Ca' Foscari
University
of Venice

Single Cycle Degree programme in
Global Development & Entrepreneurship

Final Thesis

Interpreting Machine Learning Models of Corporate Default Prediction

A Forensic Triage of the Italian Distressed Cohort

Supervisor
Ch. Prof. Lisa Crosato

Graduand
Gianmarco Bittante
Matriculation Number 975606

Academic Year
2024 / 2025

Table of Contents

Abstract 1

1 - Introduction 2
1.1 The Research Gap: The Failure of Linear Screening 2
1.2 The Objective of the Study: The Quest for Resilience 3
1.3 The "Domain-Aware" Approach 4
1.4 Structure of the Thesis: The Roadmap 5

2 - Literature Review 7
2.1 The Evolution of Distress Prediction: From Classical Statistics to Probabilistic Models . . . 7
2.1.1 Univariate and Multivariate Foundations 7
2.1.2 The Shift to Probabilistic Models 8
2.1.3 Limits of Linear Models in Acute Distress 9
2.2 Machine Learning and Cost-Sensitive Learning 10
2.2.1 Superiority of Ensemble Algorithms 10
2.2.2 Moving Beyond Statistical Accuracy 11
2.2.3 Handling Class Imbalance 12
2.3 SME Specificity and the Turnaround Narrative 13
2.3.1 Unique Decline Dynamics 13
2.3.2 The Italian Context 14
2.3.3 Liability of Newness and the Lindy Effect 15
2.4 Zombie Firms and Capital Misallocation 16
2.4.1 Zombie Lending Theory 16
2.4.2 Post-Pandemic Systemic Risk 18
2.5 The Black Box Dilemma and the XAI Requirement 19
2.5.1 Critiquing Algorithmic Opacity 19
2.5.2 Regulation and Transparency 20
2.6 The Research Gap 21

3 - Data Description 23
3.1 Data Source and Sample Selection 23
3.1.1 Data Source 23
3.1.2 SME Perimeter 24
3.1.3 The "Technical Default" Filter and Structural Rationale 24
3.1.4 Temporal Horizon 25
3.1.5 Attrition Table and Cohort Selection 26
3.1.6 Data Ingestion and Integrity 26
3.2 Target Variable Definition (`default_label`) 27
3.2.1 The Target 27
3.2.2 Hierarchical Labeling 28
3.2.3 Mitigating the Survivorship Bias 29
3.3 Independent Variables and Feature Synthesis 29
3.3.1 Firmographics and Governance 29
3.3.2 Longitudinal Accounting Panel 30
3.3.3 Standard Financial Ratios 31
3.3.4 Advanced Drivers 31
3.3.5 Consolidated Drivers 32
3.4 Preprocessing and Data Cleaning 34
3.4.1 Semantic and Technical Sanitization 34
3.4.2 Missing Values Imputation 35

3.4.3 Firewall Protocol . . . 35
3.4.4 Algorithm-Aware Outlier Policy . . . 36
3.5 Feature Optimization and Audit . . . 36
3.5.1 Multicollinearity Management . . . 37
3.5.2 Leakage Audit . . . 37
3.5.3 Class Imbalance . . . 38
3.6 Advanced Calibration for Financial Analysis . . . 38
3.6.1 Value at Risk and Exposure Mapping . . . 38
3.6.2 Economic Parameters for Cost-Sensitive Thresholding . . . 38
3.6.3 Debt Maturity Architecture and Stress-Test Metadata . . . 40
3.6.4 Stratified Partitioning and Final Preparation . . . 41
3.7 From Raw Data to a High-Fidelity Diagnostic Canvas . . . 41
4 - Methodology . . . 43
4.1 Extreme Gradient Boosting . . . 43
4.1.1 Gradient Boosting on Distress . . . 44
4.1.2 Optimization Strategy . . . 45
4.1.3 Benchmark Architecture . . . 45
4.2 Training . . . 46
4.2.1 Managing Class Imbalance . . . 46
4.2.2 Temporal Firewall & Leakage Audit . . . 47
4.3 The Gatekeeper . . . 47
4.3.1 The Asymmetry of Error: Economic vs. Statistical Thresholding . . . 48
4.3.2 The Operational Choice: Safety First and Sanitization . . . 48
4.4 Forensic Diagnostic . . . 49
4.4.1 Profiles within the Confusion Matrix . . . 49
4.4.2 Macro-Financial Simulation: The Schivardi Swap Protocol . . . 50
4.4.3 Value-at-Risk Capital Allocation . . . 50
4.4.4 The “Refinancing Cliff” Protocol . . . 51
4.5 Explainability Suite . . . 52
4.5.1 Global Interpretability . . . 52
4.5.2 Interaction Screening . . . 53
4.5.3 Local Interpretability . . . 54
5 - Results and Discussion . . . 56
5.1 Predictive Performance on Solvency Restoration . . . 56
5.1.1 Global Discriminatory Power . . . 56
5.1.2 Benchmark Comparison . . . 58
5.1.3 Model Stability and Reproducibility . . . 61
5.2 The Mechanics of Distress and Restoration: Stock, Flow, and Time . . . 61
5.2.1 “The One above All”: Solvency Ratio . . . 62
5.2.2 Momentum of Recovery . . . 63
5.2.3 Temporal Dimension . . . 65
5.3 Forensic Behavioral Dynamics . . . 67
5.3.1 Marginal Effects . . . 68
5.3.2 The Scissor Effect . . . 69
5.3.3 Structural Diversity . . . 69
5.4 Decoding Classification Errors . . . 71
5.4.1 The “Silent Killers” (False Negatives) . . . 71
5.4.2 The “Miraculous Survivors” (False Positives) . . . 71
5.4.3 Structural Biases: “Too Big to Fail” . . . 72

5.5 The Cost of Inertia: Zombie Firms and Capital Misallocation . . . 73
5.5.1 Economic Implications of Zombie Status and Market Sclerosis . . . 73
5.5.2 The Schivardi Framework: Quantifying Trapped Capital and the "Sanitization" Imperative . . . 75
5.6 The Refinancing Cliff . . . 77
5.6.1 Defining the Cliff . . . 77
5.6.2 Exposing Rate-Sensitive Zombies . . . 78
5.6.3 The "Weaning Off" Effect and Industrial Sustainability . . . 79
5.7 Threshold Analysis: The Gatekeeper Dilemma . . . 80
5.7.1 Skewness of Distress . . . 80
5.7.2 Scenario Modelling . . . 81
5.7.3 Dynamic Thresholding, a Pre-emptive Stress Test . . . 82
5.8 Synthesis, Limitations, and Theoretical Constraints . . . 82
5.8.1 The Solvency vs. Strategic Success Gap: Retrenchment and Recovery . . . 83
5.8.2 Data Constraints and Temporal Biases: The 2019–2024 Regime Shift . . . 83
5.8.3 Methodological Synthesis: Why Machine Learning Requires Stress Testing . . . 84
5.8.4 The Economic Imperative of the Gatekeeper . . . 85

6 - Conclusions . . . 87
6.1 Conceptual Summary and Results . . . 87
6.1.1 Topology of Crisis . . . 87
6.1.2 Mechanics of Cliff Effects . . . 87
6.1.3 Inertia as an Opportunity Cost . . . 88
6.2 Policy and Managerial Applications . . . 89
6.2.1 Towards a Protocol for Sanitization . . . 89
6.2.2 The Refinancing Cliff . . . 89
6.2.3 On Cost-Sensitivity . . . 90
6.3 Research Limitations . . . 91
6.4 Future Developments . . . 93
6.4.1 Beyond the Pandemic Timeframe . . . 93
6.4.2 Qualitative Data . . . 94
6.5 On Interpretability . . . 95
6.5.1 Mitigating the Survivorship Bias . . . 95
6.6 In Conclusion . . . 96

A Appendix . . . 97
A.1 Target Definition . . . 97
A.2 Dynamic Sector Specific Imputation . . . 99
A.3 Feature Engineering . . . 101
A.4 Firewall . . . 104
A.5 XGBoost Optimization & Training . . . 105
A.6 Economic Thresholds - Gatekeeper . . . 107
A.7 Counterfactual Analysis . . . 109
A.8 Institutional Entrenchment . . . 111
A.9 Post-Hoc Interpretability . . . 113
A.10 Variable List . . . 116
A.11 Relevant Tables . . . 118

Bibliography . . . 120

AI Disclosure . . . 122

Abstract

Post-pandemic monetary normalization has exposed structural fragilities in European SMEs, creating a “Refinancing Cliff” that renders linear credit scoring obsolete. This thesis departs from standard default prediction by isolating a “Distressed Cohort” of Italian firms with negative equity. The objective shifts from generic risk screening to a “Solvency Restoration” triage, distinguishing between turnaround-capable firms and “Zombie” entities destined for liquidation.

Adopting a “Domain-Aware” framework, the study employs a Cost-Sensitive XGBoost architecture. Calibrated via Basel III parameters (LGD and NIM), the model minimizes Total Expected Loss to act as a conservative gatekeeper against capital destruction. The analysis introduces a “Hybrid Target Variable” encoding “Technical Failure” alongside legal bankruptcy to mitigate survivorship bias.

To resolve the regulatory “Black Box” dilemma, the research integrates SHAP and ALE within the DALEX framework. This approach decodes nonlinear “Cliff Effects” and financial dynamics like the “Scissor Effect” of debt. Empirical results show the ensemble model significantly outperforms linear benchmarks in identifying “Silent Killers”, firms with low leverage but deep structural insolvency. Finally, the study operationalizes a “Schivardi Swap” simulation, identifying € 13 billion in “Trapped Capital.” Surgical reallocation to resilient entities is estimated to generate a € 2 billion net EBITDA delta, offering a roadmap for credit portfolio “sanitization.”

Keywords: Credit Risk, XGBoost, Italian SMEs, Zombie Lending, Financial Distress, Turnaround Prediction, Cost-Sensitive Learning, Explainable AI (SHAP/ALE).

1 - Introduction

The Macro Context: From Market Hibernation to the Crisis of Truth

The period spanning between 2019 and 2022 exceeds a mere statistical anomaly in corporate finance; it represents a natural experiment on resilience under distorted market conditions. This era created a profound "data-regime gap" between the subsidized stability of the pandemic and the volatile normalization that followed. Effectively, the suspension of standard risk dynamics made it impossible for traditional models to distinguish between inherently healthy firms and those sustained solely by external support.

For over a decade, the Eurozone's Zero Interest Rate Policy (ZIRP) created a permissive environment where the cost of debt ceased to function as a selection mechanism for efficiency. This was exacerbated by the COVID-19 pandemic, which induced a state of suspended animation. As noted by Schivardi, Sette and Tabellini (2020), while policy responses were vital to prevent the collapse of viable firms, debt moratoria and the freezing of insolvency procedures put the economy "on ice," pausing the natural process of creative destruction.

In this landscape of Market Hibernation, the link between credit risk and the cost of capital was severed. Insolvency was masked by the continuous rollover of debt, creating a solvency illusion that obscured deep structural decay. Bankruptcy filings ceased to be a proxy for economic health, signifying instead a regulatory lag or creditor forbearance.

Within the Italian context, state guarantees acted as a necessary parachute but simultaneously fueled the artificial survival of non-viable entities. SMEs especially benefitted from these maneuvers, as a primary addressee of the policy. Caballero et al. (2008) formalized this as zombie lending, where subsidized credit to insolvent firms distorts competition and depresses aggregate productivity. This created a significant cohort of Rate-Sensitive Zombies that used liquidity to service pre-existing inefficiencies. Their survival became a divergence between legal status and economic reality.

The transition to the current reality is marked by a structural disruption: the normalization of monetary policy. Aggressive interest rate hikes transform the zombie problem from a theoretical concern into a systemic emergency, creating what this study conceptualizes as the Refinancing Cliff. As the protective layer of state intervention is removed, a "weaning-off" effect becomes operative; firms that relied on public support as a permanent crutch now face an immediate, unbuffered shock to their capital structure. This restoration of market rates reveals that much of the previous decade's solvency was merely a liquidity artifact of the ZIRP era. When the cost of capital returns to historical norms, the space between operational reality and financial survival abruptly closes.

The urgency of this research lies in decoding the Crisis of Truth that follows hibernation. Traditional screening models are ill-equipped to distinguish between the Resilient Core, which used liquidity to bridge a temporary gap, and the Terminal Burners destined for collapse. By applying the frameworks of Caballero and Schivardi to granular SME data, this study provides the analytical instruments required to navigate this new paradigm, aiming to sanitize the credit system and restore the flow of capital to productive growth.

1.1 The Research Gap: The Failure of Linear Screening

Against this backdrop of systemic distortion, the transition from a generalized assessment of credit risk to the granular prediction of a turnaround requires a fundamental shift in analytical perspective. While financial literature has extensively documented the mechanics of corporate failure, there remains a critical methodological void regarding the recovery of a strictly distressed population. Traditional screening models, designed to distinguish healthy firms from unhealthy ones, suffer from structural limitations when applied to a cohort characterized by negative equity. Standard predictive models operate like a general health screening designed to detect the onset of illness in

a healthy population. However, within this distressed cohort, such an approach is insufficient; the requirement shifts to an emergency triage protocol, where the goal is not to diagnose the sickness, which is already evident, but to distinguish survivable injuries from terminal conditions. This research gap is, at its core, a failure of dimensionality: linear models are mathematically ill-equipped to navigate the non-linear "cliff effects" that define the boundary between a terminal zombie and a resilient turnaround candidate.

Standard distress prediction frameworks rely on the assumption of a steady, linear slide toward insolvency. While these tools are effective for general populations, they lose their discriminatory power when applied to the "Red Zone", a landscape where every entity is technically "underwater." In this environment, linear logic compresses the differences between marginal stress and total structural collapse into a single, indistinguishable scale. Consequently, a linear discriminator often perceives the entire cohort as statistically terminal, failing to identify the specific "Event Horizon" where a firm's financial physics fundamentally shift from recoverable distress to irreversible inertia. To navigate the Red Zone, the analytical sensor must be able to determine whether a vessel still retains buoyancy to stay afloat, and those who have already reached the bottom.

This shift is most visible in what this we will mention as "Paradox of Rigidity"[1]. In the post-pandemic SME landscape, high asset tangibility, often transforms into a source of structural inertia during a revenue crisis. Standard frameworks are structurally blind to this context-sensitive polarity reversal, where "asset heaviness" paradoxically decreases a firm's agility and survival probability. Furthermore, the trajectory of a crisis is inherently path-dependent; the likelihood of a turnaround is not a static snapshot but a function of time. A firm entering its first year of distress maintains a fundamentally different survival probability than a "zombie" entity entering its fourth, yet traditional models often overlook this temporal decay, failing to account for inertia as an accelerating force of destruction.

The core contribution of this research is thus the pivot from a screening objective to a "triage" protocol. Traditional screening becomes redundant for a cohort that has already reached technical failure. Triage, by contrast, seeks to identify the structural momentum required for regeneration. This distinction is fundamental: the dynamics of recovery differ entirely from the mechanics of decline, and a model designed merely to flag sickness is ill-equipped to spot resilience. By filtering out the "noise" of healthy enterprises, this study forces the analysis to identify subtle, non-linear signals like cash flow velocity and operational retrenchment, that standard models overlook.

Ultimately, the inability of traditional models to handle these sudden, non-linear "Cliff Effects" necessitates the adoption of high-dimensional ensemble architectures. This shift is not a mere technical upgrade but a logical requirement of the economic problem itself. Only a non-parametric approach can decode the complex structural mechanics of a distressed cohort, providing the resolution needed to differentiate between terminal inertia and genuine regenerative potential.

1.2 The Objective of the Study: The Quest for Resilience

The overarching objective of this research is to engineer a predictive framework designed for high-resolution forensic triage within a corporate population already defined by structural insolvency. Consequently, the central research question shifts from the traditional screening of default risk to the specific prediction of recovery.

Among firms already exhibiting severe financial distress, specifically those with negative equity, the study seeks to identify which entities possess the structural momentum to achieve a successful turnaround and which are destined for terminal inertia. Essentially, when a company is already in a state of failure, our interest is to distinguish between those that can be fixed and those that are effectively dead.

[1] For more insight on the polarity of tangible assets given a distressed situation, consult Crosato, Liberati and Repetto (2023)

A Specialized Classifier for the Distressed Cohort

Standard bankruptcy prediction models typically operate on vast populations of healthy firms to identify a small minority of defaults. This study departs from that convention by isolating a Distressed Cohort, in a homogeneous sample of Italian SMEs that entered the post-pandemic era already exhibiting negative shareholders' equity. Within this perimeter, the predictive challenge shifts from detecting the onset of illness to forecasting the likelihood of solvency restoration.
By focusing exclusively on this "Red Zone," the model aims to overcome the survivorship bias inherent in generalist frameworks that often mistake the artificial inertia of subsidized "zombie" firms for genuine survival. The primary objective is to identify firms possessing the structural capacity to navigate the first and most critical stage of a turnaround: operational retrenchment. This initial phase of downsizing and cost-stabilization is the necessary prerequisite for any successful recovery. The goal is to define a specialized classifier capable of seeing through "life support" to find the resilient entities prepared to undergo this rigorous survival protocol.

Identifying the Drivers of Regeneration

To distinguish a temporary liquidity crisis from irreversible decay, this study isolates the interactions between a firm's operational "flow" and its structural "stocks." The objective is to identify the specific tipping point where operational metrics, such as cash flow velocity, are overwhelmed by structural constraints, most notably the phenomenon of asset heaviness. In the context of Italian SMEs, high asset tangibility can paradoxically transform from a source of collateral strength into a "rigidity trap." While fixed assets provide stability in normal cycles, they can become a burden during a revenue crisis, limiting the agility required for survival. By identifying the threshold where the weight of these assets precludes a turnaround, the model distinguishes the Resilient Core, which are capable of operational retrenchment, from the Terminal Burners whose structures are fundamentally incompatible with recovery.

Allocative Efficiency: The Schivardi Swap

The research ultimately aims to transition from micro-prudential scoring to a mechanism for macroeconomic allocation. Rooted in theories of credit misallocation, the study posits that the precise identification of "zombie" firms is the essential prerequisite for market sanitization. The hypothesis suggests that the capital currently locked in non-viable entities represents a profound opportunity cost for the financial system. Through a counterfactual simulation termed the Schivardi Swap, this study quantifies this volume of "trapped capital" and its potential impact on the broader economy[2]. The simulation tests the premise that reallocating these resources to the resilient core would generate a significant net positive delta in aggregate earnings. Thus, the objective is to demonstrate how algorithmic precision can dismantle the congestion effects that stifle competition, restoring the physiological flow of credit to the productive economy, effectively clearing out the "dead wood" so the rest of the forest can grow.

1.3 The "Domain-Aware" Approach

To bridge the gap between computational power and prudential compliance, this research adopts a "Domain-Aware" methodological philosophy. Rather than deploying a generic machine learning classifier optimized solely for statistical accuracy, the study engineers a specialized decision-support system rooted in the economic realities of distressed debt. This approach proactively addresses the "Black Box" limitations often raised by regulatory bodies, ensuring that the algorithmic architecture adheres to the specific constraints of the banking sector through three fundamental pillars: Cost-

[2]The simulation is inspired by Schivardi, Sette and Tabellini's work: *Identifying the Real Effects of Zombie Lending*, Schivardi, Sette and Tabellini (2020)

Sensitive Learning, Algorithmic Transparency, and Data Integrity.
The first pillar, Cost-Sensitive Learning, aligns the model with the Basel III regulatory framework. Traditional algorithms often fail in credit risk because they treat all errors as equal; this study, however, rejects the simple maximization of accuracy in favor of minimizing the Total Expected Loss. By integrating an asymmetric cost function, the algorithm is forced to act as a conservative "Gatekeeper," prioritizing the identification of "Silent Killers", zombie firms that lack operational viability, over the aggressive expansion of credit. In this context, the model functions not just as a predictor, but as a mechanism for capital preservation.
To ensure this conservative logic is trustworthy, the second pillar introduces an Explainable AI (XAI) framework. The model transitions from an opaque "Black Box" to a transparent "Grey Box" architecture, utilizing interpretability tools not as post-hoc accessories, but as instruments for the validation of economic causality. This forensic audit ensures that the algorithm captures genuine financial patterns, like the non-linear "Scissor Effects" of debt, rather than spurious correlations. This transparency satisfies the regulatory "Right to Explanation" while ensuring that the predictive output remains grounded in established economic theory.
Finally, the robustness of these signals is secured through a rigorous Data Integrity Protocol. A strict "Temporal Firewall" is established to eliminate look-ahead bias, forcing the model to rely solely on the information that would have been available to a bank at the moment of the credit decision. Furthermore, the data foundation is reinforced by sector-specific strategies designed to preserve the unique financial signatures of different industries. Ultimately, this methodology ensures that the model's "engine" is as clean and realistic as the banking ledgers it aims to analyze, providing a solid bridge from raw data to actionable policy prescriptions.

1.4 Structure of the Thesis: The Roadmap

The remainder of this dissertation is organized into five chapters, structured to transition from the theoretical underpinnings of distress prediction to the empirical quantification of allocative inefficiency. The following roadmap outlines the logical progression from the "Event Horizon" of corporate failure to the "Schivardi Swap" simulation.

- **Chapter 2: Literature Review** This chapter maps the epistemological shift in credit risk modeling, tracing the evolution from the linear discriminant frameworks of Altman (1968) and Ohlson (1980) to contemporary non-parametric Machine Learning architectures. It integrates the macroeconomic theory of "Zombie Firms" (Caballero, Hoshi and Kashyap, 2008), establishing the theoretical justification for a triage approach that distinguishes between recoverable distress and the artificial survival mechanisms maintained by credit misallocation.
- **Chapter 3: Data Description** This section details the construction of the longitudinal dataset (2019–2024), defining the "Distressed Cohort" of Italian SMEs. It formalizes the "Firewall Protocol" used to eliminate look-ahead bias and introduces the "Hybrid Target Variable." This composite classification metric encodes "Technical Failure" (Zombie Stagnation) alongside formal legal bankruptcy to mitigate survivorship bias, ensuring the model identifies firms that are economically terminal even if they remain legally active.
- **Chapter 4: Methodology** This chapter outlines the computational architecture of the Cost-Sensitive XGBoost model. It defines the "Gatekeeper Protocol," a threshold optimization strategy calibrated on Basel III economic parameters (LGD/NIM) rather than pure statistical accuracy. Furthermore, it presents the DALEX-based interpretability framework designed to transition the algorithm from an opaque "Black Box" to a transparent "Grey Box" compliant with regulatory standards.
- **Chapter 5: Results & Discussion** This section presents the empirical evidence, evaluating the model's discriminatory power (AUC) and dissecting the non-linear drivers of turnaround via SHAP

and ALE analysis. Crucially, it deploys two post-hoc simulations: the "Refinancing Cliff" stress test, which exposes "Rate-Sensitive Zombies" under normalized interest rates, and the "Schivardi Swap," a counterfactual exercise quantifying the macroeconomic benefit of reallocating "Trapped Capital."

- **Chapter 6: Conclusions** The final chapter synthesizes the "Topology of Crisis," discussing the structural rigidity of assets in distressed firms. It addresses the research limitations, specifically the "Solvency Illusion" generated by the "Market Hibernation" of the pandemic era. The dissertation concludes with policy prescriptions for the banking sector regarding the sanitization of credit portfolios and the necessary "weaning off" of non-viable entities.

2 - Literature Review

The epistemological framework of corporate distress prediction has evolved from the static inspection of accounting ratios toward high-dimensional, non-parametric machine learning architectures. This trajectory reflects a broader transition in financial economics: a shift from deterministic models toward probabilistic and ensemble frameworks capable of capturing the dynamic, non-linear nature of modern insolvency.

This chapter maps the theoretical and methodological tensions that define the current state of the art, structured across five critical dimensions: From the univariate foundations of Beaver (1966) and the multivariate "Z-Score" of Altman (1968) to the probabilistic logit models of Ohlson (1980), tracing the industry's attempts to move beyond linear assumptions, the first section provides a brief view over the methodology through the years. For the transition toward Machine Learning models, an evaluation of ensemble methods, specifically XGBoost, as a solution to the "Cliff Effects" and class imbalances that traditional econometrics systematically overlook. Restricting the focus to the observed dynamics, a contextual analysis of the structural constraints facing Italian SMEs, integrating theories such as the "Liability of Newness" and the paradox of "Asset Heaviness." The theoretical conceptualization of the "Zombie Firm" and the systemic consequences of credit misallocation, providing the economic justification for a triage rather than a screening predictive objective. It is also necessary to mention transparency principles, an exploration of the "Black Box" dilemma and the regulatory necessity of Explainable AI (XAI) in reconciling computational precision with Basel-compliant transparency.

By synthesizing these perspectives, the review concludes by identifying a critical research gap: the scarcity of triage-oriented models focused exclusively on the distressed cohort to quantify the macroeconomic "Cost of Inertia."

The chapter analyzes these theoretical and methodological tensions to contextualize the contribution of this study. Section 2.1 reviews the historical evolution from MDA to probabilistic models. Section 2.2 evaluates the superiority of ensemble algorithms (XGBoost) and the challenge of class imbalance. Section 2.3 and 2.4 examine the unique distress dynamics of SMEs and the macroeconomic theory of Zombie firms. Finally, Section 2.5 addresses the "Black Box" critique, justifying the use of post-hoc interpretability tools (SHAP) to bridge the gap between algorithmic precision and economic intuition.

2.1 The Evolution of Distress Prediction: From Classical Statistics to Probabilistic Models

The evolution of distress prediction reflects a transition from isolated univariate ratios toward integrated probabilistic frameworks. Early attempts to identify signals of decline in a vacuum soon gave way to the multivariate synthesis of Altman's Z-Score and the likelihood-based estimations of logit and probit models. While these classical tools established the field's foundation, their reliance on linear assumptions leaves them ill-equipped to capture the sudden "cliff effects" and non-monotonic shifts characterizing modern insolvency. This methodological gap necessitates the shift toward the non-parametric machine learning architectures explored in this study.

2.1.1 Univariate and Multivariate Foundations

The quantitative assessment of corporate financial health traces its origins to the early 20th century, evolving from the rudimentary inspection of individual accounting ratios to sophisticated statistical frameworks. As chronicled in the comprehensive review by Bellovary[3], the initial era of distress prediction was dominated by univariate analysis, known as "ratio decomposition." Early studies,

[3] *A Review of Bankruptcy Prediction Studies: 1930 to Present*, (2007)

such as those by FitzPatrick (1932) and Winakor & Smith (1935), evaluated indicators like Net Worth to Debt in isolation to root out the early signs of decline in industrial firms. This line of research led up to the seminal work of Beaver, *Financial Ratios As Predictors of Failure* (1966), who rigorously tested the predictive power of thirty individual ratios. Beaver identified "Cash Flow to Total Debt" as a dominant univariate predictor, capable of signaling failure up to five years before the firm actually goes under. Practically speaking, the ratio showed imediately promising results, minimizing classification error to only 13%.

However, the univariate paradigm suffered from a critical operational flaw: the potential for conflicting signals. A firm might exhibit strong liquidity ratios while simultaneously suffering from deteriorating profitability, leaving the analyst with a blurred picture of solvency. This confusion pushed the field to move toward multivariate frameworks capable of analyzing the entire variable profile of the firm simultaneously. The field underwent a fundamental transformation with Altman's publication *Financial Ratios, Discriminant Analysis and the Preditction of Corporate Bankruptcy* (1968), who applied Multiple Discriminant Analysis (MDA) to finance. By synthesizing five distinct financial ratios into a single weighted index known as the Z-Score, Altman cleared up the confusion of individual signals. Essentially, by bundling multiple ratios together, the model could weigh different risks at the same time rather than looking at them in a vacuum. This architecture proved very efficient, as it reached a 95% accuracy on its initial sample, classifying 63 out of 66 events correctly. Crucially for the context of this study, Altman included Retained Earnings/Total Assets (X2) as a proxy for the age and cumulative profitability of the firm. He explicitly noted that this variable puts younger firms at a disadvantage, thereby statistically codifying the "Liability of Newness," a concept that remains central to analyzing SME survival dynamics today.

Despite establishing an enduring industry standard, the Z-Score relies on the restrictive statistical assumption of linear separability. In inherent terms, it assumes that the boundary between survival and failure follows a linear path. Ohlson later critiqued MDA for its requirement of normally distributed predictors and equal covariance matrices, conditions rarely met by the skewed distributions of distressed firms[4]. Furthermore, Barboza, Kimura, and Altman himself[5] argue that static linear models fail to capture the complex, non-linear interactions and 'cliff effects' inherent in modern financial crises. In this context, a cliff effect represents a non-monotonic tipping point where a marginal variation in a financial ratio no longer produces a proportional increase in risk, but instead triggers a sudden, structural discontinuity in the firm's solvency. While traditional models treat distress as a continuous 'slide,' financial reality often manifests as a sudden drop-off. Consequently, while the Z-Score remains a robust baseline, but its inability to model these non-linear thresholds necessitates the transition toward machine learning architectures.

2.1.2 The Shift to Probabilistic Models

While Multivariate Discriminant Analysis (MDA) established the foundational taxonomy of distress prediction, its methodological rigidity eventually brought about a paradigm shift toward probabilistic frameworks. The primary critique of the MDA approach, as articulated by Ohlson in *Financial ratios and the Probabilistic Prediction of Bankruptcy* (1980), centered on the restrictive statistical assumptions required for its validity, specifically the requirements that predictors be normally distributed and that the variance-covariance matrices of the groups be identical. Ohlson argued that these conditions are rarely met in financial datasets, rendering the output of MDA, an ordinal ranking score rather than a probability, difficult to interpret intuitively for decision-making purposes.

To address these deficiencies, Ohlson introduced the use of conditional logit analysis to the field. Unlike MDA, the logit model is computationally more robust as it does not assume multivariate normality or equal covariance matrices. To put it simply, while it still remains sensitive to extreme

[4] in (Ohlson, 1980)

[5] in Barboza, Kimura and Altman (2017)

outliers, it freed the researcher from the 'Gaussian trap', the unrealistic assumption that financial ratios must follow a perfect bell curve to be valid. More fundamentally, it transformed the output into a direct probability of failure, identifying 9 predictive variables categorized into 4 primary dimensions: size, structure, performance, and liquidity. This allowed for standard statistical testing of individual predictors, a feature often hidden away in the "black box" of discriminant scores. Its model achieved a final *Likelihood Ratio Index* of 84%, adopting an optimal cutoff of 0.038. Ohlson also identified results for an arbitrary cutoff of 0.5, with a striking but unbalanced 96.12%, and for a safe cutoff of 0.0095 which led to high Type I Errors at 47%.
Following this trajectory, Zmijewski in *Methodological Issues Related to the Estimation of Financial Distress Prediction Models* (1984) further refined the approach using Probit analysis. However, his primary contribution lay in his critique of choice-based samples, where bankrupt firms were over-represented. Zmijewski demonstrated that without correcting for this bias, the resulting models often overstated the significance of their predictors. In short, if a sample is rigged with too many bankrupt cases, the results look better than they actually are. Zmijewski's model initially showed an overall correct classification rate of 72%. Despite these advances, Shumway, in *Forecasting Bankruptcy More Accurately: A Simple Hazard Model* (2001), categorizes these as "static models" because they fail to account for the time-series nature of decline, and observed the Probit model classifying correctly only 43.2% of the cases. Furthermore, traditional probabilistic models miss the mark on the complex interactions and "cliff effects" inherent in distress. Essentially, while these models can predict a gradual decline, they often fail when a firm suddenly hits a certain limit, a limitation that has driven the shift over to machine learning techniques.

2.1.3 Limits of Linear Models in Acute Distress

While the probabilistic frameworks of Ohlson and Zmijewski advanced the field beyond simple classification, they remain bound by a fundamental econometric constraint, namely the assumption of linearity and monotonicity in the predictor variables. Standard logistic regression assumes that a unit change in a financial ratio induces a constant change in the log-odds of default across the whole board. Essentially, it assumes that an extra dollar of profit helps a failing firm just as much as it helps a healthy one, regardless of where they started. However, recent literature suggests that this assumption falls apart when applied to firms in acute financial distress.
Moscatelli, Parlapiano, Narizzano and Viggiano in *Corporate Default Forecasting with Machine Learning* (2020) pick up on this phenomenon as the failure to capture "non-monotonous relationships" inherent in credit risk data. In a healthy population, variables such as Sales Growth are generally inversely correlated with default risk. However, within a distressed cohort, this relationship often breaks down or even flips, behaving as cliff effects. For a firm with deep negative equity, aggressive sales growth may no longer signal recovery but rather "overtrading" or a desperate liquidation of inventory, thereby speeding up the path to insolvency. Linear models, constrained by fixed coefficients, require pre-specified interaction terms that are frequently unknown or omitted in traditional scoring, forcing a "one-size-fits-all" logic that leads to significant errors.
This limitation is most critical when analyzing firms with negative equity. As historically recognized by Ohlson[6], standard financial ratios effectively lose their mathematical coherence when the denominator turns negative. A linear model cannot naturally tell the difference between a "technical" deficit and a terminal solvency crisis without manual intervention. In simple terms, when the math goes into the negative, the model's "compass" stops working correctly. Ohlson attempted to fix this by introducing a binary "discontinuity correction" variable, explicitly acknowledging that the linear functional form runs into a wall when liabilities exceed assets.

