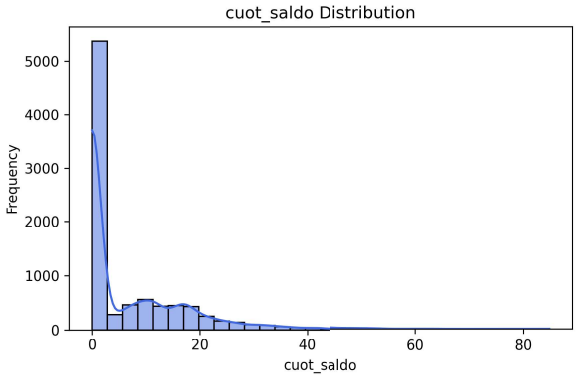
(a) prom_deudor Distribution



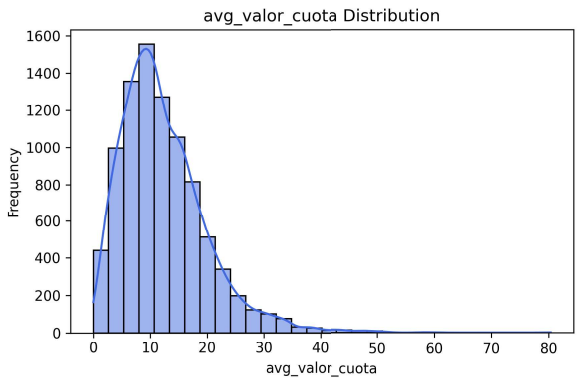
(b) std_deudor Distribution



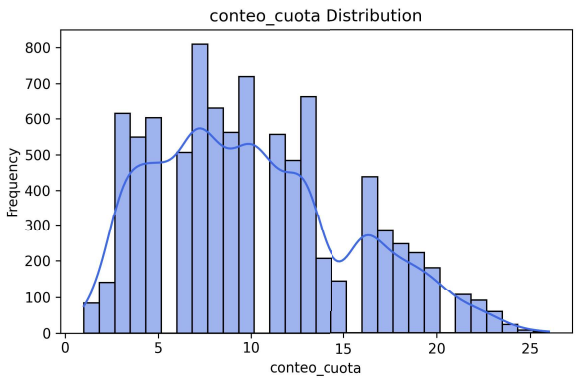
(c) cantidad_cambios_estado_civil Distribution



(d) cuot_saldo Distribution

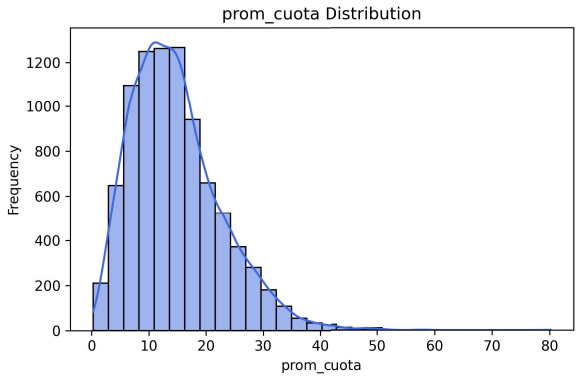


(e) avg_valor_cuota Distribution

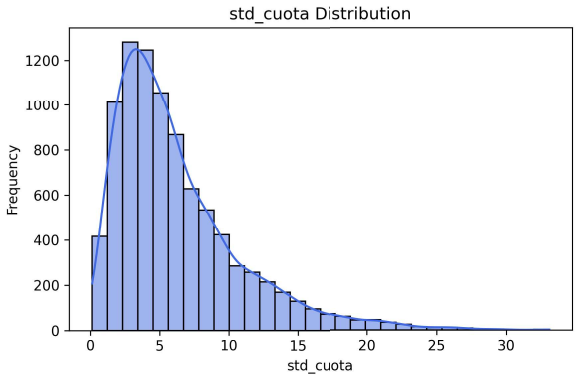


(f) conteo_cuota Distribution

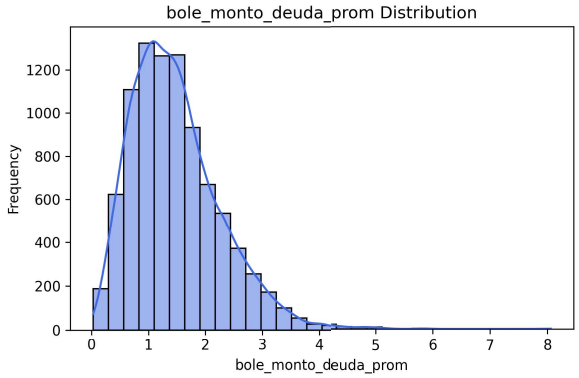
Figure 4.15: Distribution of selected numerical features (Part III



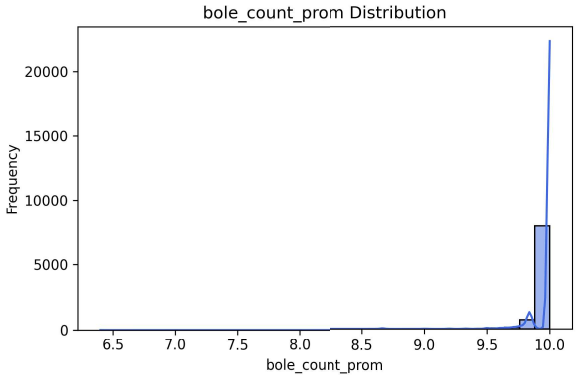
(a) prom_cuota Distribution



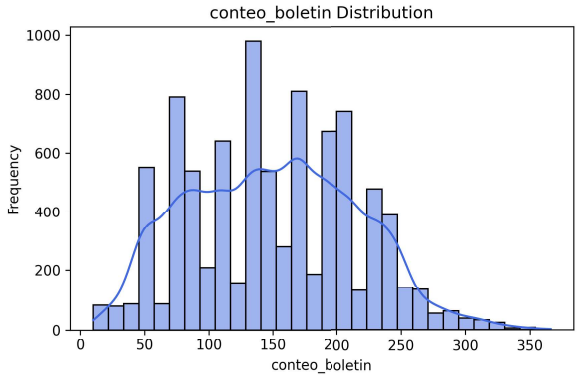
(b) std_cuota Distribution



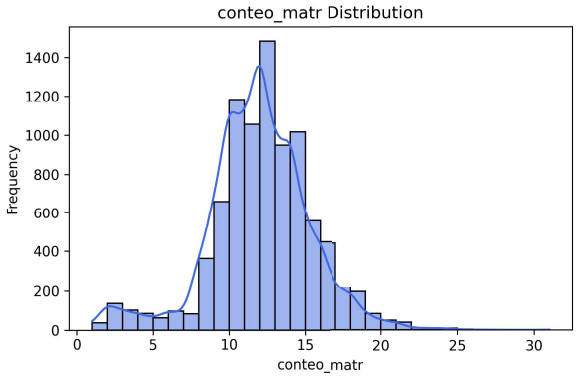
(c) bole_monto_deuda_prom Distribution



(d) bole_count_prom Distribution



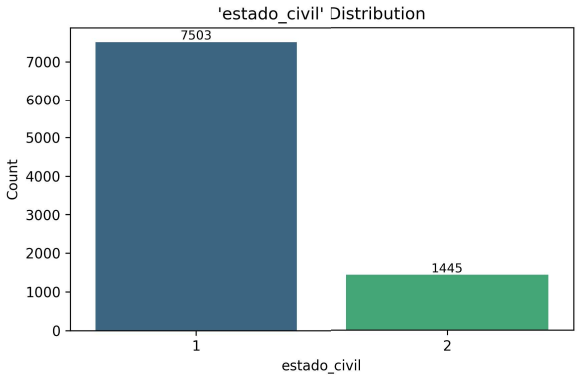
(e) conteo_boletin Distribution



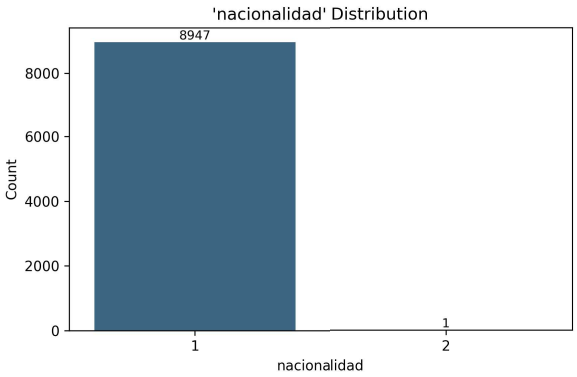
(f) conteo_matr Distribution

Figure 4.16: Distribution of selected numerical features (Part IV)

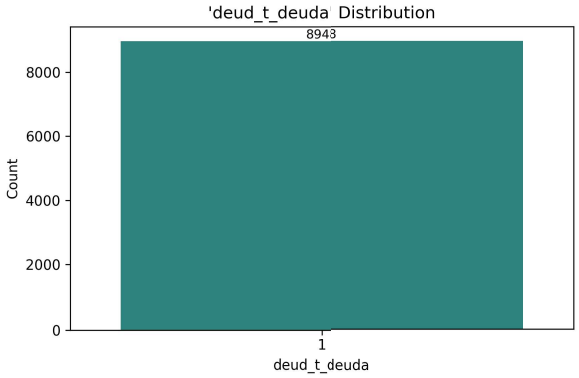
Figures 4.17 to 4.20 present the distribution of selected categorical features in the dataset. These include variables related to marital status, debt, income declaration, academic history, and institutional affiliation. The visualizations help identify dominant categories, potential imbalances, and patterns that may influence further analysis.



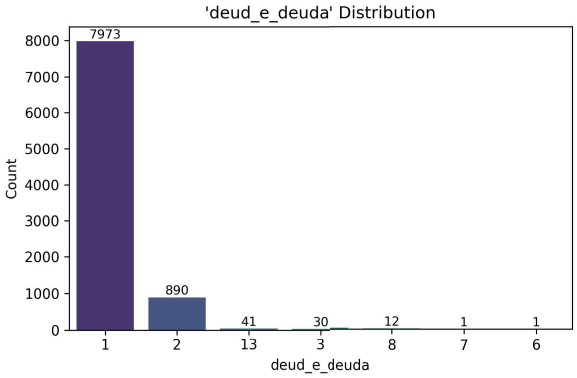
(a) estado_civil Distribution



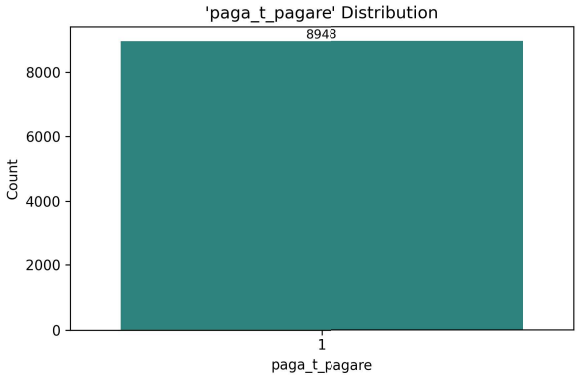
(b) nacionalidad Distribution



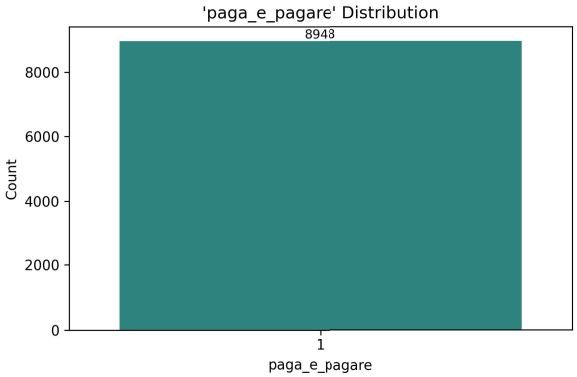
(c) deud_t_deuda Distribution



(d) deud_e_deuda Distribution

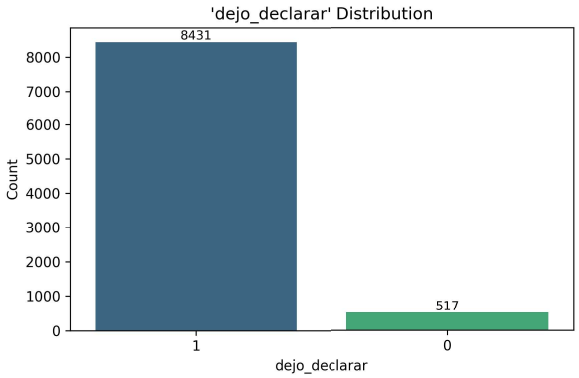


(e) paga_t_matr Distribution

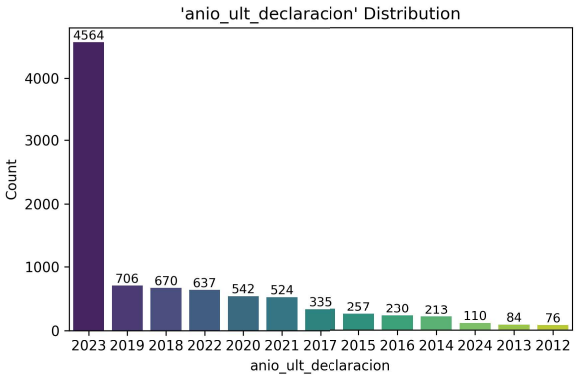


(f) paga_e_pagare Distribution

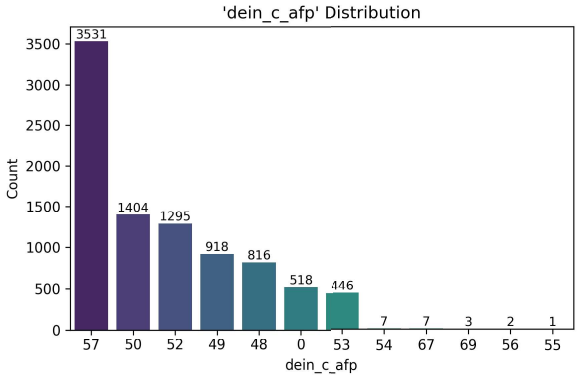
Figure 4.17: Distribution of selected categorical features (Part I)



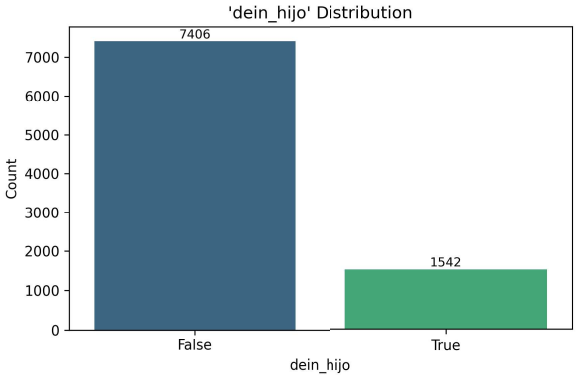
(a) deja_declarar Distribution



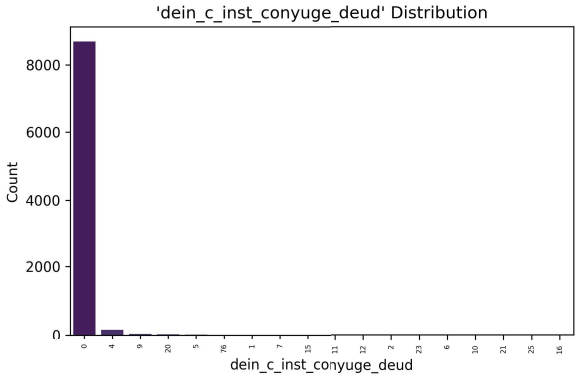
(b) anio_ult_declaracion Distribution



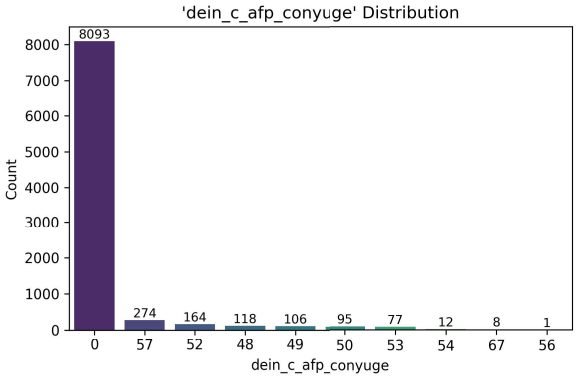
(c) dein_c_afp Distribution



(d) dein_hijo Distribution

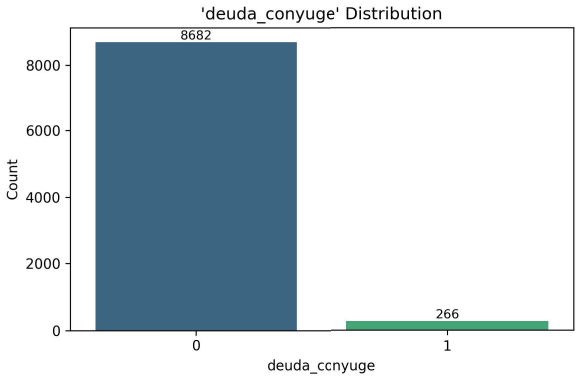


(e) dein_c_inst_conyuge_deud Distribution

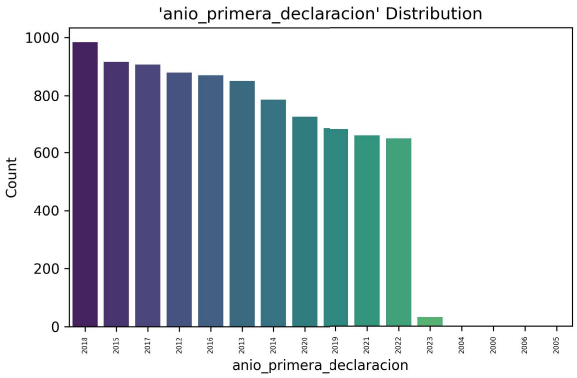


(f) dein_c_conyuge Distribution

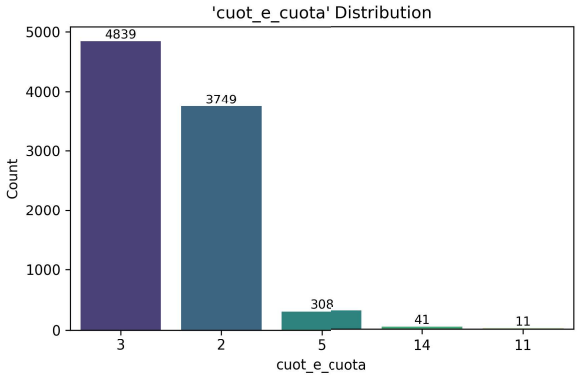
Figure 4.18: Distribution of selected categorical features (Part II)



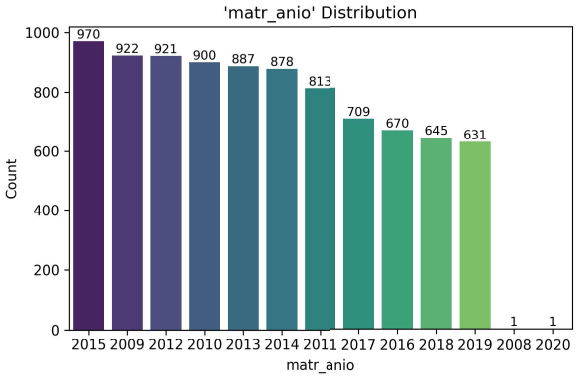
(a) deuda_conyuge Distribution



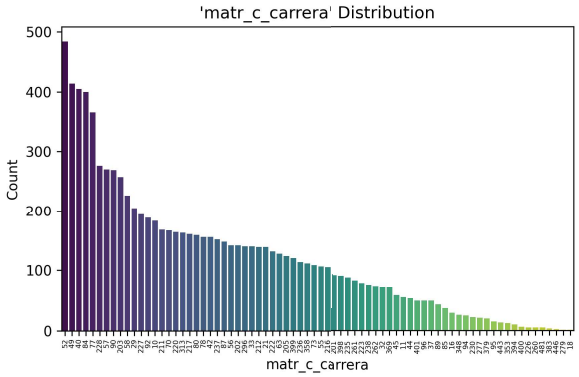
(b) anio_primera_declaracion Distribution



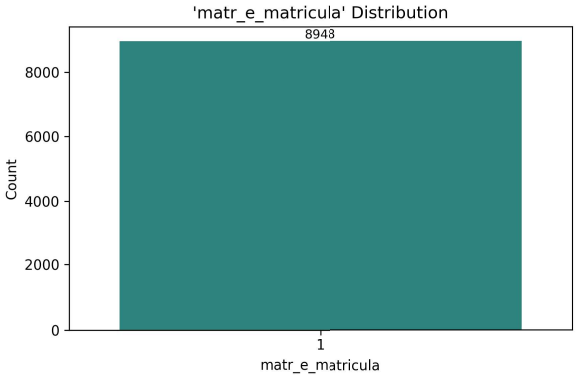
(c) cuot_e_cuota Distribution



(d) matr_anio Distribution

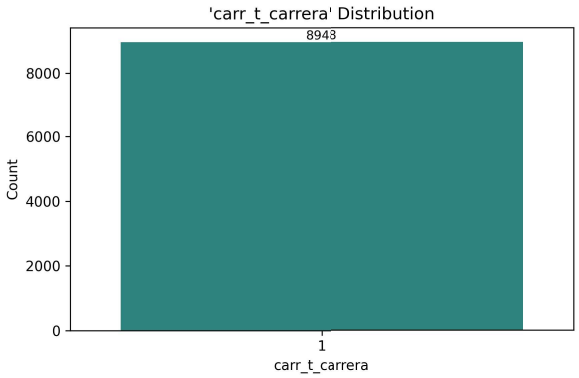


(e) matr_c_carrera Distribution

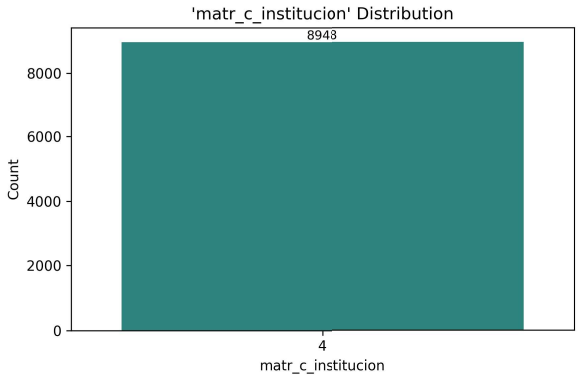


(f) matr_e_matricula Distribution

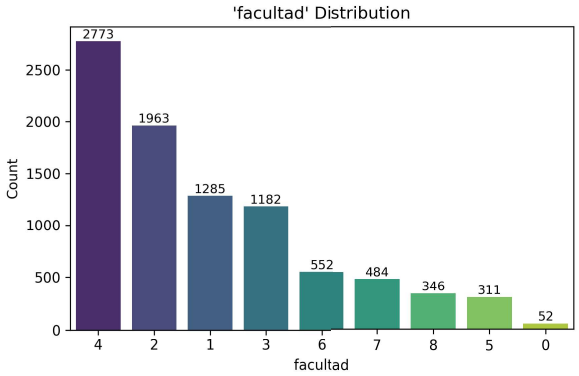
Figure 4.19: Distribution of selected categorical features (Part III)



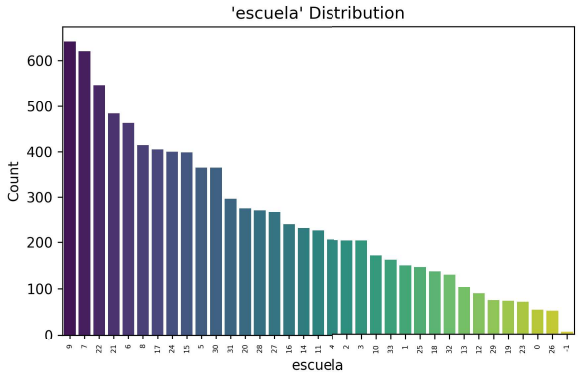
(a) carr_t_carrera Distribution



(b) matr_c_institucion Distribution



(c) facultad Distribution



(d) escuela Distribution

Figure 4.20: Distribution of selected categorical features (Part IV)

Potential redundancy among numerical variables is further analyzed using Pearson correlation coefficients, shown in Figure 4.21. A pragmatic threshold of $|r| \geq 0.70$ is applied to identify highly correlated features. From correlated groups, representative variables are retained based on interpretability and institutional relevance, while redundant counterparts are excluded to mitigate collinearity effects.