Consequently, in the zone of acute distress, the standard rules of correlation fall apart. The relationship between financial inputs and survival probability ceases to be a smooth curve and

[6] Ohlson (1980)

becomes a series of structural breaks. As argued by Barboza, Kimura and Altman (2017), because statistical models treat financial decline as a continuous slide, they systematically fail to detect the "event horizon," the specific threshold where a firm's financial physics fundamentally change from recoverable stress to irreversible inertia. This failure necessitates the shift toward non-parametric machine learning methods capable of modeling these jagged, non-linear decision boundaries.

2.2 Machine Learning and Cost-Sensitive Learning

The methodology of corporate distress prediction has witnessed a decisive migration from the rigid traditional econometrics toward the adaptive capacity of machine learning (ML) architectures. This paradigm shift is driven by the empirical superiority of ensemble methods over linear discriminants; Barboza, Kimura and Altman in *Machine Learning Models and Bankruptcy Prediction* (2017) demonstrate that techniques such as Bagging and Boosting yield an average accuracy improvement of 10% over logistic regression benchmarks by effectively modeling high-dimensional non-linearities. Crucially, this transition addresses the structural limitations of linear assumptions when modeling "cliff effects" (Moscatelli *et al.*, 2020). However, the adoption of ML introduces critical challenges regarding class imbalance. Specifically, maximizing global accuracy in a distressed cohort, where the target event is statistically rare, can lead to biased classifiers that fail to identify the minority class (Hou *et al.*, 2025). Consequently, the contemporary literature has evolved beyond simple accuracy maximization, toward Cost-Sensitive Learning. This framework integrates the asymmetric economic penalties of Type I and Type II errors directly into the algorithmic loss function to align predictive outputs with financial realities (Peykani *et al.*, 2025).

2.2.1 Superiority of Ensemble Algorithms

Contemporary literature on credit risk modelling sets apart ensemble learning as the superior methodological standard, consistently beating out individual classifiers such as single Decision Trees (CART) or Artificial Neural Networks (ANN). As demonstrated by Barboza et al. (2017), ensemble methods pulled off a 10% predictive performance improvement (measured by AUC or F1-score), over traditional benchmarks like Logistic Regression. This predictive dominance stems from the ensemble's ability to iron out the specific errors inherent in single-model predictions, specifically variance and bias, through two distinct strategies: Bagging and Boosting.

Parallel Architectures: Variance Reduction via Random Forest

Random Forest (RF) represents the pinnacle of the bootstrap aggregating (bagging) approach. RF constructs a multitude of decision trees in parallel using bootstrapped samples of the training data, which permitted to reach an accuracy of 87% (Barboza, Kimura and Altman, 2017). The algorithm irons out the overall variance of the prediction by averaging the outputs of uncorrelated trees. Essentially, by letting a crowd of trees votes on the outcome, the model prevents one "wrong" tree from skewing the final result. This "decorrelation" ensures that the ensemble does not rely too heavily on a single predictor, shielding the model from noise. Consequently, RF functions as a robust "consensus" mechanism, often prioritizing stability over granular precision in edge cases.

Sequential Architectures: Bias Reduction via Gradient Boosting

In contrast, Gradient Boosting adopts a sequential learning framework. In this process, each subsequent tree T_n is trained specifically to fix the mistakes of the previous tree T_{n-1}. In short, the algorithm doubles down on what it got wrong in the previous round, refining its prediction step-by-step. By iteratively focusing the model's attention on the "hard-to-classify" instances, Boosting actively reduces inductive bias.

The Leap from GBDT to XGBoost

XGBoost represents a structural evolution of the boosting framework. Of the architecture, we highlight two critical innovations (Chen and Guestrin, 2016)[7]:

1. **Regularized Objective Function:** XGBoost incorporates a regularization term into its objective function, which penalizes excessive model complexity and prevents the algorithm from "memorizing" noise or specific outliers. This constraint is crucial for isolating robust structural signals, such as the non-linear interaction between firm age and solvency, ensuring that these findings represent generalizable patterns rather than idiosyncratic artifacts of the training sample.
2. **Second-Order Optimization:** XGBoost utilizes a second-order Taylor expansion, of the loss function. Unlike standard Gradient Boosting which only uses the first derivative (gradient), XGBoost adopts the second derivative (Hessian). This allows the algorithm to converge faster and handle the 'Cliff Effects' with much finer granularity. Consequently, allows for a more precise estimation of loss reduction, enabling the algorithm to find the optimal tree structure with greater accuracy.

Through these innovations, Gradient Boosting architectures offer us a comparative advantage in the context of rare-event detection. In a "Distressed Cohort", the distinction between a "Zombie Firm" and a "Turnaround Candidate" often relies on subtle interactions that are easily smoothed over by the averaging logic of Random Forest. By assigning higher weights to misclassified instances, XGBoost functions as a specialized "forensic"[8] tool, thus it is exceptional at finding thresholds. For example, XGBoost can identify that a low solvency ratio is a normal 'startup' phase attribute for young firms, whereas the same ratio signals terminal distress for established entities. Basically, while Random Forest plays it safe by averaging, XGBoost behaves like an investigator that refuses to ignore the hardest clue, picking it up and unravelling its dynamic. Crosato, Liberati and Repetto in *Lost in a black-box? Interpretable machine learning for assessing Italian SMEs default* (2023) observed an advantage in sensitivity of the model, identifying 82.1% bankruptcies, compared to Logit (74.5%) and Probit (69.4%) over Italian SMEs composed sample.

2.2.2 Moving Beyond Statistical Accuracy

While the adoption of ensemble methods like XGBoost addresses the non-linear complexities of distress prediction, established literature argues that optimizing for standard statistical metrics, such as Global Accuracy, is fundamentally flawed in the context of credit risk. Hand and Till[9] assert that simple accuracy is a deceptive metric when class distributions are skewed, as it fails to set apart the types of errors committed. Consequently, a model that chases high statistical scores may inadvertently lead to significant financial loss if it fails to account for the "economic cost of errors." As highlighted by Amirshahi and Lahmiri[10], credit datasets are inherently imbalanced. In such scenarios, an algorithm can pull off deceptively high accuracy simply by predicting "survival" for every firm. Essentially, if 95% of firms survive, the model can "cheat" its way to a high score by assuming no one ever fails. Kuhn and Johnson[11] describe this as the "Accuracy Paradox", where a model achieves high specificity, identifying healthy firms, while exhibiting negligible sensitivity, identifying defaulters. In other words, the model is great at nodding along to success but blind to

[7] Chen and Guestrin are the minds behind the extreme gradient boosting architecture. Consult *XGBoost: A Scalable Tree Boosting System* Chen and Guestrin (2016) for more detail

[8] Here, "forensic" describes the capability to identify non-monotonic "Event Horizons", specific thresholds where a marginal change in financial ratios triggers a disproportionate, structural collapse in solvency. This contrasts with traditional linear assessments that perceive distress as a continuous, gradual decline rather than a series of critical discontinuities.

[9] In *A Simple Generalization of the Area Under the ROC Curve for Multiple Class Classification Problems*, Hand and Till (2001).

[10] Consult *Bankruptcy prediction using optimal ensemble models under balanced and imbalanced data*, Amirshahi and Lahmiri (2024).

[11] Consult *Applied Predictive Modelling*, Kuhn and Johnson (2013).

the early signs of a crash. In short, a model can look perfect on paper while failing to catch the very firms that will cause the most damage, rendering it useless as an early warning system.

The resolution to this paradox lies in recognizing the asymmetric fallout of the two error types. The cost of misclassification is rarely uniform in real-world applications. Within the specific domain of credit scoring, a critical disparity between the error categories is identified:

- **The False Negative (Type II Error):** This represents the "Silent Killer," a firm predicted as safe that subsequently defaults. The fallout boils down to the erasure of capital, including the loss of principal, unpaid interest, and recovery costs. Nonetheless, non the entire amount lent is completely erased, as banks often recover significant portions of the credit.
- **The False Positive (Type I Error):** Conversely, a False Positive amounts to a "Missed Opportunity," where a viable firm is incorrectly flagged and denied credit. While this incurs an opportunity cost, it is significantly lower than the direct capital destruction caused by a False Negative.

To operationalize this economic reality, Kuhn and Johnson[12] propose the use of Cost-Sensitive Learning. This approach embeds a "loss function" that penalizes False Negatives significantly more heavily than False Positives. By applying asymmetric cost ratios, such as the 5:1 ratio, or even higher risk-averse weights implied by the Basel Committee on Banking Supervision frameworks[13], the algorithm is mathematically told that letting one defaulter slip through the cracks is five times worse than turning away a healthy client. Essentially, the model is tuned to be "paranoid" enough to save the principal rather than "optimistic" enough to chase every small gain. By tuning models to minimize the Total Misclassification Cost (TMC) rather than statistical error, researchers ensure that the algorithm functions not merely as a classifier, but as a decision-support system calibrated to preserve capital. Essentially, the model acts as a practical shield in high-stakes environments where preventing capital allocation to a defaulter is the paramount goal.

2.2.3 Handling Class Imbalance

A pervasive characteristic of financial distress datasets is "Class Imbalance," where the count of healthy firms dwarfs the number of those in distress. Hou et al. [14] warn that standard machine learning algorithms, which are designed to maximize global accuracy, inherently bias predictions toward the majority class. In a credit risk context, a model can pull off deceptively high accuracy simply by predicting "solvency" for every firm, yet this leaves the model dead in the water for risk management because it misses the rare but costly default events. The literature points toward two different ways to handle this, namely Data-Level Preprocessing, or Resampling, and Algorithm-Level Adaptation, Cost-Weighting[15].

The Resampling Paradigm

Data-level solutions focus on shaking up the training distribution to artificially balance class representation. The most prominent technique is the Synthetic Minority Over-sampling Technique (SMOTE), which cooks up synthetic examples of distressed firms by interpolating between existing minority instances. Essentially, the algorithm looks at two real distressed firms and creates a "fake" one that sits somewhere between them to fill in the gaps. In their article, Hou and his colleagues demonstrate that standard SMOTE substantially bumps up the F1-score, reaching 0.73, and the Matthews Correlation Coefficient (MCC - accounts for all quadrants of the confusion matrix), hitting 0.70. Furthermore, hybrid techniques such as SMOTE-Tomek were found to boost recall even further.

[12] Consult *Applied Predictive Modelling*, Kuhn and Johnson (2013).

[13] *International Convergence of Capital Measurement and Capital Standards*, BCBS, (2005).

[14] Consult *Comparative Analysis of Resampling Techniques for Class Imbalance in Financial Distress Prediction Using XGBoost*, Hou *et al.* (2025).

[15] Originally proposed by Elkan in *The Foundations of Cost-Sensitive Learning*, (2001).

However, resampling comes with a catch. Amirshahi and Lahmiri[16] argue that for high-dimensional financial datasets, synthetic generation may introduce noise rather than signal, noting that techniques like SMOTE can actually eat away at the performance of advanced boosting algorithms. Hou et al. also note that aggressive undersampling, while achieving the highest recall at 0.85, results in catastrophic precision loss, dropping to 0.46 and generating an unacceptable volume of false alarms.

Cost-Weighting as an Economic Necessity

Given the risks of noise, recent literature has moved over to algorithm-level fixes, specifically Cost-Sensitive Learning. As articulated by Peykani et al. [17], this approach leaves the data distribution alone but tweaks the learning algorithm's loss function to penalize missing a minority class instance more heavily. In simple terms, the model is told that failing to spot a bankrupt firm is much more expensive than accidentally flagging a healthy one. Modern algorithms like XGBoost support this natively through parameters such as `scale_pos_weight`, which doesn't just "add a number"; it multiplies gradients and hessians of the minority class, thus emphasizing missed defaults more intensely. By assigning a higher weight to the minority class, the model is mathematically forced to prioritize the identification of distressed firms during gradient descent. This approach is often more methodologically conservative as it preserves the authentic multivariate distribution of the data without generating synthetic noise[18]. Consequently, this study favours cost-weighting over resampling, as it ties the model's optimization objective directly to the economic reality where the cost of a missed default far outweighs the cost of a false alarm. Essentially, the model is tuned to be "paranoid" enough to save capital rather than "optimistic" enough to chase every small gain.

2.3 SME Specificity and the Turnaround Narrative

The financial distress path of Small and Medium Enterprises (SMEs) is ruled by structural constraints and behavioural dynamics that stand apart from those of large corporations. Forstner, Vassolo, and Sevil in *Different strokes for different folks: A review of SME decline and turnaround*, (2025) argue that traditional models often miss the mark for the SME cohort due to resource scarcity, which limits the escape routes available to distressed managers. In plain terms, smaller firms have fewer places to turn when things go south. Unlike large conglomerates that can sell off non-core assets to drum up liquidity, SMEs typically lack diversified units, forcing them to lean on "retrenchment" strategies, specifically operational efficiency and strict cash flow management, to stave off failure. Essentially, the firm stops looking for growth and focuses entirely on staying afloat.
In the Italian economy, Crosato et al. (2023) emphasize that these firms deal with acute information asymmetry and a structural dependence on bank lending. To put it simply, banks typically lack a clear window into the firm's actual health, which makes trust more valuable than public data. This renders market-based signals, such as stock price volatility, irrelevant. Consequently, the "Turnaround Narrative" for an SME is defined not by strategic repositioning, but by a struggle to restore solvency through operational resilience and stakeholder support. This dynamic calls for specialized predictive variables that stand apart from those used in standard corporate finance literature.

2.3.1 Unique Decline Dynamics

The transition from financial distress to solvency in Small and Medium Enterprises (SMEs) is not just a scaled-down version of large corporate restructuring. Forstner et al. argue that SMEs respond to decline in ways that existing turnaround models fail to catch, making it necessary to use a frame-

[16] Consult *Bankruptcy prediction using optimal ensemble models under balanced and imbalanced data*, Amirshahi and Lahmiri (2024)

[17] Consult *Evaluation of Cost-Sensitive Learning Models in Forecasting Business Failure of Capital Market Firms,* Peykani *et al.* (2025).

[18] As argued by Kuhn & Johnson in *Applied Predictive Modelling*, (2013)

work that sets apart the state of decline from the act of retrenchment. Essentially, decline describes the damage already done to the firm's performance, while retrenchment refers to the active steps taken to stop the bleeding.

The Morphology of SME Decline

Unlike large diversified firms that can shift funds to support failing units, SMEs lack the "organizational slack" needed to cushion the blow of environmental shocks. Forstner et al. (2025) categorize SME decline into three distinct forms:

1. **Sudden Decline:** A rapid collapse often triggered by external shocks or "Black Swan" events. In some cases, this decline is the result of a long-standing structural crisis that has been obscured by "Window Dressing" practices, the intentional manipulation of financial statements to project health. Due to the unpredictable nature of external shocks and the data-corrupting effect of accounting fraud, this category is excluded from the current study to ensure the model trains on authentic financial distress signals.
2. **Gradual Decline:** A slow erosion of competitiveness where management may fail to spot minor issues due to a lack of sophisticated control systems. This is the primary focus of traditional bankruptcy literature and this study, as it produces the longitudinal "clues" in financial ratios that Gradient Boosting models are designed to unravel. In the Italian context, identifying this phase early is now a regulatory mandate under the "Adeguati Assetti" (Adequate Organizational Setups) requirements of the *Codice della Crisi*.
3. **Lingering Decline:** A persistent state of insolvency where firms survive through "Financial Inertia" rather than economic value creation. Often subsidized by lenient bank lending or public support, these entities exist in a suspended state between technical insolvency and formal failure. These "Zombie Firms" represent the greatest challenge for triage modelling, as they possess the structural characteristics of failure but are kept on artificial life support, effectively misallocating capital within the SME ecosystem.

Retrenchment vs. Terminal Stagnation

A critical task in predicting SME outcomes is telling the difference between a firm shrinking to die and a firm shrinking to survive. Forstner et al. identify "operational retrenchment," such as cost-cutting and asset reduction, as a primary way for SMEs to stabilize during a cash flow crisis. However, the success of this tactic hangs on how "asset-heavy" the firm is. To put it simply, a factory tied down by heavy machinery cannot pivot as easily as a lean service firm. Consequently, a reduction in total assets may signal a terminal failure in one context but a strategic pivot in another.

The Role of "Bricolage" in Recovery

Finally, Forstner et al. introduce the concept of "bricolage"[19], or the ability to improvise solutions with limited resources, as a key driver of SME recovery. Unlike large corporations that engage in formal restructuring, resilient SMEs rely on agile retrenchment and strong stakeholder ties to bridge liquidity gaps. This suggests that predictive models must see past static solvency ratios to pick up on these dynamic behavioural traits, separating firms that are merely insolvent from those actively using their limited resources to engineer a turnaround.

2.3.2 The Italian Context

The Italian economic landscape, characterized by a predominance of manufacturing SMEs, presents specific distress dynamics that stand apart from Anglo-Saxon service-oriented models. Crosato et al.

[19] Introduced in *Different strokes for different folks: A review of SME decline and turnaround*, Forstner, Vassolo and Sevil (2025)

(2023) examine this context through a phenomenon they call "Asset Heaviness." Essentially, while a bank sees collateral, a struggling firm sees an immovable weight. As traditional credit scoring models often view tangible assets solely as a safety net to lower risk, such as Loss Given Default, Crosato et al. argue that for distressed Italian SMEs, a large asset base can paradoxically trigger a default.

The Double-Edged Sword of Tangible Assets

In manufacturing, high capital intensity brings about a rigid cost structure dominated by the upkeep of physical plants. Crosato et al. observe that while micro-entities possess the agility to retrench informally, larger manufacturing SMEs run into "Structural Rigidity." To put it simply, the more physical gear a firm owns, the harder it is to downsize when the market turns. When revenue stalls, these firms cannot easily sell off machinery to preserve cash. Consequently, what appears as a strength on the balance sheet becomes an illiquid anchor, speeding up the slide toward insolvency. Another factor to account for is the higher credit demand that asset heavy firms require, making banks decide between protecting their capital or lending the requested amount and overexposing their position.

The "Recovery Bias" in Liquidation

Furthermore, Crosato et al. (2023) highlight a behavioural distortion in the credit recovery process that creates a jagged relationship between size and formal default. They posit that creditors are more likely to push asset-heavy firms into formal bankruptcy because there is tangible value to recover. Conversely, asset-light firms with no recoverable value are often ignored, leading to "phantom" defaults that stay off the legal books. This "Asset Heaviness" creates a specific vulnerability for Italian manufacturers. They are too rigid to pivot quickly, representing operational risk, and too valuable to be let off the hook by creditors, representing legal risk. In short, they are too bulky to move and too valuable to be forgiven, a double bind that standard linear scoring models often fail to capture.

2.3.3 Liability of Newness and the Lindy Effect

The relationship between firm age and financial distress is characterized by a distinct temporal asymmetry. Theoretical literature sets out two opposing forces that drive this dynamic, namely the "Liability of Newness," which dictates the high mortality rate of early-stage enterprises, and a survival premium often termed the "Lindy Effect." In short, while young firms are fragile, older firms gain strength simply by staying in the game.

The Liability of Newness

The concept of the "Liability of Newness" suggests that younger organizations face a higher risk of failure due to a lack of established routines, insufficient resources, and a deficit in legitimacy. Altman explicitly codified this in the construction of the Z-Score (Altman, 1968). He noted that the ratio of Retained Earnings to Total Assets (X2) inherently puts young firms at a disadvantage, as they have not had enough time to build up cumulative profits. Altman argued that this was not a model defect but an accurate reflection of the reality where the chance of going under is much higher in a firm's earlier years.

Forstner et al. (2025) further elaborate on this structural vulnerability in SMEs. They argue that young firms frequently wrestle with attracting high-quality talent and lack the "organizational slack" needed to cushion the blow of environmental shocks. Essentially, they have no financial "fat" to live off of when the market turns lean. Unlike mature incumbents, young SMEs cannot rely on economies of scale, making them highly sensitive to liquidity interruptions.

The Survival Premium - Lindy Effect

Conversely, literature suggests that once a firm survives its infancy, age transforms from a liability into a stabilizing factor. Rico, Pandit and Puig (2021)[20] found that firms that successfully turned things around after bankruptcy were significantly older than those that liquidated. This resilience is often attributed to "Relationship Lending." Older firms benefit from long-standing banking relationships where they can lean on "soft information," such as the trust and historical knowledge built up over decades, to secure credit access. In plain terms, a long history with a bank acts as a character reference that helps them get loans when younger firms are turned away. Consequently, age acts as a proxy for "Reputational Collateral," granting mature firms a survival buffer that stands apart from their immediate efficiency.

Age as a Proxy for Market Valuation

The statistical relevance of firm age is further reinforced by Shumway (2001), who argues that traditional accounting-based models are limited by their "static" nature. By introducing a Hazard Model, Shumway demonstrated the presence of Duration Dependence, proving that the risk of failure is intrinsically tied to the passage of time. While Shumway's empirical tests suggest that in listed markets, this "age premium" is often baked into the stock price, his findings create a critical theoretical bridge for the study of private entities.
In the absence of market-clearing prices for unlisted Italian SMEs, firm age functions as a credible signal of resilience. As Forstner et al. recommend, for these entities, longevity acts as a surrogate for the market's missing valuation. In this context, age serves as a vital proxy for "reputational collateral", the intangible assets, supply-chain trust, and institutional entrenchment that facilitate survival during acute distress. Consequently, for the "Distressed Cohort" age is not merely a chronological marker but a structural indicator of a firm's capacity to weather the "event horizon" of insolvency.

2.4 Zombie Firms and Capital Misallocation

The theoretical framework of corporate distress must move beyond binary labels to include "Zombie Firms"[21]. Caballero, Hoshi, and Kashyap (2008) define these as insolvent borrowers that stay afloat only because of subsidized credit or sham restructurings from banks that want to avoid owning up to losses. This practice, known as "evergreening," breaks the process of creative destruction by propping up inefficient firms that should have failed.
The presence of these entities leads to "congestion effects", where labor and capital are trapped in unproductive units, artificially driving up input costs and eroding the market share of healthy competitors. As Adalet McGowan, Andrews and Millot in *The walking dead? Zombie firms and productivity performance in OECD countries,* (2018) argue, this creates a state of economic "sclerosis", a systemic breakdown of the Schumpeterian "Replacement Mechanism". By lowering the barriers to survival for inefficient firms, the credit market effectively raises the barriers to entry for more productive innovators, "clogging" the flow of resources to companies with genuine growth potential. Consequently, the survival of the living dead doesn't just represent a localized failure; it acts as a structural block that prevents new growth from taking root across the SME ecosystem.

2.4.1 Zombie Lending Theory

The persistence of insolvent firms within the credit system is not merely a symptom of financial distress but often the result of credit misallocation, a phenomenon known as "Zombie Lending." Essentially, it is a system where banks choose to prop up failing firms rather than liquidating them. In their seminal analysis of the Japanese "Lost Decade," Caballero, Hoshi, and Kashyap established

[20] Further developed by Forstner, Vassolo and Sevil (2025)
[21] which are economically dead but kept alive through artificial support

that banks frequently engage in "sham loan restructurings," such as interest rate concessions, debt forgiveness, and moratoriums, to keep failing borrowers afloat. The macroeconomic scale of this misallocation is substantial: during Japan's 'Lost Decade', the share of assets held by zombie firms was observed to rise to nearly 30% in certain sectors . Reflecting this trend in the European periphery, OECD studies indicate that Italian capital 'sunk' in zombies reached 19% post-2008 (Adalet McGowan, Andrews and Millot, 2018), highlighting the systemic nature of the 'congestion effects' analysed in this study. "Evergreening" practices are driven by regulatory arbitrage. By continuing to lend to distressed firms, banks avoid classifying loans as non-performing, thereby preventing the depletion of their regulatory capital buffers (Acharya, Bharath and Srinivasan, 2003). In short, evergreening is a survival strategy for the bank as much as it is for the firm.

The Mechanism of Artificial Support

The central mechanism of zombie lending is the decoupling of credit access from economic viability. Caballero et al. define zombies as firms that are operationally unprofitable but stay in business solely because of subsidized credit that covers their interest payments. Peek and Rosengren[22] describe this as "unnatural selection," observing that credit is often handed out perversely, as struggling banks are more likely to double down on poor-performing firms to hide their own balance sheet weaknesses. Consequently, the survival of these firms is not a function of their efficiency, but of their lenders' reluctance to realize losses, a phenomenon this study identifies as the "Rate-Sensitive Zombie" cluster.

Economic Sclerosis and Congestion Effects

The persistence of these non-viable entities induces a state of industrial ossification, or "Market Sclerosis", whereby the economic bloodstream becomes choked by entities that ought to have bowed out through natural market attrition. This systemic distortion plays out as a form of structural congestion that filters through two distinct channels. In factor markets, zombie firms tend to lock up labour and capital, driving up wages and operational overheads beyond what the actual productivity of the sector can support. In simpler terms, they sit on resources that better companies could actually use to grow. Simultaneously, in product markets, cheap credit lines enable these laggards to hang onto surplus capacity and resort to fire sales, which suppresses market prices and squeezes the profit margins of healthy competitors. This creates a scrambling effect where the prospect of low profits and high costs puts off healthy firms from expanding, effectively freezing out fresh capital injections.

The long-term consequence of this environment is a widening "Productivity Gap". Industries with a higher prevalence of zombies show significantly lower Total Factor Productivity growth, as healthy firms scale back on investment while zombies remain passive, relying almost exclusively on subsidized credit without any internal restructuring. This represents a fundamental behavioural divergence. While a firm in the midst of a genuine turnaround seeks to fix its productivity through strategic realignment, a zombie simply burns through liquidity to stay in a state of terminal stagnation. It essentially acts as a drain on the system, taking in financial support without ever moving toward a sustainable recovery[23].

Implications for Distress Prediction

This framework challenges traditional models that assume a simple choice between solvency and liquidation. The normal outcome of "creative destruction" is often thwarted in a zombie-infested economy. Therefore, a predictive model must distinguish not only between "safe" and "bankrupt"

[22] Consult *Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan,* Peek and Rosengren (2005)

[23] Consult *Zombie Lending and Depressed Restructuring In Japan,* Caballero, Hoshi and Kashyap (2008) for a detailed review

firms but also identify the "living dead," active firms artificially in business as they are being propped up. In plain terms, these firms are on life support; they don't produce value, they just consume credit. This distinction is central to this study, which uses the "Schivardi Swap" simulation to quantify the capital trapped in these congested segments, responding to the need to sanitize the credit channel and restore efficiency. The swap is practically a counterfactual simulation where the credit lent to zombies is hypothetically reallocated to healthier firms of the sample. Such simulation is implicitly under the assumption of finite credit supply, and aims to provide a measurement of the congestion effect.

2.4.2 Post-Pandemic Systemic Risk

The transition from a period of liquidity abundance to a restrictive monetary environment brings about a critical discontinuity in corporate survival dynamics, best understood through the lens of systemic risk measurement. Acharya, Pedersen, Philippon, and Richardson[24] redefined systemic risk not merely as the volatility of individual institutions, but as the propensity of the financial sector to be undercapitalized during a broader market downturn. They introduced the concept of Marginal Expected Shortfall (MES) to quantify how much a single institution contributes to a systemic crisis. Crucially for this study, Acharya et al. demonstrate that when the banking sector faces capitalization shocks, it responds by "hoarding liquidity" and pulling back on the credit supply to the real economy. Essentially, when banks feel the heat, they stop lending, which triggers a domino effect of business failures. While originally applied this to large-scale financial institutions, the theoretical principle remains critical for the SME sector: a systemically undercapitalized banking sector (due to hidden Zombie losses) is less resilient to shocks, eventually triggering the 'liquidity hoarding' that starves viable SMEs during a downturn.

The Genesis of the "Refinancing Cliff"

This theoretical framework sets the stage for the post-pandemic "Refinancing Cliff." During the 2020–2022 period, state guarantees and debt moratoria effectively suppressed the transmission of systemic risk, creating a state of "Market Hibernation" where insolvency was hidden away by subsidized liquidity. However, as noted by Grajzl, Cepec, and Mörec[25], the withdrawal of public support acts as a "Weaning Off" shock. Basically, companies that used government aid as a crutch rather than a cure are now hitting the wall of market reality. Firms that leaned on public cash flows without fixing their underlying business models face a sudden realignment of their cost of capital.

From Liquidity Risk to Structural Solvency

The "Refinancing Cliff" boils down to the event horizon where the cost of debt returns to normal levels. In a low-interest environment, highly leveraged firms, defined by Caballero et al. as "Zombies," can maintain technical solvency through the "evergreening" of loans. However, a rise in interest rates exposes the "Rate-Sensitive" nature of these entities. As borrowing costs exceed the Return on Invested Capital (ROIC), the signature transitions from a state of liquidity stress to one of terminal insolvency, also corresponding to the Interest Coverage Ratio dropping below 1, its sustainability threshold. Consequently, predicting distress today requires more than just checking balance sheet ratios; it calls for a forward-looking stress test of the firm's ability to handle a normal interest rate burden. In simple terms, we are no longer looking at whether a firm is sick, but whether it can survive the end of its life-support system.

[24] Consult *Understanding the Recovery Rates on Defaulted Securities,* Acharya, Bharath and Srinivasan (2003)

[25] Consult *Weaned off public money: The effect of discontinued reception of public cash on firm outcomes,* Grajzl, Cepec and Mörec (2023)

2.5 The Black Box Dilemma and the XAI Requirement

The shift from linear econometric frameworks to non-parametric ensemble methods has brought about significant gains in predictive accuracy, allowing models to capture the complex, non-linear interactions inherent in corporate distress. Essentially, these models can map out jagged risk patterns that straight-line formulas often miss. However, this performance boost brings with it a critical operational deficit, the "Black Box" phenomenon, where the decision logic is hidden behind the recursive complexity of the algorithm.

In the highly regulated domain of credit risk management, where capital allocation decisions must pass muster under Basel standards and meet the borrower's "Right to Explanation," predictive power alone is not enough. To put it simply, a bank cannot just use a model that works; they have to be able to show why it works. Consequently, the integration of Explainable AI (XAI) frameworks has moved from a theoretical novelty to a compliance necessity, acting as the essential bridge between computational precision and the interpretability that economic stakeholders demand. It boils down to making the machine's "internal thoughts" readable to the human eye.

2.5.1 Critiquing Algorithmic Opacity

As financial distress prediction moves from linear statistical tools to non-parametric machine learning, a critical tension arises between predictive performance and algorithmic transparency. While ensemble methods like XGBoost offer superior accuracy in handling cliff effects, they inherently function as "Black Boxes," which are complex functions that are too recursive for human comprehension. Rudin, in its publication *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead* (2019), takes aim at the prevailing reliance on these opaque models, arguing that for "high-stakes decisions," a category that clearly covers credit capital allocation and insolvency proceedings, the standard practice of applying post-hoc explanations is methodologically flawed.

The Chasm Between Explanation and Interpretation

Rudin establishes a fundamental distinction between interpretability and explainability. Inherent Interpretability refers to models that are transparent by design, such as sparse decision trees, where the computation directly maps to a domain-specific understanding. In contrast, Post-Hoc Explainability refers to a second model created to approximate the predictions of the first black-box model. Rudin posits that relying on post-hoc explainability introduces a "Fidelity Gap." The logic is rigorous. Essentially, if an explanation were a perfect match for the black box, it would be the black box, making the complex model redundant. Therefore, for an explanation to be simple enough to be understood by a human, it must necessarily be an approximation, implying that it inevitably misses the mark in a certain percentage of cases. In short, the simpler the explanation, the more likely it is to smooth over the messy reality of how the algorithm actually decided. In the context of "Zombie Firm" detection, this implies that a SHAP plot might rationalize a survival prediction based on variables that the model did not actually prioritize, potentially leading to a false sense of security regarding a borrower's viability.

The Myth of the Accuracy-Interpretability Trade-off

A central tenet of the machine learning literature is that complex models are necessary to maximize accuracy on difficult datasets. Rudin aggressively critiques this "Accuracy-Interpretability Trade-off" as a myth, particularly regarding structured tabular data. She argues that the "Rashomon Set," the group of all models that achieve a similar level of accuracy, is usually big enough to include a simple model that works just as well as a complex one.

Justification for the Current Approach

Despite Rudin's compelling argument for inherently interpretable models, this study adopts the "Black Box with Explanation" approach. This methodological choice is justified by the specific nature of the "Distressed Cohort." The relationship between financial ratios and survival for firms with negative equity is characterized by extreme non-linearities and high-order interactions, such as the "Scissor Effect" where rising debt and falling revenue velocity cut through a firm's solvency. These are patterns that simple linear models often overlook. While acknowledging the "Fidelity Gap," this study uses SHAP not as a perfect map of the model's logic, but as a diagnostic tool to pick out the "tipping points" of solvency. Essentially, we are willing to put up with a small gap in fidelity if it means catching rare turnaround events that simpler models would miss.

2.5.2 Regulation and Transparency

The integration of machine learning into credit risk modeling is not merely a technological upgrade but a shift subject to stringent regulatory constraints. The regulatory landscape, primarily governed by the Basel Committee on Banking Supervision (BCBS) and European Union directives, establishes a framework where predictive accuracy is a necessary but insufficient condition for model adoption. A duality emerges between the "Prudential Mandate," aimed at protecting the bank's capital, and the "Transparency Mandate," which safeguards the borrower's rights.