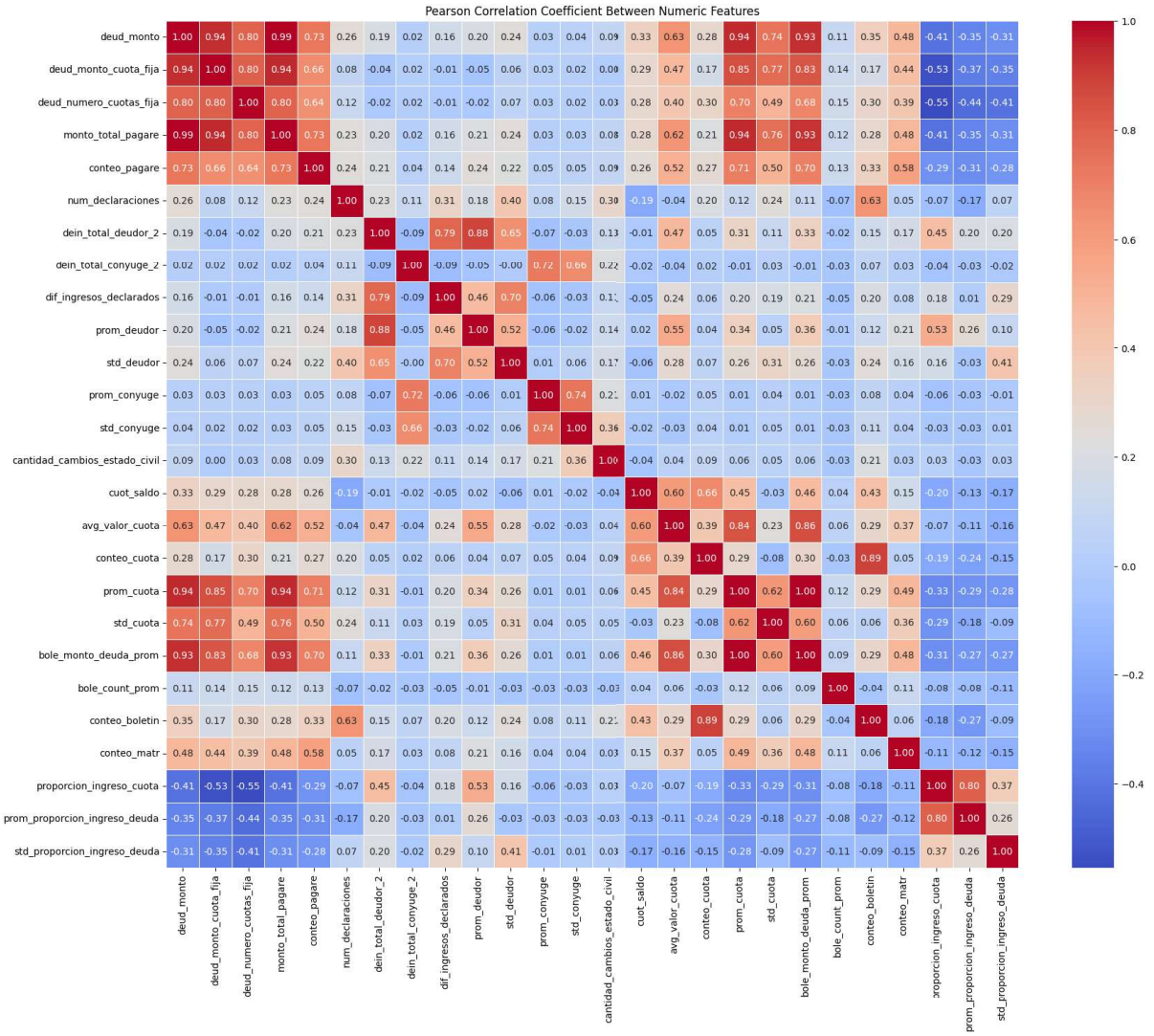


Figure 4.21: Pearson correlation heatmap among numerical features

Encoding, Scaling, and Temporal Handling.

Categorical variables are transformed using one-hot encoding, and numerical features are standardized via z-score normalization to ensure comparable scales across models. Date-related information is handled separately: the only remaining date variable (*deud_fecha_exigibilidad*) is decomposed into day and year components, as illustrated in Figures 4.22 and 4.23, while non-informative components are excluded.

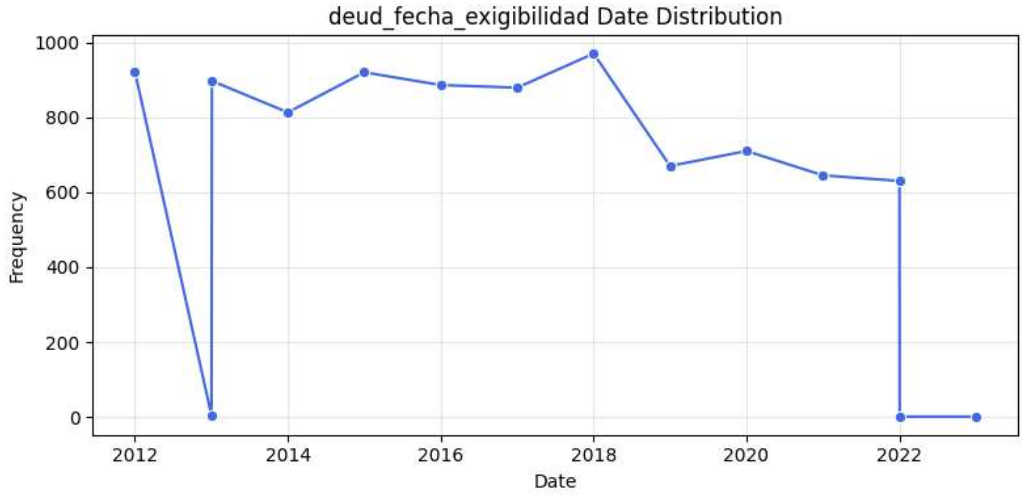
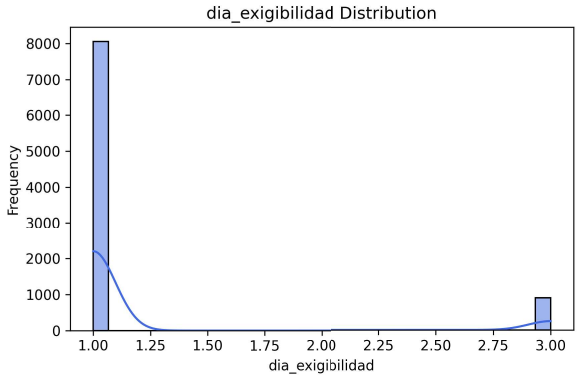
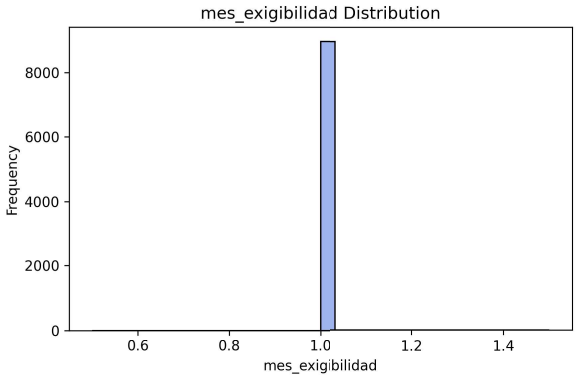


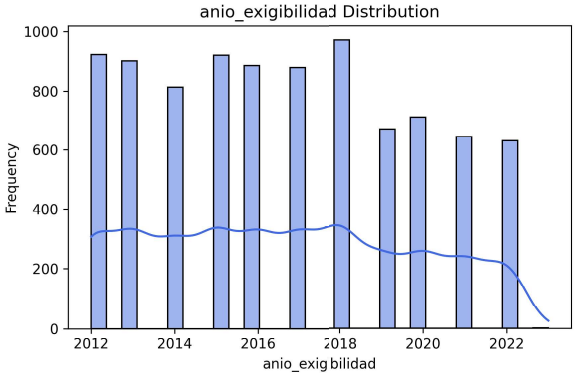
Figure 4.22: Distribution of enforceability dates



(a) Day Distribution



(b) Month Distribution



(c) Year Distribution

Figure 4.23: Decomposition of enforceability dates into day, month, and year components

Experimental Protocol and Validation Design.

The resulting feature matrix and target label are evaluated under a consistent validation protocol aligned with previous stages. Stratified data partitions are employed to preserve class proportions, and all metaheuristic optimization processes are executed under repeated runs with different random seeds. Performance metrics are aggregated across executions, enabling a robust assessment of both average predictive behavior and variability.

This experimental design ensures that the results reported in Stage III reflect stable and reproducible patterns rather than isolated favorable configurations, supporting a reliable evaluation of metaheuristic-based feature selection in a post-declarative institutional context.

4.3.4 Optimization Framework

Optimization Strategy

The optimization strategy used in this stage consists of metaheuristic algorithms.

In the literature, there are numerous metaheuristics of great interest for the feature selection problem, and in this research, we have used Particle Swarm Optimization, Grey Wolf Optimizer, and Whale Optimization Algorithm. The use of each of these metaheuristics is as follows:

- Particle Swarm Optimization (PSO): PSO uses social learning (personal best + global best) to explore globally while exploiting good subsets, often outperforming traditional sequential forward/backward methods and other metaheuristics in accuracy and subset size [108]. Overall, PSO is recommended for feature selection because it provides an effective, extensible global search framework that can be adapted to binary, high-dimensional, and large-sample settings, consistently yielding compact, accurate feature subsets with reasonable computational cost [109].

- Grey Wolf Optimizer (GWO): GWO models a small number of "elite wolves" (α, β, δ) guiding the rest (ω), naturally focusing the search around the best current subsets while still exploring other regions [110]. This hierarchy gives fast convergence with very few parameters to tune. Grey Wolf Optimizer is recommended for feature selection because its hierarchical search and many refined binary/hybrid variants reliably produce compact, high-accuracy feature subsets on challenging, high-dimensional problems [111, 112].

- Whale Optimization Algorithm (WOA): WOA simulates the hunting behavior of whales, in which they circle, attack with a spiraling bubble net, and perform a random search to alternate between exploration and exploitation in feature space [113]. Spatial-bound, altruistic, random-walk, and IoT-tailored WOA variants handle gene expression, medical, and IoT intrusion datasets, achieving very large dimensionality reductions while keeping or boosting classification accuracy above 90–98%. This makes WOA attractive, where exhaustive or greedy methods fail [114].

Particle Swarm Optimization (PSO) is a population-based metaheuristic inspired by the

collective behavior observed in natural swarms, where a set of agents (particles) cooperatively explores the search space by sharing information about good solutions [115].

Each particle encodes a candidate solution and iteratively updates its state by combining (i) a momentum term that preserves part of its previous motion, (ii) a cognitive term that attracts the particle toward its own best experience, and (iii) a social term that attracts it toward the best solution found by the swarm. This simple learning mechanism enables a dynamic balance between exploration and exploitation [116], controlled by a few interpretable parameters.

In feature selection, particles typically represent feature subsets, and PSO is commonly deployed in a wrapper setting by coupling the search with a learning algorithm and a performance metric; for binary subset representations, PSO is adapted through a probabilistic mapping that turns the "velocity" into a selection propensity for each feature [117].

Let the swarm contain N particles in a D-dimensional space. Particle i at iteration t has position $x_i^t \in \mathbf{R}^D$ and velocity $v_i^t \in \mathbf{R}^D$. Denote its personal best position by p_t^t and the global best by g^t. The standard PSO update is:

$$v_{i,d}^{t+1} = \omega\, v_{i,d}^t + c_1\, r_{1,d}^t \left(p_{i,d}^t - x_{i,d}^t\right) + c_2\, r_{2,d}^t \left(g_d^t - x_{i,d}^t\right), \quad d = 1,\ldots,D, \tag{4.1}$$

$$x_{i,d}^{t+1} = x_{i,d}^t + v_{i,d}^{t+1}, \tag{4.2}$$

where ω is the inertia weight, c_1 and c_2 are the cognitive and social acceleration coefficients, and $r_{1,d}^t, r_{2,d}^t \backsim \mathscr{U}(0,1)$ are independent random variables.

For feature selection, positions are binary vectors $X_i^t \in \{0,1\}^D$, where $x_i^t = 1$ indicates that feature d is selected. The velocity update is typically kept as in 4.1, but the position update becomes probabilistic via a Two-Step Techniques [118]:

$$s\left(v_{i,d}^{t+1}\right) = \frac{1}{1+\exp\left(-v_{i,d}^{t+1}\right)}, \tag{4.3}$$

$$x_{i,d}^{t+1} = \begin{cases} 1, & \text{if } u_{i,d}^t < s\left(v_{i,d}^{t+1}\right), \\ 0, & \text{otherwise,} \end{cases} \qquad u_{i,d}^t \sim \mathscr{U}(0,1). \tag{4.4}$$

Below is a pseudocode for Binary PSO (BPSO) as typically used in feature selection.

Algorithm 1 Binary Particle Swarm Optimization (BPSO) for Feature Selection

Require: Dataset $\mathscr{D}$, classifier $\mathscr{C}$, fitness function $f(\cdot)$, swarm size N, number of features D, max iterations T, parameters (ω, c_1, c_2), (optional) $V_{\max}$
Ensure: Best feature subset $\mathbf{g}$

Initialize each particle $i = 1, \ldots, N$:
    Position $\mathbf{x}_i^0 \in \{0,1\}^D$ randomly
    Velocity $\mathbf{v}_i^0 \in \mathbb{R}^D$ randomly
    Personal best $\mathbf{p}_i \leftarrow \mathbf{x}_i^0$
Evaluate fitness $f(\mathbf{x}_i^0)$ using $\mathscr{C}$ on subset $\mathbf{x}_i^0$
Set global best $\mathbf{g} \leftarrow \arg\min_{\mathbf{p}_i} f(\mathbf{p}_i)$ ▷ Use arg max if you maximize
for $t = 0$ to $T-1$ **do**
    for $i = 1$ to N **do**
        Evaluate $f(\mathbf{x}_i^t)$
        Update $\mathbf{p}_i$ and $\mathbf{g}$
    end for
    for $i = 1$ to N **do**
        for $d = 1$ to D **do**
            Sample $r_1, r_2 \sim \mathscr{U}(0,1)$
            Apply equation 4.1
            Apply equation 4.3
            Update $x_{i,d}^{t+1}$ by equation 4.4
        end for
    end for
end for
return g

Grey Wolf Optimizer (GWO) is a population-based metaheuristic that models the leadership hierarchy and cooperative hunting strategy of grey wolves [119].

The population is ranked into four levels: α (best solution), β (second best), δ (third best), and ω (the remaining candidates). During the search, each wolf updates its position by "encircling" and "approaching" the prey, where the prey is implicitly represented by the top-ranked wolves. This mechanism balances exploration and exploitation through adaptive coefficient vectors that shrink over iterations, enabling broad global search at early stages and progressively more exploitative movements around the current best solutions later.

In feature selection, GWO is commonly used as a wrapper metaheuristic: each wolf represents a candidate subset and is evaluated by a learning model and a target metric; when a binary representation is required, the continuous position components are subsequently mapped into $\{0,1\}$ decisions using a binarization scheme (here, the same BPSO transfer and probabilistic bit update already defined in Eqs. (4.3)–(4.4)). Stalking and surrounding behavior is modeled as follows:

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X}_1 + \overrightarrow{X}_2 + \overrightarrow{X}_3}{3} \tag{4.5}$$

where $\overrightarrow{X}_1$, $\overrightarrow{X}_2$, and $\overrightarrow{X}_3$ are the updated positions of the wolves based on the positions of the alpha wolf ($\overrightarrow{X}_\alpha$), beta wolf ($\overrightarrow{X}_\beta$), and delta wolf ($\overrightarrow{X}_\delta$)

$$\begin{aligned}\overrightarrow{X}_1 &= \overrightarrow{X}_\alpha - (\overrightarrow{A}_1 \cdot \overrightarrow{D}_\alpha)\\ \overrightarrow{X}_2 &= \overrightarrow{X}_\beta - (\overrightarrow{A}_2 \cdot \overrightarrow{D}_\beta)\\ \overrightarrow{X}_3 &= \overrightarrow{X}_\delta - (\overrightarrow{A}_3 \cdot \overrightarrow{D}_\delta)\end{aligned} \tag{4.6}$$

where $\overrightarrow{D}_\alpha$, $\overrightarrow{D}_\beta$, and $\overrightarrow{D}_\delta$ corresponds to the distance between the position of the wolf X that we are updating and the wolves $\overrightarrow{X}_\alpha$, $\overrightarrow{X}_\beta$, and $\overrightarrow{X}_\delta$.

$$\begin{aligned}\overrightarrow{D}_\alpha &= |\overrightarrow{C}_1 \cdot \overrightarrow{X}_\alpha - \overrightarrow{X}|\\ \overrightarrow{D}_\beta &= |\overrightarrow{C}_2 \cdot \overrightarrow{X}_\beta - \overrightarrow{X}|\\ \overrightarrow{D}_\delta &= |\overrightarrow{C}_3 \cdot \overrightarrow{X}_\delta - \overrightarrow{X}|\end{aligned} \tag{4.7}$$

A and C are coefficients that play different roles in metaheuristic behavior. C interferes with the direction and intensity of the position adjustment relative to the best solutions.

A plays an important role in the exploration–exploitation balance. When $|A| > 1$, the wolf moves away from the prey, allowing the exploration of new areas of the search space. On the other hand, when $|A| < 1$, the wolf approaches the prey, promoting exploitation.

The coefficient a decreases linearly from 2 to 0 during the optimization, meaning that the value of A also decreases, gradually shifting the focus of the algorithm from exploration to exploitation.

$$\begin{aligned}\overrightarrow{A} &= 2\overrightarrow{a} \cdot \overrightarrow{r}_1 - \overrightarrow{a}\\ \overrightarrow{C} &= 2 \cdot \overrightarrow{r}_2\\ \overrightarrow{a} &= 2 - \frac{2 \cdot t}{maxIter}\end{aligned} \tag{4.8}$$

Algorithm 2 shows the pseudocode of the binary version of GWO to solve the feature selection problem.

Algorithm 2 Binary Grey Wolf Optimizer (BGWO) for Feature Selection Problem

Require: Dataset $\mathscr{D}$, classifier $\mathscr{C}$, fitness function $f(\cdot)$, population size N, number of features D, max iterations T
Ensure: Best feature subset X_α

Initialize binary random population X
Evaluate the objective function of each individual
Identify the X_α, X_β, and X_δ in the population
for $t = 0$ to $T-1$ **do**
    for $i = 1$ to N **do**
        Update a, A, and C using Equation (4.8)
        Update X_1, X_2, and X_3 using Equation (4.6)
        Update the position of X_i^t using Equation (4.5)
    end for
    Binarization of population X using Eqs. 4.3 and 4.4
    Evaluate the objective function of each individual
    Update X_α, X_β, and X_δ
end for
Return (X_α)

Whale Optimization Algorithm (WOA) is a population-based metaheuristic inspired by the bubble-net hunting strategy of humpback whales [120].

The algorithm models two main behaviors to guide the search: (i) encircling and approaching the prey, which intensifies exploitation around the best solution found so far, and (ii) a spiral-shaped movement that mimics the bubble-net attack by updating candidate solutions along a logarithmic spiral around the current best.

To maintain exploration, WOA also allows whales to search for prey by moving with respect to randomly chosen individuals when a control factor indicates a non-exploitative regime. A time-varying parameter gradually shifts the dynamics from exploration to exploitation as iterations progress, enabling a coarse global search in early iterations and a refined local search later.

In feature selection, each whale represents a candidate subset evaluated by a classifier and a performance metric; when a binary representation is required, the continuous updates produced by WOA are mapped to $\{0,1\}$ decisions using the binarization mechanism already described in Eqs. (4.3)–(4.4).

Searching for prey is related to the exploration phase of the search space. The value of A is used to ensure a good exploration. When $|A| > 1$, whales move away from a known solution, allowing a global search for space exploration. This behavior is modeled as follows:

$$\begin{aligned} \overrightarrow{X}_i^{t+1} &= \overrightarrow{X_{rand}^t} - \overrightarrow{A} \cdot \overrightarrow{D} \\ \overrightarrow{D} &= |\overrightarrow{C} \cdot \overrightarrow{X_{rand}^t} - \overrightarrow{X}_i^t| \end{aligned} \tag{4.9}$$

where $\overrightarrow{A}$ and $\overrightarrow{C}$ are coefficient vectors, t denotes the current iteration, and $\overrightarrow{X^t_{rand}}$ is a random whale in the population. The coefficient vectors $\overrightarrow{A}$ and $\overrightarrow{C}$ can be computed according to Equations (4.10) and (4.11):

$$\overrightarrow{A} = 2\overrightarrow{a} \cdot \overrightarrow{r} - \overrightarrow{a} \tag{4.10}$$

$$\overrightarrow{C} = 2 \cdot \overrightarrow{r} \tag{4.11}$$

where $\overrightarrow{a}$ decreases linearly from 2 to 0 over the course of the iterations (both in the exploration and exploitation phases) and $\overrightarrow{r}$ corresponds to a random vector of values between $[0,1]$.

The first exploitation operator is related to the encircling prey. This occurs when a whale approaches the best solution found in the population so far. It is modeled using the following equations:

$$\begin{aligned} \overrightarrow{X}^{t+1}_i &= \overrightarrow{X^*} - \overrightarrow{A} \cdot \overrightarrow{D} \\ \overrightarrow{D} &= |\overrightarrow{C} \cdot \overrightarrow{X^*} - \overrightarrow{X}^t_i| \end{aligned} \tag{4.12}$$

where $\overrightarrow{X^*}$ is the best whale obtained so far and $\overrightarrow{X}^t_i$ is the i-th whale in the t-th iteration. The vectors $\overrightarrow{A}$ and $\overrightarrow{C}$ are calculated as shown in Equations (4.10) and (4.11).

The second exploitation operator is related to spiral motion. This occurs when whales move towards the best-known solution in a spiral-shaped path. The following equation models this:

$$\begin{aligned} \overrightarrow{X}^{t+1}_i &= \overrightarrow{D'} \cdot e^{bl} \cdot \cos(2\pi l) + \overrightarrow{X^*} \\ \overrightarrow{D'} &= |\overrightarrow{X^*} - \overrightarrow{X}^t_i| \end{aligned} \tag{4.13}$$

where $\overrightarrow{D'}$ is the distance of the i-th whale from the prey (the best solution obtained so far), and b is a constant to define the shape of the logarithmic spiral. l is a random number between $[-1,1]$.

In nature, humpback whales perform the simultaneous encircling of prey and a spiral movement. To model this simultaneous behavior, there is a 50% chance of choosing either the prey-encircling mechanism or the spiral model. The mathematical model is as follows:

$$\overrightarrow{X}^{t+1}_i = \begin{cases} \overrightarrow{X^*} - \overrightarrow{A} \cdot \overrightarrow{D} & \text{If } p < 0.5 \\ \overrightarrow{D'} \cdot e^{bl} \cdot \cos(2\pi l) + \overrightarrow{X^*} & \text{If } p \geq 0.5 \end{cases} \tag{4.14}$$

Algorithm 3 shows its behavior.

Algorithm 3 Binary Whale Optimization Algorithm (BWOA) for Feature Selection Problem

Require: Dataset $\mathscr{D}$, classifier $\mathscr{C}$, fitness function $f(\cdot)$, population size N, number of features D, max iterations T, constant b
Ensure: Best feature subset X^*

Initialize binary random population X
Evaluate the objective function of each individual
Identify the best individual in the population (X^*)
for *iteration* (t) **do**
    for *solution* (i) **do**
        Update a, A, C, l and p
        if $p < 0,5$ **then**
            if $|A| < 1$ **then**
                Apply Equation (4.12)
            else
                Apply Equation (4.9)
            end if
        else
            Apply Equation (4.13)
        end if
    end for
    Binarization of population X using Eqs. 4.3 and 4.4
    Evaluate the objective function of each individual
    Update X^* if there is a better solution
end for
Return (X^*)

Evaluation Protocol

We evaluate feature subsets selected by metaheuristic optimization using classification metrics computed from the confusion matrix, with a specific emphasis on the *minority* class (the default/non-filing event). In our application, the minority class corresponds to debtors who *fail to submit* the income declaration, and the most critical error is the *false negative* (FN), i.e., predicting a debtor will file when in fact they will not. This emphasis is consistent with risk-sensitive decision-making and common practice in imbalanced classification settings [121].

Let the minority (event) class be denoted as $y = 1$ and the majority (non-event) class as $y = 0$. For a set of predictions $\hat{y}$, we define:

- TP (true positives): number of instances with $y = 1$ and $\hat{y} = 1$.
- FN (false negatives): number of instances with $y = 1$ and $\hat{y} = 0$.
- FP (false positives): number of instances with $y = 0$ and $\hat{y} = 1$.
- TN (true negatives): number of instances with $y = 0$ and $\hat{y} = 0$.

In this context, FN represents missed default/non-filing cases, which is the most harmful error for the downstream process. [122].

Minority-class recall quantifies the model's ability to *detect* non-filing debtors among all truly non-filing debtors. Because it directly penalizes false negatives, it is the most suitable metric when the primary objective is to avoid missing risky cases. Mathematically, it is defined as follows:

$$\text{Recall}_{\text{min}} = \frac{\text{TP}}{\text{TP}+\text{FN}}. \tag{4.15}$$

Minority-class precision measures the *reliability* of positive predictions: among all debtors predicted as non-filers, precision indicates the fraction that truly are non-filers. While recall targets missed cases (FN), precision controls the burden caused by false alarms (FP), which is relevant when flagged cases trigger manual review, interventions, or additional administrative costs. Mathematically, it is defined as follows:

$$\text{Precision}_{\text{min}} = \frac{\text{TP}}{\text{TP}+\text{FP}}. \tag{4.16}$$

Reporting precision alongside recall is particularly important in imbalanced settings, where high recall can be trivially achieved by predicting many positives, potentially inflating FP.

The F-score provides a single summary statistic that balances the trade-off between detecting as many non-filers as possible (high recall) and maintaining credible positive predictions (high precision). This is useful for comparing feature subsets when improvements in recall may come at the cost of increased false positives. Mathematically, it is defined as follows:

The F_1-score is the harmonic mean of precision and recall:

$$\text{F1}_{\text{min}} = 2 \cdot \frac{\text{Precision}_{\text{min}} \cdot \text{Recall}_{\text{min}}}{\text{Precision}_{\text{min}} + \text{Recall}_{\text{min}}}. \tag{4.17}$$

The chosen metrics reflect the operational priorities of the task and the imbalanced nature of the target event:

- **Minority-class recall:** Given the imbalance, the most informative question is how many true non-filers are actually detected among all non-filers present in the data. $\text{Recall}_{\text{min}}$ directly quantifies this detection capability and is the metric most sensitive to false negatives, which are particularly consequential in minority-event datasets.

- **Minority-class precision:** In imbalanced datasets, even a moderate false-positive rate may translate into a large absolute number of false alarms because the majority class is prevalent. $\text{Precision}_{\text{min}}$ therefore complements recall by indicating how trustworthy positive predictions are when the base rate of the event is low, preventing an overly optimistic interpretation of a high-recall model that flags too many majority instances as positives.

- **Minority-class F-score:** Because imbalanced datasets often exhibit an intrinsic trade-off between capturing rare events and avoiding excessive false alarms, a single metric that summarizes the precision–recall balance is useful for comparative reporting across

experimental conditions. The F-score provides such a summary via a harmonic mean, and it is widely recommended when the positive class is rare, and performance must be assessed beyond accuracy.

Although the minority class is central to the problem, we also report the same metrics for the majority class by swapping the role of the positive label (i.e., treating $y = 0$ as the event). This provides a complete view of performance and highlights any degradation in the majority recognition induced by aggressive minority-focused optimization.

Objective Function

Optimization problems are formulated as an objective function subject to constraints. This formulation mathematically determines the "rules of the game" that the solutions must meet and measures their quality.

In the feature selection problem, there are different mathematical formulations, among which the use of a classifier metric as the target stands out [107]. In our research, we have used the Recall metric as the target because optimizing $\text{Recall}_{\text{min}}$ (4.15) directly minimizes false negatives, thereby reducing the number of non-filers that remain undetected. This aligns with the application's central risk-management requirement.

Thus, let n be the total number of features available, and $X \in \{0,1\}^n$ be the vector of binary decision variables to be obtained, where

$$x_i = \begin{cases} 1 & \text{if the i-}th\text{ feature in } X \text{ is selected,} \\ 0 & \text{otherwise,} \end{cases} \tag{4.18}$$

with $i \in 1,\ldots,n$. The objective function is defined as

$$\min f(X) = 1 - \text{Recall}_{\text{min}}(X), \tag{4.19}$$

subject to

$$\sum_{i=1}^{n} x_i \geq 1, \tag{4.20}$$

Where $\text{Recall}_{\text{min}}(X)$ corresponds to the recall of the minority class using the subset X.

4.3.5 Experimental Results

This subsection presents the empirical results obtained in Stage III, focusing on the impact of metaheuristic-based feature selection on the predictive performance of classification models applied to income declaration abandonment. The results are organized to highlight performance at the class level, robustness across repeated runs, and differences induced by the choice of metaheuristic and classifier.

All experiments were conducted under the same validation protocol and data partitions defined in the methodological framework, ensuring that observed performance differences can be attributed to the feature selection process rather than to changes in data handling or evaluation

strategy. Given the pronounced class imbalance of the dataset, particular attention is paid to the behavior of the models on the minority class, which represents the institutionally critical outcome of declaration abandonment.

Performance is reported in terms of Precision, Recall, and F1-score for both minority and majority classes, aggregated across multiple runs. In addition to average values, best-case performance and standard deviation are included to characterize stability and variability under stochastic optimization. This multi-dimensional evaluation enables a comprehensive assessment of the trade-offs between detection capability, robustness, and consistency across optimization strategies.

Predictive Performance after Optimization

Table 4.16, Table 4.17, and Table 4.18 jointly summarize the predictive performance achieved for the minority class after metaheuristic-driven feature selection, across different classifiers and optimization strategies. Minority-class performance is treated as the primary evaluation focus in this stage, as it reflects the ability of the models to identify declaration abandonment cases, which represent the most critical outcome from an institutional decision-making perspective.

MH	KNN			LightGBM			RandomForest		
	best	avg	std	best	avg	std	best	avg	std
GWO	0.913	0.849	0.107	0.961	0.937	0.015	0.99	0.932	0.024
PSO	0.68	0.493	0.067	0.951	0.935	0.013	0.989	0.955	0.028
WOA	0.709	0.472	0.143	0.971	0.943	0.015	0.913	0.699	0.111

Table 4.16: Recall minority class

MH	KNN			LightGBM			RandomForest		
	best	avg	std	best	avg	std	best	avg	std
GWO	0.94	0.874	0.104	0.975	0.954	0.009	0.986	0.943	0.015
PSO	0.753	0.551	0.064	0.975	0.957	0.01	0.946	0.886	0.064
WOA	0.772	0.508	0.133	0.975	0.964	0.011	0.922	0.779	0.061

Table 4.17: F-Score minority class

MH	KNN			LightGBM			RandomForest		
	best	avg	std	best	avg	std	best	avg	std
GWO	0.989	0.903	0.105	0.99	0.972	0.014	0.99	0.955	0.024
PSO	0.843	0.627	0.067	1.0	0.982	0.013	0.989	0.955	0.028
WOA	0.849	0.578	0.161	1.0	0.986	0.015	0.985	0.896	0.05

Table 4.18: Precision minority class

The results reveal a clear performance differentiation between ensemble-based classifiers and distance-based methods. LightGBM consistently achieves high and stable minority-class recall, precision, and F1-score across all optimization strategies, with average recall values exceeding 0.93 and low variability. Random Forest also exhibits strong minority detection capability, although its performance shows greater sensitivity to the selected metaheuristic, particularly under WOA optimization. In contrast, KNN-based configurations display substantially lower average performance and higher dispersion across runs, indicating sensitivity to feature subset composition in high-dimensional and imbalanced institutional datasets.

Taken together, the minority-class results indicate that metaheuristic-based feature selection substantially enhances the detection of declaration abandonment when combined with ensemble learning methods, while maintaining a favorable balance between recall and precision. These findings support the use of optimization-driven feature selection as a mechanism for improving both effectiveness and stability in complex post-declarative prediction scenarios.

Table 4.19, Table 4.20, and Table 4.21 report the corresponding results for the majority class.

MH	KNN			LightGBM			RandomForest		
	best	avg	std	best	avg	std	best	avg	std
GWO	0.999	0.994	0.006	0.999	0.998	0.001	0.999	0.997	0.002
PSO	0.992	0.982	0.004	1.0	0.999	0.001	0.999	0.998	0.001
WOA	1.0	0.978	0.018	1.0	0.999	0.001	0.999	0.995	0.004

Table 4.19: Recall majority class

MH	KNN			LightGBM			RandomForest		
	best	avg	std	best	avg	std	best	avg	std
GWO	0.996	0.993	0.006	0.999	0.997	0.001	0.999	0.997	0.001
PSO	0.986	0.976	0.003	0.999	0.997	0.001	0.997	0.994	0.003
WOA	0.987	0.973	0.008	0.999	0.998	0.001	0.995	0.988	0.003

Table 4.20: F-Score majority class

MH	KNN			LightGBM			RandomForest		
	best	avg	std	best	avg	std	best	avg	std
GWO	0.995	0.991	0.006	0.998	0.996	0.001	0.999	0.996	0.001
PSO	0.981	0.969	0.004	0.997	0.996	0.001	0.996	0.99	0.005
WOA	0.982	0.968	0.008	0.998	0.997	0.001	0.995	0.982	0.007

Table 4.21: Precision majority class

As expected, performance on the majority class remains consistently high across all classifiers and optimization strategies, with recall, precision, and F1-score values close to one

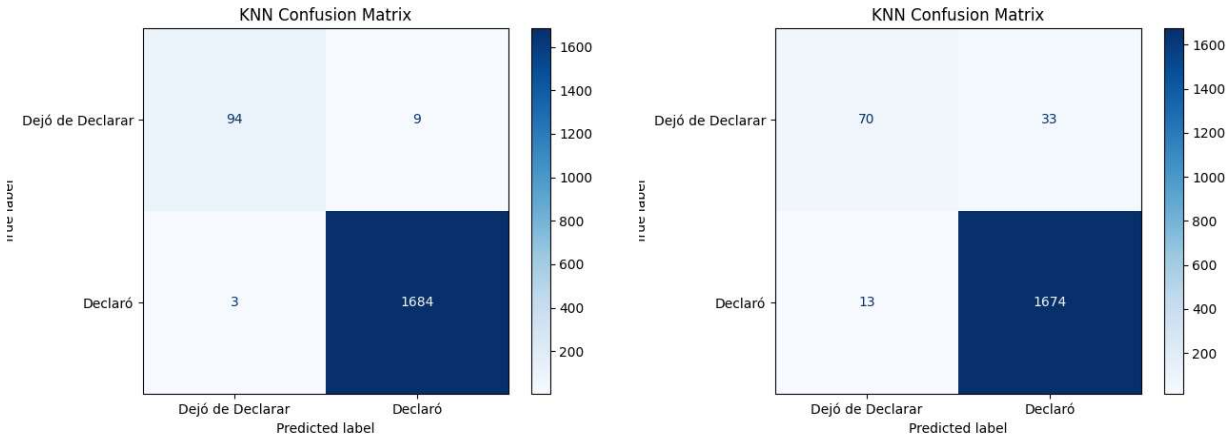
and low variability. This behavior confirms that the gains observed in minority-class detection are not achieved at the expense of a degradation in majority-class recognition, preserving overall classification balance and robustness.

Overall, the results demonstrate that metaheuristic-based feature selection constitutes an effective strategy for enhancing predictive performance in declaration abandonment modeling, particularly when coupled with ensemble classifiers. The observed improvements are achieved with acceptable variability across runs, supporting the suitability of optimization-driven approaches for high-dimensional institutional risk modeling.

Confusion Matrices Analysis

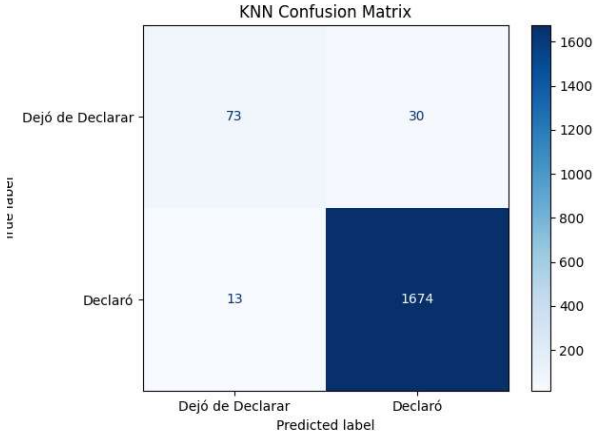
The confusion matrices presented in Figures 4.24, 4.25, and 4.26 provide a complementary, class-level perspective on the predictive behavior of the optimized models analyzed in Stage III. While aggregate metrics summarize overall performance, confusion matrices allow a direct examination of error types, particularly false negatives and false positives, which are critical in the context of declaration abandonment prediction.

Figure 4.24 illustrates the confusion matrices obtained for the KNN classifier under the three metaheuristic optimization strategies. Across configurations, KNN shows a systematic tendency to misclassify a non-negligible fraction of minority-class instances as majority-class outcomes, as reflected in the presence of false negatives in all three optimization scenarios. This behavior is particularly evident under PSO and WOA optimization, where the number of false negatives is substantially higher than under GWO, indicating increased sensitivity to the selected feature subset. These patterns are fully consistent with the lower average recall and higher variability observed for the minority class in Tables 4.16 and 4.17, and suggest limited robustness of distance-based classifiers when applied to high-dimensional and imbalanced institutional datasets.



(a) GWO

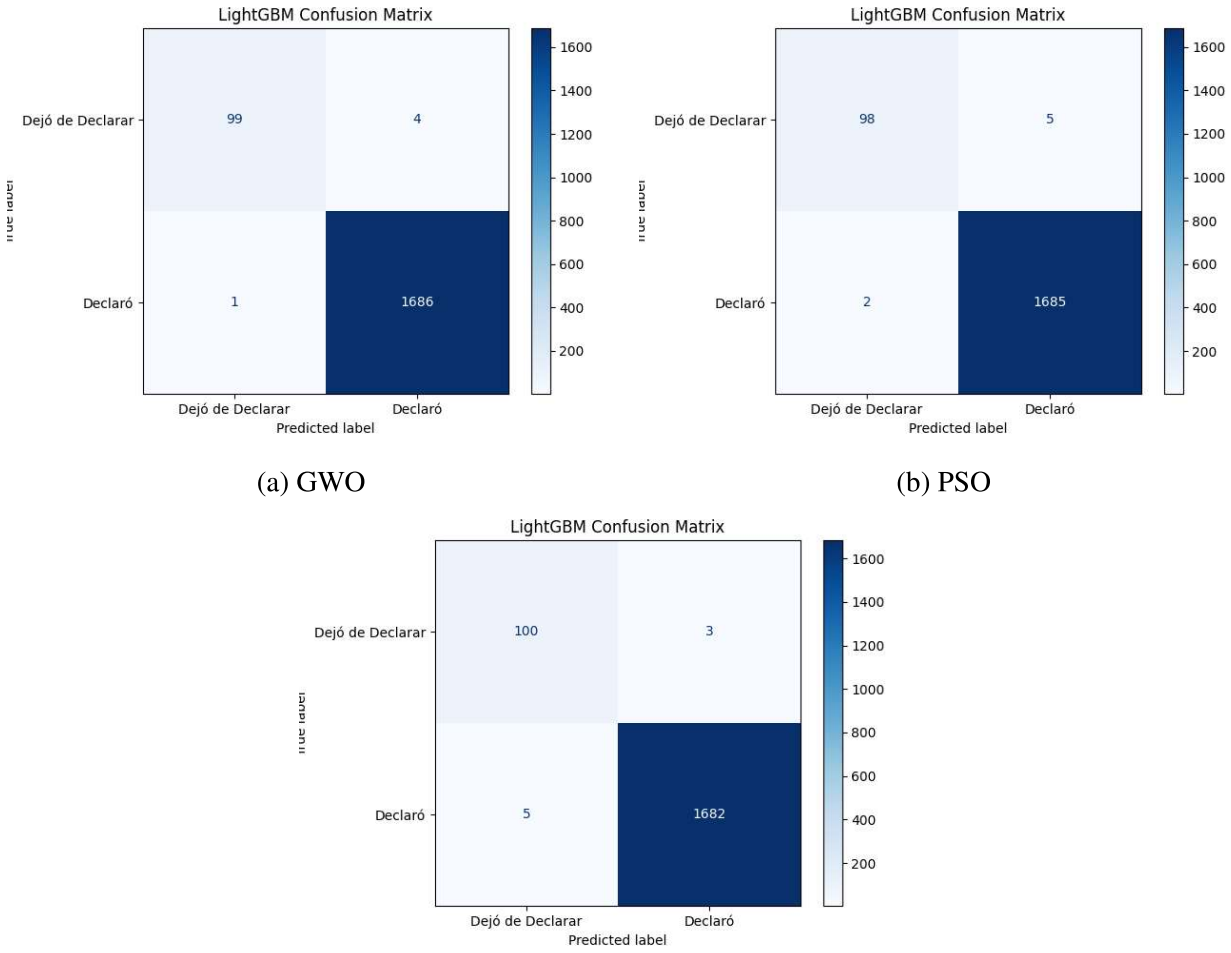
(b) PSO



(c) WOA

Figure 4.24: Confusion Matrices KNN

In contrast, the confusion matrices for LightGBM, shown in Figure 4.25, reveal a markedly improved classification pattern for declaration abandonment. Across all metaheuristic optimization strategies, the number of false negatives remains very limited, indicating a strong capacity to correctly identify minority-class instances. At the same time, false positives are minimal, reflecting stable recognition of compliant borrowers. The similarity of the confusion matrices obtained under GWO, PSO, and WOA suggests that LightGBM exhibits low sensitivity to the specific optimization strategy employed, reinforcing its robustness under metaheuristic-driven feature selection. These visual patterns are fully aligned with the high and stable minority-class recall and F1-scores reported in Tables 4.16 and 4.17.



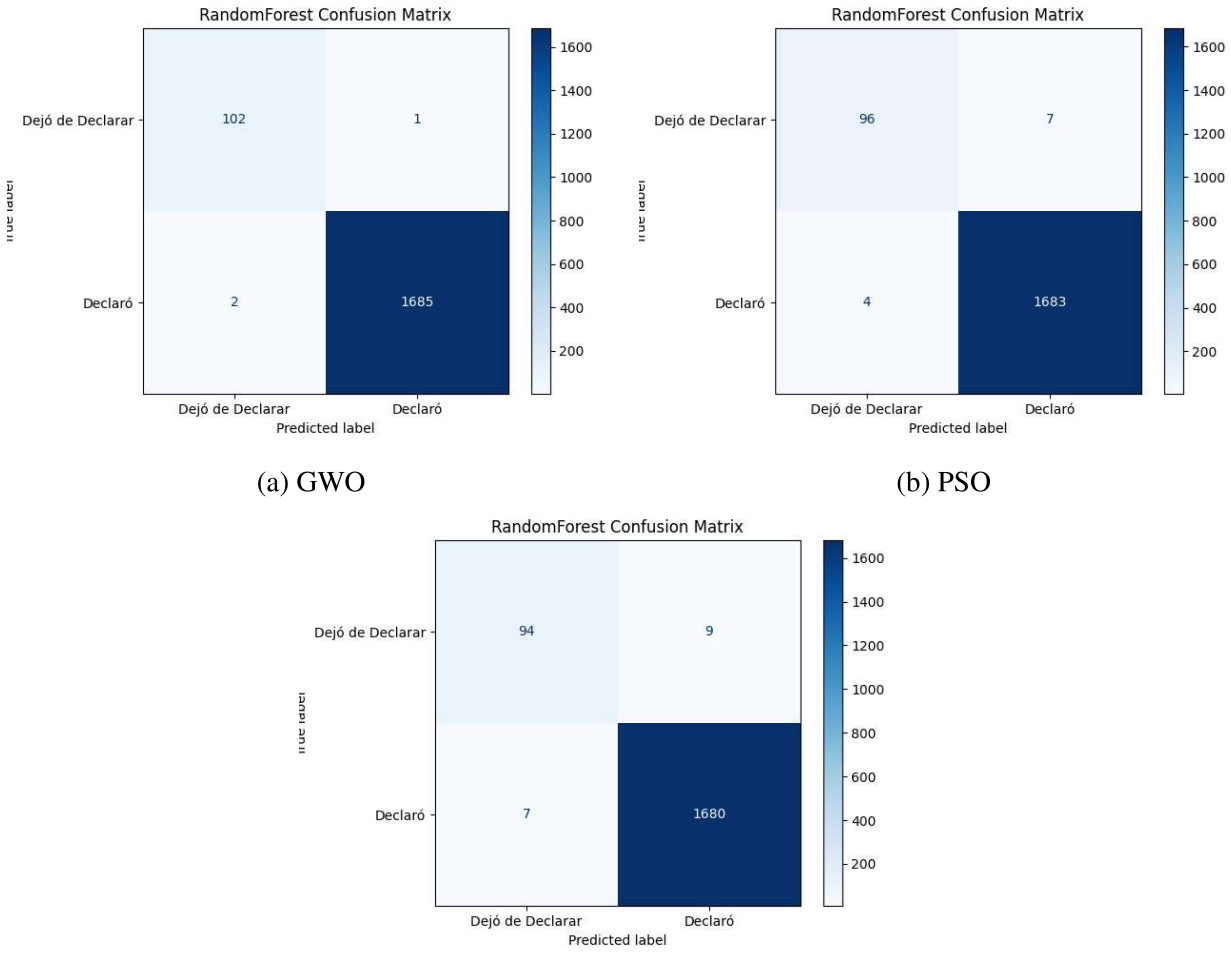
(a) GWO

(b) PSO

(c) WOA

Figure 4.25: Confusion Matrices LightGBM

Figure 4.26 presents the confusion matrices obtained for Random Forest under the three metaheuristic optimization strategies. Across configurations, Random Forest exhibits a strong ability to detect declaration abandonment cases, with a high number of correctly classified minority-class instances and a limited number of false positives. Nevertheless, compared to LightGBM, a higher degree of variability across optimization strategies can be observed. In particular, the WOA-based configuration shows an increase in false negatives relative to GWO and PSO, indicating a partial loss in minority-class detection. This variability is consistent with the dispersion reported in the metric-based analysis and suggests that Random Forest is more sensitive to the composition of the selected feature subsets, especially under certain metaheuristic search dynamics.



(a) GWO

(b) PSO

(c) WOA

Figure 4.26: Confusion Matrices Random Forest

Overall, the confusion matrix analysis reinforces the conclusions drawn from the quantitative performance metrics reported in this stage. Ensemble-based classifiers, particularly LightGBM, exhibit a favorable error structure characterized by high detection rates of declaration abandonment and controlled false alarm levels across optimization strategies. In contrast, KNN shows structural limitations in handling the complexity and class imbalance of post-declarative institutional data, as reflected in its higher incidence of false negatives. These results highlight the importance of jointly considering predictive accuracy and error composition when designing decision-support tools for institutional compliance monitoring, especially in contexts where missed detection of non-compliance carries higher operational risk.

Feature Selection Outcomes

This subsection reports the feature selection outcomes obtained in Stage III through a frequency-based analysis of the selected subsets across repeated runs. The heatmaps summarize, for each

of the 75 candidate variables, how often a given feature was selected over the 31 independent executions of the metaheuristic-driven selection process. Darker blue tones indicate higher selection frequency (features repeatedly retained across runs), whereas lighter yellow tones indicate lower frequency (features rarely retained). This representation is used to characterize the stability of the selected feature subsets and to identify consistently preferred variables under different optimization strategies and classifiers.

Figures 4.27–4.29 present the selection frequency profiles grouped by metaheuristic, showing how GWO, PSO, and WOA distribute selection pressure across the feature space for KNN, LightGBM, and Random Forest. A clear concentration pattern is observed for a reduced set of variables that are repeatedly selected across classifiers, suggesting the existence of a core predictive signal that remains stable under different learning paradigms. In particular, the dominance of specific variables (e.g., features 61 and 71 in the GWO-based heatmap) indicates that the search process consistently converges to a parsimonious subset that is repeatedly favored across runs, supporting the interpretation of these variables as robust institutional signals within the post-declarative representation.

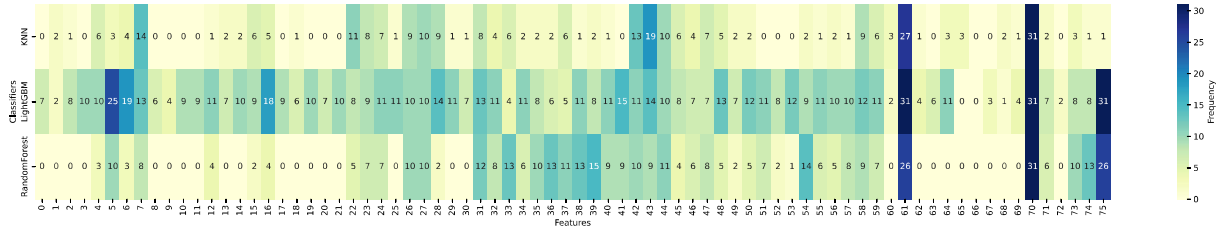


Figure 4.27: Heatmap Features Selected by GWO

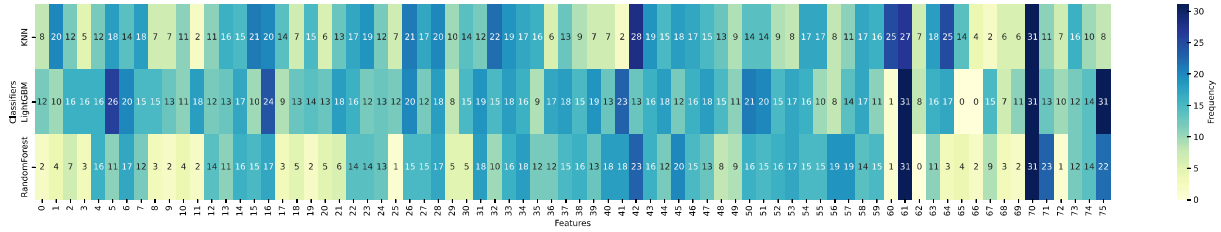


Figure 4.28: Heatmap Features Selected by PSO

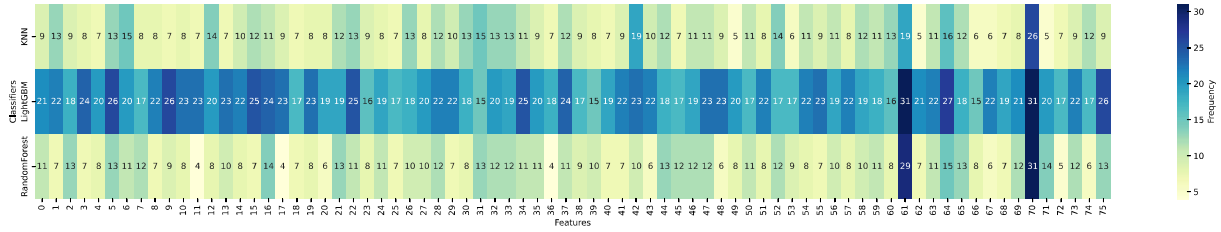


Figure 4.29: Heatmap Features Selected by WOA

Figures 4.30–4.32 reorganize the same evidence by classifier, enabling a direct comparison of how each learning model interacts with the metaheuristic selection process. In this view, ensemble-based classifiers exhibit more consistent selection patterns across metaheuristics, with recurrent high-frequency features appearing under multiple optimization strategies. Conversely, KNN shows a more heterogeneous frequency distribution across metaheuristics, reflecting greater sensitivity of the selected subsets to the optimization dynamics. This qualitative behavior is coherent with the metric-based results reported previously, where KNN displayed lower and more variable minority-class performance, while LightGBM and Random Forest achieved stronger and more stable outcomes.