The Prudential Mandate: Basel III and Model Governance

Under the Internal Ratings-Based (IRB) approach of the Basel framework, financial institutions are permitted to use internal estimates of Probability of Default (PD) to calculate capital requirements. However, this permission hangs on rigorous governance standards. Braunsberger and Aschauer[26] further highlight that regulators require models to be not only statistically sound but also appropriately conservative and regularly validated. A critical barrier for ML adoption in this context is the "Use Test," which dictates that internal ratings must play an essential role in the bank's actual risk management and credit approval processes. Consequently, a "Black Box" model that produces high accuracy but lacks an explainable logic falls short of the "Use Test." If a risk officer cannot explain how the model reached its conclusion, they cannot legally use it to pull the trigger on a loan. As noted by Moscatelli et al. (2020), while ML models cut down on credit losses, their opacity complicates the supervisory review process where banks must show that they understand the specific risk drivers of their portfolios.

The Transparency Mandate: The "Right to Explanation" for SMEs

Parallel to prudential regulation, the legal framework governing the bank-client relationship has moved toward a "Right to Explanation." Crosato, Liberati, and Repetto (2023) argue that in the specific context of the Italian credit market, interpretability is a fundamental requisite for bringing a model into production. This requirement stems from the General Data Protection Regulation (GDPR), specifically Article 22, which grants individuals the right to avoid decisions based solely on automated processing without meaningful information about the logic involved.

Crosato et al. emphasize that this right is economically vital for SMEs. Unlike large corporations with transparent statements, SMEs often deal with information asymmetry. When a credit application is rejected based on an algorithmic score, the firm has an implicit right to understand the reasons for the refusal, allowing them to fight the decision or fix their financial standing. In short, an entrepreneur needs to know if they were turned down because of a temporary cash flow dip or a deeper structural failure. A pure "Black Box" denies this possibility, effectively shutting out the borrower. Therefore, Crosato et al. posit that for ML to be viable in the Italian manufacturing sector,

[26] Consult *Corporate Failure Prediction: A Literature Review of Altman Z-Score and Machine Learning Models Within a Technology Adoption Framework*, Braunsberger and Aschauer (2025)

it must be accompanied by post-hoc interpretability tools that can translate complex interactions into actionable feedback.

The Move Toward "Grey Box" Modeling

The intersection of these regulatory forces creates a mandate for what can be termed "Grey Box" architectures, models that retain the non-linear predictive power of ensemble methods like XGBoost while providing the transparency required for regulatory scrutiny. While Rudin (2019) warns that post-hoc explanations may suffer from a "Fidelity Gap" (where the explanation risks losing the "truth" of the machine's actual logic), the integration of Explainable AI (XAI) remains a prerequisite for the systemic adoption of machine learning in distressed debt management.
To mitigate these inherent risks, this study adopts a dual-layered explanatory framework. By synthesizing Global Interpretability (identifying sector-wide risk drivers) with Local Interpretability (mapping specific "tipping points" for individual firms), the approach allows for a cross-validation of results. This ensures that the SHAP-derived narrative for a specific SME is not an isolated approximation, but is instead anchored in the broader structural logic discovered by the ensemble across the entire distressed cohort.

2.6 The Research Gap

While the transition to machine learning has significantly enhanced predictive accuracy in general bankruptcy forecasting, current literature remains constrained by a "general population" bias. The vast majority of empirical studies function as screening mechanisms designed to detect the onset of distress within a heterogeneous population of healthy and unhealthy entities. This study identifies three critical convergences where existing research falls short. First, as argued by Forstner et al. (2025), there is a distinct scarcity of research focused exclusively on the "Distressed Cohort"; traditional models trained to distinguish solvent firms from insolvent ones are structurally ill-equipped to perform the triage function required to separate terminal liquidations from recovery candidates. Second, a theoretical disconnect persists between the macroeconomic literature on "Zombie Lending"[27] and the computational frameworks of credit scoring. Finally, while cost-sensitive learning has addressed statistical class imbalance, it has rarely been adopted to quantify the "Cost of Inertia", the opportunity cost of capital trapped in terminal "Zombies", within the Italian SME sector. By applying Cost-Sensitive XGBoost to a strictly defined distressed population, this study shifts the predictive objective from preventing failure to identifying the structural capacity for recovery.

Diagnostic vs. Screening Models

The vast majority of distress prediction literature, from the seminal work of Altman (1968) to contemporary machine learning applications like Barboza et al. (2017), deals with heterogeneous samples that include both healthy and distressed firms. These models act as "screening" tools, designed to pick up on the early signs of distress in a generally healthy population. However, there is a clear scarcity of research focused exclusively on the "Distressed Cohort," namely firms that have already gone past the point of technical insolvency, such as those with negative equity. Forstner et al. (2025) argue that the dynamics of "Turnaround" stand apart from the dynamics of "Decline." A model trained to distinguish a healthy firm from a sick one is mathematically ill-equipped to tell the difference between a terminal firm and one that is bouncing back. Essentially, while screening tells you who is sick, triage tells you who can be saved. This study fills that gap by looking only at firms with negative equity, shifting the task from "avoiding failure" to "identifying recovery."

[27] For further detail, consult *Zombie Lending and Depressed Restructuring in Japan*, Caballero, Hoshi and Kashyap (2008)

AUC vs. Trapped Capital

While recent studies have introduced cost-sensitive learning to financial modelling, these applications remain largely theoretical or focused on standard settings. The literature lacks empirical work that uses Cost-Sensitive XGBoost to work out specific macroeconomic phenomena, such as "Trapped Capital." Standard studies often focus on chasing statistical metrics like Area Under the Curve (AUC). However, in high-stakes credit allocation, statistical precision comes second to economic utility, such as minimizing Loss Given Default. This study bridges this gap by calibrating the model not to chase the highest number of correct predictions, but to cut down on the "Cost of Inertia," which is the capital tied up in "Zombie" firms that fail to recover. In short, the model prioritizes mitigating high-risk exposure over achieving a perfect statistical score.

Machine Learning and Zombie Theory

Finally, there is a clear gap between the econometric literature on "Zombie Lending" and the data science literature on bankruptcy prediction. The former looks at the macroeconomic "congestion" caused by non-viable firms but falls short on predictive granularity at the firm level. The latter provides granular predictions but often misses the macroeconomic context of "unnatural selection." This study ties these distinct fields together. By using the "Schivardi Swap" simulation to interpret the model's output, it operationalizes Caballero's theory using XGBoost, creating a tool that not only predicts individual firm failure but works out the systemic efficiency gain of "sanitizing" the credit portfolio. Basically, we are bridging the gap between a high-level economic theory and the row-by-row data of actual struggling firms.

3 - Data Description

Setting up a dataset to predict a corporate turnaround calls for a fundamental shift in perspective. It means moving away from a general "screening" capability and toward a specialized triage protocol. Standard credit risk literature usually deals with mixed populations, teaching algorithms to tell the difference between healthy and failing firms. However, that binary split stops being useful when looking at a "Distressed Cohort," which is a group of firms that have already crossed the line into technical insolvency. In this setting, the challenge is no longer figuring out if a firm is in trouble, but picking out which of those firms have the hidden strength to bounce back. Essentially, we aren't asking if the patient is sick; we are asking if they can survive the surgery.

This chapter lays out the methodological architecture used to build a longitudinal panel of Italian SMEs. The data work described here isn't just a technical exercise in cleaning and fixing numbers; it is a forensic look at the corporate landscape during a time of extreme volatility. Covering the period from the "Market Hibernation" after the pandemic to the return of normal interest rates in 2023, the dataset is built to see through the "Liquidity Artifacts" created by government help. The goal is to isolate real operational resilience from firms that were just propped up by state aid. Practically speaking, we are stripping away the temporary "mirage" of government cash to see who is actually standing on their own two feet.

To get there, the chapter follows a strict plan of "Strategic Retention" and "Technical Sanitization." Unlike standard models that might throw out extreme outliers or stick to legal labels, this workflow sets up a custom "Hybrid Target Variable." By putting economic truth ahead of legal paperwork, we define failure not just as bankruptcy, but as "Zombie Stagnation," the inability to get back to positive equity. We also put "Firewall Protocols" in place to keep out look-ahead bias. This ensures the machine learning model has to guess survival based only on the financial data available at the exact moment of the decision. Essentially, the model isn't allowed to "cheat" by looking at future results; it has to make its call using only the clues available at the start.

Finally, this chapter turns raw records from the AIDA database into a clear picture of corporate distress. It marks out the boundaries of the analysis, focuses on the rigid nature of the SME sector, and gets the "Hybrid Feature Space" ready for the XGBoost algorithm. This allows the model to pick up on the jagged signals of recovery that simpler formulas would likely miss. In doing so, we move from a purely mathematical exercise to a decision-support tool designed for the high-stakes reality of credit allocation.

3.1 Data Source and Sample Selection

3.1.1 Data Source

The data for this study has been collected from AIDA (Analisi Informatizzata delle Aziende italiane) by Bureau van Dijk, part of Moody's Analytics. Its empirical robustness makes it the standard for statistical research in the Italian context. AIDA, therefore, is the most authoritative and sympathetic data provider of the segment, also seen in "Machine learning models for bankruptcy prediction in Italy: Do industrial variables count?", Bragoli et al. (2019), "Recurrent ANNs for Failure Predictions on Large Datasets of Italian SMEs", Nadali, Corazza, Parpinel & Pizzi (2020) and "Lost in a black-box? Interpretable machine learning for assessing Italian SMEs default", Crosato, Liberati, Repetto (2023), notwithstanding their use of BvD Orbis, a sister platform within the same ecosystem. The granularity of this database allows for a simultaneous analysis of balance sheet aggregates, profit and loss accounts, and official legal statuses, ensuring that the transition from financial distress to either recovery or insolvency is tracked with high precision.

3.1.2 SME Perimeter

The focus of this study revolves around Italian SMEs, and therefore, the sample was restricted exclusively to SMEs during the data selection on the platform, applying the European Commission's official thresholds for the segment. Consequently, the sample selection adheres to the criteria established by EU Recommendation 2003/361/CE:

- <250 Employees
- <50 Million Euros in Annual Turnover
- <43 Million Euros on Total Assets

The choice to isolate this segment is motivated by the structural characteristics of the Italian production fabric, where SMEs represent the vast majority of economic activity. By selecting these firms, we focus on the segment which is susceptible to liquidity constraints, exhibits a high degree of bank-dependency, and lacks the institutional 'too-big-to-fail' safety nets and state-led restructuring avenues available to large-cap corporations.
Crucially, the imposition of the €43 million asset ceiling allows this study to control for the phenomenon of "Asset Heaviness" identified by Crosato, Liberati, and Repetto (2023). Within the Italian manufacturing context, high tangibility of assets in SMEs often translates into structural rigidity rather than financial security. Unlike large diversified conglomerates, SMEs typically lack the option to divest non-core business units to generate emergency liquidity. Consequently, their "Asset Heaviness" behaves as a structural trap where capital is locked in illiquid operational infrastructure, rendering the firm prone to the specific type of structural insolvency analysed.

3.1.3 The "Technical Default" Filter and Structural Rationale

This study focuses on entities in financial distress, differentiating between firms undergoing a transitory phase and those facing terminal insolvency. The Italian context provides a unambiguous definite threshold in Art. 2482-bis/ter of the Civil Code, which states that any firm with equity below legal minimum must act a recapitalization without delay, failing which, the entity is legally compelled to initiate liquidation or restructuring proceedings. While this represents the *de jure* requirement, *de facto* evidence suggests firms often spend years with negative equity as the norm isn't promptly applied, a phenomenon exacerbated by the structural "Asset Heaviness" of Italian SMEs which creates barriers to rapid liquidation (Crosato, Liberati and Repetto, 2023).

To select the desired domain of structural recovery rather than liquidity shocks, a strict Equity < 0 filter has been applied. In support of this methodological choice, three elements were considered. Essential to the study is the distinction between Technical Insolvency and Insolvency within a Bankruptcy Consideration, as observed in Altman & Hotchkiss's (2005) framework. Specifically, technical insolvency refers to a temporary liquidity deficiency, the inability to meet immediate obligations. Insolvency in a bankruptcy context considers the firm's long-term economic viability, assessing whether total liabilities exceed the fair value of assets. By selecting firms with negative equity, the model excludes entities experiencing transitory illiquidity or working capital mismanagement to focus instead on the fundamental structural collapse of the business model. Ohlson (1980), in order to address the mathematical breakdown of financial ratios when liabilities exceed assets, also introduced a flagging variable for negative equity (OENEG)[28]. Consequently, the exclusion of positive equity also involves sudden failures of firms who showed no signs of collapse in the precedent period. Beaver (1966) suggest that firms approaching collapse often adopt "window dressing" practices, effectively maintaining an appearance of a solvent firm on financial reports, while economic reality was significantly and definitively different. Therefore, the exclusion of positive-equity firms also acts as a quality control of the signal, removing cases where accounting items might be decoupled from the true state of the firm. Finally, by focusing on negative-equity firms, we isolate

[28] Owners Equity NEGative.

entities already at the point of technical default, specifically investigate predictors that participate in a reinstatement of positive net worth. This focus effectively bypasses the arbitrary timing of legal default often criticized in literature (Balcaen and Ooghe, 2006).
This methodological approach reveals a notable insight: legal bankruptcy frequently functions as a lagging indicator rather than a real-time signal, as firms typically show "zombie" characteristics years prior to formal filings (Caballero, Hoshi and Kashyap, 2008). Acknowledging this temporal lag, means shifting away from a traditional "healthy vs. bankrupt" binary, which can involve selection bias, towards a more "dynamic recovery" vs. "terminal failure" perspective.

With this technical anchoring, we specifically investigate turnaround dynamics that help resurface to solvency against those that signal an irreversible decline toward liquidation. This defines the whole research purpose: given that a firm is already in distress, which factors can determine the likelihood of a successful restoration of technical solvency?

3.1.4 Temporal Horizon

The longitudinal panel stretches across a six-year span, ranging from 2019 to 2024, with 2022 acting as the reference year (T_0) for picking out the distressed cohort. Essentially, the model looks back at three years of history (2019–2021) to track how a firm's health was sliding before checking on its recovery over the final two years (2023–2024).
The observation window was built to catch a period marked by major structural breaks and market instability. By anchoring the data in 2019, the model sets up a pre-pandemic baseline of "business-as-usual" financial health. This baseline acts as a control for structural soundness before the world changed. As Moscatelli et al. (2020) argue, during times of intense fluctuation, the stability of standard indicators often breaks down, making it necessary to use a solid historical baseline to train machine learning models effectively.

The dataset is defined by three distinct macroeconomic phases:

1. **Market Hibernation (2020–2021):** The pandemic triggered an economic "stop-and-go" that forced markets into a state of artificial hibernation. While lockdowns halted cash flows, the risk of mass bankruptcy was held back by a massive influx of state aid. As noted by Schivardi et al. (2020), these interventions were meant to save solvent firms from temporary cash squeezes. However, this created a "Liquidity Artifact," where debt moratoria turned immediate insolvency into deferred debt. Essentially, the government put the economy on life support, which made it hard to tell who was actually healthy and who was a "Zombie" firm just hanging on through state support rather than market viability.
2. **Supply Chain Dislocation and the Bullwhip Effect (2021):** As economies tried to get back on track, the manufacturing sector struggled with logistical bottlenecks. This "bullwhip effect" meant that even as demand returned, firms were hit by a second shock, namely the inability to fill orders because global logistics had seized up. This phase brought about operational stress that standard financial ratios often miss unless they include the momentum indicators used in this study. In short, a firm could have plenty of orders but still fail because it couldn't get the raw materials to finish the job.
3. **The "Weaning Off" and the Refinancing Cliff (2022 – T_0):** The reference year, 2022, marks the point where geopolitical instability, the energy crisis, and inflation all came to a head. Crucially, it lines up with the "Weaning Off" effect described by Grajzl et al. (2023). Their research shows that cutting off public cash flows fundamentally disrupts firms that used aid as a crutch rather than a cure. Furthermore, recent articles highlight that the jump in interest rates during this time significantly pushed up credit risk (Peykani *et al.*, 2025). Basically, the "Refinancing Cliff" is the moment of truth where firms must prove they can survive without subsidized loans.

Consequently, the distress observed in 2022 is used as the T_0 anchor because it is less likely to be a temporary "mirage" and more likely to reflect deep-seated structural issues. These problems only became visible once the fiscal life support was pulled away and the cost of capital began to return to normal. The final outcome years (2023–2024) act as the verification horizon. They provide enough of a time lag to see whether firms could actually pull through the transition from a subsidized world to a high-interest market environment.

3.1.5 Attrition Table and Cohort Selection

The sequential application of the filters led to a drastically reduced final sample of the distressed cohort. To summarize the subsequent selection of the passages just mentioned, an attrition table is refined as follows:

Step	Filter Description	Remaining (n)	Reduction (%)
Raw Dataset	Initial AIDA extract (post-deduplication)	319,266	-
SME Filter	Exclusion of Large Corporations (EU Rec. 2003/361/CE)	318,458	−0.25%
Noise Removal	Removal of Administrative records and unidentified statuses	286,855	−9.93%
Distress Filter	Selection of cohort with negative equity (final)	87,311	−69.56%

Table 1: Dataset Selection Process and Attrition

This section defines the validity perimeter of the model. By focusing only on firms "below the waterline," specifically those with Negative Equity, this study sets up an explicit Domain Restriction. This boundary is critical for the fine discrimination between insolvent entities and recovering enterprises. In simple terms, we are looking past the broader market to focus exclusively on the distressed segment, filtering for companies that retain the potential for a successful turnaround.

Methodologically, this approach gets around the problem of "sampling selectivity" raised by Balcaen and Ooghe (2006). While traditional models often show high accuracy by pairing distressed firms with clearly healthy ones, a practice that makes the classification task look easier than it really is, this study keeps the focus on a homogeneous group of distressed entities. Essentially, it is easy to tell a healthy runner from someone on crutches, but it is much harder to tell which person on crutches will walk again. By narrowing the field to only the "injured," we ensure the model learns to pick up on the subtle signs of recovery.
This fits with the triage framework proposed by Forstner et al. (2025), who argue that the drivers of a turnaround stand apart from the drivers of a decline. Consequently, the "Distress Filter" is the most vital step. Even though it leads to the biggest drop in sample size, cutting it down by 69.56%, it is what turns the analysis from a general risk check into a specialized recovery prediction tool. In short, we are weeding out the noise of healthy firms to focus entirely on the high-stakes reality of financial rescue.

3.1.6 Data Ingestion and Integrity

The computational workflow was carried out within an R Markdown environment to make sure the results are solid and easy to repeat. Essentially, by using R Markdown, we create a "living" document where the code and the commentary stay locked together, making the whole process transparent. This approach moves away from the messy reality of disconnected scripts and ensures that every

step of the data analysis follows a clear, logical path. We used the `pacman` management system to keep a tight grip on package dependencies and library versions. This prevents different pieces of software from clashing with one another during the analysis. In short, it acts as an automated gatekeeper that ensures every tool used in the code is the right version, cutting down on the risk of technical glitches. This setup ensures that the entire study, from data cleaning to the final results, is built on a stable and reproducible foundation, allowing the logic of the model to stand up to outside scrutiny.
Raw data went through a manual check followed by an automated deduplication plan. We used the Tax ID (Codice Fiscale) as the primary key to weed out double entries, ensuring the dataset looks at unique companies rather than repeating administrative records. In short, we made sure each row represents a single firm to keep the data from becoming skewed. Time periods were also lined up at the annual level to make sure all financial reports match.
To handle numerous categories without losing economic meaning, we used two ways to group data. First, we cut down ATECO industrial codes to their first two digits. This picks up on broad industry trends while getting around the problem of having too few examples in very specific categories. Second, we grouped regional data into three macro-areas, North, Centre, and South. This lets us compare large spatial areas and cuts through the noise of tiny regions while still catching the major economic gaps across Italy.

The most important part of this phase was setting a bar for data quality. We put in place a strict 15% limit on missing values for each variable across the 2019–2022 window. Any variable that fell short of this mark was thrown out before we started filling in gaps. This keeps us from leaning too heavily on synthetic data, which can introduce messy noise into the model. Essentially, we only kept the data that was solid enough to trust, following the standards for data quality used in modern financial research.

3.2 Target Variable Definition (`default_label`)

3.2.1 The Target

Defining success or failure within a "Distressed Cohort" necessitates an epistemological shift away from the traditional binary outcomes used in general bankruptcy prediction. While seminal literature, such as Altman (1968) and Ohlson (1980), relies on the legal dichotomy of *Bankrupt* versus *Active* to train predictive algorithms, this classification proves insufficient when applied to a population already characterized by negative equity. In this specific setting, legal status ceases to be a reliable proxy for economic health due to the "arbitrary timing" of formal insolvency proceedings, a limitation highlighted by Balcaen and Ooghe (2006). For a firm that is already structurally insolvent, the absence of a bankruptcy filing does not imply operational viability; rather, it often signifies a regulatory lag or creditor forbearance.
The limitation of relying solely on legal status lies in its inability to detect "Zombie" inertia. As theorized by Caballero, Hoshi, and Kashyap (2008) in their analysis of the Japanese stagnation, a legally insolvent firm may remain legally active for extended periods due to "evergreening", sustained only by "sham loan restructurings" and subsidized credit. These entities create negative externalities known as "congestion effects," reducing profits for healthy competitors and deterring entry, thereby stifling aggregate productivity (Adalet McGowan, Andrews and Millot, 2018). Consequently, reliance on bankruptcy codes would introduce a systematic "Type II Error" (False Negative) into the model, classifying these parasitic entities as "survivors" simply because they have not yet been liquidated. This would mask the "Solvency Illusion" identified, where firms survive not through efficiency, but through the artificial suppression of debt servicing costs.
Therefore, a substantive economic definition of the target variable is applied, grounded in the Turnaround Process Model proposed by Robbins and Pearce (1993). This framework posits that recovery is a two-stage process: "Retrenchment" (stabilization and survival) followed by "Recovery" (*growth*

and strategic reconfiguration). For a distressed SME, the primary objective is not immediate growth, but the successful execution of Retrenchment to halt value destruction. Accordingly, the target variable is operationalized to distinguish between two qualitatively opposite economic states:

1. **Solvency Restoration (Success):** Success is rigorously defined as the restoration of structural autonomy. A firm is classified as healed only if it successfully reverses its equity deficit, transitioning from negative to positive Shareholders' Equity. This aligns with Altman and Hotchkiss (2005), who distinguish between temporary "technical insolvency" and chronic "bankruptcy insolvency" (liabilities exceeding assets). A return to positive equity signals that the firm has exited the zone of chronic insolvency and re-established the capital buffer necessary for independent operation.
2. **Terminal Inertia (Failure):** The definition of failure is expanded to encompass not only formal liquidation but also "Zombie Stagnation." This category includes active firms that fail to restore technical solvency (Positive Equity) over the observation period. Economically, these firms represent what Schivardi et al. (2020) describe as "capital traps," where resources are consumed to cover operating losses rather than productive investment. Regardless of their legal status on the business register, these entities are classified as failures because they remain in a state of value destruction, indistinguishable from terminal insolvency in terms of resource allocation efficiency.

3.2.2 Hierarchical Labeling

Unlike traditional binary classifiers that rely only on static legal codes, this study uses a Hierarchical Priority Protocol to set up the target variable. Essentially, instead of just checking if a firm is legally dead, the protocol follows a strict set of rules to figure out if it is economically viable.

Before classification, a Noise Removal filter is used to clear out records with messy administrative statuses. Records containing ambiguous administrative codes (e.g., *'Cancellata d'ufficio'* without clear insolvency markers) or undefined statuses are systematically purged. This ensures that the target variable is binary and definitive, rather than diluted by administrative churn. In essence, we strip away the bureaucratic static so the dataset only reflects real economic outcomes. Subsequently, the target is built up through three priority levels.

Priority 1: Verified Restoration (Class 0)

The algorithm first looks for positive outcomes, marking a firm as a success only if it stays "Active" and builds up positive equity after the observation period. This rule also covers Mergers and Acquisitions. By checking for positive equity after a deal, the model can tell the difference between a real recovery and a fire sale of failing assets. Crucially, this protocol applies a conditional logic to Mergers and Acquisitions (e.g., *'Fusione mediante incorporazione'*). Unlike standard models that might categorize all M&A exits as 'Success' or 'Censored,' this study classifies them as Class 0 only if the post-transaction entity exhibits positive equity. This prevents the misclassification of 'distressed sales', where a firm is absorbed solely to salvage assets, as genuine turnarounds.

Priority 2: Legal Failure (Class 1)

This level deals with Legal Failure. Firms that enter formal liquidation are automatically marked as failures. At this stage, the model stops looking at financial ratios and accepts that the firm has hit the end of the line. This is the most clear event within the target variable definition.

Priority 3: Technical Failure / "Zombie" (Class 1)

This covers what we call "Zombies." This is a critical final step. Any firm that stays legally "Active" but fails to get back to positive equity is moved over to the failure category. This fix gets around

the problem where firms stay legally alive long after they have economically failed. By flagging these entities as failures, the model picks up on chronic stagnation. It learns to tell the difference between a firm that is actually bouncing back and one that is just hanging on through artificial survival. The adoption of this class addresses the critique by Balcaen and Ooghe (2006) regarding the 'arbitrary timing' of legal bankruptcy. By incorporating firms with persistent negative equity into the Failure class, the model corrects for the temporal lag between economic insolvency and its legal recognition, effectively capturing the 'Zombie' firms defined by Caballero et al. (2008) that survive through lending inertia rather than value creation. Consequently, strictly following the script's logic, these firms are mathematically encoded as Class 1 (default_label = 1), treating chronic inertia as functionally equivalent to bankruptcy.

3.2.3 Mitigating the Survivorship Bias

By taking on a broader definition of failure instead of sticking only to formal bankruptcy filings, this study expands the target to "technical default." This move aims to cut down on distortion in the model by seeing through the "Solvency Illusion." In the distressed group, many firms stay away from legal bankruptcy but fall short of getting back to solvency. They manage to keep up with their debt payments only through "Capital Inertia" and subsidized credit, a phenomenon known as "Zombie Lending." Essentially, just because a firm hasn't been dragged into court doesn't mean it is actually healthy. This inertia is often made worse by banks holding back on flagging technical failures to protect their own capital ratios. By doing this, they keep up the appearance of a healthy loan even when the borrower is structurally broken. Basically, both the lender and the borrower are looking the other way to avoid a painful reckoning. By putting these "Zombie" entities into the Failure class, the model gets around the survivorship bias that would otherwise mark them as successes.

While this approach tilts the classifier toward a failing bias, it gives a more careful and realistic signal of a firm's long-term health. This fits with a Cost-Sensitive Learning framework, where the model is tuned to reflect the lopsided costs of credit risk. The strategic goal is to make sure that a guess of "Equity Rehabilitation" points to a real comeback. In short, we want to ensure a firm is actually standing on its own feet before we call it a success. This helps us steer clear of "Silent Killers," or those false positives that would lead to dumping capital into a sinking ship.

3.3 Independent Variables and Feature Synthesis

Corporate distress is a multidimensional "pathology", rather than a purely forensic accounting event. To capture the underlying kinetics, a set of features is implemented with the aim to capture characteristics of the latent strategic erosion and managerial stasis, taking into consideration the framework suggested by Argenti. The final dataset is composed of 383 variables, including target variable and firmographics, structured as a longitudinal panel covering a 2019-2024 timeframe. This high dimensionality is consequential to the inclusion of granular data across multiple years, capturing not only the final snapshot of distress, but a film of structural degradation. The set includes four macro-categories of variables: Firmographics and Governance, Longitudinal Accounting Panel, Standard Financial Ratios Panel, and Synthetic Drivers.

3.3.1 Firmographics and Governance

While these stay in the baseline group, these static and semi-static variables are vital for sizing up structural weak spots like the "liability of newness," sectoral risk, or regional gaps. Although these risks come from outside the firm, including these variables helps with noise reduction and sets apart fundamental risk from temporary fluctuations. Essentially, while these factors don't change every day, they set the stage for how much pressure a firm can take. We also track employee headcount over time to keep an eye on how the size of the business shifts from year to year.

To map out the firm's governance and maturity, we bring in three proxy variables: Company Age, Listing Status, and Legal Form. Company Age acts as a stand-in for operational experience and the "learning curve" effect. In short, younger firms often run into the "liability of newness," which makes them more likely to trip up than older, more established companies.

Listing Status checks on transparency and governance rigour, since public firms have to follow strict compliance rules and undergo external audits. Finally, the Legal Form is used to tell the difference between various liability setups. Even with the thin data regularly found in small business reporting, this variable picks up on the different risk profiles and liability structures that come with various corporate frameworks. The dataset is inherently biased towards more structured entities, as it excludes non-corporate forms like sole proprietorships and partnerships (S.N.C., S.A.S.), which are not mandated by Italian law to disclose their financial statements to the public.

To account for outside shocks and the broader environment, we identify two more variables: the ATECO Sectorial Map and Regional Macro-Areas. To catch industry-specific weak spots, we grouped variables by their 2-digit ATECO codes. This helps the model isolate competitive intensity and capital structure norms. Essentially, it prevents the model from blaming a firm for a high debt load if that level of debt is actually normal for its specific industry.

Meanwhile, geographical diversity is captured through three broad macro-areas: North, Centre, and South. By sticking to these large clusters instead of tiny, granular regions, the model avoids getting bogged down in local idiosyncratic spikes. Instead, it focuses on the deep-seated economic gaps and infrastructure issues that define the Italian credit market. To put it simply, we are looking at the big picture of where a firm operates to see if its environment is helping it stay afloat or dragging it down.

3.3.2 Longitudinal Accounting Panel

The dataset brings together a wide range of raw accounting and financial records, set up not just as static inputs but as a longitudinal film of how a firm changes over time. Unlike derived ratios, which squeeze information into a single number, these raw metrics act as the "footprints" of long-term distress. Essentially, they track how fast and how hard a firm is losing its capital. Instead of looking at a single snapshot, we are watching the whole movie of a company's struggle to stay afloat.

From Ratios to Raw Data

Historically, Beaver and Altman led the way in using financial ratios to predict failure, seeing distress as the end result of negative performance trends. However, Ohlson took issue with relying only on ratios, noting that accounting statements often act as "lagging indicators" that point to the past rather than the future. This is especially true when a firm is in deep trouble and the relationships between variables start to break down. Furthermore, Crosato et al. (2023) argue that for unlisted SMEs, where market data is missing, granular accounting data stays the main source for modelling risk. To put it simply, for small firms, the balance sheet is the only real map we have, as long as we know how to read the hidden patterns.

The Granular Advantage

This study makes use of the XGBoost algorithm to handle massive amounts of data without the usual limits that hold back traditional regression. By feeding the algorithm raw data instead of pre-packaged ratios, the model can pick up on "unconventional associations" and hidden patterns. For example, it can spot specific "Cliff Effects" where the mix of raw debt and falling revenue points to a total collapse much more clearly than a simple debt ratio. In short, the algorithm is smart enough to find the "tipping points" that a human analyst might overlook.

The longitudinal panel covers the period from 2019 to 2024, allowing the model to measure the "Pandemic Shock" and the "Restructuring Phase" that followed. The data structure takes in two main areas:

1. **Income Statement (Flows):** We track operational scale and margins, including Sales Revenues and Production Costs. These show how fast the firm is moving commercially. We also follow the "profitability ladder," which looks at the full drop from EBITDA down to Net Income. Basically, this lets us see exactly where the money is leaking out of the business before it reaches the bottom line.
2. **Balance Sheet (Stocks):** This includes a detailed look at assets, such as tangible and fixed assets, to test if a firm is too "heavy" to pivot. We also split up liabilities into short-term and long-term debt. This distinction is vital for spotting a "Refinancing Cliff," as it allows the model to tell the difference between a firm that needs cash today and one that is just carrying too much debt in the long run. To monitor solvency buffers, we also keep a close eye on Cash Equivalents and Shareholders' Equity.

Data Reconstruction Strategy

While standard financial ratios were in the raw data, this study chose to throw out the pre-calculated versions and build them from scratch. This ensures the math stays consistent across the whole panel and keeps us from losing data due to missing values in the pre-made fields. By doing this, we make sure that every ratio fed into the model comes from the cleaned and fixed raw items. In simple terms, we are baking the bread ourselves to make sure every ingredient is fresh and consistent.

3.3.3 Standard Financial Ratios

To sharpen the model's ability to tell firms apart, we brought in a set of predictive drivers. We started by adding fundamental ratios pulled directly from the main P&L and Balance Sheet headers. These metrics were worked out through straightforward math using raw accounting data. Key additions include the Interest Coverage Ratio (ICR), which weighs EBITDA against estimated interest expenses to measure debt serviceability. Essentially, it shows if the company makes enough money to keep up with its interest payments.

We also included Delta Debt Maturity to track shifts in the average time a firm has to pay back its loans, along with a Cash Flow Proxy to estimate asset turnover and liquidity. Other vital signals brought into the mix are EBITDA margins, the Equity to Asset ratio, and Cash to Asset ratios. These last two are crucial because they show how much of the firm is built on its own money versus how much is propped up by debt.