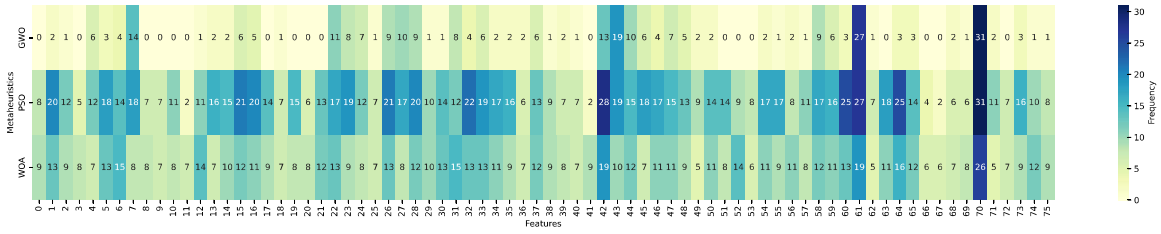


Figure 4.30: Heatmap Features Used by KNN

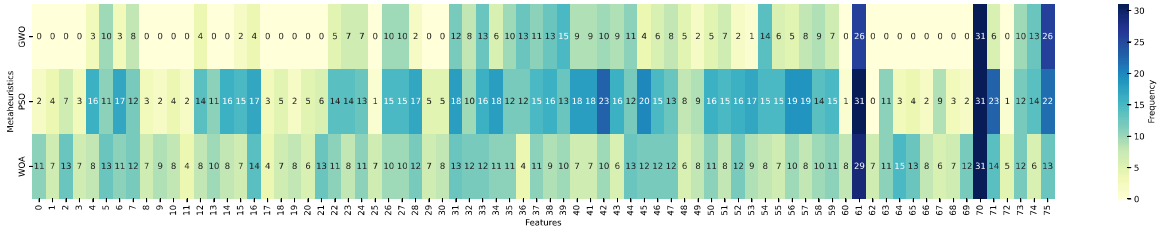


Figure 4.31: Heatmap Features Used by Random Forest

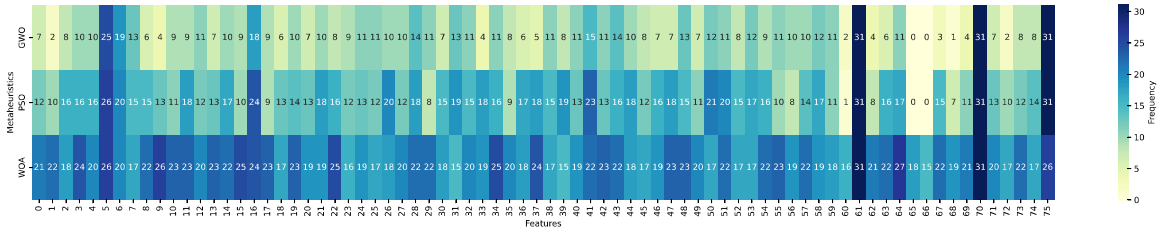


Figure 4.32: Heatmap Features Used by LightGBM

Overall, the heatmap analysis indicates that metaheuristic-driven feature selection does not operate as a purely stochastic mechanism but yields structured and recurrent feature preference profiles. These outcomes complement the predictive performance analysis by providing evidence on which variables are consistently retained across runs and configurations, and by illustrating how subset stability varies with the classifier and the metaheuristic search strategy.

Robustness and Stability Analysis

The robustness and stability of the metaheuristic-optimized models were evaluated by analyzing performance variability across repeated executions and optimization strategies. In contrast to best-case results, this analysis focuses on average values and dispersion measures reported in Tables 4.16, 4.17, and 4.18, which provide direct evidence of the consistency of the predictive behavior under repeated experimental conditions.

Ensemble-based classifiers exhibit markedly stable minority-class performance after optimization. LightGBM, in particular, maintains high average recall values for the minority class across all metaheuristics (0.937 under GWO, 0.935 under PSO, and 0.943 under WOA), with consistently low standard deviations ranging between 0.013 and 0.015. Similar stability is observed for minority-class F1-scores, whose dispersion remains at or below 0.011 across all optimization strategies. These low variability levels indicate that the metaheuristic selection process converges toward similar feature subsets and predictive behaviors, yielding reproducible results largely independent of the specific optimization strategy employed.

Random Forest also demonstrates robust average minority-class performance, although with moderately higher variability across metaheuristics. While GWO- and PSO-based configurations preserve relatively low dispersion (standard deviations below 0.03 for recall), the WOA-based configuration exhibits a noticeable increase in variability (recall standard deviation of 0.111), accompanied by a reduction in average recall. This behavior suggests that Random Forest is more sensitive to variations in the selected feature subsets, particularly under certain metaheuristic search dynamics, reflecting a trade-off between model flexibility and robustness.

In contrast, KNN shows substantially higher instability across both runs and optimization strategies. Minority-class recall variability is pronounced, with standard deviation values reaching 0.107 under GWO and 0.143 under WOA, alongside sharp drops in average recall under PSO (0.493) and WOA (0.472). A similar pattern is observed for minority-class F1-scores, with dispersion exceeding 0.13 under WOA, confirming that distance-based classifiers are particularly sensitive to fluctuations in feature subset composition when operating in high-dimensional and imbalanced institutional datasets.

Performance stability for the majority class further reinforces these observations. As reported in Tables 4.19 and 4.20, recall, precision, and F1-score values for the majority class remain consistently close to one across all classifiers and optimization strategies, with minimal dispersion. This stability confirms that the improvements observed in minority-class detection are not achieved at the expense of degraded majority-class performance.

Overall, the robustness analysis demonstrates that metaheuristic-based feature selection can be integrated into predictive modeling pipelines without introducing excessive performance volatility. The results highlight clear differences in stability across classifier families, with ensemble-based methods—particularly LightGBM—exhibiting the most reliable behavior, and underscore the importance of jointly considering average performance and variability when deploying optimization-driven decision-support models in institutional compliance monitoring contexts.

4.3.6 Comparative Analysis with Previous Stages

Stage III represents a methodological and analytical transition within the stage-based design defined in the methodological framework. Whereas Stages I and II focus on first-time income declaration compliance under strictly pre-declarative constraints, Stage III addresses a distinct predictive phenomenon: the identification of declaration abandonment among borrowers who have previously complied with their income declaration obligations. This transition involves both a change in the target population and a fundamental shift in the informational context available for prediction.

From a data perspective, Stage III is characterized by a substantial increase in informational richness. While earlier stages operate exclusively on antecedent academic and administrative variables observable prior to any declaration event, Stage III incorporates variables derived from historical declaration trajectories, repayment dynamics, and temporal patterns of compliance. This enrichment results in higher-dimensional and more heterogeneous feature spaces, consistent with the post-compliance modeling scenario defined in the methodology chapter.

Methodologically, the defining distinction of Stage III lies in the explicit introduction of metaheuristic-based feature selection as a core design component. In Stage II, feature spaces are curated through semantic and institutional reasoning and remain fixed throughout model training. In contrast, Stage III formulates feature selection as an optimization problem, jointly considering predictive performance, robustness under class imbalance, and feature compactness. As specified in the methodological framework, comparisons in this stage are performed between optimized and non-optimized models operating on identical enriched data representations, ensuring that observed differences can be attributed to the optimization process rather than to data availability.

In comparative terms, the results of Stage III reveal systematic improvements in imbalance-aware metrics relative to non-optimized baselines defined on the same feature space. While Stage II demonstrates that stable and interpretable prediction is achievable under severe informational constraints, Stage III shows that, in high-dimensional post-compliance contexts, controlled optimization becomes a necessary methodological extension rather than an optional refinement. Importantly, these improvements are accompanied by consistent reductions in feature set size, indicating that enhanced discrimination is achieved through selective refinement rather than through indiscriminate exploitation of additional variables.

From an interpretability standpoint, Stage III introduces a controlled trade-off relative to Stage II. Whereas Stage II prioritizes transparent model structures and direct explanatory mechanisms, Stage III emphasizes robustness and generalization in a more complex behavioral setting. Nevertheless, the optimized feature subsets remain institutionally interpretable and stable, consistently prioritizing variables related to declaration history, income dynamics, repayment exposure, and temporal compliance patterns. This continuity indicates that metaheuristic optimization refines the substantive drivers of declarative behavior identified in earlier stages rather than obscuring them.

Overall, the comparative analysis confirms that Stage III extends, rather than replaces, the insights obtained in Stages I and II. It demonstrates how optimization-based strategies become essential as data complexity increases, while also delineating the practical limits of performance gains achievable through feature selection within classical modeling paradigms under

interpretability constraints. This progression establishes a coherent methodological escalation and motivates the exploration of more expressive modeling strategies in the subsequent stage.

4.3.7 Summary of Stage III Results

The results of Stage III demonstrate that declaration abandonment among previously compliant borrowers constitutes a learnable and structurally distinct predictive problem within the FSCU system, directly addressing Research Question RQ2 and Hypothesis H2. By incorporating historical declaration and repayment information, this stage captures behavioral dynamics that are not observable under strictly pre-declarative regimes, enabling a more comprehensive characterization of compliance trajectories over time.

Empirically, the application of metaheuristic-based feature selection yields consistent improvements in predictive robustness relative to non-optimized models operating on identical high-dimensional data representations. These improvements are most pronounced in imbalance-aware evaluation metrics, indicating enhanced discrimination between borrowers who persist in declaring income and those who discontinue compliance. Crucially, these gains are achieved alongside substantial reductions in feature dimensionality, providing empirical support for Hypothesis H3 and confirming that performance improvements arise from redundancy reduction and noise control rather than from unrestricted feature expansion.

The optimized feature subsets exhibit strong internal coherence and institutional plausibility. Variables associated with income evolution, declaration continuity, repayment exposure, and temporal compliance patterns consistently emerge as dominant predictors across optimization runs. This stability supports the interpretation of declaration abandonment as a behaviorally driven process and confirms that metaheuristic optimization preserves, rather than degrades, the interpretability of the resulting models.

From a methodological perspective, Stage III validates the integration of metaheuristic-based feature selection as a necessary extension of interpretable modeling approaches when addressing complex, longitudinal compliance behaviors. At the same time, the results indicate that, once feature space redundancy is systematically controlled and interpretability constraints are maintained, further performance gains within the same modeling paradigm are limited.

In summary, Stage III establishes metaheuristic-driven feature selection as an effective and principled strategy for improving predictive modeling in post-compliance scenarios, while also delineating the empirical boundary of classical optimization-based approaches. These findings provide a critical conceptual and empirical bridge between early-stage, constraint-driven prediction and the exploration of more expressive modeling paradigms examined in Stage IV, and they form a solid foundation for the cross-stage synthesis developed in the discussion chapter.

4.4 Bioinspired Deep Learning for Declaration Discontinuity Prediction (Stage IV)

4.4.1 Stage IV Overview

Stage IV constitutes the final empirical modeling stage of this doctoral research and focuses on the use of bioinspired deep learning models to predict income declaration discontinuity among borrowers who previously complied with declaration requirements. This stage addresses the same operational phenomenon analyzed in Stage III, namely declaration abandonment, but deliberately relaxes interpretability and linearity constraints in order to evaluate whether increased model expressiveness yields additional predictive gains beyond those achieved through metaheuristic-optimized classical models.

In contrast to earlier stages, Stage IV does not introduce new datasets, predictive targets, or evaluation criteria. The modeling task, data representations, temporal constraints, and experimental protocol remain consistent with those defined for declaration abandonment prediction. The methodological distinction of this stage lies exclusively in the modeling paradigm. While Stage II prioritizes interpretability under strict pre-declarative constraints and Stage III incorporates metaheuristic-driven feature selection to manage high-dimensional feature spaces, Stage IV explores bioinspired deep neural network architectures capable of capturing complex non-linear relationships inherent in enriched institutional data.

The motivation for this stage is twofold. First, it enables an empirical assessment of whether deep learning models can extract additional predictive structure from high-dimensional representations of declaration behavior, providing direct evidence relevant to Research Question RQ4 and Hypothesis H5. Second, by incorporating gradient-based post-hoc interpretability mechanisms, this stage supports the examination of variable relevance within highly expressive models, complementing the interpretable analyses conducted in earlier stages without relying on inherently transparent model structures.

Methodologically, Stage IV is designed as a comparative extension rather than a replacement of previous stages. Results obtained from deep learning models are evaluated under the same validation, robustness, and temporal consistency framework adopted throughout the research, enabling direct comparison with the optimized classical models developed in Stage III. This design supports a systematic examination of trade-offs between predictive performance, stability, model complexity, and explanatory capacity under identical data representations and evaluation conditions.

Overall, Stage IV completes the progressive modeling trajectory of the thesis by situating bioinspired deep learning within the broader comparative framework of predictive paradigms explored in the FSCU context. The insights derived from this stage provide essential input for the integrated cross-stage synthesis developed later in the thesis, supporting a coherent interpretation of how successive methodological decisions shape predictive behavior, robustness, and interpretability across increasing levels of model complexity.

4.4.2 Problem Definition and Predictive Scope

Stage IV addresses the prediction of income declaration discontinuity among borrowers who had previously complied with the FSCU income reporting requirements. This predictive task corresponds to the operational phenomenon defined as declaration abandonment and is identical, in terms of target definition and temporal constraints, to that analyzed in Stage III.

Formally, the problem is defined as a binary classification task in which the outcome variable indicates whether a borrower interrupts the income declaration process after having submitted at least one prior declaration. The positive class represents declaration abandonment events, while the negative class corresponds to continued compliance. This formulation reflects a critical operational challenge for FSCU administrators, as abandonment events are typically detected only after the interruption has already occurred, limiting the institution's ability to intervene proactively.

From a predictive scope perspective, Stage IV operates under enriched informational conditions compared to pre-declarative modeling stages. In addition to academic and administrative variables available prior to the first declaration, the feature space incorporates historical information generated during the declaration process itself, including variables related to debt structure, installment behavior, and declaration history. These variables are temporally valid with respect to the prediction target and comply with the information availability constraints established in the methodological framework.

The predictive scope of this stage is explicitly constrained to evaluation rather than expansion. No additional predictive targets, auxiliary outcomes, or alternative labeling strategies are introduced. Likewise, no external data sources or derived proxy variables are incorporated. This restriction ensures that observed performance differences between Stage IV and previous stages can be attributed to modeling capacity rather than to changes in data availability or problem formulation.

By focusing on the same operational problem as Stage III while modifying only the modeling paradigm, Stage IV enables a controlled assessment of the marginal contribution of bioinspired deep learning models. This design directly supports the evaluation of whether highly expressive non-linear architectures provide systematic advantages in anticipating declaration abandonment, and whether such advantages justify the increased complexity and reduced inherent interpretability associated with deep learning approaches in institutional credit management contexts.

4.4.3 Data Preparation and Experimental Protocol

The data preparation and experimental protocol adopted in Stage IV strictly follows the methodological principles established in previous stages, ensuring full comparability of results and preventing the introduction of confounding factors. No new preprocessing strategies, data sources, or validation mechanisms are introduced at this stage. Instead, the protocol operationalizes the enriched data representation associated with declaration abandonment prediction under a controlled and reproducible experimental design.

All preprocessing steps were designed to address the intrinsic characteristics of real-world

institutional administrative data, including missing values, heterogeneous variable types, skewed distributions, and pronounced class imbalance. These steps were applied consistently across all experimental runs and partitions, preserving temporal validity and avoiding any form of information leakage between training, validation, and test subsets.

The experimental protocol is based on repeated stratified partitioning with fixed proportions and multiple random seeds, enabling a robust assessment of predictive performance, variability, and stability. This design supports statistical comparison between modeling paradigms while ensuring that observed differences reflect model behavior rather than artifacts of data partitioning or preprocessing variability.

The following subsections detail the dataset construction, preprocessing operations, imbalance handling, and partitioning strategy as applied specifically to Stage IV, all of which are inherited from the declaration abandonment modeling framework established earlier in the research.

Dataset Description and Cohort Definition

The dataset used in Stage IV corresponds to the institutional administrative records of the FSCU system previously constructed and validated in earlier stages of the research. No additional data sources or temporal extensions were incorporated. The dataset represents a real-world cohort of borrowers administered by PUCV, reflecting the operational conditions under which income declaration compliance is monitored.

The analytical cohort is restricted to borrowers who had submitted at least one valid income declaration, thereby satisfying the prerequisite condition for modeling declaration abandonment. This cohort definition ensures conceptual alignment between the predictive target and the available information, as declaration-related variables are only defined for borrowers with prior compliance history.

The resulting dataset integrates academic, administrative, financial, and declaration-related variables, capturing both antecedent characteristics and behavioral patterns emerging during the declaration process. This enriched feature space reflects the higher informational density associated with the abandonment prediction task and differentiates Stage IV from pre-declarative modeling stages.

All variables were retained in their original institutional naming convention to preserve traceability and semantic consistency with the source systems. The cohort construction process preserves temporal validity by ensuring that all predictor variables precede the abandonment event in time, thereby preventing retrospective bias.

This dataset definition provides the empirical foundation for evaluating the capacity of bioinspired deep learning models to capture complex and non-linear patterns associated with declaration discontinuity, while maintaining strict comparability with the datasets used in Stage III.

Missing Value Handling and Imputation Strategy

Missing value handling in Stage IV follows the same imputation strategy adopted for the declaration abandonment modeling framework established in Stage III. This decision ensures method-

ological consistency and prevents confounding effects when comparing predictive performance across modeling paradigms. No alternative imputation techniques or sensitivity analyses were introduced at this stage.

The presence of missing values is primarily associated with variables derived from administrative processes and historical records, reflecting data migrations and irregular institutional cases rather than systematic absence of information. Variables exhibiting missingness patterns deemed non-representative of standard FSCU operations were excluded during earlier data cleaning stages, resulting in a reduced feature space suitable for predictive modeling.

For the remaining variables with missing values, imputation was performed using the MissForest algorithm, a non-parametric method based on random forests that is well suited for datasets containing mixed numerical and categorical variables. This approach preserves multivariate relationships among variables and mitigates biases commonly introduced by univariate imputation techniques.

To preserve temporal validity and prevent information leakage, imputation models were fitted exclusively on the training subset within each stratified partition. The learned imputation mappings were then applied consistently to the corresponding validation and test subsets. This procedure was repeated independently for each of the 35 stratified partitions, ensuring that imputation variability was fully integrated into the robustness analysis.

Date-type variables were transformed into ordinal numerical representations prior to imputation to ensure compatibility with the algorithm. Following imputation, these variables were restored to their original temporal format and decomposed into calendar components, maintaining semantic interpretability while enabling effective numerical processing.

Overall, the imputation strategy adopted in Stage IV ensures that missing data are handled in a statistically principled and operationally realistic manner, supporting stable model training while maintaining strict consistency with the methodological assumptions defined in earlier stages.

Variable Transformation and Scaling

Variable transformation and scaling in Stage IV were performed following the same procedures defined for the declaration abandonment modeling framework in Stage III. No additional transformations or alternative normalization strategies were introduced, ensuring full comparability of results across stages.

Date-type variables were processed through a two-step transformation. Prior to imputation, calendar dates were converted into ordinal numerical representations to ensure compatibility with the imputation algorithm. After imputation, these variables were restored to a temporal format and decomposed into calendar components, such as year, month, and day, retaining their interpretability while enabling their use in numerical models.

Variables exhibiting highly skewed distributions were transformed to stabilize variance and reduce numerical dominance. In particular, the total debt variable was subjected to an adjusted logarithmic transformation, which preserves relative ordering while mitigating the influence of extreme values. This transformation contributes to more stable optimization dynamics during model training and prevents individual variables from disproportionately influencing the learning

process.

Given the sensitivity of deep neural networks to feature scale, all numerical variables were standardized using z-score normalization. The scaling parameters were estimated exclusively on the training subset of each stratified partition and subsequently applied to the corresponding validation and test subsets. This practice prevents information leakage and ensures that model evaluation reflects realistic deployment conditions.

By maintaining a consistent transformation and scaling strategy across all partitions and stages, Stage IV isolates the effect of model architecture and learning capacity from preprocessing-related variability. This design supports a controlled assessment of the contribution of bioinspired deep learning models relative to the optimized classical approaches evaluated in earlier stages.

Class Imbalance Management

Class imbalance management in Stage IV follows the same strategy adopted in the declaration abandonment modeling framework of Stage III. The target variable associated with declaration discontinuity exhibits a pronounced imbalance, with the minority class representing a substantially smaller proportion of observations. This imbalance reflects the empirical distribution of abandonment events within the FSCU system and constitutes an inherent characteristic of the operational problem.

To mitigate the adverse effects of class imbalance during model training, the Synthetic Minority Over-sampling Technique (SMOTE) was applied exclusively to the training subset of each stratified partition. The objective of this procedure was to balance class representation during learning, thereby reducing bias toward the majority class and improving the model's ability to detect abandonment events.

SMOTE generates synthetic instances of the minority class through interpolation between existing samples in feature space, allowing the learning algorithm to capture more representative decision boundaries without duplicating observations. In Stage IV, the oversampling ratio was adjusted to achieve a balanced 50:50 class distribution within the training set.

Importantly, the validation and test subsets were left unaltered and retained their original class proportions. This design ensures that model evaluation reflects realistic operational conditions and that reported performance metrics are not inflated by artificially balanced evaluation data.

By applying the same imbalance handling strategy across all stratified partitions and modeling approaches, Stage IV maintains methodological consistency with previous stages and enables a fair comparison between bioinspired deep learning models and the optimized classical models evaluated earlier in the research.

Stratified Partitioning and Robustness Design

The robustness design adopted in Stage IV is based on repeated stratified data partitioning, following the same experimental protocol defined for declaration abandonment prediction in previous stages. This strategy was selected to evaluate model stability and generalization capacity under different data configurations, while preserving the empirical class distribution inherent to

the FSCU dataset.

For each experimental run, the dataset was divided into three mutually exclusive subsets: training, validation, and testing. The proportions of these subsets were fixed and maintained across all runs, with stratification applied to ensure that the class distribution of the declaration abandonment outcome remained consistent in each partition. This approach prevents distortions in minority class representation and supports reliable performance estimation in imbalanced settings.

To account for variability induced by random data splitting, the partitioning process was repeated across 35 independent runs using distinct random seeds. Each run produced a new stratified split, and all preprocessing steps, including imputation, scaling, and class imbalance handling, were performed independently within each partition. This design integrates data-induced variability into the evaluation process and mitigates the risk of drawing conclusions based on favorable or atypical splits.

Model performance metrics were computed on the test subset corresponding to each partition, and aggregated statistics were subsequently derived to characterize central tendency, dispersion, and confidence intervals. By adopting a multi-seed robustness design, Stage IV ensures that reported results reflect consistent predictive behavior rather than isolated experimental outcomes.

This partitioning and robustness strategy enables a statistically grounded comparison between bioinspired deep learning models and the classical optimized models evaluated in earlier stages, supporting a controlled assessment of how modeling complexity influences predictive stability under real-world institutional conditions.

4.4.4 Modeling Framework

The modeling framework adopted in Stage IV is designed to evaluate the predictive capacity of bioinspired deep learning models under the same operational and experimental conditions defined for declaration abandonment prediction in previous stages. The framework preserves consistency in data preprocessing, validation strategy, and evaluation metrics, isolating the effect of model architecture and learning dynamics as the sole source of methodological variation.

At this stage, modeling decisions prioritize expressive capacity and non-linear representation learning over inherent interpretability. Bioinspired deep neural networks are employed as the primary modeling paradigm, enabling the capture of complex interactions among academic, administrative, financial, and declaration-related variables. These models are contrasted against a classical ensemble-based approach, implemented as a Random Forest baseline, to contextualize performance gains relative to a widely used non-linear classifier.

Hyperparameter configuration and model selection are treated as integral components of the modeling framework rather than as independent optimization objectives. Automated hyperparameter search mechanisms are used to explore architectural and training configurations, while representative configurations are subsequently fixed to ensure statistical comparability across repeated experimental runs.

Throughout this framework, strict temporal constraints are enforced to prevent information leakage, and all modeling procedures are applied independently within each stratified partition.

This design supports a controlled comparison between bioinspired deep learning and classical machine learning approaches, allowing the contribution of model expressiveness to be assessed in relation to predictive performance, robustness, and interpretability within the FSCU context.

Bioinspired Deep Neural Network Architecture

The primary modeling approach in Stage IV is based on a bioinspired deep neural network architecture designed to capture complex and non-linear patterns associated with income declaration discontinuity. The architecture is motivated by the need to model high-dimensional institutional data characterized by heterogeneous variables and intricate interdependencies that are not easily captured by linear or shallow models.

The deep neural network is implemented using a fully connected feedforward structure, composed of multiple hidden layers with non-linear activation functions. The architecture allows flexible depth and width configurations, enabling the model to adapt its representational capacity to the structure of the data. Regularization mechanisms, including dropout and L2 weight penalties, are incorporated to mitigate overfitting and promote generalization under repeated training.

Training is conducted using gradient-based optimization with binary cross-entropy as the loss function, reflecting the binary nature of the declaration abandonment outcome. Model convergence and generalization are monitored using validation-based criteria, ensuring that learning dynamics are guided by predictive performance rather than solely by training error minimization.

This bioinspired architecture deliberately relaxes interpretability constraints imposed in earlier stages, positioning Stage IV as an exploration of increased predictive expressiveness within the FSCU dataset under controlled experimental conditions. The resulting models serve as a benchmark for assessing whether deep learning architectures provide systematic advantages over optimized classical models when predicting declaration discontinuity.

Hyperparameter Optimization Strategy

The hyperparameter optimization strategy adopted in Stage IV is designed to systematically explore the architectural and training configurations of the bioinspired deep neural network, while controlling overfitting and computational cost. This strategy is treated as an internal model configuration mechanism rather than as an independent optimization objective, ensuring methodological coherence with earlier stages of the research.

Automated hyperparameter search was conducted using a bandit-based optimization framework, enabling efficient exploration of a high-dimensional configuration space. The search encompassed structural parameters, including the number of hidden layers and the number of units per layer, as well as training-related parameters such as activation functions, learning rates, optimization algorithms, dropout rates, and L2 regularization coefficients. These parameters were allowed to vary independently across layers, increasing the expressive flexibility of the model.

To prevent unnecessary computational expenditure on poorly performing configurations, early stopping and pruning mechanisms were integrated into the optimization process. Candidate

configurations exhibiting limited improvement in validation performance during early training epochs were discarded, allowing computational resources to be concentrated on more promising regions of the search space. Model selection within each trial was guided by validation-based performance, with the area under the ROC curve used as the primary optimization criterion.

Each hyperparameter configuration was trained for a bounded number of epochs under identical stopping criteria, ensuring comparability across trials. The optimization process yielded a collection of candidate models with associated configurations and validation performance records, which were subsequently analyzed to identify stable and recurrent architectural patterns across partitions.

This optimization strategy enables the identification of well-performing deep neural network configurations while mitigating overfitting risks and preserving robustness under repeated stratified evaluation, thereby supporting a fair comparison with classical baseline models in the subsequent analysis.

Representative Configuration Selection

Following the hyperparameter optimization process, a representative deep neural network configuration was selected to ensure statistical comparability across repeated experimental runs. Rather than evaluating each optimized configuration independently, the objective of this step was to identify a stable and recurrent architectural pattern that could be fixed and systematically assessed under the robustness design defined for Stage IV.

To this end, the set of optimal configurations obtained across the stratified partitions was analyzed using a normalization and grouping procedure. Continuous hyperparameter values were rounded to a fixed precision, while categorical parameters were standardized to a uniform representation. This process enabled the identification of equivalent configurations across different runs despite minor numerical variations.

The configuration exhibiting the highest frequency of occurrence among the optimized models was selected as representative. This model corresponds to a bioinspired deep neural network with three hidden layers, comprising 128, 128, and 64 units, respectively, and employing a hyperbolic tangent activation function. Layer-specific dropout rates and L2 regularization coefficients were applied to mitigate overfitting, and training was conducted using a gradient-based optimizer with a fixed learning rate and binary cross-entropy loss.

By fixing the representative configuration and retraining it across all stratified partitions, Stage IV ensures consistency in model structure and facilitates a statistically grounded comparison of performance metrics. This approach isolates variability due to data partitioning while preventing architectural differences from confounding the evaluation of predictive behavior.

Baseline Model: Random Forest

To contextualize the predictive performance of the bioinspired deep neural network, a Random Forest classifier was implemented as a classical non-bioinspired baseline under the same experimental framework. The Random Forest model serves as a point of reference, representing a widely adopted ensemble-based approach capable of modeling non-linear relationships while

maintaining relatively lower model complexity.

The baseline model was trained and evaluated using the same stratified partitions employed for the deep learning experiments. Identical preprocessing steps were applied, including variable transformations, scaling, and class imbalance handling, ensuring that performance differences could be attributed to modeling paradigms rather than to discrepancies in data preparation.

A fixed hyperparameter configuration was adopted for the Random Forest classifier, selected based on the most frequently occurring values among preliminary optimized configurations. This configuration was held constant across all experimental runs to preserve comparability and to avoid introducing additional sources of variability.

By evaluating the Random Forest model under the same robustness and validation protocol, Stage IV enables a controlled comparison between a highly expressive deep learning approach and a classical ensemble method. This comparison supports the analysis of trade-offs between predictive performance, stability, and model complexity within the context of declaration abandonment prediction.

4.4.5 Evaluation Metrics and Statistical Validation

The evaluation strategy adopted in Stage IV is consistent with that applied in previous stages and is specifically tailored to the challenges posed by imbalanced binary classification in institutional credit contexts. Model performance is assessed using a set of complementary metrics that jointly capture discrimination capacity, class-specific behavior, and agreement beyond chance.

The primary evaluation metric is the area under the receiver operating characteristic curve (AUC), which provides a threshold-independent measure of discriminative performance and is robust to class imbalance. In addition to AUC, precision, recall, and F1-score are reported to characterize the trade-off between false positives and false negatives, with particular emphasis on the correct identification of declaration abandonment events.

To further assess predictive reliability under imbalance, agreement-based metrics were incorporated, including Cohen's Kappa coefficient and the Matthews Correlation Coefficient (MCC). These metrics account for the full confusion matrix and provide a more informative assessment of classification quality than accuracy alone in skewed class distributions.

All metrics were computed on the test subset of each stratified partition. Aggregated statistics, including mean values, standard deviations, extrema, and confidence intervals, were derived across the 35 independent runs to characterize central tendency and variability. This multi-run evaluation supports robust inference regarding model stability and generalization behavior.

To assess whether observed performance differences between the bioinspired deep neural network and the Random Forest baseline are statistically significant, paired statistical tests were applied using matched test sets across partitions. This validation design ensures that statistical comparisons reflect systematic differences in model behavior rather than random variation induced by data splitting.

Together, these evaluation and validation procedures provide a rigorous and comparable basis for assessing predictive performance in Stage IV and for situating deep learning results within the broader cross-stage analysis developed later in the thesis.

4.4.6 Experimental Results

This section presents the empirical results obtained from the evaluation of the bioinspired deep neural network and the Random Forest baseline within the Stage IV experimental framework. Results are reported exclusively for the declaration abandonment prediction task and are derived from the repeated stratified evaluation protocol defined previously.

All reported outcomes correspond to test-set performance aggregated over the 35 independent stratified partitions. This design allows both central tendency and variability to be examined, providing a robust characterization of model behavior under different data configurations. The presentation of results emphasizes comparative analysis between modeling paradigms rather than isolated performance values.

The results are organized to progressively examine overall predictive performance, discrimination capacity, variability across runs, and class-wise behavior. This structure supports a systematic interpretation of how bioinspired deep learning models perform relative to a classical ensemble-based baseline when applied to enriched institutional data representations.

Importantly, no additional optimization, recalibration, or post-processing steps were applied beyond those defined in the experimental protocol. As a result, observed differences in performance can be attributed directly to the modeling approaches under consideration. The findings reported in this section provide the empirical basis for assessing the contribution of Stage IV with respect to Research Questions RQ2 and RQ4, and for contextualizing deep learning performance relative to the optimized classical models developed in earlier stages.

Predictive Performance with Confidence Intervals

The predictive performance of the models evaluated in Stage IV is summarized using aggregated test-set metrics computed across the 35 stratified partitions. Table 4.22 reports the mean values, dispersion statistics, and 95% confidence intervals for all evaluation metrics considered.

Model	Metrics	Mean	Desv. Std	Min	Max	IC 95%
DNN	AUC	0.9991	0.0003	0.9984	0.9996	0.9991 ± 0.0001
	Precision	0.9977	0.0007	0.9960	0.9989	0.9977 ± 0.0003
	Recall	0.9956	0.0012	0.9935	0.9984	0.9956 ± 0.0004
	F1 Score	0.9966	0.0006	0.9950	0.9980	0.9966 ± 0.0002
	Cohen's Kappa	0.9784	0.0040	0.9682	0.9870	0.9784 ± 0.0014
	Matthews Corr.	0.9784	0.0040	0.9683	0.9870	0.9784 ± 0.0014
RF	AUC	0.9709	0.0026	0.9647	0.9796	0.9709 ± 0.0009
	Precision	0.9714	0.0031	0.9651	0.9778	0.9714 ± 0.0011
	Recall	0.9290	0.0050	0.9202	0.9412	0.9290 ± 0.0017
	F1 Score	0.9497	0.0027	0.9464	0.9579	0.9497 ± 0.0009
	Cohen's Kappa	0.7122	0.0135	0.6939	0.7568	0.7122 ± 0.0046
	Matthews Corr.	0.7182	0.0134	0.6990	0.7621	0.7182 ± 0.0046

Table 4.22: Evaluation Metrics with 95% Confidence Intervals

The bioinspired deep neural network exhibits consistently high predictive performance across all evaluated metrics. Discrimination capacity, as measured by the AUC, remains consistently high with extremely low variability across partitions, indicating stable and robust separation between abandonment and compliance outcomes. Precision and recall values further indicate that the model effectively balances false positive and false negative errors, maintaining strong detection capability for declaration abandonment events without excessive misclassification of compliant borrowers.

Agreement-based metrics reinforce this observation. Both Cohen's Kappa and the Matthews Correlation Coefficient attain high mean values with narrow confidence intervals, suggesting a strong level of concordance between predicted and observed outcomes beyond chance. The limited dispersion observed across all metrics indicates that model performance is not driven by favorable data splits, but reflects consistent learning behavior under repeated evaluation.

In contrast, the Random Forest baseline demonstrates lower central performance values and noticeably wider confidence intervals across most metrics. While the baseline model maintains reasonable discrimination ability, its recall and agreement-based metrics indicate a reduced capacity to correctly identify abandonment events under the same experimental conditions.

Paired statistical comparisons between models, reported in Table 4.23, confirm that the observed performance differences are statistically significant across all evaluated metrics. These results provide quantitative evidence that the performance gains achieved by the bioinspired deep neural network are systematic rather than incidental.

Metric	t-value	p-value
AUC	62.71	0.00
Precision	47.56	0.00
Recall	83.48	0.00
F1 Score	107.66	0.00
Cohen's Kappa	118.66	0.00
Matthews Corr.	116.56	0.00

Table 4.23: Results of the paired Student's t-test between the bioinspired DNN and RF baseline models

Overall, the confidence interval analysis supports the conclusion that deep learning models offer a substantial and robust improvement in predictive performance for declaration abandonment when compared to classical ensemble-based approaches under identical data and evaluation conditions.

ROC Curve Analysis

The discrimination behavior of the evaluated models is further examined through receiver operating characteristic (ROC) curve analysis. Figures 4.33 and 4.34 present the mean ROC curves obtained across the 35 stratified partitions for the bioinspired deep neural network and the Random Forest baseline, respectively.

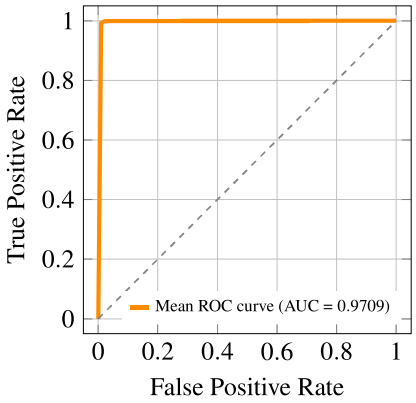


Figure 4.33: Mean ROC Curve of the bioinspired DNN model

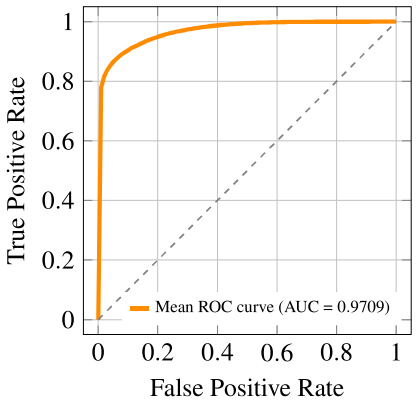


Figure 4.34: Mean ROC Curve of the RF baseline model

The ROC curve associated with the bioinspired deep neural network remains consistently close to the upper-left corner of the ROC space across the entire range of false positive rates. This behavior reflects a high true positive rate even under stringent decision thresholds, indicating strong and stable discrimination between declaration abandonment and continued compliance. The limited deviation observed across partitions is consistent with the narrow confidence intervals reported for the AUC metric.

In contrast, the ROC curve of the Random Forest model, while clearly outperforming random classification, exhibits a visibly lower trajectory throughout the ROC space. This pattern indicates reduced sensitivity to abandonment events at comparable false positive rates, particularly in regions where decision thresholds are operationally relevant for institutional intervention.

The separation between the two ROC curves provides a visual complement to the quantitative performance analysis, reinforcing the conclusion that the bioinspired deep neural network achieves superior discrimination capacity under identical evaluation conditions. Importantly, these differences are observed without altering class distributions in the evaluation data, ensuring that the ROC analysis reflects realistic operational scenarios.

Overall, the ROC curve analysis corroborates the statistical performance results and highlights the ability of deep learning models to maintain high discrimination performance across a broad range of decision thresholds in the declaration abandonment prediction task.

Distribution and Variability of Performance Metrics

The stability and variability of model performance across repeated experimental runs are examined through the distributional analysis of evaluation metrics. Figures 4.35 to 4.36 present boxplots summarizing the dispersion of metric values obtained over the 35 stratified partitions for both the bioinspired deep neural network and the Random Forest baseline.

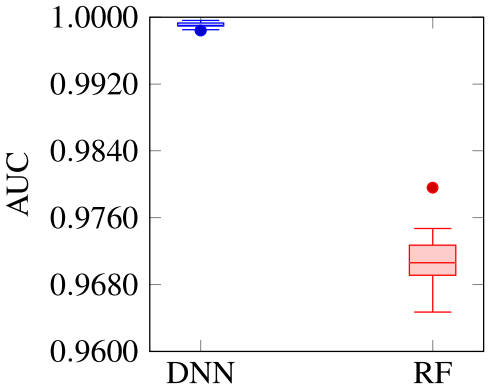


Figure 4.35: Boxplot AUC

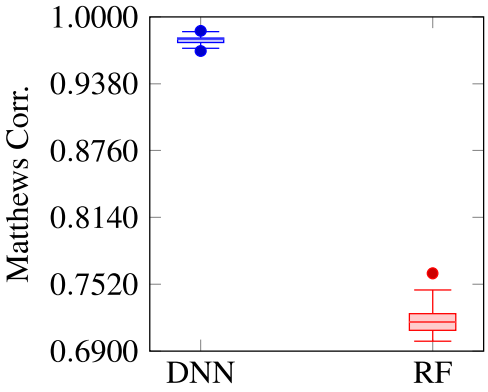


Figure 4.36: Boxplot Matthews Correlation

Across all evaluated metrics, the bioinspired deep neural network exhibits highly concentrated distributions with minimal dispersion. The boxplots corresponding to AUC, precision, recall, and F1-score are tightly clustered toward high metric values, indicating consistent performance across different data partitions. The limited interquartile ranges and short whiskers suggest that model behavior is largely invariant to the specific composition of the training and test subsets.

Agreement-based metrics display a similar pattern. The distributions of Cohen's Kappa and the Matthews Correlation Coefficient for the deep neural network remain narrow and centered at high values, reinforcing the conclusion that predictive consistency extends beyond threshold-dependent measures and reflects robust agreement between predicted and observed outcomes.

In contrast, the Random Forest baseline shows noticeably wider distributions across all metrics. The broader interquartile ranges and increased variability observed in the boxplots indicate greater sensitivity to data partitioning, particularly for recall and agreement-based metrics. This variability suggests less stable detection of declaration abandonment events under changing data configurations.

Overall, the distributional analysis confirms that the performance advantages of the bioinspired deep neural network are not confined to average metrics, but are consistently maintained across repeated experimental runs. This stability is particularly relevant in institutional contexts, where predictive reliability under varying operational conditions is a critical requirement.

Class-wise Classification Analysis

Class-wise classification behavior is examined through aggregated confusion matrix statistics computed across the 35 stratified partitions. Tables 4.24 and 4.25 report the mean values, dispersion, and observed ranges for each element of the confusion matrix corresponding to the bioinspired deep neural network and the Random Forest baseline, respectively.

Element	Mean	Desv. Est.	Minimum	Maximum
True Negative (TN)	676.37	2.74	670	681
False Positive (FP)	8.63	2.74	4	15
False Negative (FN)	16.49	4.64	6	24
True Positive (TP)	3692.49	4.63	3685	3703

Table 4.24: Confusion matrix of the bioinspired DNN model

Element	Mean	Desv. Est.	Minimum	Maximum
True Negative (TN)	583.60	11.58	560	606
False Positive (FP)	101.40	11.58	79	125
False Negative (FN)	263.51	18.46	218	296
True Positive (TP)	3445.46	18.44	3413	3491

Table 4.25: Confusion matrix of the RF baseline model

For the bioinspired deep neural network, the confusion matrix statistics indicate a strong balance between sensitivity to declaration abandonment events and control of false alarms. The number of true positives remains consistently high across partitions, while false negatives are limited and exhibit low variability. This pattern reflects the model's ability to correctly identify borrowers at risk of abandoning the declaration process without systematically missing positive cases. At the same time, the relatively small number of false positives indicates that compliant borrowers are rarely misclassified as abandonment cases.

In contrast, the Random Forest baseline shows a markedly different class-wise behavior. While true positives remain substantial, the number of false negatives is considerably higher, indicating reduced sensitivity to abandonment events. Additionally, the number of false positives increases significantly relative to the deep learning model, suggesting a higher likelihood of incorrectly flagging compliant borrowers. Both error types exhibit greater dispersion across partitions, pointing to increased instability in class-specific predictions.

The class-wise analysis highlights the operational implications of the observed performance differences. In the context of FSCU administration, false negatives correspond to missed opportunities for early intervention, while false positives may lead to unnecessary administrative actions. The reduced incidence of both error types achieved by the bioinspired deep neural network underscores its practical advantage for supporting anticipatory decision-making under real-world institutional constraints.

Overall, the confusion matrix analysis complements global and distributional performance metrics by revealing how predictive gains translate into concrete improvements in class-specific decision outcomes, directly addressing the predictive objectives associated with declaration abandonment.

4.4.7 Model Interpretability Analysis

Given the limited inherent interpretability of bioinspired deep neural networks, Stage IV incorporates post-hoc interpretability mechanisms to examine whether the learned representations align with institutionally meaningful variables. The objective of this analysis is not to establish causal relationships, but to assess the internal consistency of the model and to identify variables that systematically contribute to predictions of declaration abandonment.

Interpretability analysis is conducted exclusively for the bioinspired deep neural network, as the classical baseline models evaluated in previous stages rely on more transparent decision mechanisms. The analysis focuses on global attribution patterns derived from repeated experimental runs, enabling the identification of stable explanatory signals across partitions rather than instance-specific explanations.

By integrating interpretability analysis into Stage IV, the research maintains continuity with the explanatory objectives defined in earlier stages and contributes empirical evidence relevant to Research Question RQ4, extending variable relevance analysis to highly expressive modeling paradigms.

Feature Attribution using Integrated Gradients

Feature attribution was assessed using the Integrated Gradients method applied to the bioinspired deep neural network. This technique quantifies the contribution of each input variable to the model's output by integrating gradients along a path from a neutral reference point to the observed input, enabling consistent attribution in fully differentiable architectures.

Attributions were computed across all stratified partitions and aggregated to obtain stable global importance estimates. Figure 4.37 presents the variables exhibiting the highest absolute mean differences in attribution between the abandonment and compliance classes, reflecting their

differential contribution to model predictions.

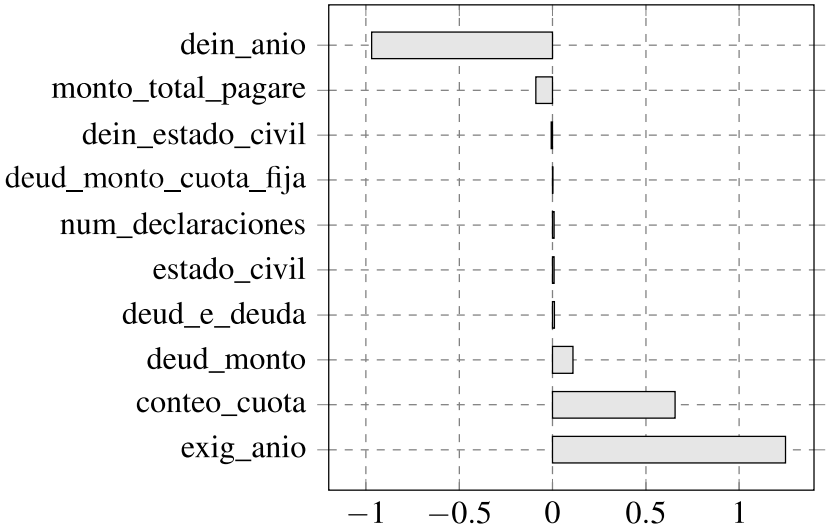


Figure 4.37: Feature Importance (Integrated Gradients)

The attribution results indicate that variables associated with debt structure, installment progression, and administrative timelines dominate the explanatory profile of the model. Temporal indicators related to debt exigibility and declaration history exhibit particularly strong contributions, suggesting that the deep neural network leverages longitudinal and structural information to distinguish between continued compliance and abandonment behavior.

Conversely, demographic variables display minimal attribution magnitude, indicating a limited influence on predictive outcomes within the model. This pattern is consistent with the interpretability findings reported in earlier stages and suggests that predictive performance is primarily driven by financial and administrative variables rather than by personal characteristics.

Overall, the Integrated Gradients analysis provides evidence that the bioinspired deep neural network focuses on institutionally relevant variables when predicting declaration discontinuity, supporting the interpretability and credibility of the model despite its high representational complexity.

4.4.8 Comparative Analysis with Previous Stages

Stage IV is positioned as the final empirical extension of the stage-based methodological framework, introducing a fundamentally different modeling paradigm while preserving strict methodological continuity with previous stages. The comparative analysis with earlier stages is therefore grounded in differences in modeling strategy rather than in changes in data availability, predictive targets, or evaluation design.

From a methodological standpoint, Stage IV differs from Stages II and III primarily in the class of predictive models employed. Whereas Stage II relies on interpretable classical machine learning models operating under strictly pre-declarative constraints, and Stage III

introduces metaheuristic-based feature selection to manage high-dimensional post-declarative representations, Stage IV adopts bioinspired deep learning architectures designed to learn complex non-linear representations directly from enriched institutional data.

Importantly, Stage IV maintains full consistency with Stage III in terms of predictive task, dataset composition, preprocessing pipeline, class imbalance treatment, and validation strategy. The declaration abandonment target, feature availability, temporal constraints, and robustness design remain unchanged. As a result, differences observed between Stages III and IV can be attributed exclusively to the modeling paradigm itself, rather than to confounding methodological or informational factors.

In contrast with Stage III, where dimensionality reduction and performance gains are achieved through explicit optimization of feature subsets guided by metaheuristic search, Stage IV shifts the locus of complexity from the feature space to the model architecture. Feature selection is no longer an explicit methodological component; instead, representational learning is embedded within the deep neural network structure. This distinction enables a direct methodological comparison between optimization-driven parsimony and representation-driven expressiveness under identical experimental conditions.

Relative to Stage II, Stage IV represents a deliberate relaxation of interpretability and transparency constraints. While Stage II establishes the earliest feasible and most interpretable predictive reference under informational scarcity, Stage IV explores predictive modeling under conditions where richer representations and increased model complexity are admissible. The comparison between these stages therefore illustrates the methodological trajectory of the thesis, from early transparent prediction to advanced modeling under enriched data regimes.

This stage is motivated by the empirical boundary identified in Stage III, where optimization-driven feature selection under interpretability constraints exhibits diminishing marginal returns. Stage IV explicitly examines whether additional predictive improvements can be obtained only by shifting the modeling paradigm toward higher representational expressiveness, rather than through further refinement within classical or optimization-based frameworks.

Overall, the comparative positioning of Stage IV clarifies its role within the methodological architecture of the research. Rather than replacing earlier approaches, this stage complements them by empirically characterizing the conditions under which increased model expressiveness yields substantive predictive and robustness benefits, thereby informing evidence-based methodological choices for institutional deployment.

4.4.9 Summary of Stage IV Results

Stage IV evaluates the applicability of bioinspired deep learning models to the prediction of income declaration abandonment among previously compliant borrowers under the same operational and methodological constraints defined in Stage III. By preserving identical datasets, preprocessing procedures, class imbalance handling, and validation protocols, this stage enables a controlled assessment of deep learning models within the established experimental framework, directly addressing Research Question RQ4 and Hypothesis H5.

The results indicate that the bioinspired deep neural network achieves high and stable predictive performance across repeated stratified evaluations. Discrimination metrics and class-

wise error analyses show that the model effectively identifies declaration abandonment events while maintaining controlled misclassification rates for compliant borrowers. Performance dispersion across runs remains limited, supporting the robustness of the observed outcomes under the adopted evaluation design.

When contrasted with classical ensemble-based models evaluated under identical conditions, the deep learning approach exhibits systematic performance improvements. These gains are reflected not only in average metric values but also in reduced variability across validation partitions, indicating improved stability in complex and imbalanced prediction scenarios. Such behavior is particularly relevant in institutional contexts where consistent decision support is prioritized over isolated peak performance.

Interpretability analysis based on Integrated Gradients reveals that the deep learning model predominantly relies on financial, administrative, and temporal variables associated with debt structure and declaration dynamics. Demographic attributes play a secondary role in the predictive process, a pattern that aligns with explanatory findings reported in earlier stages. This convergence indicates that increased model expressiveness does not fundamentally alter the substantive drivers of prediction, but rather refines their interactions within a more flexible representational framework.

From a methodological perspective, the results of Stage IV demonstrate the effects of relaxing explicit feature selection and interpretability constraints in favor of representation learning. While this shift yields measurable gains in predictive performance and stability, it also introduces increased model complexity. These findings provide a concrete empirical basis for evaluating the trade-offs between predictive accuracy, robustness, transparency, and operational applicability, which are synthesized in the cross-stage analysis developed in the subsequent chapter.

4.4.10 Empirical Summary

This chapter presented the empirical results of the proposed stage-based methodological framework, progressively addressing predictive modeling for income declaration compliance under increasing levels of data availability and modeling complexity. Across stages, the results illustrate how early-stage interpretable models establish a causally grounded baseline, how metaheuristic optimization enhances predictive robustness in high-dimensional post-compliance contexts, and how bioinspired deep learning further improves performance and stability when interpretability constraints are relaxed.