We went for ratios rather than raw numbers because they are scale-invariant. This allows the model to compare firms of completely different sizes on a level playing field. In short, using ratios ensures that the "financial physics" of a small family business can be measured against a medium-sized manufacturer without the larger firm's raw volume drowning out the smaller one's efficiency signals. By focusing on these proportions, the model can pick up on the structural health of the business regardless of its total turnover.

3.3.4 Advanced Drivers

Momentum

Supplementing static ratios by leveraging the multi-year panel structure, drivers encapsulating the trajectory and velocity of these financial changes, momentum indicators have been included. Drivers like two-year Sales growth and EBITDA margin variations, act as dynamic vectors, distinguishing between operational retrenchment, terminal deterioration, and stagnation. In accordance with Shumway's (2001) work, a firm's health is a dynamic process with cumulative path dependency

where the *direction* of movement is as critical as its current state. Momentum Indicators representing the year-on-year growth rates for key drivers were implemented. For a given financial variable X, the momentum at time t is defined as:

$$\Delta\%X_t = \frac{X_t - X_{t-1}}{\mid X_{t-1} \mid + \epsilon} \tag{1}$$

where ϵ is a small constant (10^{-6}) included to ensure numerical stability in cases where X_{t-1} approaches zero. These variables function as "velocity vectors" capable of detecting early signals of solvency restoration.

Scissor Effect

To quantify structural divergence between liability accumulation and revenue generation, the Scissor Effect metric is introduced.

$$\text{Scissor Effect}_t = \Delta\%\text{Total Debt}_t - \Delta\%\text{Sales}_t \tag{2}$$

This metric is rooted in the Capital Consumption Theory. A positive Scissor Effect signals that the company is "consuming" its future cash flows to service a liability base that is expanding faster than its commercial viability, a classic sign of an unsustainable distress trajectory, identified by Hillegeist et al. (2004). Particularly pertinent to 2022, this metric is significant as the surge of gas prices and the consequent energy crisis precipitated a generalized inflationary surge. This variable behaves like a proxy for Pricing Power, differentiating between firms possessing the pricing power necessary to safeguard margins, and demand-sensitive enterprises forced to erode their capital. The price elasticity of demand became a primary determinant of resilience as opposed to accelerated liability accumulation.

Distressed History Score

To account for the cumulative burden of capital erosion and the resulting erosion of resilience to endure a crisis, the Distressed History Score is added. The score accounts for each year the firm spent with negative equity, reflecting a state of entrenched financial distress. Considering such path-dependent distressed cumulative score, helps to profile between an acute shock, a single year, and a chronic distress, persistent through the years. The persistence of technical default typically triggers a deterioration of credit ratings. A high score suggests that the enterprise structure is structurally compromised. In the context of the turnaround stages defined by Robbins and Pearce (1993), standard '*retrenchment*' strategies, such as cost containment, often prove ineffective if the firm has already depleted its internal resources through years of capital erosion. The DHS serves as a proxy for structural insolvency and terminal financial mortality. Furthermore, the score acts as proxy for what Caballero, Hoshi and Kashyap (2008)defined as Zombie firms : enterprises with prolonged negative equity and surviving through "financial inertia", typically with subsidized loans or credit forbearance. In this study, Shareholders' Equity was preferred over more granular interest-based metrics like the Interest Coverage Ratio (ICR) for two primary reasons. First, the legal relevance of the negative path equity in art. 2482 c.c., which explicitly identifies the event as legal trigger for liquidation/recapitalization. Second, during the pandemic period (2020-2021), extensive public moratoria on interest payments rendered the ICR artificially low, making it an unreliable signal of structural health compared to the accumulated equity deficit.

3.3.5 Consolidated Drivers

To ensure methodological comparability with the Machine Learning applications for Italian SMEs, this study integrates a specific subset of nine predictors identified by Crosato, Liberati, and Repetto

(2023). In their seminal work on interpretable machine learning (XAI), these authors demonstrated that this specific configuration of variables optimizes the trade-off between predictive accuracy and economic interpretability within the Italian manufacturing and service sectors.

While the *Synthetic Drivers* described in Section 3.2.4 capture the *derivative* dynamics (momentum and velocity) of the distress, this feature set captures the *structural* state of the firm across three fundamental dimensions. As implemented in the computational workflow, these variables were engineered to test the persistence of non-linear phenomena, such as the "Asset Heaviness" paradox, specifically within the Distressed Cohort:

Size Metrics (Input/Output/Resources)

- **Total_Size (Total Assets):** Used to test the "Too-Big-To-Fail" hypothesis versus the structural rigidity of asset-heavy firms. Crosato et al. (2023) suggest that in the Italian context, high asset tangibility can paradoxically increase default risk due to lack of flexibility. This variable is also mentioned as the "stock" aspect of the firm.
- **Sales_Volume (Revenues):** Represents the market output dimension, quantifying the commercial scale of the entity. This variable is also adopted to assess the "flow" aspect of the firm.
- **Num_Employees:** Represents the input dimension of firm size. This study utilizes this metric to investigate non-linear risk thresholds related to organizational complexity and labor rigidity, testing the "Liability of Growth" hypothesis.

Profitability Ratios

- **ROE (Return on Equity):** Calculated as *Net Income / Shareholders' Equity*. Within this study, specific exception handling (`safe_ratio`) was applied to manage the sign inversion and mathematical instability caused by negative equity denominators.
- **ROCE (Return on Capital Employed):** Estimated as *EBIT / (Total Assets - Short Term Debt)*. This ratio isolates operational efficiency from tax and leverage structures, distinguishing between firms that are simply large and those that efficiently deploy long-term capital.
- **Profit_Margin:** Defined as *Net Income / Sales Revenues*, capturing the final residual value generation per unit of revenue.

Structural and Liquidity Ratios

- **Solvency_Ratio:** Defined as *Shareholders' Equity / Total Assets*. As the cohort is characterized by negative equity, this ratio is inherently negative. Crucially, a "Hard Cap" at −5 (representing an equity deficit five times larger than assets) to prevent extreme outliers from destabilizing the gradient boosting algorithm is applied.
- **Gearing_Ratio:** Calculated as *Total Payables / Total Assets*, serving as a proxy for the overall leverage burden independent of equity value. This metric explicitly quantifies the "Asset Heaviness" burden described by Crosato, identifying firms where the asset base is entirely financed by third-party liabilities.
- **CF_Generative (Cash Flow Proxy):** Unlike standard definitions dependent on cash flow statements (often unavailable for smaller SMEs), a synthetic operating cash flow is reconstructed. It is calculated as *Net Income + (EBITDA - EBIT)*, effectively adding back depreciation and amortization proxies to the net result. This aligns with the emphasis in recent literature on "Cash Flow" as a superior predictor of survival compared to static liquidity ratios like the Current Ratio.

Finally, it is essential to note that the feature space is not composed solely of synthetic ratios. As detailed in the computational workflow, the model strategically retains critical original variables directly from the raw dataset, specifically *Total Assets* (Total_Size), *Sales Revenues* (Sales_Volume), and the *Number of Employees*. Preserving these raw structural dimensions is not merely a data

retention strategy; it is a methodological prerequisite for valid external benchmarking. By adopting this hybrid feature set, blending calculated efficiency ratios with original dimensional metrics, the model establishes a common ground for comparing Internal Work findings against established External Literature.
This approach is particularly vital for the application of Accumulated Local Effects (ALE), discussed in Chapter 5. Using raw scales allows the ALE visualization to map specific non-linear distress dynamics that purely synthetic ratios would obscure, such as the "Liability of Growth" threshold often observed near the 50-employee mark or the absolute scale of "Asset Heaviness" in manufacturing firms. Thus, the feature set is designed not just for predictive maximization, but to function as a bridge between algorithmic outputs and observable economic reality.

Feature Category	Core Concept	Temporal Depth	Total Var.	Base Conc.	Inflation
Firmographics & Governance	Structural, legal and geographic metadata defining the entity	1 (19-24 Empl.)	67	9	Dummyfication of categorical variables introduces multiple variables for iterations
Longitudinal Accounting Panel	Raw absolute values from Balance Sheet and P&L	2019-2024	182	17	Multiple years introduce several iterations
Financial Ratios	Standardized performance metrics	2019-2024	119	18	Multiple years introduce several iterations
Advanced Drivers	Synthetic indicators for Momentum, Stress, and History	at T_0	24	24	Different depths produce several iterations
Target Variable	The binary classification outcome	2023-2024	1	1	-
		Total	**393**	**69**	

Table 2: Feature Breakdown and Variable Inflation

3.4 Preprocessing and Data Cleaning

In order to ensure suitable data for a machine learning pipeline out of a highly dimensional financial dataset, rigorous preprocessing was included in the workflow. The main objective to address during this stage were technical redundancy, integrity of the data and imputation of missing data through a robust industry-specific logic.

3.4.1 Semantic and Technical Sanitization

As raw items from Balance Sheet and Income Statement are often redundant, if left unaddressed, they would introduce significant noise and inflationary bias to the feature space. Therefore, accounting synonyms like *Utile Netto* and *Profit/Loss* were consolidated. In order to find all the semantic aliases, a systematic collinearity analysis was performed, identifying variables with near-perfect overlap ($r > 0.98$). Furthermore, redundant iterations were excluded from the dataset, mitigating the risks associated with multicollinearity (e.g., *Sales Revenues* and *Revenue on Sales of Goods and Services*). Moreover, as raw data from AIDA were named in Italian, each variable was renamed in english language with a snake_case nomenclature, in order to facilitate reproducibility and interoperability with modern libraries, while also ensuring clarity over subsequent outputs.

3.4.2 Missing Values Imputation

Given that financial variables exhibit high structural interdependence, the protocol adopts a Ratio Imputation approach derived from De Waal, Pannekoek, and Scholtus (2008). As defined in the data description, this method imputes missing values (y_i) not as static averages, but as estimates of a sector-specific median ratio ($\tilde{R}$) applied to the firm's specific scale (x_i) .

$$y_i = \tilde{R}_{t,s} * x_i \tag{3}$$

Crucially, to preserve accounting coherence, the protocol adheres to a strict Stock-Flow Dichotomy implemented in the computational workflow:

1. **Stock Variables (Balance Sheet):** Items such as Inventory, Payables, or Liquidity are anchored to Total Assets (x stock). This ensures that the imputed values respect the structural proportions and capitalization of the firm, preventing a small entity from receiving a median debt value derived from larger peers.
2. **Flow Variables (Income Statement):** Operational items such as Costs, EBITDA, or Services are anchored to Sales Revenues (x flow). This captures the "operational velocity" of the firm, ensuring that imputed margins reflect the entity's actual commercial volume.

By linking the median multiplier ($\tilde{R}$) dynamically to the specific year (t) and sector (s), and scaling it by the firm's actual drivers (x_i), the imputation ensures size-integrity. A micro-enterprise and a medium enterprise within the same sector will receive the same *ratio* imputation, but vastly different *absolute* values, thereby preserving the variance required for the XGBoost model to detect distress signals accurately. The drivers of the firm are strictly not imputed, and removed from the dataset. In fact, these are very basic and significant data that should not be missing, therefore the entire observation is removed. Sector-specific stratification finds significance in Bradley, Jarrell, and Kim (1984), who demonstrated significant variations across industrial sectors, discouraging a "one-size-fits-all" imputation. Finally, as argued by Kennedy, Lakonishok, and Shaw (1992), the mean is an inefficient estimator in the presence of financial ratios that can present extreme outliers. Even more so within a distressed cohort, which is notoriously characterized by skewed distributions and heavy tails. Consequently, this study adopts the Median within each adjustment cell (Sector/Year), aligning with Eurostat guidelines for robust estimation in the presence of outliers.

The imputation and feature engineering protocols were exclusively applied to the 2019–2022 observation period, ensuring that no data was imputed for years preceding an entity's incorporation.

3.4.3 Firewall Protocol

To stop target leakage and protect the model's integrity, a strict Temporal Firewall was put in place. This protocol cleared out all features containing information generated after the reference year T_0 (2022). By getting rid of any indicators from the outcome period, such as future-dated financials or updated legal markers, the workflow sets up a realistic prediction environment. Essentially, we made sure the model cannot peek at the answers before it takes the test.

This setup ensures the classifier stays away from look-ahead bias. Ohlson (1980) took issue with early studies for "back-casting," where models used reports that were actually made public after a firm had already failed, labelling it as a major flaw. By sticking to this firewall, the model leans only on information available to a creditor at the exact moment of assessment. This prevents "perfect predictions" based on artifacts that would otherwise pump up the performance scores.

This approach also fits with Shumway (2001), who argued that models must avoid bias by using only the data history available up to the point of prediction. In short, we are forcing the algorithm to play by the rules of the real world, where the future is always hidden.

3.4.4 Algorithm-Aware Outlier Policy

To preserve the data sample and avoiding extensive trimming, this study adopts a policy of Strategic Retention and Technical Sanitization, effectively bypassing the constraints of Gaussian assumptions. Traditional econometric models like Altman's Z-Score or Logistic Regression, often exclude extreme distress signals to ensure mathematical stability, while this procedure aims to preserve their informative value.

The initial phase of selection is the mentioned filter for only SMEs, applying EU Recommendation 2003/361/CE criteria (Turnover<€50M, Assets<€43M, Employees<250). While this aligns with the primary scope of the research, it also serves a technical purpose: large corporations typically present as statistical outliers within a distressed population, and their presence usually skews feature distribution.

To ensure the numerical stability of the high-dimensional feature space, a suite of 'safe functions' was applied during the transformation of raw accounting data into ratios. In the context of distressed SMEs, where negative equity or zero-turnover are common, standard calculations frequently generate mathematical artifacts, such as infinite values (inf) or undefined results (NaN).

If left unmanaged, these artifacts can compromise the gradient boosting process, leading to erratic split decisions and model destabilization. To mitigate this, an exception-handling protocol was implemented: whenever a zero-denominator is encountered, the function intercepts the operation and sanitizes the output to a neutral value (zero). This ensures that the XGBoost algorithm processes a continuous and bounded distribution, preventing 'explosive' gradients and maintaining the integrity of the feature importance rankings.

Strategic Retention and Technical Sanitization

Traditional *Statistical Winsorization* (e.g., clipping financial ratios at the 1st and 99th percentiles) has been intentionally avoided to preserve the integrity of the distress signal. In a standard Gaussian framework, extreme outliers are viewed as noise; however, within a Distressed Cohort, extreme values, such as a Leverage Ratio exceeding 50x or deep negative margins, are often accurate representations of "Zombie" inertia rather than reporting errors,.

This choice leverages the internal logic of the XGBoost algorithm. Unlike linear models (e.g., Logistic Regression) or distance-based algorithms (e.g., KNN) that are sensitive to Euclidean distance and require normalized distributions, gradient boosting relies on recursive partitioning. The algorithm searches for optimal split points based on rank order, making it inherently invariant to monotonic transformations of the feature space. Consequently, preserving extreme values allows the model to learn localized "Cliff Effects", non-linear thresholds where a firm's financial physics fundamentally shift from recoverable stress to irreversible insolvency.

However, to ensure numerical stability during gradient descent, a protocol of "Technical Sanitization" was prioritized over statistical trimming. Specifically, structural ratios susceptible to mathematical explosion due to near-zero denominators (e.g., Equity ≈ 0) were subjected to "Hard Capping" rather than distribution-based winsorization. As detailed in the computational script, variables such as the *Leverage Ratio* were capped at a ceiling of 50, and the *Solvency Ratio* was floored at −5. This "safety rail" approach prevents infinite gradients from destabilizing the loss function while ensuring that the model continues to treat these observations as extreme high-risk signals at the very edge of the "Event Horizon".

3.5 Feature Optimization and Audit

While the previous stage aimed to clean the dataset from synonyms and identical items, we now look to optimize for statistical stability, checking for overwhelming redundancy in the model.

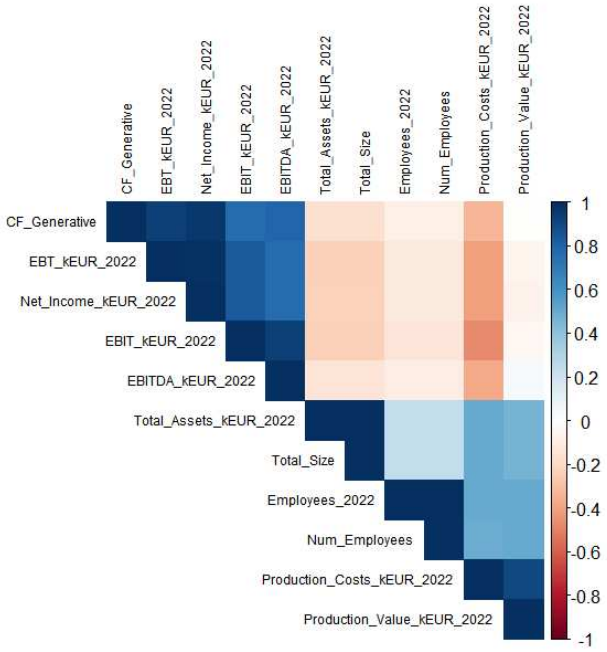

Figure 1: The featured displayed have shown significant correlation between them, some due to technical reasons, like the different stages of revenues (EBT, EBIT, EBITDA, Net Income and CF Generative), others by semantic repetition, and others showing significant resemblance.

3.5.1 Multicollinearity Management

Beyond semantic overlap, certain variables within the financial dataset exhibit high collinearity, which can mask individual feature contributions. Unlike the previous audit for semantic aliases where we looked for near-identical overlaps, this protocol adopts a Pearson pairwise correlation matrix to identify feature pairs and visualized through a correlation heatmap.

The visual inspection provided by the Correlation map reveals two informational clusters, one of the production value and production cost, which are intuitively related, and another cluster formed by EBT, EBIT, EBITDA and Net Income, as they share the same meaning, just at different stages of accounting. Following the framework by Kuhn & Johnson (2013), visualizing the correlation matrix, we can disclose groups of variables that express the same information. Excluding them prevents feature dilution within the model.

Cut-off Threshold has been set at $r > 0.90$, a conservative limit to preserve informational nuances for specific "edge case" turnarounds, and trusting the inherent architectural strength of Gradient Boosted Decision Trees. Since XGBoost adopts recursive partitioning, it's particularly robust to multicollinearity, allowing the model to pick the variable with the highest predictive gain when considering between a moderately correlated pair.

In order to preserve key fundamental risk drivers, as the `cash_flow_proxy`, a list of features were shielded from removal, ensuring the final feature space remains aligned with the economic frameworks underpinning the model. With this protocol, if two features showed a coefficient of 92%, one was to be removed, particularly the one not included in the "shielded list".

3.5.2 Leakage Audit

To validate the efficacy of the Temporal Firewall, a post-processing Leakage Audit was conducted. This statistical scan systematically evaluated the point-biserial correlation between each finalized

independent variable and the binary `default_label`. The objective was to identify any 'informational artifacts' that might have bypassed the manual filtering process. Features exhibiting anomalously high correlation with the target, which would suggest the presence of post-T_0 information, were flagged for re-evaluation. This secondary defensive layer ensures that the model's predictive power is derived from genuine early-warning signals rather than ex-post realization bias, thereby securing the validity of the subsequent machine learning training phase.

3.5.3 Class Imbalance

The distribution of the target variable is a core characteristic of the dataset, from which the model's loss function and calibration is based off. Within the distressed cohort, there is a significant imbalance between entities characterized by terminal stagnation (class 1), and those capable of a successful restoration (class 0). If not addressed, classifiers are inherently biased toward the majority class, optimizing for global accuracy at the expense of the minority class's predictive power. To address this issue, a ratio of negative samples to positive samples is computed and applied, operationalized as weighting factor.

$$scale_pos_weight = \frac{N_{\text{majority samples}}}{N_{\text{minority samples}}} \tag{4}$$

This weight will be used by the XGBoost algorithm for the `scale_pos_weight` hyperparameter. The parameter will mathematically penalize errors on the minority class more heavily, sensitizing the model toward the successful turnaround class and minimizing the omission of viable firms.

3.6 Advanced Calibration for Financial Analysis

To advance the model from a predictive tool to a decision-support system, final parameters aiding evaluation and interpretation, not only from a statistical perspective, but a portfolio management and regulatory perspective as well.

3.6.1 Value at Risk and Exposure Mapping

In the context of distressed credit, a turnaround probability lacks economic significance if not accompanied by exposure data. To implement this dimension, and following Basel III Internal Ratings-Based (IRB) frameworks, Total Payables (2022) is adopted as a proxy for Systemic Exposure at Default (EAD). While regulatory EAD is typically calculated at the facility level, this study utilizes Total Payables to capture the aggregate credit exposure of the firm toward the entire financial system at the moment of assessment (T0).

The integration of EAD transforms the model from a simple binary classifier into a pragmatic Expected Loss (EL) framework. By combining the predicted Probability of Default (PD) with the Exposure (EAD) and a standard Loss Given Default (LGD) of 45% (consistent with F-IRB unsecured standards), the study estimates the systemic cost of inertia. This weighting ensures that performance evaluation prioritizes 'Capital at Risk', allowing the model to distinguish between high-frequency, low-impact defaults and those rare but consequential high-exposure failures that drive systemic instability.

3.6.2 Economic Parameters for Cost-Sensitive Thresholding

To transition from a purely statistical classification (maximizing global accuracy) towards a pragmatic business-oriented risk framework, this study implements specific cost parameters aligned with the Basel III Regulatory Framework and the operational realities of SME lending. As argued by Hand and Till (2001), simple accuracy is often a deceptive metric in credit scoring because it fails

to distinguish between the asymmetric economic consequences of error types. Consequently, the model's decision threshold is calibrated not to minimize the error rate, but to minimize the Expected Loss (EL) based on two fundamental banking parameters defined in the computational workflow.

Loss Given Default (LGD)

The parameter for the cost of a "Missed Default" (False Negative) is established at 45%. This value is not arbitrary but strictly adheres to the Foundation Internal Ratings-Based (F-IRB) approach stipulated by the Basel Committee on Banking Supervision. Specifically, paragraph 287 of the Basel III accord mandates that "senior claims on corporates... not secured by recognized collateral will be assigned a 45% LGD". By adopting this regulatory standard rather than a lower internal estimate, the model enforces a "Prudential Floor." This ensures that the penalty for failing to identify a terminal "Zombie" firm reflects the definitive capital destruction recognized by supervisory authorities, treating the missed detection of insolvency as a direct erosion of Tier 1 Capital.

Net Interest Margin (NIM)

The parameter for the cost of a "False Alarm" (False Positive) is calibrated at 4.5%. In economic terms, this represents the Net Interest Margin (NIM), the forgone profit incurred when a viable turnaround candidate (Class 0) is erroneously rejected. This figure proxies the risk-premium pricing characteristic of the distressed SME lending market. Theoretically, this aligns with the Type I Error cost in credit scoring: the opportunity cost of rejecting a "good" loan, which differs significantly from the direct loss of principal associated with a bad loan (West, 2000; Kuhn & Johnson, 2013).

Debt Load

Debt Load serves as a flow-based leverage metric, defined as the ratio of Total Payables to Sales Revenues:

$$Debt\ Load = \frac{Total\ Payables_{kEUR}}{Sales\ Revenues_{kEUR}} \tag{5}$$

This metric measures the operational sustainability of a firm's debt. It essentially gauges how many years of top-line revenue a firm would need to cough up to wipe out its liabilities. In practice, a firm with a high debt load relative to its sales is often running on fumes, even if its balance sheet appears stable at a glance. In the computational script, this metric is technically capped at 20x to account for cases with near-zero revenue, ensuring that extreme outliers do not skew the visual analysis.

This metric is kept in the toolkit exclusively as a post-hoc forensic tool for visualization and error profiling, specifically within the Event Horizon map and the subsequent Forensic Autopsy. Its primary role is to help researchers spot the Silent Killers, which are firms that manage to maintain deceptively low debt loads relative to their sales but still fail due to deep capital insolvency. By looking at the tension between structural solvency, which is the equity-to-assets relationship, and operational sustainability, the analysis can pinpoint exactly where a firm's financial physics begin to break down.

Despite its diagnostic value, Debt Load is ruled out for the primary training feature set due to severe numerical instability. In a distressed cohort, many firms exhibit revenues that are close to or equal to zero, which ends up sending the ratio through the roof as it approaches infinity. This mathematical volatility is a significant hurdle because dividing by a zero-sum denominator creates values that the model simply cannot process. Including such unstable figures would trigger explosive gradients that throw the XGBoost algorithm off balance during the backpropagation phase. To get around this, the model is trained on the Gearing Ratio instead:

$$Gearing\ Ratio = \frac{Total\ Payables}{Total\ Assets} \tag{6}$$

The Gearing Ratio provides a mathematically stable proxy for leverage because Total Assets rarely, if ever, bottom out at zero in this dataset. By anchoring the leverage calculation to assets rather than fluctuating revenue, the model maintains a steady learning rate while still capturing the underlying risk of high indebtedness. This ensures the algorithm focuses on the structural reality of the company's obligations rather than being misled by temporary sales collapses.

Asymmetric Cost Ratio and Threshold Optimization

To operationalize these economic realities within the XGBoost loss function, a Cost Ratio is introduced to weigh the relative impact of errors. Adopting the framework of Elkan (2001) for cost-sensitive learning, the optimal decision threshold (Pth) depends on the ratio of the cost of false negatives (CFN) to the cost of false positives (CFP):

$$Cost\ Ratio = \frac{Cost_{TypeI}(LGD)}{Cost_{TypeII}(NIM)} \tag{7}$$

This calculation reveals a structural asymmetry: failing to detect a terminal firm (LGD) is 15 times more costly than rejecting a viable one. Consequently, optimizing the model implies shifting the classification threshold to prioritize Precision within the Success Class (Class 0). By adopting this conservative weighting, the model effectively insulates the lender from high-exposure capital erosion. As structured in the methodology applied, this calibration forces the classifier to accept a higher rate of missed opportunities (lower Recall on Class 0) as the necessary insurance premium to avoid the catastrophic inclusion of "Silent Killers" (Class 1) in the loan portfolio.

3.6.3 Debt Maturity Architecture and Stress-Test Metadata

To enable a dynamic behavioural analysis of corporate resilience that transcends static solvency ratios, a granular segregation of the debt structure at the reference time T0 (2022) is adopted. Unlike traditional distress prediction models that often aggregate liabilities into a single scalar (Total Financial Debt), the dataset was partitioned to isolate two distinct vectors of financial exposure:

1. **Short-Term Financial Debt:** Defined as liabilities maturing within 12 months (`Short_Term_Debts_kEUR`), this component serves as the proxy for immediate Rollover Risk. In the context of the Basel Committee on Banking Supervision guidelines, short-term exposures represent the portion of the capital structure most vulnerable to the exogenous shock of monetary tightening. Consequently, this vector identifies the critical pressure point for liquidity during interest rate hikes.
2. **Long-Term Financial Debt:** These consolidated liabilities (`Long_Term_Debts_kEUR`) are utilized to measure structural leverage and the firm's long-term capital inertia. While high long-term exposure indicates significant leverage, it is not inherently a negative indicator; rather, it often reflects the firm's capacity for capital investment and long-term strategic growth.

The rationale for this architectural split is rooted in the necessity to identify "Rate-Sensitive" entities. As argued by Caballero, Hoshi, and Kashyap (2008), "Zombie" firms are frequently sustained not by operational efficiency, but by the receipt of subsidized credit or "sham loan restructurings" that artificially depress debt service costs. By isolating short-term debt, the necessary baseline to calculate a Simulated Interest Coverage Ratio (ICR) under normalized market conditions is upheld.

This metadata allows for the execution of the "Refinancing Cliff" simulation (detailed in Chapter 5), a deterministic stress test that subjects the Distressed Cohort to a hypothetical market interest

rate of 6.5%. By projecting the cost of servicing short-term liabilities at this normalized rate, the model can distinguish between firms possessing genuine operational margins and those whose solvency is an artifact of the low-interest-rate environment (ZIRP).

Crucially, these debt maturity components are structured as validation metadata rather than direct training features for the primary predictive algorithm. In strict adherence to the Firewall Protocol previously outlined, this separation ensures that the stress-test logic remains a post-hoc diagnostic tool. This design prevents "data leakage" and ensures that the XGBoost classifier predicts turnaround probability based on structural latent variables, while the debt maturity metadata serves to economically qualify those predictions, specifically identifying "Silent Killers" without biasing the ex-ante learning process.

3.6.4 Stratified Partitioning and Final Preparation

The final stage of data preparation involved the partitioning of the dataset into training and testing subsets. To provide a robust learning environment while maintaining a rigorous verification base, this study applied a 70/30 split (70% Training, 30% Testing). This specific ratio aligns with the experimental protocols of Crosato, Liberati, and Repetto (2023) and Kuhn and Johnson (Kuhn and Johnson, 2013), who suggest this configuration as an optimal balance for capturing complex non-linear distress patterns within finite financial datasets.

Crucially, the partitioning employed Stratified Random Sampling anchored to the `default_label`. As argued by Kuhn and Johnson (2013), simple random sampling in the presence of severe class imbalance risks creating non-representative subsets where the minority class is under-populated or entirely absent. By strictly preserving the prevalence of "Terminal Failures" (Class 1) across both subsets, this study ensures that the test set remains a statistically valid reflection of the training distribution. This stratification prevents the "optimistic bias" that often arises when validation sets fail to reflect the true rarity of the target event.

Finally, to maximize computational efficiency during the gradient descent optimization, the resulting subsets were converted into `xgb.DMatrix` objects. As detailed in the computational workflow, this specialized data structure optimizes memory usage for sparse matrices, facilitating the rapid processing of the high-dimensional feature space required by the XGBoost algorithm.

3.7 From Raw Data to a High-Fidelity Diagnostic Canvas

The methodological architecture laid out here has turned a raw pile of administrative data into a "Forensic Dataset" specifically built for the triage of corporate distress. By moving away from the traditional "Screening" approach, this study has set up a strict "Distressed Cohort." Essentially, we are focusing only on the firms that have already hit the wall of negative equity.

This rigorous boundary fundamentally changes the predictive challenge. The dataset is no longer built to ask if a firm is in trouble, but to pick out which specific structural setup sets apart a "Terminal Burner" from a firm that can actually pull through. In short, we are trying to tell the difference between a company that is just burning cash and one that has the engine to start back up. The way we set up the "Hybrid Target Variable" is the main innovation of this phase. By marking "Zombie Stagnation," where a firm is active but insolvent, as a failure event, the model learns that economic death is not about a legal stamp, but about the inability to build back value. Essentially, a firm is dead when it can no longer stand on its own feet, regardless of what the business register says.

Furthermore, the truth of this "Bio-Image" of distress is backed up by "Firewall Protocols" and the "Dynamic Imputation" strategies we used. By locking out future information and keeping the size-integrity of financial ratios intact, we have made sure the analysis stays clear of look-ahead bias

and mathematical errors. Basically, the model is forced to make a call without peeking at the future, ensuring that the results reflect real-world decision-making.

Ultimately, this work has prepared the "Fuel" for the algorithmic "Engine." The resulting data space, packed with "Scissor Effects," "Distress History Scores," and "Solvency Ratios," holds the hidden, non-linear signals needed to see through the "Liquidity Artifacts" of the pandemic era. In simple terms, we have stripped away the government-backed mirage to find the real operational health underneath. The stage is now set to decode these complex patterns and find the "Refinancing Cliff" hidden deep within the balance sheets.

4 - Methodology

This chapter outlines the analytical framework and the computational techniques employed to develop the turnaround prediction model, anchoring the study at the nexus of traditional financial econometrics and contemporary algorithmic risk assessment. The methodology is designed to address the inherent limitations of static, linear predictors, such as those pioneered by Altman in 1968, by introducing a dynamic, non-parametric approach specifically calibrated for the forensic triage of the Distressed Cohort. This strategy boils down to creating a digital filter that can distinguish between firms with recovery potential and those destined for failure within a highly volatile environment.

The analytical workflow is structured into three integrated phases. First, we detail the implementation of a Cost-Sensitive Extreme Gradient Boosting (XGBoost) architecture. This section establishes that the primary objective of the model is not the maximization of descriptive statistics, such as Accuracy, but rather the minimization of economic risk, defined here as the Expected Loss. Deviating from the standard maximization of statistical accuracy, this phase focuses on a parsimonious Shallow Learner topology optimized to prevent overfitting on noisy SME data, and employs a weighted loss function to heavily weight the cost of missing Zombie firms. By doing so, the model moves away from chasing raw percentages and instead focuses on the economic gravity of the classification errors.

Second, we establish the Gatekeeper Protocol, a decision framework that replaces the arbitrary default threshold of 0.5 with an economically optimized cut-off based on Basel III parameters, including Loss Given Default and Net Interest Margin. This phase targets a sensitivity of 80% to ensure the model prioritizes capital preservation over aggressive lending. In practice, this means the system is tuned to be cautious, preferring to flag a potential risk early rather than overlooking a terminal decline.

Finally, we deploy an Explainable AI framework using the DALEX ecosystem. This choice is vital for positioning the model not as an infallible oracle, but as a robust decision support system that operates within a transparent Grey Box architecture. This framework functions as a clear bridge between algorithmic rigour and clinical credit assessment, making the internal logic of the machine accessible to human auditors. The diagnostic suite goes beyond simple feature importance by utilizing Accumulated Local Effects to visualize Cliff Effects, where small changes in financial ratios can lead to sudden collapses in creditworthiness. Furthermore, we use Interaction Screening to untangle the non-linear synergies between structural and operational variables, ensuring that every prediction can be backed up by a logical narrative during committee reviews.