Taken together, these findings delineate the empirical capabilities and limitations associated with each modeling paradigm, highlighting the trade-offs between predictive accuracy, robustness, transparency, and operational applicability. The evidence reported in this chapter provides the empirical foundation for the cross-stage synthesis developed in the following chapter, where the implications of these trade-offs are examined in relation to the research questions, hypotheses, and the broader institutional decision-making context.

5 Discussion

Chapter Overview

This chapter synthesizes and interprets the empirical findings obtained across the methodological stages of the research. Rather than re-reporting results, the discussion integrates evidence from Stages II to IV in order to identify cross-stage patterns, performance boundaries, and methodological trade-offs associated with progressively increasing data richness, feature dimensionality, and model complexity.

The stage-based design adopted in this thesis enables a structured interpretation of how predictive capacity, interpretability, robustness, and methodological choices evolve across distinct yet interrelated modeling scenarios. In this sense, the discussion chapter functions as an integrative and interpretative stage of the research, contextualizing the results within the proposed methodological framework and the broader body of related literature.

5.1 Cross-Stage Interpretation of Results

This section provides an integrated interpretation of the empirical evidence obtained across Stages II to IV, adopting a cross-stage analytical perspective rather than a stage-by-stage recounting of results. The objective is to synthesize how predictive behavior, methodological effectiveness, and modeling trade-offs evolve as progressively richer information, higher feature dimensionality, and more expressive modeling paradigms are introduced.

The analysis is grounded in the stage-based methodological design of the research, which intentionally structures the experimentation as a sequence of increasingly complex yet interrelated scenarios. As such, this section does not seek to identify a single optimal model or configuration, but rather to characterize the conditions under which different methodological choices become appropriate, necessary, or limiting from both a predictive and institutional standpoint.

Specifically, the discussion focuses on four complementary dimensions that emerge consistently across stages. First, it examines the feasibility and reliability of predictive modeling under strict informational constraints, as established in early-stage scenarios. Second, it analyzes the impact of feature space enrichment and optimization mechanisms on predictive robustness and parsimony. Third, it explores the trade-offs between predictive performance, stability, and interpretability that arise as methodological complexity increases. Finally, it contrasts the empirical behavior of distinct modeling paradigms, ranging from interpretable classical models to optimization-enhanced approaches and deep learning architectures.

By integrating these dimensions, the cross-stage interpretation delineates empirical performance boundaries and methodological transition points that are not observable when stages are analyzed in isolation. This synthesis provides the conceptual foundation for revisiting the research questions, situating the findings within the existing literature, and deriving theoretical, methodological, and practical implications in the subsequent sections of this chapter.

5.1.1 Predictive Feasibility under Information Constraints

A central finding emerging from the cross-stage analysis is that meaningful predictive modeling is feasible even under strict informational constraints. The results of Stage II demonstrate that first-time income declaration compliance can be anticipated using exclusively pre-declarative academic and administrative data, without reliance on behavioral feedback, post-declarative variables, or optimization-based feature selection. This finding establishes the empirical viability of early-stage prediction at the earliest feasible decision point within the institutional process.

From a methodological perspective, this result is significant because it confirms that predictive signal exists prior to any observed declaration behavior, thereby preserving causal precedence and preventing information leakage. The ability to achieve stable discrimination under these constraints supports the methodological choice of isolating Stage II as a transparency- and causality-oriented baseline rather than as a performance-maximizing exercise. In this context, predictive feasibility is not defined by peak accuracy, but by the reliability and interpretability of the signal extracted under realistic operational conditions.

The cross-stage comparison further reveals that the feasibility observed in Stage II is not an artifact of model choice or experimental configuration. Instead, consistent predictive behavior across multiple classical and interpretable model families indicates that early-stage compliance is structurally associated with antecedent institutional characteristics, such as academic trajectory, financial exposure, and enrollment history. This stability suggests that early compliance behavior reflects underlying institutional and socioeconomic patterns rather than transient or purely stochastic effects.

Importantly, the results of subsequent stages do not invalidate the feasibility established under information constraints, but rather contextualize it. While Stages III and IV demonstrate that predictive performance can be improved when richer behavioral information and more complex modeling paradigms are introduced, the cross-stage analysis indicates that such improvements build upon, rather than replace, the foundational predictive capacity identified in Stage II. In this sense, early-stage feasibility constitutes a necessary methodological prerequisite for later-stage refinement.

Overall, the evidence supports the conclusion that predictive modeling under strict information constraints is both empirically viable and methodologically justified in institutional credit systems. This finding has direct implications for early intervention and policy design, as it indicates that meaningful risk assessment can be conducted before declaration behavior is observed, thereby enabling proactive decision-making while maintaining transparency and causal coherence.

5.1.2 Impact of Feature Space Enrichment and Optimization

The cross-stage analysis reveals that the enrichment of the feature space and the introduction of systematic optimization mechanisms have a substantive impact on predictive behavior once informational constraints are relaxed. In Stage III, the incorporation of declaration history, income dynamics, and repayment-related variables transforms the predictive task from a constrained, antecedent-based problem into a high-dimensional and behaviorally informed scenario. This

transition marks a qualitative shift in both data representation and methodological requirements.

From an empirical standpoint, the enriched feature space introduces a level of dimensionality and heterogeneity that exceeds the practical limits of purely semantic or institutionally driven feature curation. While Stage II demonstrates that early-stage prediction is feasible using a compact and interpretable set of antecedent variables, the results of Stage III indicate that such manual curation becomes increasingly insufficient as behavioral information accumulates. Under these conditions, redundancy, collinearity, and noise emerge as critical challenges that directly affect predictive robustness.

The introduction of metaheuristic-based feature selection in Stage III addresses these challenges by explicitly formulating feature selection as an optimization problem. Rather than maximizing predictive performance in isolation, the optimization process jointly considers discrimination capacity, stability under class imbalance, and feature parsimony. The cross-stage comparison shows that this approach leads to systematic improvements relative to non-optimized baselines operating on the same enriched data representations, confirming the methodological necessity of controlled optimization in high-dimensional post-compliance contexts.

Importantly, the benefits of feature space enrichment are not realized through indiscriminate exploitation of additional variables. Instead, the optimized feature subsets consistently reduce dimensionality while preserving, and in many cases enhancing, predictive robustness. This pattern indicates that performance gains arise from selective refinement of informative features rather than from the mere availability of richer data. As such, feature space enrichment and optimization function as complementary mechanisms rather than as independent contributors.

At the same time, the cross-stage evidence suggests that the impact of optimization exhibits diminishing returns once redundancy is effectively controlled and interpretability constraints are maintained. While Stage III establishes that metaheuristic feature selection is essential for managing enriched representations, the results also delineate an empirical boundary beyond which further optimization within the same modeling paradigm yields limited additional benefit. This boundary directly motivates the exploration of alternative modeling strategies examined in subsequent stages.

Overall, the cross-stage interpretation confirms that feature space enrichment fundamentally alters the nature of the predictive problem and necessitates the adoption of optimization-driven methodologies. The combined use of enriched representations and systematic feature selection enhances predictive robustness and parsimony, while also clarifying the methodological limits of optimization-based refinement under controlled interpretability requirements.

5.1.3 Trade-offs between Performance, Stability, and Interpretability

A key insight emerging from the cross-stage interpretation is that predictive performance, stability, and interpretability are inherently interdependent objectives whose balance shifts as methodological complexity increases. The stage-based design of the research makes these trade-offs explicit by progressively relaxing informational and modeling constraints while preserving a common evaluation framework.

In Stage II, interpretability and causal coherence are prioritized by design. Predictive models operate on a compact and semantically curated set of antecedent variables, yielding stable

and transparent behavior across experimental configurations. Although absolute performance is constrained by limited information availability, the results demonstrate that reliable discrimination can be achieved with minimal variance, reinforcing the suitability of this stage for early institutional decision-making under strict transparency requirements.

Stage III introduces a controlled rebalancing of these objectives. The enrichment of the feature space and the application of metaheuristic-based feature selection lead to measurable performance gains and improved robustness under class imbalance. However, these gains are achieved through increased methodological complexity, including iterative optimization procedures and higher-dimensional representations. Importantly, interpretability is not abandoned but redefined: rather than relying on direct feature inspection alone, explanatory coherence is preserved through the stability and institutional plausibility of the optimized feature subsets.

In Stage IV, the trade-off shifts more decisively toward predictive expressiveness and stability. By adopting bioinspired deep learning models, this stage relaxes explicit interpretability constraints in favor of representational learning, enabling the capture of complex non-linear relationships that are inaccessible to classical or optimization-enhanced models. While this approach yields additional gains in performance and consistency across partitions, it also necessitates post hoc interpretability techniques and increased computational resources, highlighting the cost of enhanced expressiveness.

The cross-stage comparison indicates that no single modeling strategy simultaneously optimizes performance, stability, and interpretability across all contexts. Instead, each stage occupies a distinct position within a continuum defined by these competing objectives. Early-stage models favor transparency and causal alignment, intermediate-stage models balance robustness and parsimony through optimization, and advanced-stage models prioritize expressive capacity under relaxed interpretability constraints.

Overall, these findings underscore that methodological choices in predictive modeling should be guided by explicit consideration of the intended decision context and institutional requirements. The stage-based framework demonstrates that increases in performance and stability are accompanied by corresponding trade-offs in interpretability and complexity, reinforcing the need for evidence-based selection of modeling strategies rather than universal optimization.

5.1.4 Comparative Behavior across Modeling Paradigms

The stage-based design of this research enables a direct and methodologically controlled comparison across distinct predictive modeling paradigms, ranging from interpretable classical machine learning models to optimization-enhanced approaches and bioinspired deep learning architectures. By progressively introducing model complexity while preserving consistent data representations and evaluation protocols within each stage, the cross-stage analysis reveals systematic differences in predictive behavior that extend beyond isolated performance metrics.

Classical machine learning models, as employed in Stage II, exhibit stable and transparent behavior under strict informational constraints. Their predictive capacity relies on linear or moderately non-linear relationships between antecedent institutional variables and declaration outcomes, resulting in consistent discrimination and low variance across experimental configurations. This paradigm demonstrates particular suitability for early-stage decision-making scenarios

where transparency, causal alignment, and ease of interpretation are critical.

The introduction of metaheuristic-based feature selection in Stage III modifies the behavior of classical models by enabling adaptive exploration of high-dimensional feature spaces. Rather than altering the underlying model class, optimization reshapes the effective representation on which prediction operates. The resulting models exhibit improved robustness and parsimony, with reduced sensitivity to redundancy and noise. This paradigm occupies an intermediate methodological position, balancing interpretability and stability while leveraging optimization to manage enriched behavioral information.

Bioinspired deep learning models, introduced in Stage IV, represent a qualitative shift in predictive behavior. These models learn hierarchical and non-linear representations directly from the data, enabling them to capture complex interactions among financial, administrative, and temporal variables. As a result, deep learning exhibits superior expressive capacity and improved stability in imbalanced and behaviorally rich scenarios. However, this paradigm also introduces increased computational complexity and relies on post hoc interpretability techniques, reflecting a departure from the transparency-oriented design of earlier stages.

The comparative analysis indicates that differences across modeling paradigms are not merely incremental but structural. Each paradigm embodies distinct assumptions about representation, optimization, and explanation, leading to characteristic strengths and limitations. Importantly, the results suggest that paradigm selection should be guided by the informational context, interpretability requirements, and operational constraints of the target application, rather than by absolute performance considerations alone.

Overall, the cross-stage comparison demonstrates that predictive modeling in institutional credit systems benefits from a principled progression across paradigms. Interpretable models establish feasibility under constraints, optimization-enhanced models extend robustness in high-dimensional settings, and deep learning models provide an empirical reference for the highest performance observed under the same data and evaluation protocol. Together, these paradigms form a coherent methodological spectrum that underpins the stage-based design of the research.

5.2 Discussion by Research Question

This section revisits each research question in light of the integrated empirical evidence obtained across the methodological stages of the thesis. Rather than addressing individual experimental outcomes in isolation, the discussion emphasizes how the results collectively contribute to answering the research questions, clarifying their scope, limitations, and methodological implications. The analysis draws on the cross-stage interpretation developed in the previous section to ensure consistency between empirical findings, methodological design, and research objectives.

5.2.1 Discussion of RQ1

Research Question 1 investigates whether machine learning models can accurately anticipate first-time income declaration non-compliance using exclusively pre-declarative institutional data. The evidence obtained in Stage II provides a clear and affirmative answer to this question,

demonstrating that meaningful predictive signal is present prior to any observed declaration behavior.

The results indicate that antecedent academic, administrative, and financial variables contain sufficient information to support reliable discrimination between compliant and non-compliant borrowers at the earliest feasible decision point. Importantly, this predictive feasibility is achieved under strict causal and temporal constraints, without reliance on post-declarative information, behavioral feedback, or optimization-driven feature selection. This finding confirms that early-stage prediction is not only technically possible but also methodologically sound within realistic institutional settings.

From a methodological standpoint, the findings associated with RQ1 highlight the value of prioritizing interpretability and causal coherence over maximal performance in early decision-making contexts. The stability observed across classical and interpretable model families suggests that early non-compliance is structurally linked to underlying institutional characteristics rather than to model-specific artifacts. This reinforces the robustness of the conclusion and supports the use of transparent models for initial risk assessment and policy design.

The cross-stage analysis further contextualizes the answer to RQ1 by showing that subsequent improvements observed in later stages do not negate the feasibility established at this stage. While richer information and more complex modeling paradigms enhance predictive performance in post-compliance scenarios, these advances build upon the foundational signal identified under pre-declarative constraints. As such, RQ1 defines a necessary baseline that underpins the methodological progression of the thesis.

Overall, the discussion of RQ1 confirms that early prediction of first-time income declaration behavior is both empirically viable and institutionally relevant. This conclusion provides a critical justification for proactive intervention strategies and establishes the methodological foundation upon which the remaining research questions are addressed.

5.2.2 Discussion of RQ2

Research Question 2 examines how effectively predictive models can anticipate income declaration abandonment among borrowers who have previously complied with declaration requirements. The empirical evidence from Stage III confirms that abandonment constitutes a learnable and behaviorally distinct predictive problem when historical declaration and repayment information is incorporated. This supports the rationale that post-compliance scenarios expose longitudinal dynamics that are not observable under strictly pre-declarative regimes, and it provides direct evidence in favor of Hypothesis H2 under real-world institutional data conditions.

From a methodological standpoint, the cross-stage synthesis indicates that the shift from first-time compliance (Stage II) to abandonment prediction (Stage III) is not a simple change of label but a substantive change in behavioral signal. In this setting, the predictive task becomes more sensitive to temporal patterns, debt structure, and prior declaration trajectories. Consequently, predictive feasibility remains high, but the conditions for robustness and generalization become more demanding, particularly under class imbalance and high-dimensional representations.

Overall, the answer to RQ2 is affirmative: abandonment can be anticipated with robust discrimination when enriched institutional data is available. At the same time, the cross-stage

analysis clarifies that the quality of this anticipation depends on controlling redundancy and instability introduced by richer feature spaces, which motivates the methodological choices evaluated in RQ3 and RQ4.

5.2.3 Discussion of RQ3

Research Question 3 investigates how metaheuristic-based feature selection influences predictive performance, stability, and interpretability in the prediction of declaration abandonment. The evidence from Stage III indicates that feature selection functions as a methodological necessity in high-dimensional post-compliance contexts rather than as a marginal enhancement. In particular, optimization-driven selection improves imbalance-aware performance and robustness while reducing feature dimensionality, thereby supporting Hypothesis H3.

The cross-stage interpretation suggests that the main contribution of feature selection is not merely performance maximization but the stabilization of learning under heterogeneous and redundant representations. By refining the effective representation used by classical models, the optimization process reduces sensitivity to noise and collinearity and yields more compact predictors that remain institutionally interpretable. This aligns with the stage-based design objective of preserving explanatory plausibility while introducing methodological escalation only where empirically justified.

At the same time, the integrated evidence points to diminishing returns once redundancy is controlled and interpretability constraints remain in place. This empirical boundary is methodologically relevant because it defines a transition point: beyond this point, additional gains are more likely to require increased representational capacity rather than further optimization within the same modeling family.

5.2.4 Discussion of RQ4

Research Question 4 assesses whether bioinspired deep learning models improve predictive performance in the prediction of income declaration abandonment when compared to optimized classical models under identical data representations and evaluation protocols. Under this controlled comparison, Stage IV provides evidence that deep learning can yield systematic gains in discrimination and stability, supporting Hypothesis H5 within the operational constraints of the study.

The cross-stage synthesis indicates that these gains are best interpreted as a consequence of representation learning rather than as an artifact of data availability or evaluation design, which are held constant relative to Stage III. In this sense, Stage IV clarifies that once feature redundancy is controlled and classical models approach a performance plateau, additional improvements can emerge from models capable of capturing higher-order non-linear interactions among financial, administrative, and temporal variables.

However, the discussion also highlights that these improvements come with increased computational cost and a shift in the interpretability approach, from inherent transparency to post hoc explanation. Therefore, the contribution of Stage IV is not to replace earlier paradigms but to delineate when the additional complexity of deep learning becomes empirically justified for

abandonment prediction.

5.2.5 Discussion of RQ5

Research Question 5 examines which academic, financial, and administrative variables are consistently associated with declaration compliance and non-compliance behaviors across different stages of the income declaration process. Across Stages II–IV, the integrated evidence suggests a stable semantic pattern: variables reflecting financial exposure and institutional trajectory repeatedly emerge as dominant drivers of predictive behavior, while purely demographic attributes tend to play a secondary role. This supports Hypothesis H4 by indicating that a reduced and institutionally interpretable set of variables can capture the dominant explanatory structure across stages.

Methodologically, this cross-stage consistency is important because it indicates that increases in model expressiveness do not necessarily change the substantive determinants of prediction. Instead, richer models appear to refine how these determinants interact. In practical terms, this supports the feasibility of designing institutional decision support around a compact explanatory core, while reserving highly expressive models for contexts where incremental predictive gains justify higher complexity.

Overall, RQ5 is answered by identifying a consistent explanatory backbone across stages, which strengthens both the interpretability claims and the operational plausibility of the stage-based framework.

5.3 Comparison with Prior Studies

Rather than pursuing a direct metric-by-metric comparison with individual prior studies, this section contrasts the empirical and methodological patterns observed in the present research with dominant approaches identified in the state of the art. The comparison is grounded in the systematic review and critical synthesis developed in Chapter 2, and focuses on differences in problem formulation, data usage, modeling paradigms, optimization strategies, and evaluation practices. This perspective enables situating the contributions of the thesis within the broader literature while accounting for institutional constraints, stage-based design, and decision-support requirements that are rarely addressed jointly in prior work.

5.3.1 Comparison with Machine Learning-Based Credit Risk Studies

The majority of machine learning-based credit risk studies reported in the literature focus on predicting default, delinquency, or repayment failure using cross-sectional or short-horizon datasets, typically framed as static classification problems. Within this dominant paradigm, supervised learning approaches prevail, with frequent use of Logistic Regression, Decision Trees, Random Forests, Support Vector Machines, and Gradient Boosting variants, often evaluated through aggregate discrimination metrics such as AUC or accuracy [15, 41, 53, 54, 57].

In contrast, the present research departs from this dominant framing in several fundamental ways. First, the predictive target addressed in this thesis is not default per se, but income

declaration compliance within an income-contingent credit system. This distinction is not merely semantic: declaration compliance represents an institutional behavior embedded in a longitudinal administrative process rather than a terminal financial outcome. While a limited number of prior studies have explored student loan or education-related credit contexts, these works typically focus on default or repayment outcomes and do not explicitly model declaration behavior as a distinct compliance process [43, 51].

Second, whereas prior work frequently leverages extensive financial and behavioral variables without strict temporal separation, the stage-based design adopted here explicitly enforces causal and informational constraints. In particular, Stage II demonstrates that meaningful prediction is feasible using exclusively pre-declarative institutional data, an aspect that is rarely isolated or evaluated explicitly in existing credit risk literature. Most reviewed studies implicitly combine antecedent and post-event variables, potentially conflating predictive signal with behavioral feedback [15, 70]. By contrast, the controlled early-stage scenario examined in this thesis provides empirical evidence that predictive signal can be extracted prior to any observed compliance behavior.

Third, existing machine learning-based credit risk research often emphasizes performance maximization through increasingly complex models, with limited attention to interpretability, stability, or institutional plausibility of the predictors. Although ensemble methods and hybrid frameworks dominate the state of the art due to their predictive power [41, 48, 51], comprehensive assessments that jointly consider performance, robustness, and interpretability remain relatively uncommon. The present research addresses this gap by explicitly structuring the empirical analysis around these dimensions and by examining how their balance evolves across stages.

Finally, while deep learning approaches have gained visibility in credit risk modeling as competitors to classical and ensemble-based methods [36, 42, 55], they are often evaluated without controlled comparison against optimized classical baselines or under limited experimental protocols. In contrast, Stage IV positions bioinspired deep learning as an an empirical upper bound of predictive expressiveness under the given data and evaluation protocol, preserving identical data representations, validation strategies, and evaluation metrics. The results indicate that deep learning provides incremental gains over optimized classical models in behaviorally rich post-compliance scenarios, while maintaining consistency with explanatory patterns identified in earlier stages [15].

Overall, the comparison highlights that the contribution of this thesis lies not in outperforming individual credit risk studies on isolated benchmarks, but in reframing compliance prediction as a staged, institutionally grounded process. By integrating causal constraints, optimization mechanisms, and paradigm-level comparisons within a unified methodological framework, the present research extends machine learning-based credit risk modeling beyond static default prediction toward a more nuanced and operationally relevant understanding of borrower behavior.

5.3.2 Comparison with Feature Selection and Optimization Approaches

Feature selection and optimization techniques have been widely explored in the literature as mechanisms for improving predictive performance, reducing dimensionality, and mitigating overfitting in machine learning models applied to credit risk and related domains. Prior studies

commonly adopt filter, wrapper, or embedded strategies, frequently combined with heuristic or metaheuristic search methods such as Genetic Algorithms, Particle Swarm Optimization, Differential Evolution, or hybrid evolutionary frameworks [15, 41, 53, 54]. In most cases, the effectiveness of these techniques is assessed primarily through aggregate performance gains relative to non-selected baselines.

While the present research builds upon this methodological tradition, it departs from much of the existing literature in both the role assigned to optimization and the manner in which its effects are evaluated. In a large proportion of prior studies, feature selection is introduced as an auxiliary enhancement applied to a fixed predictive task, often simultaneously with changes in model architecture, hyperparameter tuning, or data preprocessing. As a result, the individual contribution of optimization is frequently conflated with other methodological factors, making it difficult to attribute observed improvements specifically to feature selection mechanisms [48, 51].

In contrast, Stage III of this thesis positions feature selection as a core methodological necessity rather than as a secondary refinement. Optimization is explicitly motivated by the transition to enriched, high-dimensional post-declarative feature spaces, where manual or semantically driven feature curation becomes impractical. The experimental design isolates the effect of metaheuristic-based feature selection by comparing optimized and non-optimized models operating on identical data representations, evaluation protocols, and validation strategies. This design choice directly addresses a limitation identified in prior optimization-based studies, where the lack of controlled comparisons often obscures causal interpretation of optimization effects [15].

Empirically, the results of Stage III show that optimization-driven feature selection yields consistent improvements in imbalance-aware metrics while substantially reducing feature dimensionality. Importantly, these gains are achieved through selective refinement rather than through indiscriminate exploitation of additional variables, a distinction that is not always made explicit in existing studies. The stability of the selected feature subsets across optimization runs further contrasts with prior work reporting highly variable or unstable feature selections, particularly in high-dimensional financial datasets [41, 54].

Another point of divergence concerns interpretability. Although feature selection is often justified in the literature as a mechanism for enhancing interpretability, selected feature subsets are rarely examined for institutional, behavioral, or domain-level coherence. In the present research, the optimized feature sets identified in Stage III remain consistently aligned with substantive drivers of declarative behavior, including income evolution, repayment exposure, and temporal compliance patterns. This continuity suggests that metaheuristic optimization refines, rather than obscures, the explanatory structure identified in earlier stages, addressing concerns raised in prior studies regarding the opacity of optimization-based selection processes [53].

Finally, whereas many optimization-based studies implicitly assume that predictive performance scales monotonically with increased search depth or algorithmic sophistication, the results of this thesis delineate a practical boundary for optimization under interpretability constraints. Once redundancy and noise are effectively controlled, further optimization within the same modeling paradigm yields diminishing returns. This empirical boundary is rarely discussed explicitly in benchmark-oriented literature, yet it plays a central role in motivating the paradigm shift explored in Stage IV, where representational learning is introduced as an alternative to

continued optimization [15].

Overall, this comparison indicates that the contribution of the present research lies in redefining the role of feature selection and optimization within a staged and institutionally grounded framework. By treating optimization as a response to data complexity rather than as a generic performance booster, and by evaluating its effects under controlled comparative conditions, this work advances a more principled and operationally meaningful use of optimization techniques in predictive modeling for institutional credit systems.

5.3.3 Methodological Contributions beyond Existing Benchmarks

The majority of machine learning-based credit risk studies reported in the literature focus on predicting default, delinquency, or repayment failure using cross-sectional or short-horizon datasets, typically framed as static classification problems. These studies commonly employ classical supervised learning models, including logistic regression, decision trees, random forests, gradient boosting, and ensemble-based methods, with performance primarily evaluated through aggregate discrimination metrics such as AUC or accuracy [18, 23, 41, 54]. However, these benchmark-oriented formulations typically abstract away from institutional timing, causal precedence, and staged information availability, limiting their applicability in operational credit systems.