4.1 Extreme Gradient Boosting

The choice of XGBoost is motivated by its ability to isolate intricate non-linear dependencies and multi-order interactions within SME financial data that traditional parametric models, such as Logistic Regression or Multiple Discriminant Analysis, often fail to detect. Unlike linear models that rely on the independence of predictors, XGBoost is designed to model non-monotonic relationships, where the impact of a financial ratio may reverse direction as a firm nears insolvency. This capability is vital for navigating the Cliff Effects frequent in the Distressed Cohort, such as the sudden exponential rise in risk when negative equity breaches a specific depth. By accounting for these sudden shifts, the model moves toward a more fluid decision support system.

Crucially, this architecture offers a forensic advantage over parallel ensemble methods like Random Forest. While Bagging algorithms prioritize stability by averaging predictions to reduce variance, Gradient Boosting adopts a sequential correction strategy, essentially chipping away at

the residual errors left behind by previous iterations. Each subsequent tree in the ensemble is explicitly trained to correct the mistakes of its predecessors. This allows the model to double down on hard-to-classify edge cases, specifically the False Negatives that often act as silent killers within a credit portfolio. This iterative refinement is what aligns the model with the primary objective of minimizing Expected Loss, as it relentlessly pursues the errors that are most costly to the lender.

4.1.1 Gradient Boosting on Distress

The selection of XGBoost over traditional models, such as Logistic Regression or Random Forest, is rooted in its superior ability to navigate high-dimensional noise and capture the intricate non-linear interactions characteristic of SME financial data. Essentially, the model filters out the static of irrelevant variables to focus on the signals that actually precede a collapse. XGBoost belongs to the family of Tree Boosting algorithms, where a strong learner is set up through the sequential addition of shallow decision trees. While Bagging methods like Random Forest prioritize stability by averaging predictions to reduce variance, they often smooth over the very Cliff Effects that signal terminal distress (Chen and Guestrin, 2016). In contrast, XGBoost minimizes bias by training each subsequent tree to correct the residual errors of the previous ensemble. It essentially chips away at the mistakes of its predecessors, allowing it to double down into hard-to-classify edge cases that linear models typically overlook.

The objective function minimized during training is the regularized Log-Loss, also known as binary cross-entropy, which is defined as follows:

$$L(\theta) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1-y_i)\log(1-p_i)] + \Omega(\theta) \tag{8}$$

In this formula, y_i represents the actual label, where 1 indicates Failure and 0 indicates Turnaround, while p_i is the predicted probability. The $\Omega(\theta)$ term represents the regularization, which acts as a mathematical tether to keep the model from running away with the noise or overfitting the training data. This choice of objective function establishes that the primary goal of the model is not the maximization of descriptive statistics, such as Accuracy, but rather the minimization of economic risk, defined here as the Expected Loss.

Crucially, unlike standard Gradient Boosting Machines that utilize only the first derivative, or the Gradient, XGBoost approximates the loss function using a second-order Taylor expansion that incorporates the Hessian. By taking into account the curvature of the loss surface rather than just its immediate slope, the algorithm can steer through the Cliff Effects of financial distress with much finer granularity than first-order methods. This mathematical depth ensures the system functions not as an infallible oracle, but as a nuanced decision support system that can provide granular evidence for its classifications.

A critical technical advantage for this specific study is the sparsity-aware split finding algorithm. In the context of SME data, missing data points are rarely distributed at random, or what practitioners call purely stochastic gaps. As discussed regarding Administrative Failures, firms on the brink of insolvency often hold off on, or stop the filing of, financial statements entirely. XGBoost addresses this by learning a default direction for missing values at each node, effectively leaning into the absence of data as a valuable clue.

This capability allows the model to differentiate between a healthy firm with a missing minor ratio and a distressed firm whose lack of transparency, frequently referred to in the literature as Information Asymmetry, is a proxy for organizational decay. This means the algorithm can account for how a lack of reporting might amplify other risk factors in ways a simple formula cannot. Consequently, this mechanism ensures that the model does not require the aggressive imputation of every

variable, preserving the informational void as a distinct predictor of insolvency risk. This approach ultimately transforms the model into a Grey Box, providing a transparent link between algorithmic rigor and clinical credit assessment, ensuring that the final output is both mathematically sound and operationally relevant for credit committees.

4.1.2 Optimization Strategy

The model strikes a balance between predictive accuracy and structural parsimony through the regularization term, which distinguishes XGBoost from standard Gradient Boosting Machines. This trade-off is codified in the complexity function $\Omega(f)$, which is mathematically expressed as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \parallel w \parallel 2 \tag{9}$$

In this formulation, T represents the number of leaves in a given tree and w denotes the leaf weights. The term γ, acting as the Lagrangian multiplier, serves as a complexity cost that essentially functions as a gatekeeper for tree growth. To put this into practice, the algorithm is operationalized through a Shallow Learner configuration. This setup restricts the trees to a maximum depth of 4 and utilizes a conservative learning rate of 0.02. By keeping the trees shallow, the system only signs off on a split if the resulting gain in the loss function exceeds a rigorous threshold, ensuring that the model does not become needlessly intricate.

This pruning mechanism is vital for weeding out the noise associated with the Triple Shock period of 2019 to 2024, as it prevents the model from assigning structural importance to temporary macroeconomic fluctuations. By setting a high bar for new branches, the framework forces the algorithm to pick up on robust economic signals, such as fundamental solvency thresholds, instead of memorizing the noise inherent in distressed balance sheets. This ensures that the model focuses on the broad strokes of corporate recovery rather than the erratic data points characteristic of acute SME distress. Furthermore, we employ a histogram-based tree method to scale up computational efficiency across the high-dimensional dataset, allowing for rapid iteration without sacrificing the precision of the split-finding process.

Complementing this structural constraint, the L2 regularization term, λ, penalizes extreme leaf weights to ensure that no single financial ratio, such as a massive spike in Debt to Equity, dominates the prediction. By reining in these weights, the model avoids the trap of pinning its entire decision on a single outlier, favouring instead a more holistic view of the firm's financial health. Taking small, deliberate steps via the 0.02 learning rate further stabilizes the ensemble, preventing the model from overreacting to the "Silent Killers" in a credit portfolio. This technical setup transforms the final output into a transparent Grey Box, providing a clear link between the regularized parameters and the clinical assessment of credit risk, ultimately anchoring the predictions in generalizable trends that a credit committee can trust.

4.1.3 Benchmark Architecture

To rigorously validate the discriminatory power of the XGBoost architecture within the Distressed Cohort, a comparative benchmark protocol was implemented. The objective was to test whether the non-linear shallow learner topology offers a genuine forensic advantage over established statistical and machine learning alternatives. To ensure a scientifically valid comparison, a strict *Ceteris Paribus*[29] condition was applied: all models were trained and tested on the exact same stratified data partitions used for the primary model, ensuring that performance divergences stem solely from algorithmic architecture rather than data variances .

[29] all as equal

The Statistical Baselines (Logit & Random Forest)

To address class imbalance without altering the data distribution, we implemented a weighted General Linear Model (Logit). The observation weights were dynamically assigned based on the inverse class prevalence ($W_{pos} = N_{neg}/N_{pos}$), mirroring the scale_pos_weight logic used in XGBoost. This ensures that the Logit model is penalized equally for False Negative errors, preventing it from defaulting to a naive all-safe prediction strategy. It serves as a test to see if a simple linear boundary can separate the data once costs are balanced.

In parallel, an ensemble of 500 decision trees was trained using the Random Forest implementation. Unlike XGBoost's sequential correction, this model utilizes bootstrap aggregating (Bagging) to reduce variance. It serves as a control to determine if the boosting mechanism is specifically required to detect rare turnaround signals that might otherwise be averaged out in a Random Forest.

The Theoretical Baseline (Altman Z′-Score)

Recognizing that the standard Z-Score coefficients from 1968 are ill-suited for private non-manufacturing firms, we reconstructed the Altman Z'-Score for Private Firms. Crucially, because the feature engineering phase transformed raw accounting items into specialized ratios, the standard dataset could not be directly plugged into the Altman formula. Therefore, a specific Data Retrieval Protocol was executed to fetch the raw historical variables, specifically Working Capital, Retained Earnings, and EBIT, directly from the source dataset for the test split:

$$Z' = 0.717X_1 + 0.847X_2 + 3.107X_3 + 0.420X_4 + 0.998X_5 \tag{10}$$

This reconstruction allows for a fair test of the Financial Physics hypothesis: we need to see if a linear combination of static ratios can hold up in a cohort where X_4, the Book Value of Equity, is inherently negative. This comparison helps determine if modern machine learning is actually adding value or if traditional accounting formulas still reign supreme in extreme distress scenarios.

Comparative Thresholding (0.40)

To maintain consistency with the Gatekeeper logic defined in Section 4.3, the decision threshold for the machine learning benchmarks was aligned with the High-Sensitivity Operational Target of approximately 0.40. This overrides the default 0.50 cut-off, forcing all models to operate under the same Safety First economic constraint. By pinning the models to the same operating point, we ensure that differences in Sensitivity and Specificity reflect true architectural capability rather than mere threshold calibration. This setup forces a direct "head-to-head" comparison to the operational threshold adopted by this study.

4.2 Training

The training protocol outlines the systematic approach used to bridge the gap between processed financial data and a deployable predictive model. This phase addresses the twin hurdles of severe class imbalance and potential data leakage, ensuring the learning process remains both statistically robust and economically meaningful. By setting up these specific guardrails, the methodology moves away from simple curve fitting and toward a structural understanding of corporate distress.

4.2.1 Managing Class Imbalance

To manage class imbalance, a Cost-Sensitive Learning approach was adopted to ensure the model remains grounded in actual observed performance. The examined sample exhibits a class imbalance

ratio of approximately 0.3955, where $N_0 : N_1$[30]. To preserve the integrity of accounting identities ($Assets = Liabilities + Equity$), the study explicitly ruled out synthetic oversampling techniques like SMOTE (Synthetic Minority Over-sampling Technique)[31]. In essence, creating fake data could lead to models that learn financial relationships which simply cannot exist in the real world, thereby undermining the reliability of the predictive outcomes.

Class imbalance has been addressed via the scale_pos_weight hyperparameter within the XGBoost objective function. This parameter scales gradients and Hessians associated with the positive class during backpropagation:

$$scale_pos_weight = \frac{Count(Y = 0)}{Count(Y = 1)} = \frac{\sum \text{Survival Instances}}{\sum \text{Failure Instances}} \tag{11}$$

By setting this weight to the ratio of negative to positive instances, this modification of the objective function mathematically penalizes False Negatives. The penalty for misclassifying a minority instance becomes proportionally higher than misclassifying a majority instance (Amirshahi and Lahmiri, 2024). This setup ensures that the model places a higher premium on correctly identifying terminal failure than on maintaining global accuracy. This forces the XGBoost algorithm to prioritize the distinctive characteristics of the Failure group, aligning the algorithm with the banking reality where a failure to identify a defaulting firm is significantly more costly than a false alarm. The specific weighting strategy ensures that the computational objective is inseparable from the economic imperative of risk mitigation in lending.

4.2.2 Temporal Firewall & Leakage Audit

To ensure a strictly ex-ante environment, the model utilizes a temporal separation between the feature space (2019–2022) and the outcome horizon (2023–2024). All independent variables are calculated using data available up to the Reference Year ($T_0 = 2022$).

Independent variables were restricted to data points preceding the outcome window. During feature selection, raw variables directly defining the initial distress status, specifically `Shareholders_Equity_kEUR_2022`, were removed from the feature pool. This ensures the model identifies risk through latent performance drivers (e.g., Solvency and EBITDA margins) rather than the absolute magnitude of the equity deficit at T_0.

Sector-specific median imputation was strictly confined to 2022 cross-sectional and historical (2019–2021) data. No information from the 2023–2024 outcome period was permitted within the imputation algorithm, preventing the infiltration of success or failure markers into the missing value estimates.

The 2023–2024 window was utilized exclusively for the definition of the target variable (Y). Prior to model training, a Statistical Leakage Audit was conducted via point-biserial correlation. Any independent variable exhibiting a correlation coefficient of $r > 0.95$ with the target was systematically excluded. This automated filter identifies and removes informational artifacts, such as updated Procedure Status flags, that contain ex-pot realizations of insolvency.

4.3 The Gatekeeper

The transition from a continuous probability score to a binary credit decision calls for a carefully calibrated decision boundary. In standard machine learning applications, this threshold is often set at 0.5 as a matter of course. However, in the context of distressed debt, where the cost of a missed

[30] $Y = 0$ represents survival and $Y = 1$ represents failure. The $Y = 1$ Majority class represents 71.66% of the population

[31] As recommended by Amirshahi and Lahmiri (2024)

default implies capital destruction and a false alarm boils down to a mere opportunity cost, a static threshold is methodologically insufficient. As noted by Peykani et al. (2025), general metrics like accuracy can be misleading, as sensitivity frequently remains dangerously low, paving the way for credit disasters. To address this, the study implements a Gatekeeper Protocol, which is a dynamic thresholding strategy derived from the computational workflow that prioritizes economic utility over raw statistical accuracy.

4.3.1 The Asymmetry of Error: Economic vs. Statistical Thresholding

To pin down the optimal operating point for the classifier, this study weighed up the trade-offs between three distinct cut-off strategies calculated during the validation phase. Each strategy represents a different philosophy of risk, ranging from theoretical neutrality to extreme caution. The first approach considered was the Academic Baseline, or the Youden Index, which is calculated to maximize the statistic $J = Sensitivity + Specificity - 1$. While common in econometric literature for its theoretical balance, this approach assigns equal weight to survivors and failures, which fails to account for the lopsided risks in lending. Experimental results indicate this yields a sensitivity of only approximately 66%. Allowing one-third of terminal failures to pass as safe is economically reckless, as it leaves the lender wide open to capital destruction. This aligns with the Accuracy Paradox described by Kuhn and Johnson (2013), where a model achieves high specificity by correctly identifying the majority class of survivors while failing to pick up on the critical minority event of default. In a distressed environment, a model that misses every third bankruptcy is a liability, not an asset.

Conversely, we computed a cost-sensitive threshold aligned with the Basel III Foundation Internal Ratings-Based (F-IRB) approach, often referred to as the Prudential Bunker. By integrating specific cost parameters, specifically an LGD of 45% and a Net Interest Margin (NIM) of 4.5%, we minimized the Total Expected Loss function defined as:

$$Total\ Loss = (False\ Negatives \times LGD) + (False\ Positives \times NIM) \tag{12}$$

Minimizing this function forces the threshold dangerously low, to approximately 0.06. While this creates near-perfect sensitivity exceeding 99%, the results show that specificity bottoms out at approximately 4.5%. This triggers a Credit Crunch scenario where nearly all applicants are rejected. As Amirshahi and Lahmiri (2024) argue, while keeping default risk down is crucial, there is a clear trade-off between minimizing risk and maximizing revenue by extending credit to a broader range of customers.

4.3.2 The Operational Choice: Safety First and Sanitization

Rejecting both the recklessness of the Youden Index and the paralysis of the Basel-optimized threshold, the final methodology settles on a High-Sensitivity Operational Target. As codified in the computational script, the global decision threshold was strictly calibrated to ensure a Sensitivity, or Recall, of at least 80% on the test set.

The rationale for this Safety First rule is twofold and begins with establishing an acceptable benchmark. This choice aligns with the turnaround mandate defined by Altman and Hotchkiss (2005), who established that for distressed prediction models, a Recall of roughly 80% is considered quite acceptable for identifying deterioration. By capturing four out of five failures, the protocol ensures the model acts as an early warning system for capital preservation without severing credit entirely. In essence, it provides enough warning to act without overreacting to every signal.

Furthermore, the strategy facilitates a process of forensic sanitization. The 10:1 asymmetry in costs between LGD and NIM dictates this path. Given that the cost of a missed default is approximately 10 times higher than the margin gained from a successful loan, trading off specificity to lock

in an 80% capture rate is the only mathematically sound way to weed out the Silent Killers from the portfolio. As demonstrated in the Refinancing Cliff simulation, this threshold acts as a pre-emptive stress test. It is aggressive enough to filter out Rate-Sensitive Zombies, which are firms appearing solvent only due to subsidized rates, while maintaining the granularity required to recognize truly resilient entities.

4.4 Forensic Diagnostic

The evaluation of a turnaround prediction model transcends mere statistical accuracy; it requires a granular assessment of how algorithmic errors translate into economic losses. As noted by Ohlson (1980) and reinforced by modern cost-sensitive learning frameworks (Peykani *et al.*, 2025), the utility of a distress model is inherently tied to the asymmetric costs of misclassification.

4.4.1 Profiles within the Confusion Matrix

In a Distressed Cohort where the Terminal Failures significantly outnumbers the Turnover class, global accuracy becomes a deceptive metric. This phenomenon, described by Kuhn and Johnson (2013) as the Accuracy Paradox, allows a naive model to achieve high scores simply by predicting survival for every firm, thereby failing its primary risk management function. To get around this, this study evaluates performance through a Forensic Confusion Matrix, which redefines standard statistical errors into specific economic profiles characterized by divergent cost structures.

The Silent Killer (Type II Error / False Negative)

Technically defined as a firm predicted as Solvency Restoration that subsequently deteriorates into Terminal Failure, these entities represent the model's most dangerous oversight. Forensic analysis reveals that these entities successfully camouflage their structural rot, often masking deep insolvency with temporary liquidity or low debt-to-sales ratios, deceiving the algorithm into assigning a safe rating. For a lender, this error results in the direct destruction of principal. Consistent with Basel II/III regulatory standards for unsecured corporate exposures, this incurs a definitive cost equal to the Loss Given Default (LGD) of 45%. Missing a Silent Killer effectively locks in a capital loss that wipes out the gains from dozens of successful loans.

The Miraculous Survivor (Type I Error / False Positive)

Conversely, a Miraculous Survivor is technically defined as a firm predicted as Failure that manages to restore positive equity. Forensic analysis suggests these are often rational failures of the algorithm. The model correctly identifies them as structurally insolvent based on the laws of financial physics, such as equity deficits exceeding five times total assets, yet they survive.

This survival is frequently attributable to unmodeled external interventions, such as shareholder recapitalizations or Ex-Machina bailouts. Unlike the catastrophic loss of principal, misclassifying these firms incurs only an opportunity cost. This is defined by the Net Interest Margin (NIM), calibrated at 4.5% in this study. This asymmetry justifies the model's conservative calibration, making Type I errors significantly more tolerable than their counterparts. The setup ensures that the model errs on the side of caution, prioritizing capital preservation over the pursuit of marginal interest income.

To transition the analysis from micro-prudential risk assessment to macro-financial utility, this study implements a Counterfactual Reallocation Simulation. This simulation is theoretically grounded in the credit misallocation framework proposed by Schivardi, Sette, and Tabellini (2020). While the XGBoost classifier identifies who is likely to fail, this post-processing module quantifies the systemic cost of keeping them alive versus the benefit of reallocating their capital to more productive uses.

4.4.2 Macro-Financial Simulation: The Schivardi Swap Protocol

The first phase of the simulation focuses on quantifying Trapped Capital and the associated Burn Rate. The algorithm aggregates the financial exposure of all firms classified as Terminal Failures by the Gatekeeper model. This aggregate represents the Trapped Capital (K_{trap}), which is the liquidity currently locked in non-productive entities. Simultaneously, the study calculates the Aggregate Burn Rate, defined as the summation of the negative EBITDA generated by these terminal firms.

The Liquidation Dividend and Conservative LGD

To simulate the sanitization of the credit channel, the model calculates the Liquidation Dividend (D_{liq}), representing the net recoverable capital available for redeployment after support to terminal firms is cut off. Crucially, this simulation deviates from the standard Basel III LGD (45%) used in the training phase. Instead, it applies a Conservative Liquidation LGD of 60% (LGD_{cons}). This steeper penalty reflects the economic reality of a Fire Sale scenario.

As argued by Shleifer and Vishny (1992), the forced liquidation of distressed assets during an industry-wide downturn yields significantly lower recovery rates because industry peers, the natural buyers, are themselves liquidity constrained. Furthermore, Altman and Hotchkiss (2005) demonstrate that recovery rates are endogenous to the default cycle, dropping significantly when default rates rise. This parameter ensures the simulation is robust against optimistic recovery assumptions:

$$Liquidation\ Dividend\ (D_{liq}) = \sum_{i \in Terminal} (Total\ Debt_i) \times (1 - 0.60) \tag{13}$$

The Reallocation Logic ("The Swap")

Finally, the simulation projects the economic impact of redeploying this dividend. The recoverable capital is hypothetically transferred to the Resilient and Rate-Sensitive clusters, firms identified by the model as structurally sound but credit-constrained. The Net Economic Delta is calculated by estimating the potential EBITDA generation of this reallocated capital, based on the average Return on Invested Capital (ROIC) of the resilient cohort.

This formalizes the Schivardi Swap: trading a one-time stock loss through a capital write-off for a permanent restoration of flow via positive EBITDA generation. By moving capital away from the "Burners" and toward the "Builders," the simulation demonstrates how algorithmic precision can be harnessed to repair the aggregate efficiency of the credit market.

4.4.3 Value-at-Risk Capital Allocation

The transition from a raw probability score (P_d) to an actionable credit decision requires a bridge between statistical theory and the harsh realities of capital allocation. In the domain of distressed debt, focusing solely on the likelihood of failure is economically insufficient. A 90% default risk attached to a micro-enterprise implies a negligible systemic impact, whereas a 10% risk associated with a highly leveraged industrial player represents a catastrophic capital threat. As noted by Altman and Hotchkiss (2005), credit risk is a tripartite function of Probability of Default (PD), Loss Given Default (LGD), and Exposure at Default (EAD); ignoring the magnitude of potential capital destruction effectively blinds the lender to the true scale of the risk. Consequently, this study bridges the gap between statistical prediction and portfolio management by implementing a Value-at-Risk (VaR) Allocation protocol. This phase translates the algorithmic output from a dimensionless scalar into a monetary Expected Loss (EL) metric.

To operationalize this framework, the model utilizes Total Payables (2022) as the proxy for Exposure at Default (EAD). Unlike standard banking models which calculate EAD at the single-facility level, this study adopts a systemic perspective. Total Payables captures the aggregate credit

demand placed by the firm on the entire financial ecosystem, including banks, suppliers, and tax authorities, at the moment of assessment (T_0). This approach aligns with the Basel Committee on Banking Supervision guidelines, which recognize that in distressed scenarios, off-balance sheet commitments and trade payables often crystallize into immediate liabilities.

To pin down the specific capital at risk, the model combines the predicted probability (P_{XGB}) with the exposure (EAD) and a fixed Loss Given Default (LGD) parameter. By the application of the same prudential parameters calibrated previously in Section 4.3 (LGD 45%, NIM 4.5%), the modulation for Expected Loss (EL) is defined as:

$$Expected\ Loss_i = P(Default)_{XGB,i} \times EAD_{Total\ Payables,i} \times 0.45 \tag{14}$$

The computation of EL generates a Risk Map, which plots Exposure on the x-axis against Probability on the y-axis. This visualization facilitates a strategic segmentation of the portfolio that raw accuracy metrics simply cannot provide. It helps in weeding out the noise by identifying high-probability, low-exposure entities as statistically relevant but economically marginal.

More importantly, the map isolates Systemic Carriers. These are entities characterized by moderate probabilities of failure, perhaps between 30% and 40%, but massive balance sheet exposures. In the context of Caballero, Hoshi, and Kashyap (2008), these firms often represent Large Zombies, companies that are too large to fail but too inefficient to survive without constant support. This transformation enables a risk-based allocation of monitoring resources, directing the attention of Workout Units away from high-frequency micro-failures and toward the high-exposure Grey Zone cases. As argued by Elkan (2001) regarding cost-sensitive learning, intervention in these high-value segments yields the highest return on capital preservation. This ensures that limited restructuring resources are deployed where they can prevent the greatest destruction of systemic value.

4.4.4 The "Refinancing Cliff" Protocol

While the XGBoost architecture excels at identifying structural insolvency based on historical patterns, it remains bound by the specific macroeconomic regime of its training data. To overcome this Regime-Shift Wall and evaluate the cohort's resilience against the normalization of monetary policy, this study implemented a Deterministic Stress Test, as a post-hoc diagnostic tool.

The Mechanism of Simulation

Unlike the predictive model, which outputs a probability of failure (P_d), this module functions as a binary logic gate based on financial physics. It recalculates the sustainability of the firm's specific debt structure under a hypothetical stress scenario: a Refinancing Rate (r_{stress}) of 6.5%. To simulate a back-to-normal monetary policy, a deterministic recalculation of the Interest Coverage Ratio (ICR) is executed using a counterfactual base rate of 6.5% applied to the T_0 short-term debt stock.". The simulation re-evaluates the Interest Coverage Ratio (ICR) using the stock of short-term debt held at T_0 (2022):

$$Simulated\ ICR_i = \frac{EBITDA_{i,2022}}{(Short\text{ - }Term\ Financial\ Debt_{i,2022}) \times 6.5\%} \tag{15}$$

This methodology aligns with the Basel III regulatory framework, which mandates that banks conduct rigorous and comprehensive stress-testing to assess vulnerability to worst-case losses under severe market conditions. Furthermore, the use of the ICR as the primary filter draws directly from Adalet McGowan, Andrews, and Millot (2018), who operationalize the definition of Zombie firms as entities persistently unable to cover interest payments from operational income, where $ICR < 1$.

This deterministic calculation serves as a validation layer for the subset of firms classified as "Safe" (Class 0) by the XGBoost model. The procedure is designed to partition these survivors into two distinct economic clusters based on their projected resilience to the withdrawal of monetary support.The first category, the Resilient Core, is defined by firms maintaining a Simulated $ICR > 1$. This threshold is employed to isolate entities capable of generating sufficient operational margins to service liabilities at normalized rates. This methodological choice follows the framework of Caballero, Hoshi, and Kashyap (2008), who define healthy firms by their ability to meet interest obligations at market prime rates without implicit subsidies.Conversely, the Rate-Sensitive Zombies are identified as firms whose Simulated ICR falls below the unit threshold under the 6.5% stress scenario. While these firms may display positive EBITDA in the 2022 training data, the test is intended to detect solvency profiles that are contingent upon subsidized credit. By applying this filter, the methodology operationalizes the "unnatural selection" mechanism described by Peek and Rosengren (2005), where credit misallocation allows inefficient firms to avoid exit.Finally, this classification is designed to capture the "Weaning Off" shock theorized by Grajzl, Cepec, and Mörec (2023). By simulating the removal of pandemic-era "monetary life support," the model seeks to identify firms that, despite healthy contemporary balance sheets, remain structurally vulnerable to the upcoming credit cycle.

4.5 Explainability Suite

To estabilish the XGBoost architecture as a transparent grey box from an opaque black box to compliant with the Right to Explanation, this study implements a dual-layered global interpretability framework using the DALEX ecosystem. This phase is designed to decode the financial physics of the Distressed Cohort, identifying the precise structural laws that govern survival. By peeling back the algorithmic layers, the methodology moves beyond mere prediction to offer forensic insights into why specific firms are flagged for failure.

4.5.1 Global Interpretability

First step of the interpretability framework is the interpretation of the global variables, investigating the general effect of the variables within the prediction model. This phase is designed not merely to rank variables, but to decode the "Financial Physics" of the Distressed Cohort, identifying the precise structural laws that govern survival.

The Hierarchy of Risk (SHAP Importance)

First, to establish the dominance hierarchy among predictors, we utilize SHAP (SHapley Additive exPlanations) Feature Importance. Unlike standard gain-based metrics which can be biased toward high-cardinality features, SHAP aggregates the marginal contribution of each variable across all possible coalitional combinations. As codified in the script, we calculate the Mean Absolute SHAP Value for the test set:

$$Global\ Importance_j = \frac{1}{n}\sum_{i=1}^{n} \mid \phi_j^{(i)} \mid \tag{16}$$

This metric provides a robust ranking of the "drivers of distress," isolating which variables, such as the *Solvency Ratio* versus *EBITDA Margin*, act as the primary "dictators" of the model's decision-making process. Furthermore, SHAP Summary Plots (Beeswarm) are generated to visualize the *directionality* of these impacts, confirming whether high values of a feature systematically push the prediction toward failure.

Decoupling Correlation via Accumulated Local Effects (ALE)

While SHAP identifies *which* features matter, it does not fully explain *how* they behave across their range, particularly given the high multicollinearity typical of financial ratios. Standard Partial Dependence Plots (PDP) fail here because they average predictions over unrealistic data combinations, such as high debt paired with zero interest expenses, which simply do not exist in the real world. To resolve this, we employ Accumulated Local Effects (ALE). ALE calculates the accumulation of local gradients within small neighbourhoods of the feature distribution, isolating the effect of a specific variable while controlling for correlated covariates. This method allows us to see how a firm's risk profile changes as a single ratio evolves, without letting the noise from other variables thwart.

Magnitude vs. Necessity

It is critical to note a methodological distinction in the feature selection process for this diagnostic phase. While the SHAP summary plots rank variables based on the *magnitude* of their contribution (`mean_absolute_SHAP`), the selection of drivers for ALE visualization is governed by Permutation Importance (`dropout_loss`), as codified in the computational workflow. This metric measures the specific degradation in the model's AUC when a feature is randomized, effectively isolating the variables that are statistically indispensable for predictive performance. This dual-metric strategy offers a forensic advantage: it ensures that the ALE analysis focuses on the 'core structure of the model, variables without which the classifier fails, rather than merely those with the loudest average impact, thereby filtering out redundant features that may appear important solely due to multicollinearity.

Visualizing Cliff Effects and Thresholds

ALE plots allow for the forensic identification of Cliff Effects. By mapping ALE curves on the raw scale of the variables (preserved via the "Strategic Retention" policy), the methodology detects exact tipping points where the risk profile changes dramatically. Specifically, the model highlights the Scissor Effect, identifying the zero-threshold where cost growth outpaces revenue growth and begins to eat away at capital. It also identifies the Event Horizon, detecting the specific point in the *Solvency Ratio* (typically an equity deficit exceeding 5x assets) where linear stress transforms into terminal insolvency. Finally, ALE curves are centered, allowing for the superposition of distinct sub-cohorts. This enables us to visualize sectoral risk premiums for *Manufacturing* versus *Services* independent of firm size, ensuring that the interpretation of distress is tailored to the specific industry context.

4.5.2 Interaction Screening

Standard credit scoring models, such as Logistic Regression or the Z-Score, operate on an additive assumption, implying that the impact of a variable like the Solvency Ratio is constant regardless of the state of other variables. However, financial distress is characterized by non-linear synergies where risk factors amplify each other, for instance, high leverage might be manageable for a mature incumbent but fatal for a startup. To decode these complex dependencies, this study implements a three-stage Interaction Screening Protocol.

To move beyond the limitations of linear association, we utilize the SHAP Interaction Value decomposition provided natively by the XGBoost engine. Mathematically, this decomposes the total prediction into main effects and interaction effects for every feature pair (i, j):

$$\Phi_{total} = \Phi_i + \Phi_j + \Phi_{ij} \tag{17}$$

Where Φ_{ij} captures the joint contribution, the additional risk generated solely because features i and j co-occur at specific values, minus the sum of their individual effects. This approach allows the model to pick up on nuanced relationships that a simple correlation matrix would miss.

To quantify these systemic synergies, we aggregate these local values into a Global Interaction Matrix. By summing the absolute interaction effects across the entire test set, we derive a global Interaction Strength score for every possible pair of top predictors:

$$Strength_{ij} = \sum_{n=1}^{N} \mid \Phi_{ij}^{(n)} \mid \tag{18}$$

This matrix functions as a radar, visualizing which variables are mathematically talking to each other. Darker cells in the resulting heatmap indicate strong non-linear coupling, signalling where the model has learned complex conditional logic, such as a rule that triggers only when both debt is high and cash flow is low.

To avoid capturing trivial autocorrelations, such as the inherent link between sequential yearly sales, a Domain-Aware filtering protocol is implemented. This stage is designed to isolate economically significant crossovers, specifically interactions between distinct structural dimensions like capital structure and firm demography.The primary mechanism for analyzing these non-linearities is the SHAP Dependence Plot, configured with interaction coloring. In this framework, the x-axis represents a primary financial driver (e.g., Solvency Ratio), while the y-axis measures its marginal impact on the predicted default risk ($f(x)$). Methodologically, the presence of vertical dispersion at a single point on the x-axis indicates an interaction effect driven by a secondary feature, such as Company Age. For example, by examining how the risk penalty for low solvency varies across different age cohorts, the methodology seeks to determine if the algorithm implicitly weighs "organizational seasoning" as a moderating variable. Therefore, the setup allows for a systematic assessment of whether the model treats firm longevity as a form of reputational collateral, potentially offsetting the risk typically associated with technical insolvency.

4.5.3 Local Interpretability

While global interpretability methods reveal the rules of the game for the entire cohort, they often fail to satisfy the specific Right to Explanation required for individual credit decisions. Under Article 22 of the GDPR and the Use Test requirements of Basel III, financial institutions must justify credit refusals to specific clients. To bridge the gap between algorithmic probability and case-specific rationale, this study implements a Local Fidelity Protocol using iBreakDown Waterfall Plots.