In contrast, the present research departs from this dominant framing in several fundamental ways. First, the predictive target addressed in this thesis is not default per se, but income declaration compliance within an income-contingent credit system. As emphasized in the literature on income-contingent loans, declarative compliance constitutes a procedural and behavioral requirement embedded in a longitudinal administrative process rather than a terminal financial outcome [19, 20, 24]. Consequently, the modeling task addressed here is inherently process-oriented and temporally structured, differing from the static formulations commonly adopted in credit risk studies.

Second, while prior work frequently leverages extensive financial and behavioral variables without strict temporal separation, the stage-based design adopted in this thesis explicitly enforces causal and informational constraints. In particular, Stage II demonstrates that meaningful prediction is feasible using exclusively pre-declarative institutional data, a condition that is rarely isolated or evaluated explicitly in existing credit risk literature [15, 18]. This controlled early-stage scenario provides empirical evidence that predictive signal can be extracted prior to any observed compliance behavior, an aspect largely absent from traditional default prediction studies.

Third, existing machine learning-based credit risk research often prioritizes performance maximization through increasingly complex models, with limited attention to interpretability, stability, or institutional plausibility of predictors [24, 26]. By contrast, the present research systematically examines how predictive performance evolves under increasing methodological intervention, explicitly balancing interpretability, robustness, and complexity across stages. This structured progression enables the identification of empirical boundaries that remain obscured in single-stage or single-paradigm studies.

Finally, whereas deep learning approaches in credit risk are frequently introduced as direct performance competitors to classical models, Stage IV positions bioinspired deep learning as an

upper bound of predictive expressiveness within a controlled comparative framework. Consistent with recent findings in student loan and institutional credit contexts [15, 58], the results show that deep learning provides incremental gains over optimized classical models in behaviorally rich post-compliance scenarios, while preserving alignment with explanatory patterns identified in earlier stages.

Overall, this comparison indicates that the contribution of the present thesis lies not in outperforming individual credit risk studies on isolated benchmarks, but in reframing compliance prediction as a staged, institutionally grounded process. By integrating causal constraints, optimization mechanisms, and paradigm-level comparisons within a unified methodological framework, this research extends machine learning-based credit risk modeling beyond static default prediction toward a more operationally and institutionally relevant understanding of borrower behavior.

Taken together, the comparative analysis with prior studies highlights that the contributions of this research are not limited to incremental performance improvements, but reside in the way predictive modeling is methodologically structured, evaluated, and interpreted across evolving institutional scenarios. By moving beyond benchmark-centric comparisons and emphasizing staged methodological reasoning, the thesis provides a clearer basis for understanding when specific modeling strategies are appropriate and what trade-offs they entail. These insights naturally extend beyond empirical evaluation and motivate a discussion of their theoretical, methodological, and practical implications, which are addressed in the following section. In this sense, the proposed framework complements existing benchmarks by shifting the analytical focus from isolated algorithmic performance to structured methodological reasoning under institutional constraints.

5.4 Implications

This section discusses the implications of the findings from theoretical, methodological, and practical perspectives. Rather than extrapolating beyond the empirical evidence, the implications are derived directly from the observed cross-stage patterns and the methodological structure of the research, emphasizing their relevance for both academic inquiry and institutional decision-making.

5.4.1 Theoretical Implications

From a theoretical standpoint, this research contributes to the understanding of compliance-related behavior in institutional credit systems by demonstrating that such behavior is structured, learnable, and progressively revealed across different informational regimes. The stage-based results indicate that early compliance decisions are not purely reactive or stochastic, but are systematically associated with antecedent academic, administrative, and financial characteristics observable prior to any declaration event. This finding aligns with prior conceptualizations of income-contingent loan systems as institutionally mediated environments in which borrower behavior emerges from the interaction between socioeconomic conditions and administrative

processes rather than from isolated financial outcomes [18, 20].

The findings further suggest that declaration abandonment among previously compliant borrowers constitutes a behaviorally distinct phenomenon rather than a simple extension of first-time non-compliance. The differential predictive structures observed across stages support the theoretical distinction between initial compliance decisions and longitudinal disengagement processes, reinforcing arguments in the literature that emphasize the need to model compliance as a dynamic and multi-stage behavioral process rather than as a static binary outcome [21–23].

In addition, the cross-stage analysis contributes to theoretical discussions on model complexity and explanatory sufficiency in institutional decision-support systems. The results indicate that increased predictive expressiveness does not fundamentally alter the substantive drivers of compliance behavior, but rather refines how their interactions are represented within the model. This observation supports theoretical perspectives that emphasize structural continuity in behavioral determinants, even as modeling paradigms evolve from interpretable classical approaches to more expressive architectures [20, 24].

Overall, the research advances a theoretically grounded view of institutional compliance prediction as a multi-stage phenomenon, in which different informational and behavioral layers become salient at different points in time. This perspective extends existing theoretical frameworks by explicitly linking behavioral staging, information availability, and modeling strategy, thereby providing a coherent basis for analyzing compliance-related risk in income-contingent student loan systems.

5.4.2 Methodological Implications

Methodologically, the primary implication of this research lies in the validation of a staged modeling framework as a principled alternative to single-shot or benchmark-driven predictive studies. By explicitly aligning modeling strategies with information availability, interpretability requirements, and institutional timing, the proposed framework enables more coherent methodological reasoning and more meaningful empirical comparison. This approach directly addresses limitations identified in prior studies, where heterogeneous data regimes and modeling choices are often combined without explicit consideration of causal precedence or deployment constraints [23, 24].

The results further demonstrate that optimization techniques, particularly metaheuristic-based feature selection, should not be treated as universally beneficial enhancements. Instead, their utility emerges conditionally, becoming essential only when feature dimensionality and heterogeneity exceed the limits of semantic or institutionally guided curation. This finding refines conclusions reported in prior optimization-focused studies, which frequently emphasize performance gains without explicitly accounting for data regime or methodological saturation [21, 22]. The evidence presented here cautions against the indiscriminate application of optimization and highlights the importance of aligning methodological tools with the structural complexity of the data.

Another methodological implication concerns the treatment of interpretability. Consistent with recent methodological discussions in applied machine learning, the results confirm that interpretability is not a binary attribute but a spectrum that evolves across modeling stages [20].

Maintaining explanatory coherence requires different strategies depending on model complexity, ranging from transparent model structures and direct feature inspection in early stages to stability-based explanations and post hoc attribution techniques in later stages. Explicitly managing this evolution improves both methodological clarity and empirical validity.

Finally, the consistent use of imbalance-aware metrics, robustness analysis, and repeated stratified validation across stages underscores the importance of evaluation design as a methodological component in its own right. The findings reinforce observations from prior empirical studies indicating that conclusions about model superiority or suitability are highly sensitive to evaluation choices, particularly in imbalanced and institutionally constrained settings [18]. As such, carefully controlled and context-aware experimental protocols are essential for producing reliable and operationally meaningful results.

5.4.3 Practical and Institutional Implications

From a practical perspective, the findings of this research provide actionable insights for institutional decision-making within income-contingent credit systems. The demonstrated feasibility of early-stage prediction under strictly pre-declarative conditions indicates that institutions can perform meaningful risk assessment before declaration behavior is observed, enabling proactive intervention strategies that preserve transparency and causal coherence. This result is consistent with the institutional literature reviewed in Chapter 2, which emphasizes the importance of anticipatory monitoring mechanisms in administratively mediated credit systems where compliance precedes observable repayment outcomes [19, 20, 24].

The results further suggest that different operational objectives warrant different modeling strategies. Interpretable models based on pre-declarative academic and administrative data are particularly suitable for early warning systems, policy design, and communication with stakeholders, where transparency and institutional plausibility are central. In contrast, optimization-enhanced and deep learning models are better suited for backend risk monitoring, portfolio-level analysis, and resource allocation once richer behavioral information becomes available. This differentiated deployment strategy aligns with prior evidence highlighting the need to adapt predictive tools to specific institutional functions and stages of borrower engagement [18, 23].

Importantly, the staged findings highlight that increased model complexity should be introduced selectively and purposefully. While advanced models offer improved performance and stability in complex and imbalanced scenarios, they also entail higher computational cost, reduced transparency, and greater implementation complexity. In institutionally governed credit systems, such trade-offs directly affect auditability, longitudinal deployment, and stakeholder trust, reinforcing the need to balance predictive gains against operational and governance constraints [20, 22, 24].

Overall, the research supports an evidence-based approach to institutional deployment in which modeling strategies are selected not solely on the basis of peak predictive performance, but in alignment with decision timing, data availability, administrative capacity, and governance requirements. By providing empirical evidence across multiple stages and modeling paradigms, this thesis offers a practical foundation for designing adaptive decision-support systems that evolve coherently with institutional data maturity and operational needs.

5.5 Limitations

Despite the contributions of this research, several limitations must be acknowledged. First, the analysis is based on data from a single institutional credit system, which may limit the direct generalizability of the findings to other contexts with different regulatory frameworks, borrower populations, or administrative processes.

Second, the temporal structure of the data introduces potential cohort and period effects. While the staged design explicitly accounts for information availability over time, changes in institutional policy, economic conditions, or borrower behavior across cohorts may influence the observed patterns and warrant further longitudinal investigation.

Third, the use of metaheuristic optimization and deep learning models entails non-negligible computational costs. Although these methods yield robustness and performance gains in high-dimensional settings, their scalability and operational feasibility may be constrained in resource-limited environments.

Finally, while interpretability is addressed through multiple mechanisms across stages, post hoc explanation methods employed in advanced models do not provide the same level of transparency as inherently interpretable structures. This limitation reflects a broader trade-off in contemporary predictive modeling and underscores the importance of aligning model choice with institutional accountability requirements.

These limitations do not undermine the validity of the findings, but rather delineate the scope within which the conclusions should be interpreted. They also motivate future research directions aimed at extending the proposed framework to other institutional settings, exploring adaptive and resource-aware optimization strategies, and further refining explainability mechanisms for complex models.

6 Conclusion and future work

Chapter Overview

This chapter concludes the doctoral research by revisiting the general objective of the study and synthesizing the main outcomes obtained across the full set of methodological stages. Rather than reiterating empirical results, the chapter integrates the evidence generated throughout the research to articulate its overall contributions and significance.

Specifically, the chapter summarizes the scientific, methodological, and practical contributions of the thesis, critically reflects on the scope and limitations of the work, and outlines structured and coherent directions for future research. In doing so, it positions the proposed stage-based framework as a transferable foundation for predictive modeling in institutional and compliance-driven contexts.

6.1 Global Synthesis of the Research

This doctoral research set out to design, validate, and apply predictive models capable of anticipating income declaration compliance behavior within an institutional income-contingent credit system, under realistic informational, methodological, and operational constraints. To achieve this objective, the study adopted a stage-based methodological framework that progressively increased data richness, feature dimensionality, and model expressiveness while preserving strict temporal validity and institutional plausibility.

The research process unfolded through a sequence of empirically and conceptually connected stages. Early stages focused on establishing predictive feasibility under strictly pre-declarative conditions, demonstrating that meaningful and stable predictive signal exists prior to any observed declaration behavior. These stages emphasized interpretability, causal coherence, and transparency, thereby validating early-stage prediction as both technically viable and institutionally defensible.

Subsequent stages extended the analysis to post-declarative scenarios characterized by higher data dimensionality and behavioral complexity. The incorporation of declaration history and repayment dynamics transformed the predictive problem into a high-dimensional setting in which manual or purely semantic feature curation proved insufficient. In this context, metaheuristic-based feature selection emerged as a necessary methodological response, enabling systematic control of redundancy and noise while preserving institutional interpretability.

The final empirical stage explored the limits of predictive expressiveness by introducing bioinspired deep learning architectures under controlled experimental conditions. By maintaining identical datasets, preprocessing pipelines, and validation protocols, this stage isolated the effect of model paradigm on predictive behavior. The results showed that deep learning provides additional gains in performance and stability in behaviorally rich scenarios, thereby establishing an empirical upper bound within the proposed framework.

Across all stages, the research did not pursue isolated performance maximization, but rather a principled examination of how predictive capacity, robustness, interpretability, and methodological complexity interact under evolving informational regimes. The global synthesis of results demonstrates that no single modeling strategy dominates across all contexts. Instead, the appropriateness of a given approach depends on decision timing, data availability, institutional constraints, and explanatory requirements.

Taken together, the stage-based framework enabled a coherent and cumulative exploration of income declaration behavior, transforming a complex institutional process into a structured analytical problem. This synthesis confirms that predictive modeling for compliance in income-contingent credit systems is not a monolithic task, but a multi-stage endeavor in which different methodological choices become relevant at different points in time. This insight provides the foundation for articulating the scientific, methodological, and practical contributions of the thesis, which are detailed in the following section.

6.2 Summary of Contributions

This section synthesizes the main contributions of the doctoral research, organized into practical dimensions, scientific and methodological. These contributions emerge from the cumulative evidence generated across the stage-based framework and reflect advances that go beyond isolated empirical findings, encompassing conceptual, analytical, and operational insights relevant to predictive modeling in institutional credit systems.

6.2.1 Practical and Institutional Contributions

From an institutional perspective, this research provides concrete and operationally viable decision-support capabilities for universities administering income-contingent credit systems such as the FSCU. In the context of institutions such as PUCV, which administers the FSCU in practice, the proposed predictive frameworks enable the early identification of borrowers at risk of income declaration non-compliance and abandonment, under strictly realistic informational constraints. This supports proactive, evidence-based management strategies that improve financial planning, reduce administrative uncertainty, and strengthen the long-term financial sustainability and regulatory compliance of income-contingent credit administration.

From a practical standpoint, the study demonstrates the feasibility of deploying predictive analytics under pre-declarative and post-declarative regimes, preserving causal coherence and interpretability while achieving meaningful predictive performance.

The findings further support a differentiated deployment strategy aligned with operational objectives. Interpretable models based on antecedent data are well suited for early warning systems, policy design, and stakeholder communication, while optimization-enhanced and deep learning models are more appropriate for backend risk monitoring and resource allocation once richer behavioral information becomes available. This staged perspective enables institutions to match modeling complexity to decision timing and organizational capacity.

Importantly, the research highlights that increased model complexity entails trade-offs

related to computational cost, transparency, and implementation effort. By empirically characterizing these trade-offs across stages, the thesis provides institutions with an evidence-based basis for selecting modeling strategies that balance predictive gains against operational constraints and explainability requirements.

Overall, the practical contribution of this research lies in demonstrating that effective institutional deployment of predictive models is not achieved by selecting a single best-performing algorithm, but by adopting a structured and context-aware modeling strategy. The stage-based framework offers a transferable blueprint for integrating predictive analytics into institutional decision-making processes where data availability, regulatory considerations, and operational objectives evolve over time.

6.2.2 Scientific Contributions

From a scientific perspective, this research advances the understanding of compliance-related behavior in income-contingent credit systems by demonstrating that such behavior is structured, learnable, and progressively revealed across different informational regimes. The results provide empirical evidence that income declaration compliance is not a purely reactive or stochastic outcome, but is systematically associated with underlying academic, administrative, and financial characteristics observable at different stages of the borrower lifecycle.

A key scientific contribution lies in the explicit differentiation between first-time declaration compliance and declaration abandonment among previously compliant borrowers. The findings show that these phenomena exhibit distinct predictive structures and behavioral dynamics, supporting the need to model them as separate processes rather than as variants of a single static outcome. This distinction contributes to a more nuanced theoretical characterization of compliance behavior in longitudinal institutional settings.

In addition, the research demonstrates that increasing model expressiveness does not fundamentally alter the substantive drivers of compliance behavior, but rather refines the representation of their interactions. Across stages, core financial, academic, and administrative factors remain consistently relevant, even as modeling paradigms evolve. This structural continuity contributes to theoretical discussions on explanatory sufficiency and supports the view that behavioral determinants retain coherence across different analytical resolutions.

Overall, the thesis contributes a scientifically grounded interpretation of institutional compliance as a multi-stage process, in which different layers of information become salient over time and require differentiated analytical treatment.

6.2.3 Methodological Contributions

Methodologically, the central contribution of this research is the design and validation of a stage-based predictive modeling framework that aligns methodological choices with information availability, institutional timing, and interpretability requirements. Rather than adopting a single-shot or benchmark-driven approach, the proposed framework structures predictive modeling as a sequence of progressively complex yet interrelated stages, enabling controlled and meaningful comparison across modeling strategies.

A major methodological contribution concerns the principled integration of optimization techniques. The research demonstrates that metaheuristic-based feature selection should not be treated as a universally beneficial enhancement, but as a conditional methodological response to increased feature dimensionality and heterogeneity. By isolating the effects of optimization through direct comparisons between optimized and non-optimized models operating on identical data representations, the thesis clarifies when and why optimization becomes necessary, and where its benefits reach diminishing returns.

Another contribution lies in the explicit treatment of interpretability as a continuum rather than a binary property. Early stages prioritize transparent model structures and direct explanations, intermediate stages preserve explanatory coherence through stable optimized feature subsets, and later stages rely on post hoc attribution mechanisms within highly expressive models. This explicit management of interpretability across stages provides a realistic and methodologically sound approach to balancing explanation and performance in applied predictive modeling.

Finally, the research contributes a robust and replicable evaluation framework. The consistent use of imbalance-aware metrics, repeated stratified validation, and robustness analysis across stages ensures that observed differences in performance and stability are attributable to methodological choices rather than to evaluation artifacts. This design strengthens the internal validity of the findings and supports methodological reasoning beyond isolated performance comparisons.

6.3 Scope and Limitations of the Study

This section complements the interpretative limitations discussed in Chapter 5 by delineating the global scope and structural constraints within which the thesis as a whole should be interpreted.

The findings of this doctoral research should be interpreted within the scope defined by its institutional context, data availability, and methodological design. The study is grounded in real-world administrative data from a single income-contingent credit system, which enabled a detailed and operationally realistic analysis of income declaration behavior, but also constrains the external generalizability of the results. While the proposed stage-based framework is conceptually transferable, empirical outcomes may vary when applied to credit systems operating under different regulatory, socioeconomic, or administrative conditions.

A second limitation relates to the temporal structure of the available data. Although the research explicitly distinguishes between pre-declarative and post-declarative informational regimes and incorporates longitudinal elements in later stages, the observation window remains bounded by the institutional records available at the time of analysis. Longer follow-up periods could reveal additional behavioral patterns, delayed effects, or structural changes in compliance dynamics that are not fully captured in the present study.

Methodological limitations are also acknowledged. The modeling strategies adopted in each stage were selected to balance interpretability, robustness, and predictive performance under realistic deployment constraints. As a result, certain classes of models or optimization techniques were deliberately excluded when they conflicted with causal validity, transparency requirements, or computational feasibility. While this design choice strengthens institutional plausibility, it also

implies that alternative modeling configurations could yield different empirical trade-offs.

Finally, computational considerations impose practical limits on the breadth of experimentation. Metaheuristic optimization and deep learning models entail substantial computational cost, which constrained the exploration of larger hyperparameter spaces, additional architectures, or more extensive ensemble strategies. Although robustness was ensured through repeated stratified evaluation, further experimentation could refine or extend the observed performance boundaries.

These limitations do not undermine the validity of the conclusions drawn, but rather contextualize them. They highlight the importance of interpreting the results as empirically grounded within a specific institutional and methodological setting, and they motivate several of the future research directions outlined in the following section.

6.4 Future Research Directions

Building on the findings and limitations of this research, several coherent and complementary directions for future work emerge. These directions are aligned with the stage-based framework proposed in the thesis and aim to extend its applicability, depth, and relevance across broader institutional and analytical contexts.

6.4.1 Cross-Institutional and Contextual Validation

A natural extension of this research involves validating the proposed framework across multiple institutions and credit systems. Applying the stage-based modeling approach to income-contingent loan programs operating under different regulatory regimes, socioeconomic conditions, or administrative structures would enable assessment of its external validity and robustness. Such cross-institutional studies could identify which behavioral patterns and predictive structures are system-specific and which are transferable across contexts.

In addition, comparative analyses across countries or educational systems could shed light on how policy design, repayment rules, and enforcement mechanisms influence compliance behavior. This line of research would contribute to a more general theory of institutional compliance prediction and support evidence-based policy adaptation.

6.4.2 Temporal and Longitudinal Extensions

Future work could extend the temporal scope of the analysis by incorporating longer longitudinal trajectories and finer-grained temporal modeling. This includes examining delayed compliance effects, cyclical declaration patterns, and transitions between compliance states over extended time horizons. Advanced time-aware modeling techniques, such as survival analysis or sequence-based learning approaches, could complement the existing framework while preserving temporal validity.

Such extensions would deepen the understanding of compliance as an evolving process rather than a sequence of isolated events, enabling more precise identification of critical intervention points throughout the borrower lifecycle.

6.4.3 Adaptive and Hybrid Optimization Strategies

Another promising direction concerns the development of adaptive and hybrid optimization strategies that dynamically adjust their behavior during the modeling process. Building on the evidence that optimization exhibits diminishing returns under certain conditions, future research could explore meta-optimization schemes in which optimization intensity, search strategy, or objective weighting adapts in response to data characteristics or convergence behavior.

Hybrid approaches combining metaheuristic feature selection with representation learning or adaptive control mechanisms could further refine the balance between parsimony, robustness, and predictive performance. Such strategies would extend the current framework by introducing an additional layer of methodological adaptivity.

6.4.4 Governance, Fairness, and Policy-Oriented Modeling

Finally, future research should address governance, fairness, and policy-oriented dimensions of predictive modeling in institutional credit systems. This includes systematic evaluation of potential bias across demographic or socioeconomic groups, assessment of fairness-aware modeling constraints, and integration of explainability requirements aligned with regulatory and ethical standards.

Extending the framework to incorporate explicit fairness objectives or policy-driven constraints would enhance its relevance for real-world deployment, particularly in public or semi-public institutions. This direction would bridge predictive analytics with responsible AI practices and support the development of transparent, accountable, and socially sustainable decision-support systems.

Together, these future research directions position the proposed framework as a foundation for continued investigation into predictive modeling for institutional compliance, enabling both theoretical advancement and practical impact beyond the specific context examined in this thesis.

6.5 Final Remarks

The findings of this doctoral research confirm the relevance of data-driven approaches for addressing complex operational challenges in higher education finance, particularly within income-contingent credit systems. By leveraging real-world institutional data and adopting a staged methodological perspective, the thesis bridges the gap between advances in predictive analytics and their practical application in compliance-oriented decision-making processes.

Beyond empirical performance, the research demonstrates that effective institutional deployment of predictive models requires careful alignment between data availability, methodological choices, interpretability requirements, and decision timing. The proposed framework shows that predictive modeling in such contexts is not a matter of identifying a single optimal algorithm, but of selecting and integrating modeling strategies that are appropriate to evolving informational and operational conditions.

While the framework was developed and validated within a specific institutional setting, the principles underlying its design are transferable to comparable credit and compliance scenarios

characterized by longitudinal processes, regulatory constraints, and heterogeneous data sources. In this sense, the thesis contributes a structured and adaptable approach to predictive modeling that supports both scientific inquiry and responsible institutional practice.

7 References

[1] Biblioteca del Congreso Nacional de Chile. Ley 19.287. establece normas sobre fondos solidarios de crédito universitario, 1994. Base legal del FSCU (creación de fondos, reglas básicas).

[2] Biblioteca del Congreso Nacional de Chile. Decreto n° 225 (1994) del ministerio de educación. reglamento de la ley n° 19.287, 1994. Reglamento: otorgamiento y devolución de créditos FSCU.

[3] Ministerio de Educación de Chile. Fondo solidario de crédito universitario (fscu), 2025. Requisitos socioeconómicos y académicos; cobertura; universo CRUCH.

[4] Ministerio de Educación de Chile. ¿qué es el fuas? (presentación 2025), 2025. Umbral PAES 485 para acceso a crédito.

[5] ChileAtiende, Gobierno de Chile. Fondo solidario de crédito universitario (fscu), 2025. Modalidad en UTM, tasa 2%; información operativa.

[6] Subsecretaría de Educación Superior (MINEDUC) and University College London. Reforma a los créditos estudiantiles en chile, 2025. Síntesis oficial: tasa 2%, tope pago anual 5% del ingreso, condonación temporal.

[7] Consejo de Rectoras y Rectores de las Universidades Chilenas (CRUCH). Deudores morosos de fondo solidario de crédito universitario (publicaciones anuales), 2025.

[8] Superintendencia de Educación Superior. Norma de carácter general n°3: Registros y contabilidad del fscu, 2024.

[9] Subsecretaría de Educación Superior, MINEDUC. Primer informe cae: Características de la población deudora e impactos, 2022.

[10] Comisión Ingresa (Administrador CAE). Estadísticas cae: Disminuyen las nuevas asignaciones y se observan aumentos en la morosidad (2024). `https://portal.ingresa.cl/estadisticas/estadisticas/`, 2024.

[11] Consejo de Rectoras y Rectores de las Universidades Chilenas. Universidades cruch a lo largo de chile. `https://consejoderectores.cl/el-consejo/universidades-cruch/`, 2025. Last accessed: 2025-10-16.

[12] Biblioteca Nacional del Congreso de Chile. Ley fondos solidatiros de crédito universitario. `https://www.bcn.cl/leychile/navegar?idNorma=30654`, 1994. Last accessed: 2025-10-16.

[13] Biblioteca Nacional del Congreso de Chile. Modificación ley fondos solidatiros de crédito universitario. https://www.bcn.cl/leychile/navegar?idNorma=1036996&idParte=9235355&idVersion=2012-02-04, 2012. Last accessed: 2025-10-16.

[14] Pontificia Universidad Católica de Valparaíso. Fondo solidario de crédtio universitario. https://estudiantespucv.cl/fscu/, 2025. Last accessed: 2025-10-16.

[15] Álex Paz, Broderick Crawford, Eric Monfroy, José Barrera-García, Álvaro Peña Fritz, Ricardo Soto, Felipe Cisternas-Caneo, and Andrés Yáñez. Machine learning and metaheuristics approach for individual credit risk assessment: A systematic literature review. *Biomimetics*, 10(5):326, 2025.