Technically, this phase utilizes the predict_parts() function from the DALEX ecosystem to decompose the model's prediction for a single observation (x_i) into additive contributions. Unlike the black box output, which provides a silent probability score, the Waterfall Plot visualizes the prediction as a cumulative sum starting from the baseline average risk. This intercept represents the average failure rate of the cohort, which the model then adjusts based on individual firm characteristics:

$$f(x) = \phi_0 + \sum_{j=1}^{M} \phi_j(x) \tag{19}$$

Here, ϕ_0 represents the cohort's baseline failure rate and ϕ_j denotes the marginal contribution of feature j. Methodologically, the Waterfall Plot is utilized as the primary visualization tool to map this transition. The resulting plot will display positive and negative attributions, shifting the prediction toward either failure or turnaround, respectively. By adopting this sequential decomposition, the methodology establishes a local explanation layer. This allows for the translation of high-dimensional, non-linear model interactions into a structured format suitable for granular credit risk assessment, ensuring that each prediction is auditable and tied to specific financial indicators.

The Forensic Error Audit

Crucially, this methodology moves beyond random case explanations to perform a systematic forensic autopsy of the model's errors. As codified in the computational script, the test set is segmented to isolate the most egregious misclassifications, ensuring the model's weaknesses are as transparent as its strengths.

Specifically, we perform an autopsy of Silent Killers, or False Negatives. We generate waterfall plots for firms predicted as safe that subsequently failed. This diagnostic reveals the Masking Effect, identifying specific variables, such as deceptively low debt loads or temporary liquidity spikes, that fooled the algorithm into overlooking deep structural insolvency. In these cases, the model's logic is laid bare, showing exactly which "green" signals outweighed the underlying distress.

Conversely, we analyse Miraculous Survivors, the False Positives flagged as terminal that managed to restore solvency. The waterfall visualization allows us to determine if the model was factually wrong or rationally correct but overridden. It highlights cases where a firm survived despite deep equity holes due to unmodeled external factors, such as Ex-Machina shareholder bailouts. This instance-level granularity ensures the model functions not as an opaque oracle, but as a transparent support system capable of audit. This approach satisfies both the regulatory mandate for transparency and the operational need to understand the "why" behind every credit score.

5 - Results and Discussion

This chapter presents the empirical evidence derived from the XGBoost model, focusing on the application of the computational framework established in Chapter 4. predicting solvency restoration, the primary milestone of a corporate turnaround, within a sample of distressed SMEs.

Traditionally, we are used to the prediction of failure within a heterogeneous sample of healthy and failing firms, adopting a broader scope of analysis to discern general credit risk. This study tries to delve deeper, investigating the structural variances inherent in financial crises within a sample composed of already distressed firms. By adding granularity to the topic, the study has a diagnostic and economic target: to distinguish between firms capable of restoring their statutory solvency from those characterized by distress inertia, identified as "Zombies", or surrender to formal liquidation ("Terminal Failure").

The empirical findings are articulated across four analytical dimensions within the topic. The first fundamental result is the discriminatory power of the model, yielding a cross-validated average AUC of 0.7843. While this represents a lower nominal value compared to models trained on heterogeneous samples, it demonstrates significant discriminatory power within a high-noise, distressed-only environment. Further attention is dedicated to the operative threshold, providing an analysis not only to the optimization aspect, but in consideration of its economic implications on the credit market.

To mitigate the inherent opacity of ensemble architecture, we employ an interpretability protocol to extract transparent causal insights, revealing what structural aspects drive the prediction. This protocol responds to transparency legal necessities, revealing not only the drivers, but the interactions between factors, "opening the Black-Box" and surpassing traditional "White-Box" model interpretation.

Finally, given the implications of the results, a quantification of the economic utility via a "Schivardi Swap" simulation, modelled on the credit allocation framework. Data reveal within the sample, more than 10 billion euros is allocated to terminal firms, evaluating the cost of over-lending during pandemic practices and where more efficiency can be found.

The interpretations of evidence presented are not just the output of an extensively validated algorithm, but an empirical evaluation on evaluation of creditworthiness grounded in a methodological framework of dynamic architecture that accounts for restructuring kinetics.

5.1 Predictive Performance on Solvency Restoration

5.1.1 Global Discriminatory Power

The XGBoost architecture is deployed to discriminate between two divergent outcomes within a scenario of financial distress, between the reinstatement of solvency and the bankruptcy of the firm. This optic differentiates from conventional literature of bankruptcy prediction over a broader sample of healthy and distressed firms. The difference is significant when considering the inherent homogeneity of the distressed cohort, which naturally obscures the structural variances between potential survivors and terminal entities.

The observed positive class of the prevision (Class 0) is the "Solvency Restoration", defined by the capacity of a firm to restore positive net equity, having originated from a deficit position during the reference period (T_0, 2022), maintaining operability over the mid- to long-term. Consequently, the opposed outcome is firms that remain characteristic of the state of insolvency (Class 1), identifying firms that declare bankruptcy or that experience a progressive deterioration of structural integrity, eventually categorized as 'Zombie' firms.

The primary metric of evaluating the model's classifying performance is the Area Under the Receiver Operating Characteristic Curve (AUC-ROC), which quantifies the model's aggregate capacity to distinguish between positive and negative outcomes. While the AUC-ROC is a standard diagnostic, its magnitude is intrinsically sensitive to the underlying data distribution. In this case, datasets characterized by skewed class distribution and homogeneous observations, can significantly inflate or suppress the score, as the AUC reflects the "Ranking Ability" of the classifier and may be susceptible to bias. The imbalanced nature of the outcome, as survival rate is extremely low, involves a structural bias, tilting the prediction toward the majority class.

The primary performance evaluation of the model results in a Balanced Accuracy of 0.7699, stable across multiple random seeds.

It is important to note that this metric is marginally different from the statistically balanced accuracy initially found over a 5-fold cross-validation of 0.7843 during training.

While conventional studies on bankruptcy prediction got us accustomed to high AUC scores, sometimes exceeding 0.90, these results are typically derived from heterogeneous samples. Specialized research on solvency restoration and turnaround prediction fluctuates around a score of 0.80 (logit 0.766, gradient boosting 0.832, Altman's Z-Score 0.799), however, more specific values computed on the study's dataset provide a clearer comparison.

Providing the distressed-only nature of the sample, across-validated AUC of 0.7699 is a robust discriminatory outcome, capable of intercepting underlying signals of a complex and non-linear event.

With a training AUC of 0.82 and a testing AUC of 0.79, the resulting Generalization Gap ($TrainAUC - TestAUC$) results less than 0.05 which, in conjunction with the adoption of stratified cross-validation, confirms that the model successfully mitigated overfitting risks.

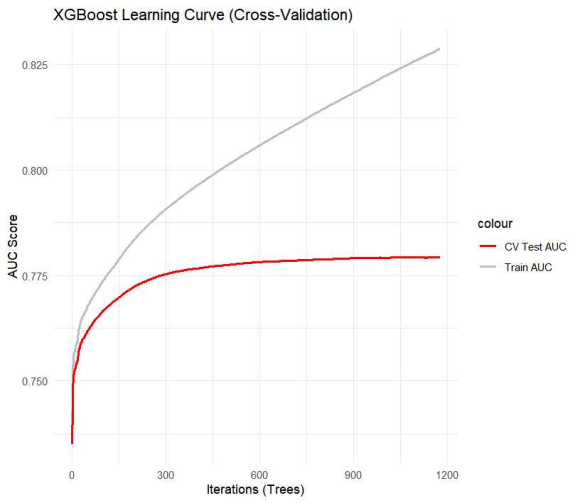

Figure 2: The training curve of XGBoost, displaying the rate of overall correct classifications. Grey curve for training and subjected to overfitting, and red for testing

Beyond the aggregated AUC-ROC, auxiliary metrics are adopted to discern more granular classification characteristics of the model, especially significant over samples with skewed class distribution.

These metrics investigate the classifying errors that a model might produce: False Positives and False Negatives. A False Positive ($TypeIError$) occurs when a 'Terminal Failure' (Class 1) is incorrectly classified as 'Solvency Restoration' (Class 0), representing a defaulting firm that erroneously passes the model's viability threshold. Conversely, a False Negative ($TypeIIError$) represents a

'Solvency Restoration' (Class 0) wrongfully categorized as 'Terminal Failure' (Class 1), effectively flagging a viable firm as a 'Silent Killer' destined for collapse.

Sensitivity (or recall) is the measure for correct identification of True Negatives (Class 1) events related to Type II Errors. Within the framework of creditworthiness assessment, sensitivity surpasses the aggregate AUC in importance due to its direct role in capital protection. Prioritizing sensitivity serves to minimize the Loss Given Default (LGD) by providing a robust early warning for potential portfolio impairment. The model was tuned to achieve a sensitivity of ~80% effectively identifying four out of five failures, with an actual sensitivity score of 0.7924. To achieve this, the threshold was adjusted to 0.40 representing a 'pragmatic investor' configuration that balances safety with operational viability. The threshold of 80% is set in view of Altman's work, which considered an 80% prediction accuracy on defaults, to be *"quite acceptable",* and allowing us significant gains on sensitivity, at the cost of a slightly lower specificity (see more at 5.4).

Specificity, conversely, measures the efficacy of the model in identifying firms capable of 'Solvency Restoration' (Class 0), and relates inversely to Type I Errors. A low sensibility rate implies a high rejection rate of promising lenders, and an excessively conservative classification threshold induces significant opportunity costs, leaving potential profits unresolved. The model achieved a specificity of 0.57 successfully capturing most turnaround opportunities while maintaining a prudential focus on capital preservation.

The F1-score is defined as the harmonic average of precision and recall. The metric balanced assessment of a classifier's performance, as it penalizes models that achieve high predictive performance on the majority class at the expense of the minority class. In this context, a naive classifier that assigns all observations to the majority class (Class 1) would yield a nominal accuracy of 71.66 corresponding to the sample prevalence, while resulting in a 100% sensitivity but 0% specificity, producing a balanced accuracy of 50%. Such model, despite a high nominal accuracy, lacks any real discriminatory power and its efficiency is relatable to a coin toss.

The resulting model achieved an F1-Score of 0.8044, and a Balance Accuracy of 0.6868 ensuring an actual balance between conservative capital protection and commercial lending viability.

5.1.2 Benchmark Comparison

To validate the classifier's robustness, we benchmarked its predictive discriminatory power against consolidated statistical and machine learning architectures. Benchmark metrics have been computed on the research's dataset, instead of relying on literature-based metrics overlooking the sample's characteristics. In other words, we tested the model against standard industry tools to see if they offer a sharper edge in predicting survival. Given the specific nature of the Distressed Cohort, this comparison serves as a diagnostic tool. The goal is not merely to identify the highest AUC, but to determine which architecture can navigate the 'Negative Equity Trap' without succumbing to the heuristic bias of blanket default prediction. The goal is to rule out models that simply 'give up' on firms with negative equity and instead identify those that can spot the subtle signs of a turnaround.

Performance Comparison Table

Model	AUC	Sensi-tivity	Speci-ficity	F-Score	Bal. Acc.	PPV	NPV
XGBoost	0.79	0.79	0.60	0.81	0.69	0.82	0.55
Random Forest	0.77	0.90	0.38	0.84	0.64	0.78	0.62
Logit	0.72	0.80	0.44	0.79	0.62	0.77	0.48
Z-Score	0.65	0.79	0.35	0.77	0.57	0.74	0.41

Table 3: Comparative Performance Metrics across Models

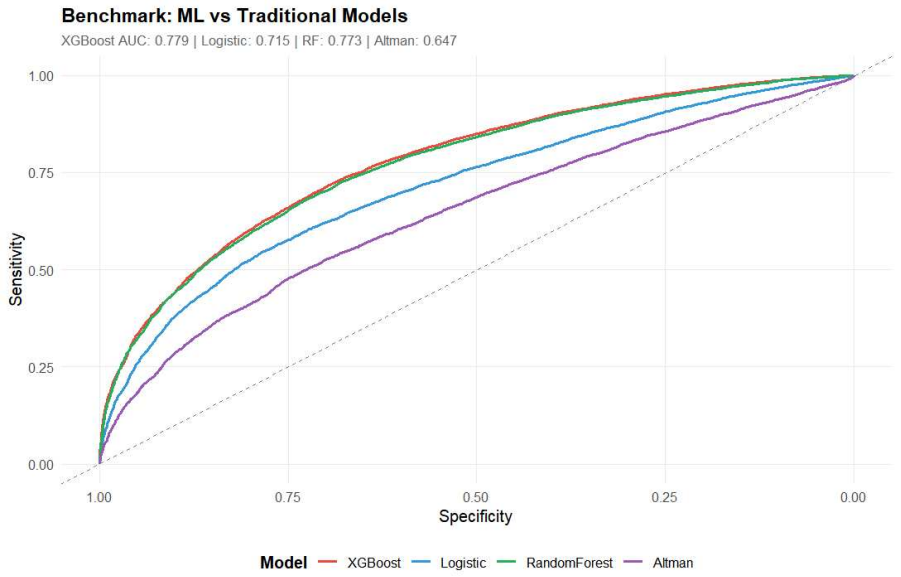

Figure 3: Comparison of the selected models learning curves. XGB (in red) and RF (in green) curves are almost overlapped, with the sharpest rates.

Linear Discriminant - Altman Z′-Score

The Altman Z-Score serves as our theoretical baseline. However, its poor performance (AUC 0.65) on our primary dataset is not merely an artifact of obsolescence; it is a mathematical inevitability for this specific cohort. Essentially, the model is being asked to solve a problem it wasn't built for.

The Z-Score formula relies heavily on the X_4 component (Book Value of Equity / Total Liabilities). In a standard population, this ratio distinguishes levered firms from solvent ones. However, in a Distressed Cohort, every firm has a negative numerator. Consequently, the linear combination flattens "deep insolvency" and "marginal insolvency" onto a single, continuous scale. To put it simply, the Z-score can't tell the difference between a firm that is slightly underwater and one that is at the bottom of the ocean.

The model also suffers from theoretical constraints. While Altman(2005) acknowledges the model's reliance on "healthy" vs. "bankrupt" samples, our empirical testing proves that these coefficients fail when applied to a homogenous distressed population. The model essentially loses its grip on the data, failing to pick up on the non-linear "turnaround signals" (like cash flow velocity) that separate recovery from collapse. The low Balanced Accuracy (0.57) confirms that the Z-Score performs only marginally better than a random guess in this specific domain. This observation is

not a critique on the model, but the mere recognition of its limits given its use over an unintended case dimension of distressed enterprises.

The "Paranoid Predictor" Bias - Logistic Regression & Random Forest

A critical finding of this study is the behaviour of the Logistic Regression and Random Forest benchmarks. While their F1 Scores appear competitive (0.71 and 0.77 respectively), a forensic examination of their Specificity reveals a structural collapse on our dataset. Essentially, the high F-Score hides the fact that the models are failing to distinguish between survivors and failures.

Despite the weight balancing, the Logit model yields a Specificity of only 0.44. This reveals a persistent "Pessimism Bias"; even when penalized for misclassifying survivors, the model still leans toward a "Failure" label for most of the cohort. Our results corroborate with the "Cliff Effects" identified by Crosato et al (2023). In practice, this means the Logistic model ignores the temporal "Hockey Stick" of distress we observed in our data, treating every insolvent firm as a lost cause. The relationship between financial ratios and turnaround success is not monotonic; there are thresholds of debt and operational decay beyond which risk explodes exponentially. Because the Logistic regression is constrained by a monotonic sigmoid curve, it cannot account for the sudden "breaking points" in the data.

More striking is the Random Forest's (RF) collapse, with a Specificity of 0.38. In this specific context, the RF's non-linearity looks powerful on paper but fails to protect the minority class from being averaged out. However, the RF model achieved the highest Sensitivity in the study (0.90), indicating it is extremely effective at identifying failing firms, albeit at the cost of many false alarms.

The performance gap between XGBoost and Random Forest stems from a structural divergence. As highlighted by Barboza (Barboza, Kimura and Altman, 2017), both ensemble methods outperform classic linear models due to their capacity to model non-linearities. However, Amirshahi & Lahmiri suggest that Gradient Boosting holds a comparative advantage in imbalanced datasets due to its sequential refinement[32].

The Boosting Advantage

The XGBoost model pulls ahead as the only architecture capable of genuine discrimination, yielding the highest AUC (0.79) and a superior Balanced Accuracy (0.69) Essentially, while other models were guessing based on the majority class, XGBoost learned the underlying patterns of survival.

Unlike Random Forest, XGBoost employs Gradient Boosting, where each new tree is trained specifically to correct the errors of its predecessors. This enables the model to "double down" on the edge cases, the "Miraculous Survivors" that maintain operational viability despite technical insolvency.

The true victory of XGBoost is found in its Specificity (0.5762). While the benchmarks acted as "paranoid gatekeepers," predicting failure for almost the entire cohort, XGBoost successfully filtered out 57% of the viable firms. It managed to spot the survivors without losing its grip on the failing ones (0.7947 Sensitivity). This validates the hypothesis that turnaround prediction requires a model capable of learning non-linear, conditional logic, a complexity that linear and bagging models simply couldn't get a handle on. To put it simply, XGBoost recognizes that a bad balance sheet isn't always a death sentence if the operations are still humming.

Banks have also found significant marginal advantages in the adoption of machine learning models for default prediction over proprietary datasets[33].

[32] Consult *Bankruptcy prediction using optimal ensemble models under balanced and imbalanced data,* (Amirshahi and Lahmiri, 2024).

[33] Consult *Corporate default forecasting with machine learning,* (Moscatelli *et al.*, 2020).

5.1.3 Model Stability and Reproducibility

A fundamental requirement for credit risk models in a regulatory context (e.g., Basel III) is stability. To verify the reliability of the XGBoost architecture, the model was stress-tested across three different random seeds for data partitioning, with the main study conducted on seed 123, and two more seeds for stability checking, 456 and 789. Essentially, we wanted to ensure the results weren't just a lucky slice of data, but a consistent reflection of the underlying risks.

The results remained strikingly stable, with fluctuations in balanced accuracy not exceeding ± 0.015, respectively at 0.6976, 0.6859 and 0.6921. This indicates that the signal captured by the model is structural to the dataset rather than a byproduct of stochastic noise. It boils down to this: the model is finding real patterns, not just chasing shadows in a specific split.

Across all iterations, the hierarchy of risk drivers remained steady. Whether we reshuffled the data or changed the initialization, the same variables kept rising to the top of the feature importance. `Solvency_Ratio` consistently emerged as the primary predictor, followed by `ROE`, `Company_Age`, `Gearing_Ratio` and `distress_history_score`. This stability in feature attribution (XAI) is a prerequisite for 'White-Box' interpretability. It ensures the economic narrative, anchored by Solvency and the 'Hockey Stick' of distress history, is reproducible and not dependent on algorithmic jitter.

5.2 The Mechanics of Distress and Restoration: Stock, Flow, and Time

To overcome the intrinsic opacity of the gradient boosting algorithm used in this study, which often trades off explanatory power for predictive performance, we must address the "Black-Box" nature of the model. These high-performance engines typically possess logic that remains largely obscured unless specific protocols are adopted to extract their decision logic. To this end, the study employs an Explainable AI (XAI) framework grounded in cooperative game theory, specifically Shapley Values, applied from both a global and local perspective. Essentially, this approach allows us to "deconstruct" the model's complexity, turning an opaque algorithm into a transparent decision-support tool where the contribution of every financial variable is quantified.

The predictive hierarchy of the model is characterized by a stark dominance of structural variables over operational ones. This hierarchy validates what we term the "Physics over Momentum" hypothesis: the reality that for a firm starting from a position of negative equity, the structural "weight" of the balance sheet dictates the probability of survival far more than the temporary "speed" of its operations.

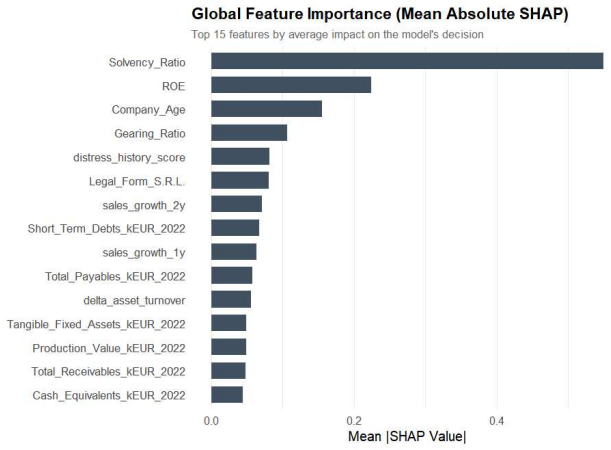

Figure 5: This plot ranks the predictors by their mean absolute SHAP value, effectively establishing the model's "hierarchy of risk."

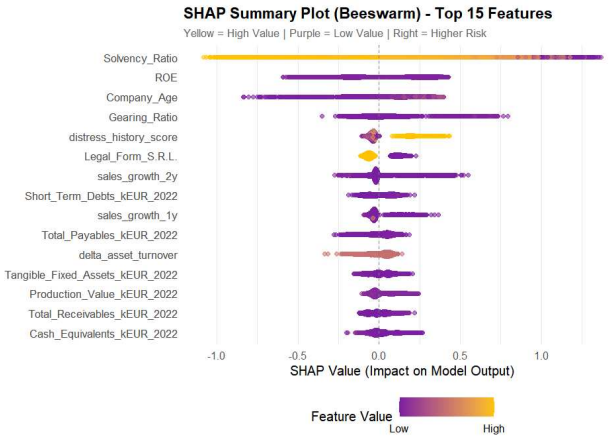

Figure 6: This visualization maps the directionality and distribution of feature impacts. Each dot represents a firm; its position on the x-axis indicates the impact on default probability (Right = Higher Risk, Left = Higher Likelihood of Turnaround), while the color represents the feature value (Yellow = High, Purple = Low).

5.2.1 "The One above All": Solvency Ratio

At the apex of this hierarchy sits the `Solvency_Ratio`[34], which acts as the undisputed "dictator" of the predictive engine. In the specific context of this study, where every subject enters the dataset with negative equity, the model identifies the magnitude of the deficit as the primary determinant of life or death. Practically, the feature functions as an "anchor" to the risk magnitude itself.

Empirically, the SHAP contribution of Solvency dwarfs its operational counterparts by nearly four-fold. This suggests that the "physics" of the balance sheet creates a financial Event Horizon. While traditional turnaround literature often emphasizes strategic reorientation and "momentum" (sales growth), the XGBoost model identifies a point of no return. If the Solvency Ratio drops below a critical threshold (typically ≈ -0.5), the structural decay becomes so profound that even aggressive operational "flows", such as double-digit sales growth, are insufficient to pull the firm back from insolvency. In essence, no matter how fast the engine is running, it cannot escape the gravity of a collapsed hull.

Further inspection of the interaction between structural insolvency and operational sustainability, showed that `Debt_Load`[35] and `Solvency_Ratio` interact through a non-linear "severity threshold".[36].

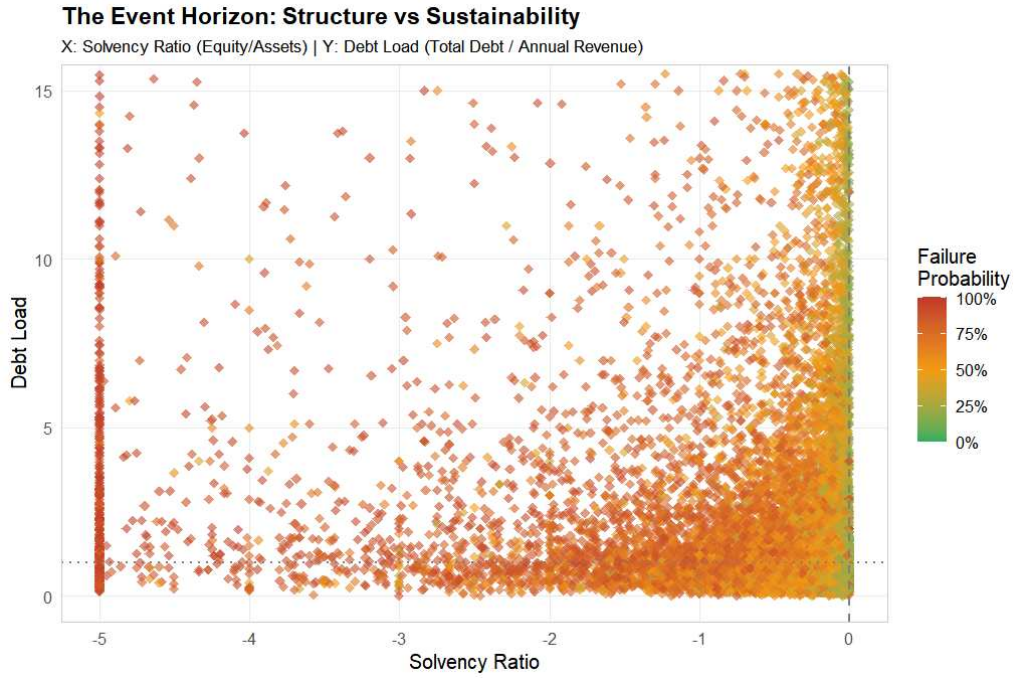

Figure 7: This visualization maps the critical non-linear interaction between Solvency Ratio (X-axis: Structural Stock) and Debt Load (Y-axis: Operational Flow), empirically defining the "physics" of the decision boundary

The model identifies a clear "Event Horizon" regarding structural damage. At extreme levels of insolvency, specifically when the equity deficit exceeds five times the assets ($x = -5$)[37], a firm's Debt Load becomes entirely irrelevant. In this "Red Zone," the model flags every company for failure regardless of its revenue-to-debt ratio. This evidence validates the "Stock Hypothesis": structural solvency acts as the ultimate gatekeeper, effectively overriding operational performance metrics once the balance sheet is fundamentally broken.

[34] The ratio between the Shareholders' Equity to Total Assets.

[35] Debt Load is the rate of total debt over sales revenues

[36] Debt Load is an *ex-post* variable obtained from Total Debt over Annual Revenue, for explanatory reasons, and the model learned its value by its close proxy: the Gearing Ratio. The measurement is very unstable given the denominator close to zero

[37] Solvency Ratio is capped at −5, as further values do not improve the model, only introducing unnecessary noise

For firms positioned outside this terminal zone, the model reveals a shifting margin for error. A company with a solvency ratio near zero can often withstand significantly higher debt loads, up to five times its annual revenue, while maintaining a lower probability of failure. However, as that ratio drops toward -3, the tolerance for debt collapses entirely; even a modest debt load of twice the annual revenue is sufficient to trigger a high-risk classification. Essentially, the deeper the equity hole, the less room a company must maneuver with financial leverage.

Finally, the data warns of a "Low Debt" trap located at the bottom of the risk landscape. This is the segment where the model is most likely to be "fooled," leading to False Negatives. Because these firms maintain manageable debt levels (below the dotted line on the interaction map), they appear operationally safe on the surface. However, this facade often hides the fact that they are already structurally dead from a "Stock" perspective and likely to fail due to sudden liquidity shocks or operational "burn".

5.2.2 Momentum of Recovery

While Solvency determines the baseline risk, acting as the "anchor" of the predictive engine , specific operational and longitudinal variables function as the critical "tiebreakers" for firms hovering in the borderline zone. This is where the non-linear power of machine learning is most evident: for firms near the decision boundary that have not yet crossed the terminal "Event Horizon," ROE, Company Age, and Distress History emerge as the decisive discriminators.

Operational Lifeline

The model pins down a high-sensitivity zone around the zero-mark for profitability, where Return on Equity (ROE) functions as the primary operational lifeline for firms in distress. In the SHAP hierarchy, ROE comes in second only to Solvency, confirming that for a firm with a compromised balance sheet, the capacity to churn out immediate returns is the decisive factor in predicting a successful turnaround. This finding ensures that recovery momentum carries more weight in the final score than the static volume of balance sheet activity, as the algorithm looks for signs of life rather than just the size of the deficit.

The interaction between these two variables reveals a "Vertical Divergence" in risk. At any given level of ROE, the model maintains a lower risk prediction for firms whose Solvency Ratio is closer to zero compared to those in deep insolvency. This confirms that while profitability is a powerful mitigant, the "Stock" of solvency acts as a persistent anchor on the firm's ultimate probability of recovery

This prioritization matches up with the Reservoir Theory established by Beaver (Beaver, 1966), who conceptualized the firm as a reservoir of liquid assets, essentially a tank that must be refilled to prevent the engine from seizing. While the reservoir itself provides a buffer against shocks, the net liquid-asset flow is what determines whether the firm remains solvent over the long haul. In this context, even if the reservoir is nearly empty due to low solvency, a high flow rate from profitability suggests that the firm can eventually stabilize itself.

This distinction also rings true with the work of Caballero, Hoshi, and Kashyap (2008), who define zombies as entities whose profitability is simply too low to cover their interest payments. By putting a premium on ROE, the model effectively weeds out operationally insolvent firms that are being kept on artificial life support through credit forbearance or evergreening practices. These results further back up the findings of Crosato, Liberati, and Repetto (2023), who noted that profitability indicators significantly drive down default probability in Italian SMEs. By capturing these non-linear turnaround signals, the XGBoost model identifies resilient firms that possess the operational heat required to grow their way out of a solvency crisis

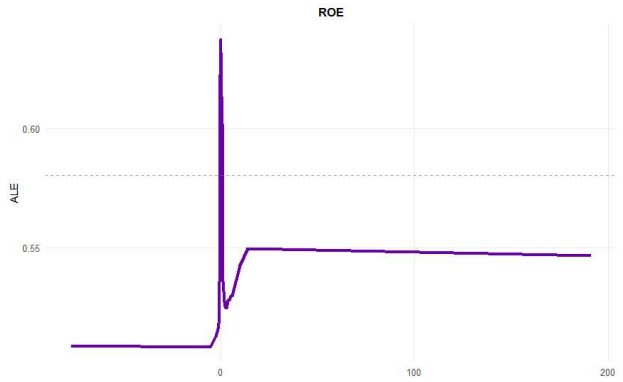

Figure 9: This visualization isolates the marginal effect of Return on Equity (ROE) on the probability of default, holding all other variables constant. As the second most influential predictor in the model hierarchy, ROE functions as the primary "operational lifeline" for firms in distress.

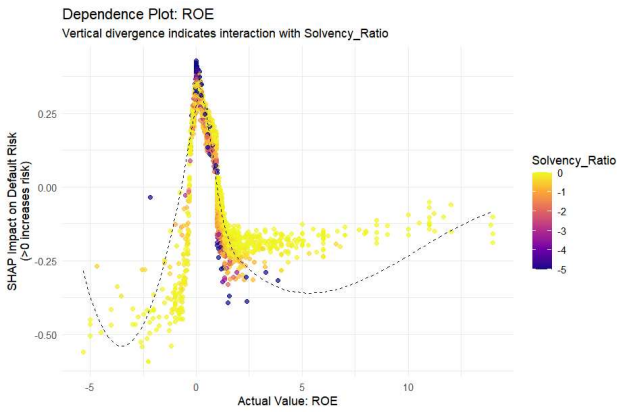

Figure 10: This Dependence Plot maps ROE on the x-axis against its SHAP impact on the y-axis, with the color scale representing the Solvency Ratio (Yellow = High/Better, Purple = Low/Deep Insolvency). It empirically validates Beaver's Reservoir Theory: the interaction between the Flow of new wealth (ROE) and the remaining Stock of capital (Solvency).

The Complexity Trap

The model pins down a non-linear risk trajectory for the variable $Num_Employees$, revealing a perilous middle ground where risk ramps up as firms scale toward the 15 to 50 employee mark. Growing SMEs in this cohort often run into sudden structural rigidity, accumulating fixed labour costs and organizational layers before they can get past the threshold to achieve the protection usually afforded to those that are too big to fail. While micro-entities appear more vulnerable on paper, their inherent agility allows them to survive shocks that would otherwise sink a larger, more rigid competitor.

The resilience of micro-entities identified by the model is well-supported by the work of Senyard et al. (2014), who describe bricolage as a key survival trait. This refers to the ability of small teams to make do with existing resources through creativity, essentially fixing the engine while the car is still moving. As firms grow, this flexibility often gets lost. The resulting complexity trap mirrors the findings of Crosato et al. (2023) regarding Asset Heaviness. In this context, high organizational complexity sets off a trap during periods of distress, as these firms fall short on the flexibility required to pivot quickly compared to lighter, smaller entities.

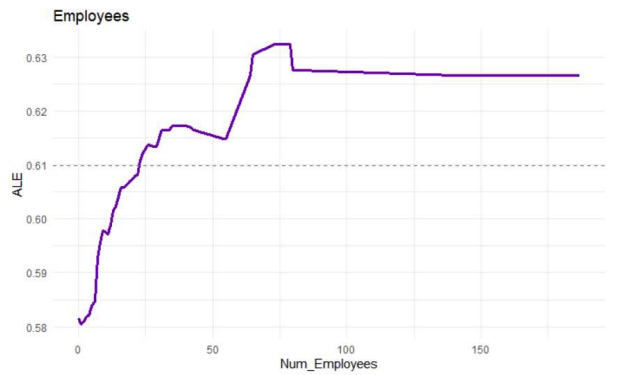

Figure 12: This visualization challenges the traditional "bigger is safer" assumption by isolating the marginal effect of workforce size on default probability. The curve empirically identifies a "Complexity Trap" in the transition zone between micro-enterprises and structured SMEs (typically 15–50 employees).

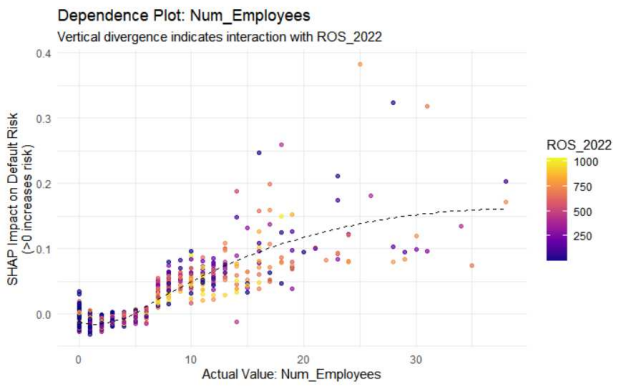

Figure 13: This SHAP Dependence Plot visualizes the interaction between Organizational Scale (x-axis: Num_Employees) and Efficiency (Color Scale: ROS). It confirms the topology of crisis described in Section 6.1.1: Complexity is an accelerator, not a driver).

Furthermore, the vulnerability of the middle zone bears out the theories of El Kalak and Hudson (2016), who suggest that scaling SMEs frequently lean on short-term bank debt that can be withdrawn abruptly during a crisis. This creates a high-risk transition zone where the firm has outgrown its informal, agile roots but has not yet established the diversified capital base or political weight of a large corporation. This structural gap makes the 15 to 50 employee segment particularly prone to terminal failure, as they bear the burden of a high burn rate without the safety net of systemic importance.

5.2.3 Temporal Dimension

The Lindy Effect

The model confirms `Company_Age` as a top three predictor. This variable acts as a proxy for industrial durability, essentially "discounting" the risk for firms that have stayed in the game for decades. While simple on the surface, the SHAP dependence plots reveal a nuanced, inverse relationship: as a firm matures, its predicted probability of default decreases, *ceteris paribus*. However, this variable functions as a double-edged sword within the predictive engine, acting simultaneously as a proxy for deep-seated resilience and a potential mask for operational inefficiency.

To dismantle the model's internal logic, we must address the "Paradox of Longevity." Company Age does count for a firm's longevity, however, is not merely a chronological count; it functions as a nuanced proxy for a firm's latent structural integrity. SHAP dependence plots reveal a complex, non-linear relationship where age serves as both a badge of resilience and a potential mask for deep-seated inefficiency. This creates what we might call a "survival premium," where the algorithm treats the sheer fact of a firm's historical persistence as an empirical justification for future optimism.

Liability of Newness vs. The Lindy Effect

The model's behaviour here provides validation of the "Liability of Newness", the empirical observation that younger entities are inherently fragile, lacking the established routines and market credibility necessary to weather a crisis. The risk trajectory is steeply non-linear: the most perilous "infancy" for an SME occurs within the first five to ten years. During this window, any sign of financial distress is treated by the model with extreme prejudice. Yet, once a firm crosses the twenty-year threshold, it begins to benefit from the Lindy Effect, where its past longevity is interpreted as a predictor of future persistence. The algorithm assumes the existence of intangible resilience factors, such as brand equity, specialized know-how, or deep-seated creditor relationships, that allow these

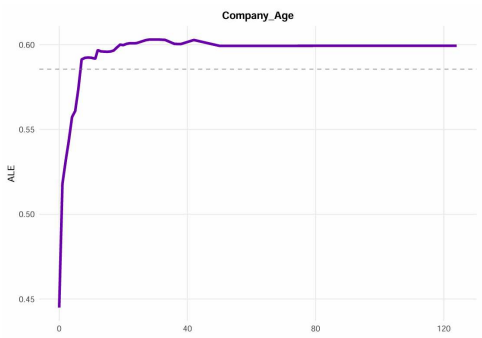

Figure 15: The visualization isolates the marginal effect of Company_Age on the probability of default, holding other financial variables constant. It empirically maps the "Liability of Newness," revealing a steeply non-linear risk trajectory that challenges the assumption that risk is uniform across corporate lifecycles.

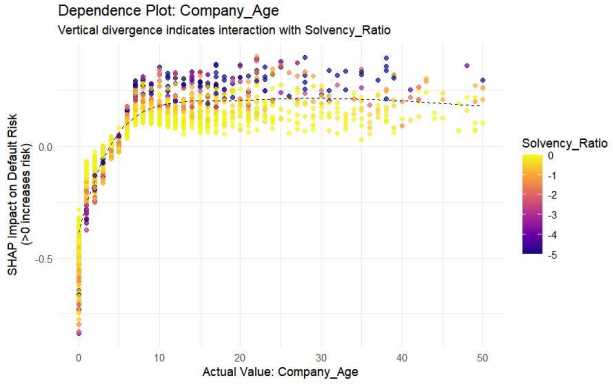

Figure 16: Its SHAP Dependence Plot visualizes the interaction between Firm Maturity (x-axis: Company_Age) and Structural Health (Color Scale: Solvency_Ratio). It reveals a massive "Vertical Divergence" in how the model penalizes insolvency based on the firm's age.

firms to survive levels of distress that would be fatal for a younger startup. The algorithm essentially recognizes that a firm that has survived decades of shifting economic cycles has likely developed an industrial durability that a new entrant simply cannot claim.

The true predictive depth of this variable emerges through its interaction with the Solvency Ratio. The Global Interaction Matrix reveals that the model essentially maintains two different sets of rules: one for the "unproven" and another for the "established." For a young firm, a negative solvency position triggers an immediate and aggressive escalation in default probability, as the model sees no historical evidence to justify a recovery. In stark contrast, mature firms are granted a form of "institutional leniency." The model tolerates significantly deeper equity deficits in a forty-year-old company than it would in a five-year-old startup before flagging it as terminal. This suggests that the algorithm has identified an unrecorded "Survival Buffer", a reservoir of intangible value, such as brand equity or legacy relationships, that never appears on a modern balance sheet. In effect, the model treats age as a form of "hidden equity," assuming that if an old firm is still standing despite a collapsed solvency ratio, it must be supported by an accounting ghost that the balance sheet is failing to capture.

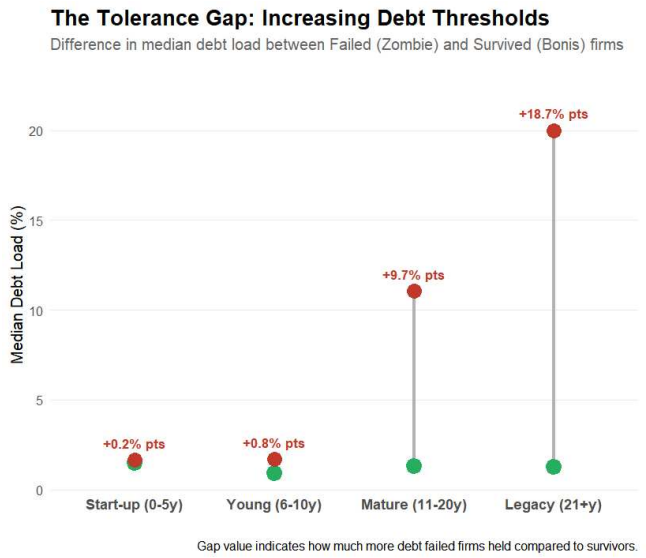

Figure 17: The visualization tests the "Institutional Entrenchment" hypothesis by mapping the Debt Load (Total Payables / Total Assets) of firms at the moment of failure, stratified by age

This visual intuition is statistically corroborated by a Mann-Whitney U Test ($p < 0.05$), performed on the distinct failure clusters. The test confirms that 'Legacy Zombies' (21+ years) sustain significantly higher median debt loads at the moment of failure compared to 'Startup Zombies'. This disparity provides empirical evidence of differential tolerance: lenders permit established firms to erode their asset base far deeper than younger entrants before triggering liquidation, effectively validating the hypothesis of Institutional Entrenchment.

Institutional Entrenchment over Viability

However, this reliance on age demands a more cynical interpretation through the lens of Zombie Lending. While the Lindy Effect assumes that survival implies competence, the model's strong negative correlation between age and default may reflect institutional entrenchment rather than industrial quality. This introduces the "Evergreening Hypothesis": the reality that distressed firms are often kept on life support by creditors eager to avoid the immediate pain of realizing losses.

Older firms, possessing decades of long-standing relationships with lenders, become the primary beneficiaries of this forbearance.

Consequently, there is a risk of Algorithmic Complicity. The XGBoost model may have learned to predict bank behaviour rather than pure economic viability. It correctly identifies an older, insolvent firm as "Safe" not because the entity is undergoing a structural turnaround, but because the banking system is statistically less likely to liquidate a legacy player compared to a younger, politically or socially insignificant startup. In this light, the model identifies "Survivability" as a function of continued credit access rather than internal efficiency. It highlights how institutional inertia can mask terminal insolvency, effectively predicting that a firm will stay alive simply because the system is too entrenched to let it die.

The Chronicity Law

Perhaps the most sobering insight emerges from the `distress_history_score` ALE plot, which reveals a non-linear, compounding escalation in default probability as years of negative equity accumulate. This finding effectively dismantles the "Mean Reversion" myth, the hope that distress is a temporary, cyclical dip. Instead, the data suggests a "Point of No Return" dynamic.

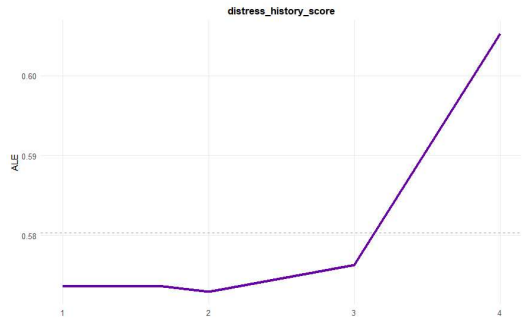

Figure 18: This Accumulated Local Effects (ALE) plot empirically validates the "Chronicity Law," illustrating the non-linear relationship between the duration of insolvency (years with Negative Equity) and the probability of terminal failure.

While the increase in risk is relatively moderate between the first and second year of distress, the slope becomes aggressively steep after year two. This threshold suggests that while a firm might survive a brief shock to its equity, a third, or fourth year of negative equity represents a terminal "Zombie" state where the probability of recovery doesn't just stall, it effectively collapses. In this framework, time acts as a catalyst for structural decay; with every passing year of inertia, the internal resources and external credibility required for a turnaround evaporate at an accelerating rate, making late-stage intervention statistically futile.

5.3 Forensic Behavioral Dynamics

To further unmask the non-linear logic of the XGBoost algorithm, the study transitions from feature importance to the visualization of marginal effects through Accumulated Local Effects (ALE) plots. Crucially, the selection of predictors for this forensic phase is steered by Permutation Importance, or Dropout Loss, rather than relying on simple SHAP magnitude. This distinction ensures that we zero in on the variables most statistically indispensable to the model's accuracy. By measuring how much the model's performance degrades when a specific feature is shuffled, we identify the structural backbone whose removal would trigger the largest drop in AUC, essentially the features the model cannot afford to lose. In financial datasets, where ratios often overlap, permutation importance acts as a filter that separates true predictive drivers from those that are merely hitchhiking on the signals of others. As established by Crosato (2023), these plots allow us to transcend the static coefficients

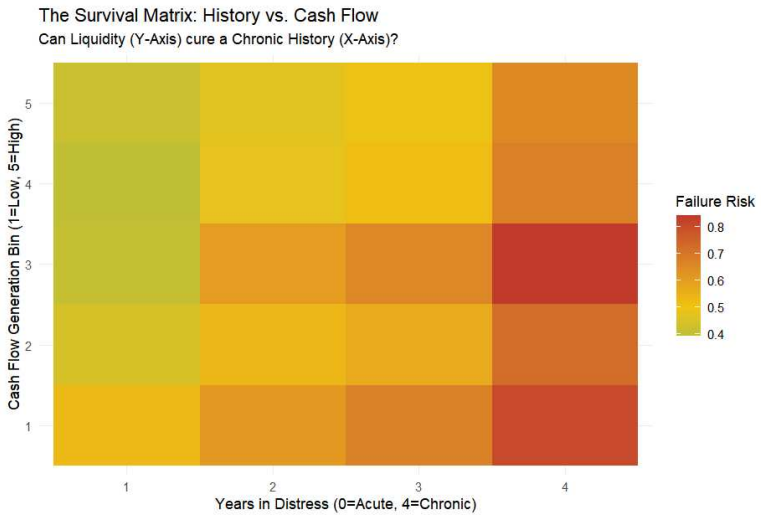

Figure 19: This heatmap visualizes the interaction between the duration of insolvency (X-axis: Distress History Score) and the capacity for cash generation (Y-axis: Cash Flow Generation Bin). It empirically defines the limits of operational recovery against the weight of historical inertia.

of traditional linear models, pinning down the specific financial mechanisms and tipping points that govern the distressed cohort. The resulting ALE Grid Plot provides a granular roadmap of these critical thresholds. Unlike linear models that assume a constant risk-to-ratio relationship, where one unit of change always equals one unit of risk, the grid reveals complex Step-Function dynamics.

5.3.1 Marginal Effects

First, the model highlights Threshold Sensitivity. Variables like the solvency ratio and gearing ratio do not show a smooth gradient; instead, they exhibit sharp cliffs where a minor change in the ratio triggers a disproportionate spike in default probability, essentially reaching a point of no return. Second, the grid identifies a Plateau Effect. It maps out specific ranges where improvements in financial health, such as increasing liquidity, yield diminishing returns. This suggests that for

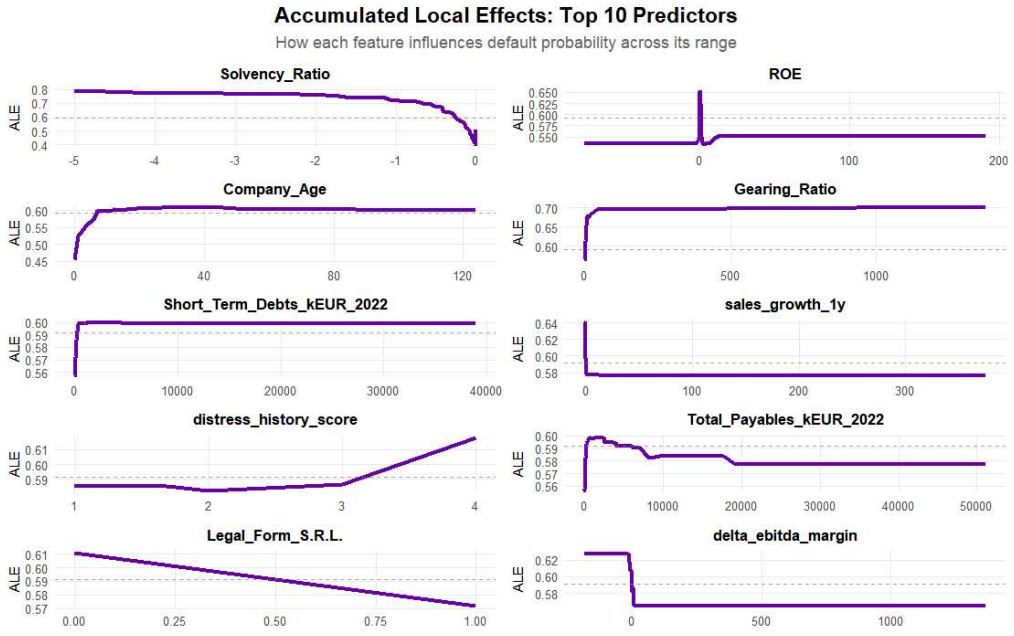

Figure 20: This grid visualizes the Accumulated Local Effects (ALE) for the top 10 predictors selected via Permutation Importance. Unlike linear coefficients, these curves map the non-linear "Financial Physics" of the distressed cohort, exposing critical "Cliff Effects" where risk does not grow proportionally but explodes vertically upon breaching specific thresholds

certain SMEs, more of a positive attribute does not necessarily lower risk once a baseline of stability is reached.

Finally, the visualization reveals Variable Asymmetry. By visualizing these effects side-by-side, the model demonstrates that distress is not a monolithic state. A decay in one dimension, like equity history, can be exponentially more damaging than a decay in another, such as short-term profitability. This allows the model to prioritize structural insolvency over temporary operational hiccups, ensuring the risk assessment remains focused on long-term viability rather than seasonal volatility.

5.3.2 The Scissor Effect

The ALE plot for `scissor_debt_sales`, the differential between debt growth and sales growth, uncovers a distinct "Cliff Effect" that serves as a vital diagnostic for corporate health. The empirical curve remains remarkably stable if the metric is negative, indicating that the firm's commercial activity is successfully outpacing its debt accumulation. However, the logic shifts violently at zero threshold.

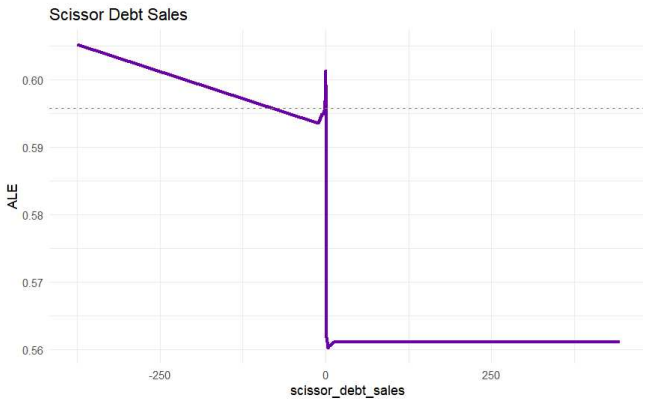

Figure 21: This Accumulated Local Effects (ALE) plot visualizes the marginal impact of the Scissor Effect (defined as Δ%Debt−Δ%Sales) on the probability of default. It reveals a structural "Cliff Effect" at the zero-threshold, marking the boundary between sustainable leverage and terminal erosion.

The curve exhibits a sharp vertical spike the moment debt growth overtakes revenue generation. This behaviour provides a clear theoretical validation of the Capital Consumption Hypothesis: when a distressed firm's debt expands faster than its sales, it signals that new financing is being diverted to "plug holes" and cover operational losses rather than funding productive growth. The model effectively identifies a "Viability Threshold," where the inversion of this trend becomes a terminal signal for survival. In simpler terms, once a firm begins to borrow more just to sell less, the algorithm recognizes that the entity has entered a death spiral.

5.3.3 Structural Diversity

Asset Heaviness

When the ALE plots are segmented by sector and geography, the model reveals the structural drivers of failure that define the Italian industrial landscape. A prominent finding is the "Asset Heaviness Penalty" observed in manufacturing.

Manufacturing firms carry a consistently higher risk profile than service-based entities for identical levels of financial distress. This reflects the structural difficulty of a turnaround when capital is locked in illiquid machinery and inventory. While an asset-light service firm can pivot or downsize

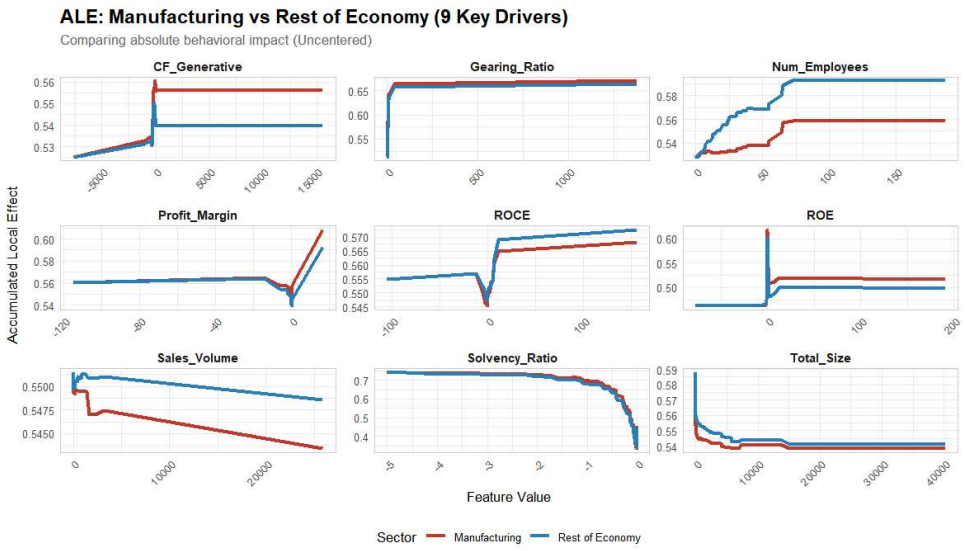

Figure 22: This Accumulated Local Effects (ALE) plot segments the marginal impact of financial drivers by sector, comparing Manufacturing (Red) against the Rest of the Economy (Blue). It empirically identifies a structural "Rigidity Trap" inherent to industrial firms.

with relative agility, the "heavy" manufacturer is often anchored by its own physical infrastructure, making it harder to escape the gravity of insolvency.

Regional Advantage

This structural logic is further complicated by a phenomenon we might call "Regional Darwinism." Surprisingly, firms located in Northern Italy exhibit a higher probability of default than those in the South for the same financial ratios. This suggests a more aggressive and efficient ecosystem in the North; in a more competitive market, inefficient firms are not afforded the luxury of slow decay. Instead, a more proactive network of creditors and competitors ensures that "Darwinian Efficiency" takes hold, liquidating failing entities faster than in regions where institutional inertia might allow them to linger.

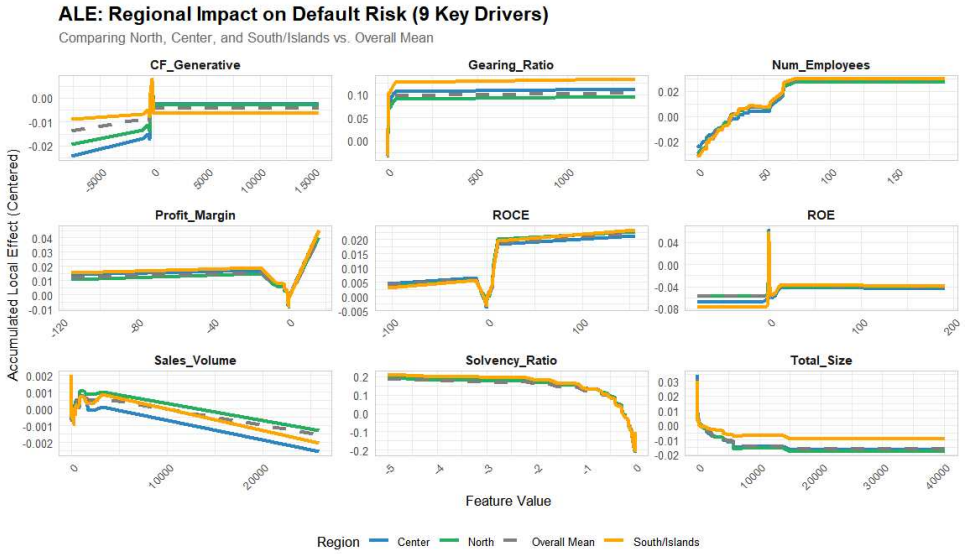

Figure 23: This Accumulated Local Effects (ALE) plot segments the marginal impact of financial drivers by Macro-Area (North, Center, South/Islands), revealing a phenomenon of "Regional Darwinism".

5.4 Decoding Classification Errors

While global metrics like the AUC of 0.79 confirm the model's overall discriminative power, a true "forensic autopsy" of the results requires us to look beyond the successes and into the structural blind spots of the algorithm. By partitioning the test set into False Negatives, the "Silent Killers" of our lending program, and False Positives, the "Miraculous Survivors" who succeeded despite being classified as failures, we can uncover the specific financial configurations that manage to deceive even a high-performing XGBoost classifier.

5.4.1 The "Silent Killers" (False Negatives)

The most dangerous errors in credit modeling are the "Silent Killers": firms labeled as "Safe" by the model's operational threshold that subsequently collapsed into failure. Error profiling reveals a fascinating "Debt Load Paradox" among these entities. Their median debt-to-revenue ratios were compressed between 0.5 and $1.0x$, figures nearly identical to those of genuine turnaround successes.

A SHAP Waterfall analysis of specific cases demonstrates how the model was "rationally fooled." It assigned a negligible probability of default because manageable debt-to-sales ratios and a temporarily positive ROA provided a convincing alibi for structural rot. In these instances, the model interpreted low leverage as a sign of health, failing to recognize a subset of "Terminal Burners." These are firms that appear stable on paper but lack any financial buffer; they are so lean that a single, sudden liquidity shock can prove fatal, leaving no room for the model to signal a gradual decline.

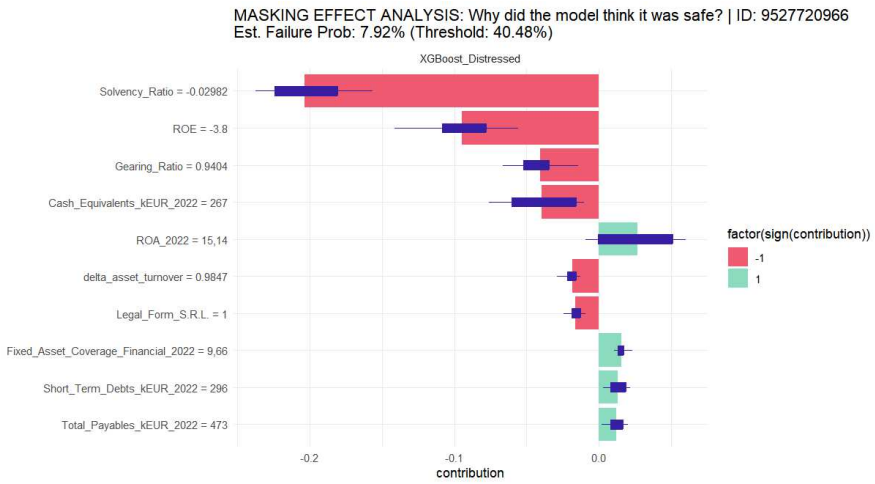

Figure 24: This SHAP Waterfall plot illustrates the "Masking Effect", a Type II error where the algorithm erroneously classifies a terminal firm as a "Turnaround Candidate."

5.4.2 The "Miraculous Survivors" (False Positives)

At the other end of the spectrum lie the "Miraculous Survivors", firms flagged by the model as "dead" (Collateral Damage) that nonetheless managed to reinstate solvency. Interestingly, these False Positives often possessed solvency ratios significantly more negative (median -2.0 to -4.0) than the true successes.

In a sense, these errors represent a vindication of the model's logic rather than a failure of its intelligence. In extreme cases, the model assigned a 98% default probability to a firm whose equity deficits were 4.6 times its assets. No algorithm based on financial physics would, or should, have predicted survival for such a hollowed-out hull. These outcomes highlight a "Zone of Ignorance": survival here is typically driven by "Ex-Machina" rescues, shareholder cash injections, private debt

write-offs, or political interventions, that are entirely invisible to accounting-based datasets. The model was directionally correct about the risk; external support simply overrode the reality.

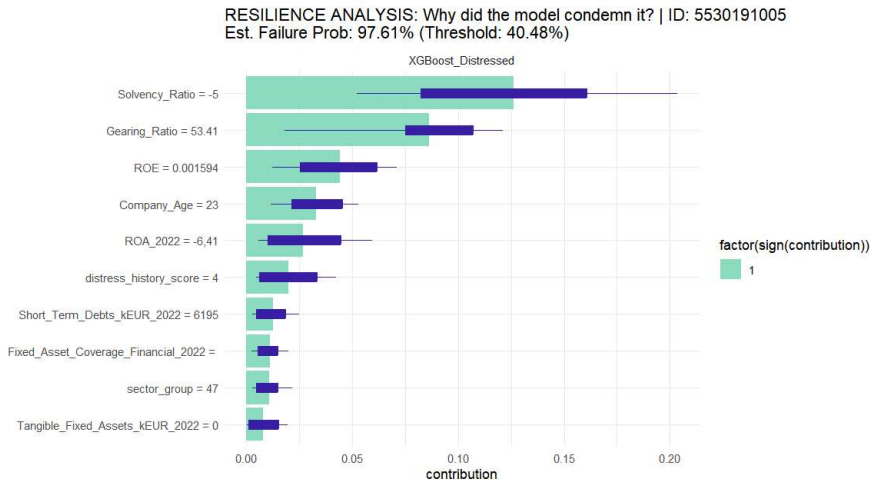

Figure 25: SHAP Waterfall plot of a "Miraculous Survivor", a Type I error where the model predicted specific failure (e.g., >90% probability) for a firm that successfully restored solvency.

5.4.3 Structural Biases: "Too Big to Fail"

Beyond individual cases, the forensic audit reveals a nuanced size bias that sets apart operational burden from financial scale. While the algorithm assigns a "Stability Premium" to larger enterprises, this protection is not uniform across all metrics.

The model identifies a distinct "Liability of Growth" regarding human capital: as the headcount creeps toward the 50-unit mark, the predicted risk rises sharply, reflecting the operational drag and overhead that come with scaling up. However, this trend shifts when accounting for Total Size. Here, the algorithm confirms the "Soft Budget Constraint" theories prevalent in the Italian market. Larger asset bases act as a systemic shield, and as they grow, the SHAP impact on default risk tails off, mirroring the "Too Big to Fail" safety net of the banking sector.

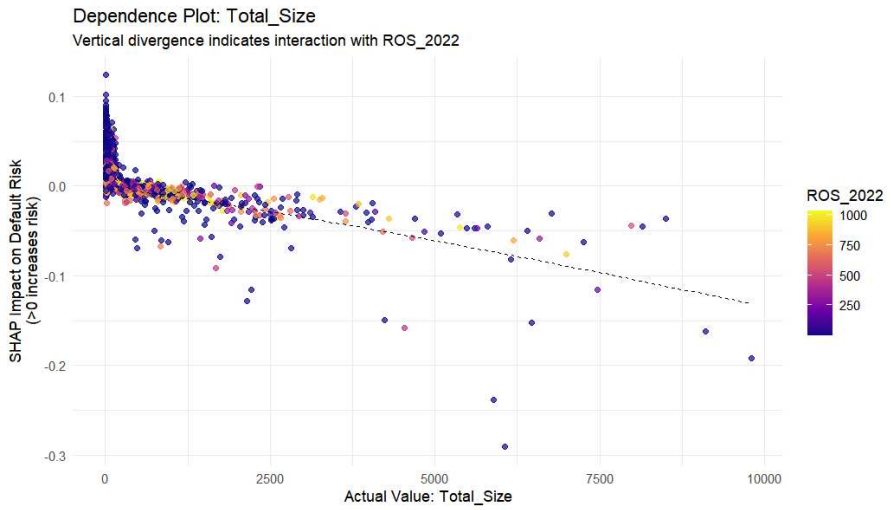

Figure 26: This SHAP Dependence Plot visualizes the interaction between Financial Scale (x-axis: Total_Size / Total Assets) and Operational Efficiency (Color Scale: ROS / Return on Sales). It exposes a structural "Stability Premium" granted to asset-heavy firms, validating the Soft Budget Constraint hypothesis.

Ultimately, the model recognizes that "Survivability" is often a function of institutional inertia and credit network density rather than pure internal efficiency. While expanding a workforce creates immediate operational vulnerability, a massive balance sheet lets firms hang on in a state of insolvency far longer than an agile micro-SME. Essentially, while more workers create a heavier management burden, a massive balance sheet acts as a structural insurance policy against failure.

Synthesis of the Error Audit

The following table summarizes the systematic nature of these errors, proving they are not random noise but reflections of deep market behaviours.

Error Category	Model Diagnosis	Forensic Cause	Theoretical Support
Silent Killers – FN	Low Risk	Low Debt/Sales masked structural rot	"Operational Burner profile"
Miraculous Survivor – FP	High Risk	Deep Equity holes; survival due to external bailout	Ex-Machina rescues
Systemic Bias	Size Preference	"Stability Premium"	Evergreening of large debtors

Table 4: Qualitative Error Analysis and Forensic Diagnosis

This forensic audit dictates a clear path for implementation: the model should function as a sophisticated support system where "Safe" predictions for firms with high turnover volatility or high employee counts require manual verification. This ensures we rule out the "Silent Killer" profile before capital is committed.

5.5 The Cost of Inertia: Zombie Firms and Capital Misallocation

While the preceding analyses confirm that the XGBoost model effectively distinguishes between viable turnaround candidates and terminal failures, the true utility of this "Gatekeeper" mechanism lies in its ability to map the systemic risk and aggregate economic inefficiency plaguing the Italian credit market. By applying the analytical framework of Schivardi to these classifications, this section develops a comprehensive estimate of the "Cost of Inertia." This metric represents the capital presently trapped in unproductive entities, resources that, if liberated, could facilitate a broader economic recovery. Ultimately, this inertia is not a mere statistical byproduct; it is a structural impediment that actively distorts the natural mechanisms of market selection. In essence, the model moves beyond simple prediction to act as a diagnostic tool, revealing how institutional forbearance prevents capital from flowing toward its most productive uses.