[16] Paz, Álex and Crawford, Broderick and Monfroy, Eric and Rodriguez-Tello, Eduardo and Barrera-García, José and Cisternas-Caneo, Felipe and Cortés, Benjamín López and Lazo, Yoslandy and Yáñez, Andrés and Peña Fritz, Álvaro and Soto, Ricardo. Interpretable binary classification under constraints for financial compliance modeling. *Mathematics*, 14(3), 2026.

[17] Yoslandy Lazo, Álex Paz, Broderick Crawford, Carlos Valle, Eduardo Rodriguez-Tello, Ricardo Soto, José Barrera-Garcia, Felipe Cisternas-Caneo, and Benjamín López Cortés. Bioinspired deep neural networks for predicting income-reporting discontinuities in the chilean student loan program. *Biomimetics*, 11(2), 2026.

[18] Narcisa Roxana Mosteanu. Improving quality of online teaching finance and business management using artificial intelligence and backward design. *Quality - Access to Success*, 23(187), April 2022.

[19] William F Crittenden, Isabella K Biel, and William A Lovely, III. Embracing digitalization: Student learning and new technologies. *J. Mark. Educ.*, 41(1):027347531882089, December 2018.

[20] Mateus De Oliveira Fornasier. Legal education in the 21st century and the artificial intelligence. *Rev. Opinião Juríd. (Fortaleza)*, 19(31):1, May 2021.

[21] Guillermo Alexander Quezada Castro, María del Pilar Castro Arellano, and María del Pilar Quezada Castro. Inteligencia artificial y enseñanza del derecho: su incorporación durante la pandemia de la covid-19. *Rev. Venez. Gerenc.*, 27(Especial 8):750–764, November 2022.

[22] Alfonso Renato Vargas-Murillo, Ilda Nadia Monica de la Asuncion Pari-Bedoya, and Francisco de Jesús Guevara-Soto. Challenges and opportunities of AI-assisted learning: A systematic literature review on the impact of ChatGPT usage in higher education. *Int. J. Learn. Teach. Educ. Res.*, 22(7):122–135, July 2023.

[23] Raluca-Giorgiana (Popa) Chivu, Ionuț-Claudiu Popa, Mihai-Cristian Orzan, Cristian Marinescu, Margareta Stela Florescu, and Anca-Olguța Orzan. The role of blockchain

technologies in the sustainable development of students' learning process. *Sustainability*, 14(3):1406, January 2022.

[24] Mark Treve. What COVID-19 has introduced into education: challenges facing higher education institutions (HEIs). *High. Educ. Pedagog.*, 6(1):212–227, January 2021.

[25] Douglas Newton, Yuqian (linda) Wang, and Lynn Newton. 'allowing them to dream': fostering creativity in mathematics undergraduates. *J. Furth. High. Educ.*, pages 1–13, May 2022.

[26] Marcela Hernandez-de Menendez, Carlos Escobar Díaz, and Ruben Morales-Menendez. Technologies for the future of learning: state of the art. *Int. J. Interact. Des. Manuf. (IJIDeM)*, 14(2):683–695, June 2020.

[27] Valentin Kuleto, Rocsana Bucea-Manea-Țoniş, Radu Bucea-Manea-Țoniş, Milena P Ilić, Oliva M D Martins, Marko Ranković, and Ana Sofia Coelho. The potential of blockchain technology in higher education as perceived by students in serbia, romania, and portugal. *Sustainability*, 14(2):749, January 2022.

[28] Jing (bill) Xu, Pimtong Tavitiyaman, Hyun Jeong Kim, and Siu-Kam Jamie Lo. Hospitality and tourism higher education in the post-COVID era: Is it time to change? *J. Hosp. Tour. Educ.*, 34(4):278–290, October 2022.

[29] Xiaolin Xia and Xiaojun Li. Artificial intelligence for higher education development and teaching skills. *Wirel. Commun. Mob. Comput.*, 2022:1–10, April 2022.

[30] Shaofeng Wang, Zhuo Sun, and Ying Chen. Effects of higher education institutes' artificial intelligence capability on students' self-efficacy, creativity and learning performance. *Educ. Inf. Technol.*, October 2022.

[31] Sasha Nikolic, Scott Daniel, Rezwanul Haque, Marina Belkina, Ghulam M Hassan, Sarah Grundy, Sarah Lyden, Peter Neal, and Caz Sandison. ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity. *Eur. J. Eng. Educ.*, 48(4):559–614, July 2023.

[32] Amin Al Ka'bi. Proposed artificial intelligence algorithm and deep learning techniques for development of higher education. *International Journal of Intelligent Networks*, 4:68–73, 2023.

[33] Ismail Dergaa, Karim Chamari, Piotr Zmijewski, and Helmi Ben Saad. From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biol. Sport*, 40(2):615–622, April 2023.

[34] Ravil I Mukhamediev, Yelena Popova, Yan Kuchin, Elena Zaitseva, Almas Kalimoldayev, Adilkhan Symagulov, Vitaly Levashenko, Farida Abdoldina, Viktors Gopejenko, Kirill Yakunin, et al. Review of artificial intelligence and machine learning technologies: Classification, restrictions, opportunities and challenges. *Mathematics*, 10(15):2552, 2022.

[35] Parimal Kumar Giri, Sagar S De, Sachidananda Dehuri, and Sung-Bae Cho. Biogeography based optimization for mining rules to assess credit risk. *Intelligent Systems in Accounting, Finance and Management*, 28(1):35–51, 2021.

[36] Vadipina Amarnadh and Nageswara Rao Moparthi. Prediction and assessment of credit risk using an adaptive binarized spiking marine predators' neural network in financial sector. *Multimedia Tools and Applications*, pages 1–37, 2023.

[37] Feng Zhao, Yan Lu, Xinning Li, Lina Wang, Yingjie Song, Deming Fan, Caiming Zhang, and Xiaobo Chen. Multiple imputation method of missing credit risk assessment data based on generative adversarial networks. *Applied Soft Computing*, 126:109273, 2022.

[38] Ricardo Muñoz-Cancino, Cristián Bravo, Sebastián A Ríos, and Manuel Graña. On the dynamics of credit history and social interaction features, and their impact on creditworthiness assessment performance. *Expert Systems with Applications*, 218:119599, 2023.

[39] Juanjuan Tian, Li Li, et al. Digital universal financial credit risk analysis using particle swarm optimization algorithm with structure decision tree learning-based evaluation model. *Wireless Communications and Mobile Computing*, 2022, 2022.

[40] Khalil Masmoudi, Lobna Abid, and Afif Masmoudi. Credit risk modeling using bayesian network with a latent variable. *Expert Systems with Applications*, 127:157–166, 2019.

[41] Lu-an Dong, Xin Ye, and Guangfei Yang. Two-stage rule extraction method based on tree ensemble model for interpretable loan evaluation. *Information Sciences*, 573:46–64, 2021.

[42] Xiaogang Wang et al. Analysis of bank credit risk evaluation model based on bp neural network. *Computational intelligence and neuroscience*, 2022, 2022.

[43] Henry Villarreal-Torres, Julio Ángeles-Morales, William Marín-Rodriguez, Daniel Andrade-Girón, Edgardo Carreño-Cisneros, Jenny Cano-Mejía, Carmen Mejía-Murillo, Mariby C Boscán-Carroz, Gumercindo Flores-Reyes, and Oscar Cruz-Cruz. Development of a classification model for predicting student payment behavior using artificial intelligence and data science techniques. *EAI Endorsed Transactions on Scalable Information Systems*, 10(5), 2023.

[44] Neepa Biswas, Anindita Sarkar Mondal, Ari Kusumastuti, Swati Saha, and Kartick Chandra Mondal. Automated credit assessment framework using etl process and machine learning. *Innovations in Systems and Software Engineering*, pages 1–14, 2022.

[45] Shousong Cai and Jing Zhang. Exploration of credit risk of p2p platform based on data mining technology. *Journal of Computational and Applied Mathematics*, 372:112718, 2020.

[46] Dangxing Chen, Jiahui Ye, and Weicheng Ye. Interpretable selective learning in credit risk. *Research in International Business and Finance*, 65:101940, 2023.

[47] Pragya Pandey and Kailash Chandra Bandhu. A credit risk assessment on borrowers classification using optimized decision tree and knn with bayesian optimization. *International Journal of Information Technology*, 14(7):3679–3689, 2022.

[48] Dalia Atif and Mabrouka Salmi. The most effective strategy for incorporating feature selection into credit risk assessment. *SN Computer Science*, 4(2):96, 2022.

[49] SB Coşkun and M Turanli. Credit risk analysis using boosting methods. *Journal of Applied Mathematics, Statistics and Informatics*, 19(1):5–18, 2023.

[50] Nisha Arora and Pankaj Deep Kaur. A bolasso based consistent feature selection enabled random forest classification algorithm: An application to credit risk assessment. *Applied Soft Computing*, 86:105936, 2020.

[51] Yuan Wang, Yanbo Zhang, Mengkun Liang, Ruixue Yuan, Jie Feng, and Jun Wu. National student loans default risk prediction: A heterogeneous ensemble learning approach and the shap method. *Computers and Education: Artificial Intelligence*, 5:100166, 2023.

[52] Xiaoming Zhang, Lean Yu, Hang Yin, and Kin Keung Lai. Integrating data augmentation and hybrid feature selection for small sample credit risk assessment with high dimensionality. *Computers & Operations Research*, 146:105937, 2022.

[53] Zeinab Hassani, Mohsen Alambardar Meybodi, and Vahid Hajihashemi. Credit risk assessment using learning algorithms for feature selection. *Fuzzy Information and Engineering*, 12(4):529–544, 2020.

[54] Giandomenico Cornacchia, Vito Walter Anelli, Fedelucio Narducci, Azzurra Ragone, and Eugenio Di Sciascio. A general architecture for a trustworthy creditworthiness-assessment platform in the financial domain. *Annals of Emerging Technologies in Computing (AETiC)*, 7(2), 2023.

[55] Andro Merćep, Lovre Mrčela, Matija Birov, and Zvonko Kostanjčar. Deep neural networks for behavioral credit rating. *Entropy*, 23(1):27, 2020.

[56] Fisnik Doko, Slobodan Kalajdziski, and Igor Mishkovski. Credit risk model based on central bank credit registry data. *Journal of Risk and Financial Management*, 14(3):138, 2021.

[57] Tianhui Wang, Renjing Liu, and Guohua Qi. Multi-classification assessment of bank personal credit risk based on multi-source information fusion. *Expert systems with applications*, 191:116236, 2022.

[58] Huan Li and Weixing Wu. Loan default predictability with explainable machine learning. *Finance Research Letters*, 60:104867, 2024.

[59] Davide Chicco, Luca Oneto, and Erica Tavazzi. Eleven quick tips for data cleaning and feature engineering. *PLOS Computational Biology*, 18(12):e1010718, 2022.

[60] Alberto Fernández, Salvador Garcia, Francisco Herrera, and Nitesh V Chawla. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, 61:863–905, 2018.

[61] Joffrey L Leevy, Justin M Johnson, John Hancock, and Taghi M Khoshgoftaar. Threshold optimization and random undersampling for imbalanced credit card data. *Journal of Big Data*, 10(1):58, 2023.

[62] Su Özgür and Mehmet Orman. Application of deep learning technique in next generation sequence experiments. *Journal of Big Data*, 10(1):160, 2023.

[63] Renuka Agrawal, Aaditya Ahire, Dimple Mehta, Preeti Hemnani, and Safa Hamdare. Optimizing stroke risk prediction using xgboost and deep neural networks. *International Journal of Advanced Computer Science & Applications*, 15(11), 2024.

[64] Philip Buczak, Jian-Jia Chen, and Markus Pauly. Analyzing the effect of imputation on classification performance under mcar and mar missing mechanisms. *Entropy*, 25(3):521, 2023.

[65] Tlamelo Emmanuel, Thabiso Maupong, Dimane Mpoeleng, Thabo Semong, Banyatsang Mphago, and Oteng Tabona. A survey on missing data in machine learning. *Journal of Big data*, 8(1):140, 2021.

[66] Ahmad R Alsaber, Jiazhu Pan, and Adeeba Al-Hurban. Handling complex missing data using random forest approach for an air quality monitoring dataset: a case study of kuwait environmental data (2012 to 2018). *International Journal of Environmental Research and Public Health*, 18(3):1333, 2021.

[67] Shangzhi Hong and Henry S Lynn. Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction. *BMC medical research methodology*, 20(1):199, 2020.

[68] Damien Dablain, Bartosz Krawczyk, and Nitesh V Chawla. Deepsmote: Fusing deep learning and smote for imbalanced data. *IEEE transactions on neural networks and learning systems*, 34(9):6390–6404, 2022.

[69] Sebastián Maldonado, Carla Vairetti, Alberto Fernandez, and Francisco Herrera. Fw-smote: A feature-weighted oversampling approach for imbalanced classification. *Pattern Recognition*, 124:108511, 2022.

[70] Justin M Johnson and Taghi M Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of big data*, 6(1):1–54, 2019.

[71] Mohamed Elhoseny, Noura Metawa, Gabor Sztano, and Ibrahim M El-Hasnony. Deep learning-based model for financial distress prediction. *Annals of operations research*, 345(2):885–907, 2025.

[72] Y Srinivasa Rao, Karaka Ramakrishna Reddy, Janjhyam Venkata Naga Ramesh, Elangovan Muniyandy, MVAL Narasimha Rao, Yousef A Baker El-Ebiary, B Kiran Bala, et al. Improving financial forecasting accuracy through swarm optimization-enhanced deep learning models. *International Journal of Advanced Computer Science & Applications*, 16(3), 2025.

[73] Bernd Bischl, Martin Binder, Michel Lang, Tobias Pielok, Jakob Richter, Stefan Coors, Janek Thomas, Theresa Ullmann, Marc Becker, Anne-Laure Boulesteix, et al. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(2):e1484, 2023.

[74] Prachi Agrawal, Hattan F. Abutarboush, Talari Ganesh, and Ali Wagdy Mohamed. Metaheuristic algorithms on feature selection: A survey of one decade of research (2009-2019). *IEEE Access*, 9:26766–26791, 2021.

[75] Maha Nssibi, Ghaith Manita, and Ouajdi Korbaa. Advances in nature-inspired metaheuristic optimization for feature selection problem: A comprehensive survey. *Computer Science Review*, 49:100559, 2023.

[76] Sangeeta Kurman and Sumitra Kisan. An in-depth and contrasting survey of meta-heuristic approaches with classical feature selection techniques specific to cervical cancer. *Knowledge and Information Systems*, 65(5):1881–1934, 2023.

[77] Tin H. Pham and Bijan Raahemi. Bio-inspired feature selection algorithms with their applications: A systematic literature review. *IEEE Access*, 11:43733–43758, 2023.

[78] Zohre Sadeghian, Ebrahim Akbari, Hossein Nematzadeh, and Homayun Motameni. A review of feature selection methods based on meta-heuristic algorithms. *Journal of Experimental & Theoretical Artificial Intelligence*, 0(0):1–51, 2023.

[79] R. Arun Kumar, J. Vijay Franklin, and Neeraja Koppula. A comprehensive survey on metaheuristic algorithm for feature selection techniques. *Materials Today: Proceedings*, 64:435–441, 2022. International Conference on Advanced Materials for Innovation and Sustainability.

[80] Olatunji O Akinola, Absalom E Ezugwu, Jeffrey O Agushaka, Raed Abu Zitar, and Laith Abualigah. Multiclass feature selection with metaheuristic optimization algorithms: A review. *Neural Computing and Applications*, 34(22):19751–19790, 2022.

[81] Tansel Dokeroglu, Ayça Deniz, and Hakan Ezgi Kiziloz. A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing*, 494:269–296, 2022.

[82] Ruba Abu Khurma, Ibrahim Aljarah, Ahmad Sharieh, Mohamed Abd Elaziz, Robertas Damaševičius, and Tomas Krilavičius. A review of the modification strategies of the nature inspired algorithms for feature selection problem. *Mathematics*, 10(3):464, Jan 2022.

[83] Li Yu Yab, Noorhaniza Wahid, and Rahayu A. Hamid. A meta-analysis survey on the usage of meta-heuristic algorithms for feature selection on high-dimensional datasets. *IEEE Access*, 10:122832–122856, 2022.

[84] Esther Omolara Abiodun, Abdulatif Alabdulatif, Oludare Isaac Abiodun, Moatsum Alawida, Abdullah Alabdulatif, and Rami S Alkhawaldeh. A systematic review of emerging feature selection optimization methods for optimal text classification: The present state and prospective opportunities. *Neural Computing and Applications*, 33(22):15091–15118, 2021.

[85] Qasem Al-Tashi, Said Jadid Abdulkadir, Helmi Md Rais, Seyedali Mirjalili, and Hitham Alhussian. Approaches to multi-objective feature selection: A systematic literature review. *IEEE Access*, 8:125076–125096, 2020.

[86] Heda Song, Isaac Triguero, and Ender Özcan. A review on the self and dual interactions between machine learning and optimisation. *Progress in Artificial Intelligence*, 8(2):143–165, 2019.

[87] Dan Simon. Biogeography-based optimization. *IEEE transactions on evolutionary computation*, 12(6):702–713, 2008.

[88] Guoqing Yu, Zhiyong Meng, Hongtao Ma, and Lei Liu. An adaptive marine predators algorithm for optimizing a hybrid pv/dg/battery system for a remote area in china. *Energy Reports*, 7:398–412, 2021.

[89] Mohammad Zaeimi and Ali Ghoddosian. Color harmony algorithm: an art-inspired metaheuristic for mathematical function optimization. *Soft Computing*, 24:12027–12066, 2020.

[90] Said M Mikki and Ahmed A Kishk. Quantum particle swarm optimization for electromagnetics. *IEEE transactions on antennas and propagation*, 54(10):2764–2775, 2006.

[91] Xin-She Yang. *Nature-inspired metaheuristic algorithms*. Luniver press, 2010.

[92] James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. ieee, 1995.

[93] Kalyanmoy Deb, Amrit Pratap, Sameer Agarwal, and TAMT Meyarivan. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE transactions on evolutionary computation*, 6(2):182–197, 2002.

[94] John H Holland. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[95] Lyn Thomas, Jonathan Crook, and David Edelman. *Credit scoring and its applications*. SIAM, 2017.

[96] Hans Hofmann. Statlog (German Credit Data). UCI Machine Learning Repository, 1994.

[97] Ross Quinlan. Statlog (Australian Credit Approval). UCI Machine Learning Repository, 2019.

[98] UCI Machine Learning Repository. South German Credit. UCI Machine Learning Repository, 2019.

[99] I-Cheng Yeh. Default of Credit Card Clients. UCI Machine Learning Repository, 2016.

[100] Will Cukierski Arnaud de Servigny. Loan default prediction - imperial college london, 2014.

[101] Zaur Begiev. My dataset. `https://www.kaggle.com/datasets/zaurbegiev/my-dataset`, 2018. Accessed: 2024-06-05.

[102] Anna Montoya, inversion, KirillOdintsov, and Martin Kotek. Home credit default risk, 2018.

[103] Will Cukierski Credit Fusion. Give me some credit, 2011.

[104] Petr Berka. Pkdd'99 discovery challenge website. `https://sorry.vse.cz/~berka/challenge/pkdd1999/chall.htm`, 1999. Accessed: 2024-06-05.

[105] China UnionPay Co., Ltd. China UnionPay Credit Dataset. `https://open.chinaums.com/saas-web/`, 2019. Accessed: 2024-06-05.

[106] Jinsung Yoon, James Jordon, and Mihaela Schaar. Gain: Missing data imputation using generative adversarial nets. In *International conference on machine learning*, pages 5689–5698. PMLR, 2018.

[107] José Barrera-García, Felipe Cisternas-Caneo, Broderick Crawford, Mariam Gómez Sánchez, and Ricardo Soto. Feature selection problem and metaheuristics: A systematic literature review about its formulation, evaluation and applications. *Biomimetics*, 9(1):9, 2023.

[108] An-Da Li, Bing Xue, and Mengjie Zhang. Improved binary particle swarm optimization for feature selection with new initialization and search space reduction strategies. *Applied Soft Computing*, 106:107302, 2021.

[109] Sani Tijjani, Mohd Nadhir Ab Wahab, and Mohd Halim Mohd Noor. An enhanced particle swarm optimization with position update for optimal feature selection. *Expert Systems with Applications*, 247:123337, 2024.

[110] E. Emary, Hossam M. Zawbaa, and Aboul Ella Hassanien. Binary grey wolf optimization approaches for feature selection. *Neurocomputing*, 172:371–381, 2016.

[111] Pei Hu, Jeng-Shyang Pan, and Shu-Chuan Chu. Improved binary grey wolf optimizer and its application for feature selection. *Knowledge-Based Systems*, 195:105746, 2020.

[112] Dan Chang, Congjun Rao, Xinping Xiao, Fuyan Hu, and Mark Goh. Multiple strategies based grey wolf optimizer for feature selection in performance evaluation of open-ended funds. *Swarm and Evolutionary Computation*, 86:101518, 2024.

[113] Abdelazim G Hussien, Diego Oliva, Essam H Houssein, Angel A Juan, and Xu Yu. Binary whale optimization algorithm for dimensionality reduction. *Mathematics*, 8(10):1821, 2020.

[114] Kunpeng Zhang, Yanheng Liu, Xue Wang, Fang Mei, Geng Sun, and Jindong Zhang. Enhancing iot (internet of things) feature selection: A two-stage approach via an improved whale optimization algorithm. *Expert Systems with Applications*, 256:124936, 2024.

[115] R. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pages 39–43, 1995.

[116] Y. Shi and R. Eberhart. A modified particle swarm optimizer. In *1998 IEEE International Conference on Evolutionary Computation Proceedings. IEEE World Congress on Computational Intelligence (Cat. No.98TH8360)*, pages 69–73, 1998.

[117] J. Kennedy and R.C. Eberhart. A discrete binary version of the particle swarm algorithm. In *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, volume 5, pages 4104–4108 vol.5, 1997.

[118] Broderick Crawford, Ricardo Soto, Gino Astorga, José García, Carlos Castro, and Fernando Paredes. Putting continuous metaheuristics to work in binary search spaces. *Complexity*, 2017(1):8404231, 2017.

[119] Seyedali Mirjalili, Seyed Mohammad Mirjalili, and Andrew Lewis. Grey wolf optimizer. *Advances in engineering software*, 69:46–61, 2014.

[120] Seyedali Mirjalili and Andrew Lewis. The whale optimization algorithm. *Advances in engineering software*, 95:51–67, 2016.

[121] Haibo He and Edwardo A Garcia. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*, 21(9):1263–1284, 2009.

[122] Charles Elkan. The foundations of cost-sensitive learning. In *International joint conference on artificial intelligence*, volume 17, pages 973–978. Lawrence Erlbaum Associates Ltd, 2001.

Titre : Apprentissage Automatique et Sélection de Caractéristiques par Métaheuristiques pour la Prédiction du Comportement de Déclaration de Revenus Dans les Systèmes de Prêts Étudiants

Mots clés : Prêts étudiants, Modélisation prédictive, Apprentissage automatique, Optimisation, Métaheuristiques, Sélection de caractéristiques.

Résumé : La viabilité des prêts étudiants conditionnels au revenu dépend de la conformité des emprunteurs aux exigences de déclaration. Anticiper le non-respect est complexe face aux contraintes de données et aux trajectoires variables des diplômés. Cette thèse propose un cadre de modélisation prédictive par étapes pour anticiper la conformité et l'abandon déclaratif via les données réelles du Fonds Solidaire de Crédit Universitaire (FSCU). La méthodologie comprend cinq phases. Les étapes initiales évaluent des modèles d'apprentissage automatique pour la conformité initiale en utilisant des données académiques et administratives pré-déclaratives. Les phases suivantes traitent de l'abandon de la déclaration via une sélection de caractéristiques par métaheuristiques, optimisant la haute dimensionnalité et l'interprétabilité. Des approches bio-inspirées capturent ensuite des motifs non linéaires complexes, suivies d'une analyse intégrative. Les résultats montrent que ces comportements sont prévisibles avec une performance stable. La sélection par métaheuristiques renforce la robustesse et identifie les variables clés de conformité. Ce cadre offre des outils scientifiques et pratiques pour la décision institutionnelle.

Title : Machine Learning and Metaheuristic Feature Selection for Predicting Income Declaration Behavior in Student Loan Systems

Keywords : Student loans, Predictive modeling, Machine learning, Optimization, Metaheuristics, Feature Selection.

Abstract: Income-contingent student loan systems rely on borrowers' compliance with income declaration requirements to ensure financial sustainability and equitable repayment. Anticipating non-compliance is challenging due to strict informational constraints and the evolving trajectories of graduates. This doctoral research addresses this through a stage-based predictive modeling framework designed to anticipate compliance and abandonment behaviors using real-world data from the University Credit Solidarity Fund (FSCU). The methodology is structured into five interconnected stages. Initial stages focus on constructing a reproducible data pipeline and evaluating interpretable machine learning models for first-time compliance using exclusively pre-declarative information. Subsequent stages address declaration abandonment among previously compliant borrowers, incorporating metaheuristic-based feature selection to manage high-dimensional spaces while preserving interpretability. Advanced bio-inspired modeling is explored to capture complex non-linear patterns, followed by a final integrative analysis for cross-stage comparison. Empirical results demonstrate that declaration behaviors can be anticipated with stable performance using institutional data alone. Metaheuristic-driven selection improves robustness and parsimony, identifying a reduced subset of academic and financial variables that consistently explain compliance patterns. This framework offers both scientific insights and practical relevance for institutional decision-making in education finance.

.