5.5.1 Economic Implications of Zombie Status and Market Sclerosis

In this study, "Zombie Status" is not just a descriptive label, but a rigorous construction rooted in the model's ability to predict outcomes for firms starting with negative equity. Consequently, "Technical Failure" is redefined: it is no longer just the formal start of insolvency proceedings, but the demonstrable inability of an active firm to claw back toward positive net worth. Rather than recovering, these entities persist in a state of structural stagnation, validating the hypothesis that time acts as a decay factor rather than a healer for the distressed cohort. Essentially, this definition shifts the

focus from legal status to economic viability, identifying firms that are "functionally dead" despite remaining operationally active.

The empirical evidence generated facilitates a granular disaggregation of the 71.66% of firms classified as "Failure/Zombie", moving beyond traditional binary definitions to reveal two economically distinct subcategories. The most prominent group within this cluster, comprising $18,715$ entities, is defined as "Terminal Burners". These firms do not merely suffer from excessive leverage; they suffer from fundamental operational insolvency (analysed under stress in Section 5.6). From a macroeconomic perspective, they function as"liquidity black holes", where every unit of credit extended is consumed to offset recurrent operating losses rather than to finance productive investment. This effectively validates the "Trapped Capital" theory (Caballero, Hoshi and Kashyap, 2008; Schivardi, Sette and Tabellini, 2022), identifying these firms as the primary drivers of value destruction within the cohort. To put it plainly, these are businesses that lose money just by opening their doors, making them unsalvageable through financial restructuring alone.
Conversely, a significant minority, representing 14% of the exposed capital consists of "Rate-Sensitive Zombies". While these 3594 firms generate positive EBITDA and may appear "alive" on a standard income statement, their operational engine is too weak to sustain market-rate debt. Specifically, their Interest Coverage Ratio (ICR) collapses below unity when subjected to a 6.5% market-rate stress test. These firms survive exclusively by leveraging subsidized rates or pandemic-era debt moratoria, functioning as interest-rate arbitrageurs rather than viable businesses. They represent a latent systemic risk: entities buoyed by temporary subsidies that mask an underlying inability to survive normalized financial conditions. Essentially, these firms are interest-rate hostages; they remain solvent only as long as "cheap money" persists, making them highly vulnerable to the "weaning off" of state support.

The persistence of these entities triggers a "Sclerosis Effect"(Caballero, Hoshi and Kashyap, 2008), to describe how zombies effectively choke off market vitality. Within this framework, propping up unproductive firms is far from a zero-sum game; instead, it generates tangible negative externalities, or "Congestion Effects," that actively penalize healthy enterprises. Specifically, zombies artificially inflate the demand for labour and raw inputs, driving up operational costs for their viable competitors. By keeping these resources tied up, they effectively squeeze profit margins across the broader industrial sector. In other words, by refusing to fail, these firms act as a "clog" in the economic system, forcing healthy companies to pay a premium for resources that are being wasted elsewhere.
This allocative distortion is further worsened by a noticeable "Size Bias" identified in the model's error profile. The data reveals that credit is often funnelled into mature, larger entities at the direct expense of younger, more dynamic firms, a phenomenon that throws a wrench in the Schumpeterian process of "Creative Destruction". By forestalling the entry of more efficient operators who would otherwise fill the market space, this bias keeps the ecosystem in a state of hibernation. Essentially, the credit market ends up protecting the past instead of financing the future, keeping established "ghosts" on life support while starving the next generation of potential winners and innovators.

The 2019–2022 temporal scope of this study exposes the lingering fallout of the "Market Hibernation" induced by pandemic-era policies. While state guarantees, such as the MCC[38] in Italy, successfully averted an immediate credit crunch, the model's findings suggest that for many firms, emergency liquidity did little more than kick the can down the road. In these cases, temporary liquidity gaps were transformed into deep-seated structural insolvency. Essentially, the safety net became a trap; it preserved the firm's existence without addressing the underlying business model decay.

[38] MedioCredito Centrale, grants loans and financial instruments, regularly in collaboration with the EIB, and acts as a public guarantee fund for investments. Its financial activity is usually often directed to the support of SMEs, Green Transition or Specialized Financing for innovative startups

This aligns with the "Weaning Off" phenomenon, proposing that the drying up of public cash flows triggers an immediate shock for firms that failed to restructure during the subsidized grace period. By penalizing the `Distress_History_Score`, the XGBoost model precisely identifies entities that squandered their window of opportunity.
These firms accumulated debt without building the repayment capacity needed to survive a return to ordinary market conditions. In other words, the model uses the chronicity of distress as a proxy for operational failure, flagging firms that have stayed underwater for too long to ever resurface.

5.5.2 The Schivardi Framework: Quantifying Trapped Capital and the "Sanitization" Imperative

Setting a pragmatic "Gatekeeper" threshold at approximately 0.40 inevitably triggers the debate over credit rationing. To shift this from an ethical dispute to an economic imperative, this study applies the framework of Schivardi, viewing the liquidation of terminal firms not as a destruction of value, but as a mandatory "sanitization"(Caballero, Hoshi and Kashyap, 2008) of the credit channel. The model essentially performs a forensic autopsy on the balance sheet, pinpointing exactly where capital is being used to mask a "mortal wound" rather than to drive a genuine turnaround. Put simply, we are cleaning the pipes; by removing the obstructions of the past, we allow liquidity to flow back to the productive economy

When we aggregate the financial exposure of firms marked as "Terminal," the sheer scale of the system's allocative inefficiency comes into sharp focus. This "Sclerotic Burden" represents an estimated € 12.991 billions in credit exposure, a staggering 68% of the total portfolio currently held by firms with no statistical path back to solvency. While it is difficult to discuss the termination of support for so many entities, the data suggests that this credit is effectively "trapped" in a cycle of decay rather than fuelling a turnaround.
From a regulatory and prudential standpoint, this trapped capital imposes a "double cost" on the financial system. First, these exposures carry an exceptionally high Probability of Default, which forces banks to set aside substantial loan loss provisions that eat into their profitability. Second, under the Basel III framework, these high-risk positions attract punitive Risk-Weighted Asset (RWA) calculations. By keeping these firms on "life support" through institutional forbearance, banks end up tying their own hands, artificially inflating their capital requirements and reducing the liquidity available to fund healthy, viable SMEs. In essence, the choice to avoid the pain of liquidation today creates a silent credit crunch for the healthy companies of tomorrow, effectively shrinking the economy's productive capacity.

The case for active intervention is strengthened when we look at the operational efficiency, or lack thereof, within this distressed group. Our analysis reveals that the "Terminal" cluster is not just standing still; it is actively bleeding value. The total Annual EBITDA for these firms hits € 1.559 billion.
This figure represents a critical "Burn Rate": for every fiscal year these entities remain on life support through credit expenditures, the economy suffers a net loss exceeding € 1.5 billion. This phenomenon validates the "Congestion Theory", as these "Terminal Burners" burn through labour and capital to produce output that the market values at less than its cost of production. Consequently, credit extended to them acts as a subsidy for inefficiency, throwing a wrench in the market's ability to clear and serving as a "hidden tax" on the competitiveness of the "Resilient" cluster. In simple terms, these firms aren't just stagnant; they are value-shredders that actively drain the resources needed by healthier companies to grow.

A Counterfactual Reallocation Analysis

To reverse this systemic destruction, the model proposes a strategic pivot known as the "Schivardi Swap." This exercise moves beyond theoretical risk to demonstrate the transformative potential of active credit selection grounded in predictive modelling. By accepting a conservative Loss Given

Default (LGD) of 60% upon the immediate liquidation of terminal firms, the simulation acknowledges a severe "sanitization cost": the write-off of approximately €7.74 billion. While this figure is undeniably high, the model frames such loss recognition as a mandatory prerequisite for systemic health rather than a mere destruction of value.

Despite this substantial haircut, the liquidation process yields a "Liquidation Dividend" of approximately € 5.16 billion in recoverable capital. The true power of the swap is realized when this capital is redeployed exclusively to the "Hidden Gems", those viable turnaround candidates identified by the model that still possesses positive operational momentum. As argued by Caballero, being transparent about these losses is economically superior to the indefinite "evergreening" of dead-end loans.

By shifting resources from value-destroyers to these resilient targets, the simulation estimates the generation of € 473.4 million in new, positive EBITDA annually. In essence, the "Schivardi Swap" proves that a surgical strike on the balance sheet today is the only way to fund the industrial recovery of tomorrow; it trades a one-time write-off for a permanent restoration of productive capacity. From a macroeconomic perspective, this is far from a zero-sum redistribution. It yields a "Net Economic Delta", a systemic performance jump of approximately € 2 billion. This figure provides the empirical "smoking gun" for the model's broader utility. Utilizing machine learning as an active "Gatekeeper" does more than just protect the balance sheets of individual lenders; it maximizes the collective return on capital by moving liquidity away from resource-sinkholes and toward entities capable of true resource multiplication. Essentially, the model proves that the most "socially responsible" path for a lender is not forbearance for the unviable, but the surgical reallocation of credit toward the survivors.

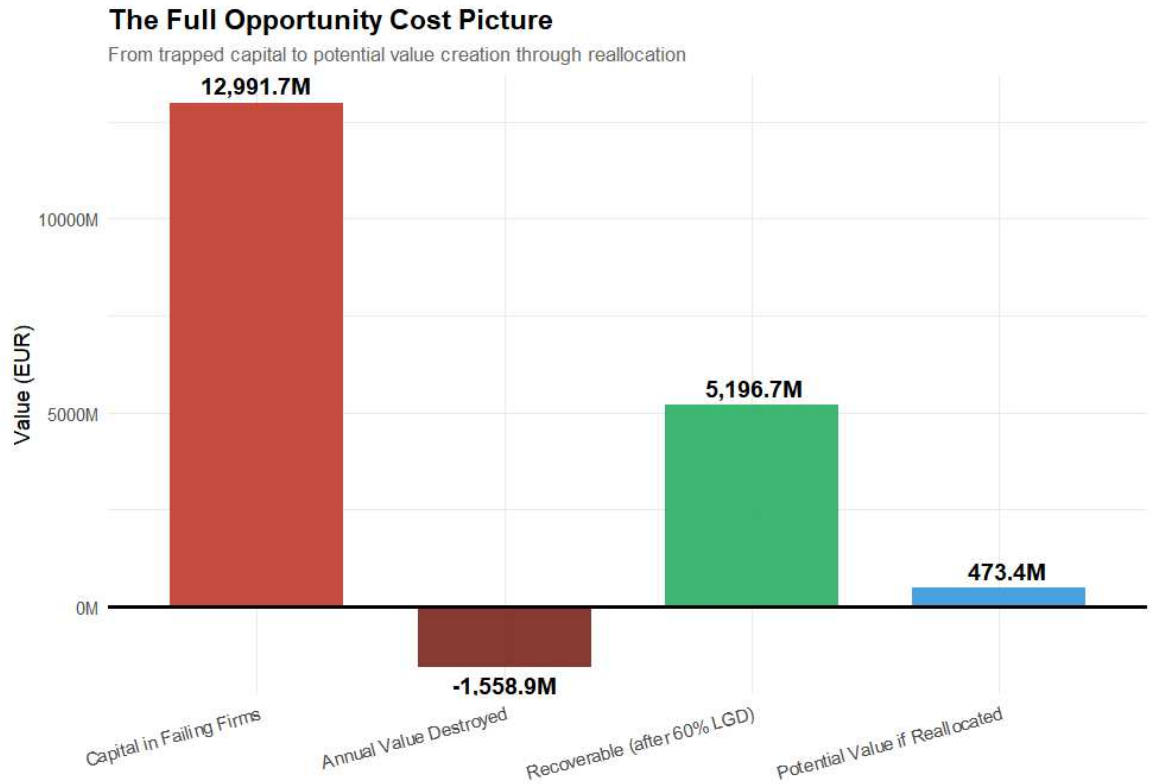

Figure 27: The counterfactual analysis shows the credit misallocation's burden, and its hypothetical effects over a reallocation favouring resilient firms.

5.6 The Refinancing Cliff

This final evaluative phase marks a critical transition: moving from static historical classification toward a dynamic, forward-looking stress test. While the XGBoost model successfully identifies firms that restored solvency under the idiosyncratic conditions of 2019–2022, the aggressive normalization of monetary policy now demands a more rigorous assessment. We must determine whether these turnarounds are industrially resilient or merely buoyed by temporary, subsidized liquidity. To this end, this section utilizes the "Refinancing Cliff" simulation as a forensic tool to separate genuine operational recoveries from"Rate-Sensitive Zombies", entities that may appear healthy on paper but are effectively living on borrowed time. Essentially, we are stripping away the artificial safety net of the pandemic era to see which firms can actually stand on their own in a high-interest reality.

5.6.1 Defining the Cliff

To transcend the limitations of static accounting ratios, which often function as lagging indicators of distress rather than real-time predictors (Ohlson, 1980; Shumway, 2001), this study implements a dynamic behavioural simulation. This procedure evaluates corporate resilience under normalized monetary conditions, moving beyond the "Solvency Illusion" created by the low-interest-rate environment of the pandemic era. By leveraging the granular partition of short-term and long-term debt established at T_0, the model defines a 'Refinancing Cliff': the event horizon where legacy liabilities, previously sheltered by state guarantees or moratoria, must be dropped at competitive market costs. The stress test can help the identification for facilitated financing conditions. A deterministic refinancing scenario was simulated by subjecting the entire distressed cohort to a hypothetical 6.5% market interest rate. As detailed in the experimental code, this rate also serves as a proxy for the post-tightening cost of capital for SMEs, effectively stripping away the "subsidy veneer" described by Caballero, Hoshi, and Kashyap (2008), who argue that firms surviving solely due to below-market financing distort the mechanism of creative destruction. By recalibrating the Interest Coverage Ratio (ICR) under this stress condition, the simulation acts as a forensic filter, reclassifying firms into four distinct clusters based on their structural sensitivity to the cost of debt:

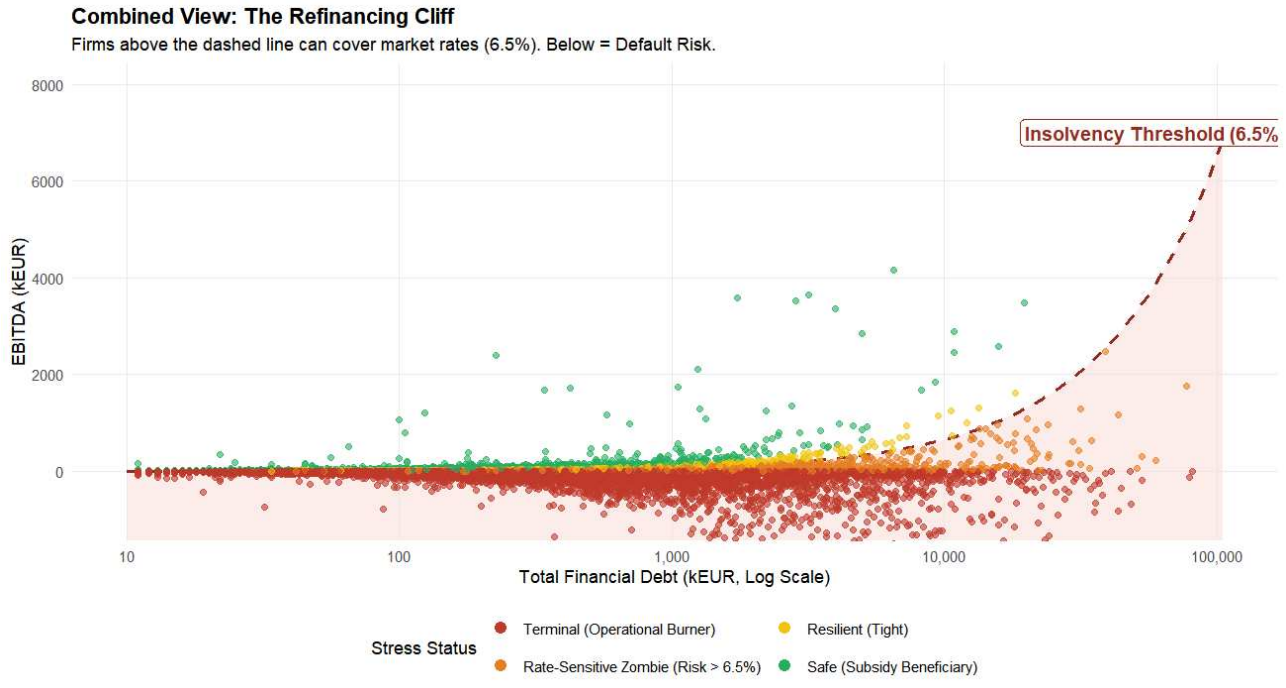

Figure 28: This scatter plot simulates a deterministic stress test on the entire Distressed Cohort, subjecting firm balance sheets to a normalized market interest rate of 6.5%.

1. **Safe (Subsidy Beneficiaries)**: Operationally robust entities whose margins are sufficient to absorb the rate hike while maintaining a healthy coverage ratio ($ICR \geq 2.0$). The research identifies these firms as genuine turnaround successes; their survival is driven by operational efficiency rather than financial aid, aligning with the "Resilient Core" identified in the empirical analysis. They amount to $1,978$ firms.
2. **Resilient (Tight)**: Firms capable of servicing debt at 6.5% through structural adjustments, though with reduced buffers ($1.0 \leq ICR < 2.0$). While vulnerable to liquidity shocks, these entities possess positive economic value, consistent with the definition of viability where operating income exceeds interest obligations (Adalet McGowan, Andrews and Millot, 2018). A quantity of $1,907$ have been identified as resilient.
3. **Rate-Sensitive Zombies ('ZIRP Zombies')**: Entities that appeared solvent only due to legacy subsidized rates. While operationally active, their EBITDA is insufficient to cover interest expenses at the normalized 6.5% rate ($0 \leq ICR < 1.0$). This classification operationalizes the OECD definition of "Zombie Firms" as those with persistent interest coverage ratios below unity (Adalet McGowan, Andrews and Millot, 2018). This study classifies these firms as "Interest-Rate Arbitrageurs" rather than value creators; their solvency is an artifact of the Zero Interest Rate Policy (ZIRP) era, making them highly susceptible to the "Weaning Off" effect as public support is withdrawn. In this study, $3,594$ firms are considered rate-sensitive zombies.
4. **Terminal Failure (Operational Burners)**: Non-viable entities generating negative EBITDA, rendering them insolvent regardless of the interest rate environment. Unlike zombies, which are kept alive by cheap credit, these firms actively destroy capital. This study categorizes them as "Terminal Burners," representing the capital misallocation described by Schivardi et al. (2020) where credit extended merely covers operating losses rather than productive investment. An amount of $18,715$ operational burners have been spotted.

This simulation provides the model with a dynamic view of future solvency, ensuring that turnaround predictions are not based on the obsolete "free money" paradigm of the pandemic era (Market Hibernation), but on the 'new normal' of the 2023–2025 monetary cycle. By distinguishing between firms suffering from structural insolvency and those merely addicted to low rates, the model enforces a "Sanitization Imperative," identifying capital that must be reallocated to prevent market sclerosis.

5.6.2 Exposing Rate-Sensitive Zombies

To quantify the true fragility of the "Safe" cluster, this study subjects these firms to a deterministic stress test. We recalibrate the Interest Coverage Ratio (ICR) of each entity by applying the 6.5% refinancing rate. By stripping away the subsidy, we find a critical divergence in the path to recovery.

The "Resilient Core" consists of $1,907$ firms whose recovery is authentic. Their EBITDA growth is robust enough to cover normalized financial charges, proving that even highly leveraged portfolios can contain "Hidden Gems" if their cash flows are rate insensitive. Conversely, the "Rate-Sensitive Segment", comprising $3,594$ firms, stand as the "False Negatives" of a regime-blind model. These are operationally stagnant businesses that only "survived" the pandemic because the cost of their inefficiency was socialized during the hibernation phase. Practically, the 6.5% simulation acts as the ultimate filter: it separates the firms that actually took advantage the crisis to reinvent themselves from those that merely used it to hide their financial deficiencies.

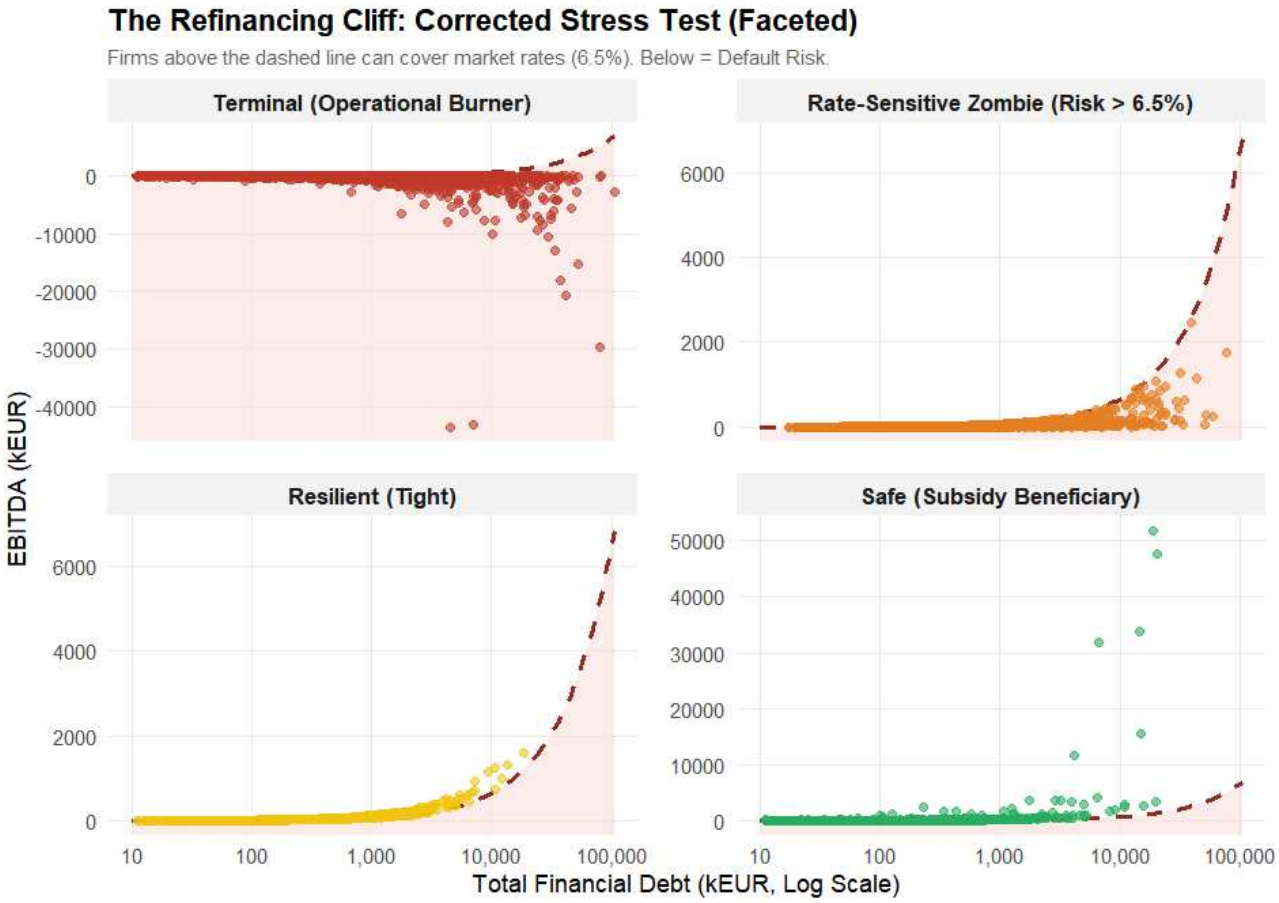

Figure 29: This faceted grid decomposes the cohort into four distinct economic clusters based on their resilience to the 6.5% stress test.

5.6.3 The "Weaning Off" Effect and Industrial Sustainability

The stress test results offer a striking empirical echo of the "Weaning Off" effect, documented by Grajzl et al.. While the state-mandated "grace period" was intended to facilitate restructuring, the model reveals a stark bifurcation between genuine Liquidity Support, a bridge to future viability, and Insolvency Masking, the artificial preservation of terminal decay.

By penalizing firms where debt accumulation outpaced revenue growth (scissor_debt_sales), the model flags those that spent their "Hibernation" period. In these cases, the liquidity buffer was never funnelled into productive capacity; instead, it was devoured by inert liabilities. The "Refinancing Cliff" simulation confirms this diagnosis: the 3,594 "Rate-Sensitive Zombies" are precisely those entities that piled on debt without generating the EBITDA margin necessary to service it at a normalized 6.5% rate. In short, they didn't use the state's money to fix the engine; they used it to pay for the fuel they were already burning.

Furthermore, the great significance of the `Distress_History_Score` shatters the comforting myth that corporate distress is merely a fleeting, cyclical phase. The model instead uncovers a "Chronicity Penalty", the deep-seated financial scar tissue that accumulates when a firm lingers in negative equity throughout the 2019–2022 window. These firms exhibit a non-linear spike in default probability that defies simple recovery logic. This discovery aligns perfectly with the "Sclerosis" framework of Caballero, suggesting that for this sub-segment, the "Weaning Off" event does not interrupt a genuine turnaround; it merely pulls the plug on a life-support system that was never going to be enough. In these cases, time was not a healer, it was simply a stay of execution.

Consequently, the "Refinancing Cliff" simulation serves as the ultimate validation of the Gatekeeper's role. It proves that a passive reliance on accounting solvency, which appeared stable only because of state-sponsored subsidies, creates a dangerous "Solvency Illusion." By identifying and isolating firms whose survival is purely a function of subsidized pricing, the model provides the necessary tool for active credit selection.

This mechanism ensures that capital is no longer "parked" in rate-sensitive dead-ends, where it effectively goes to die but is instead funnelled toward resilient enterprises capable of thriving in a normalized, 6.5% interest rate environment.

5.7 Threshold Analysis: The Gatekeeper Dilemma

Calibrating the acceptable risk tolerance is not merely a technical optimization, but a decision with profound economic implications for capital allocation and credit availability. When do we cut credit availability?
Conventional credit-scoring models typically establish preemptive thresholds well before a firm enters a state of negative equity, but the scope of the study is exactly to venture into the distressed cohort who show an ex ante negative position and to obtain further granularity of the event, identifying and sustaining resilient entities that would otherwise be relegated to terminal liquidation.
This section specifically discusses the variation of different thresholds on risk/return profiles for investors, credit lenders and public institutional support, like incentives and grants provided during the pandemic crisis.

5.7.1 Skewness of Distress

To understand the gravity of this calibration, one must first consider the sample distribution. In our distressed cohort, the prevalence of "Class 1" (Failure or Zombie stagnation) stands at a dominant 71.66% This confirms that for a firm in severe distress, the modal outcome is not recovery, but rather legal failure or a permanent state of operational inertia. Conversely, only approximately 28.34% of the sample eventually manages to reinstate debt viability. Essentially, the restoration of solvency within this environment is a statistical long shot, a rare event that requires the model to find the "needles in the haystack" of the Italian SME landscape.

A naive classifier might simply label every firm as "Terminal" and achieve a deceptively high accuracy score due to this imbalance. However, such a model would offer zero informative value for active credit selection. The true challenge lies in navigating the asymmetrical costs of error. We are constantly weighing the risk of a Type II error (a False Negative that results in lending to a "Zombie") against the risk of a Type I error (a False Positive that leads us to reject a firm with genuine turnaround potential).

Strategy	Thresh.	Specificity	Sensitivity	PPV	NPV	F1	Detection Rate	Balanced Accuracy
Academic	0.5086	0.7469	0.6800	0.8717	0.4800	0.7640	0.5590	0.7134
Cost-Sensitive	0.0964	0.0137	0.9995	0.7193	0.9189	0.8366	0.9958	0.5066
High Sensitivity	0.4036	0.5952	0.8000	0.8333	0.5406	0.8163	0.5733	0.6976

Table 5: Performance Comparison by Threshold Strategy

5.7.2 Scenario Modelling

To explore the practical consequences of this dilemma, we delineate three distinct thresholding paradigms. These are not merely different mathematical cut-offs; they represent divergent methodological frameworks and theoretical objectives that fundamentally alter the credit-market landscape.

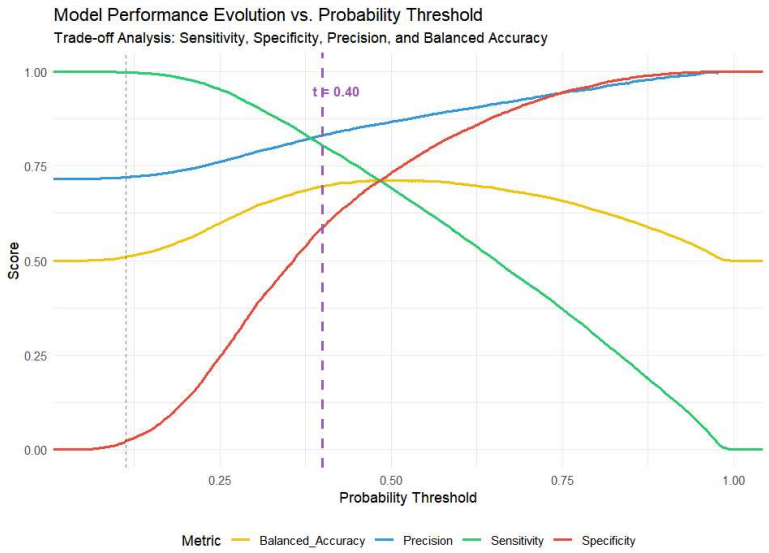

Figure 30: The Optimal Economic Threshold set at 0.40 finds a balanced compromise between the various parameters.

Academic Baseline

This configuration represents the theoretical statistical optimum, assigning equal weighting to the classification of both outcomes. By maximizing the Youden Index, the model identifies the "Point of Parity" on the ROC curve where the arithmetic sum of misclassifications is minimized. For our model, this parity-optimized threshold is established at 0.5086 yielding a statistical equilibrium between sensitivity (68%) and specificity (75%).

While the resulting Balanced Accuracy of 71% suggests a statistically capable model, this configuration remains economically reckless for institutional risk management. Within a distressed cohort, a sensitivity of 68% implies that approximately one-third of terminal failures would bypass the model's filters. To put this in perspective, the Loss Given Default (LGD) typically outweighs the opportunity cost of a rejected loan by a magnitude often cited as a 10:1 ratio. Permitting 32% of terminal entities to access further liquidity would result in capital impairment. While this scenario serves as an essential laboratory baseline, it lacks the defence required for real-world solvency management.

Prudential Cut-Off

Where the Youden Index seeks total information gain, the Prudential Cut-Off prioritizes absolute safety. This scenario establishes a threshold at 0.0964 functioning as a preemptive, highly protective signal. At this level, Sensitivity achieves a perfect score of 99.95%, ensuring that not a single terminal failure is misclassified as a survivor. However, this absolute safety comes at a devastating cost: specificity collapses to 1.37%.

Derived from the most conservative interpretations of the Basel III framework, this threshold operates as a "Predictive Bunker." It effectively intercepts nearly all terminal failures, but it does so by indiscriminately rejecting almost the entire distressed cohort. The classifier becomes so

overwhelmed by risk-aversion that it loses the capacity to discern firms with genuine potential. In other words, this configuration stops the bleeding by effectively halting all lending activity. This perfectly illustrates the phenomenon of "Market Sclerosis": while standard risk-averse practices prevent defaults, they simultaneously impede economic recovery by starving viable, resilient firms of the capital they need to survive.

Turnaround Private Equity Fund

The third paradigm emulates the mandate of a specialized operator, such as a turnaround private equity fund. This configuration is calibrated to maintain a baseline safety margin while ensuring a viable flow of investment into the market. By artificially setting the operative threshold to 0.4036, we create a model that is significantly more inclusive than the prudential bunker, yet far more defensive than the academic baseline.

In this scenario, the model successfully identifies 80% of the failing entities while still classifying over half of the non-failing firms correctly. This targeted sensitivity of ~80% is grounded in the theoretical framework of Altman and Hotchkiss (2005), who established this as the "quite acceptable" benchmark for Type I accuracy in credit risk modelling. This standard is corroborated by contemporary Italian research, who achieved similar recall rates using the same AIDA database. Ultimately, the 0.40 threshold represents the only pragmatic path forward; it is tight enough to preserve institutional capital, yet loose enough to provide the "liquidity lifeline" necessary for resilient firms to achieve a successful turnaround.

5.7.3 Dynamic Thresholding, a Pre-emptive Stress Test

The divergence between these three scenarios confirms that the "Gatekeeper" mechanism cannot remain a static statistical artifact; it must function as a dynamic valve adjusted to the prevailing macroeconomic climate. In the current Italian landscape, this calibration is not occurring in a vacuum. As we move from the absolute risk aversion of a 'Regulator' to the calculated tolerance of an 'Investor,' the choice of the Operational Threshold (0.40) serves a specific forensic purpose.
By prioritizing the sensitivity of 80% this threshold effectively filters out the "Rate-Sensitive Zombies" whose survival has been artificially prolonged by pandemic-era liquidity. Within this cohort, the 0.40 cut-off acts as a pre-emptive stress test: it identifies those entities with the operational EBITDA and structural solvency required to withstand the withdrawal of state support.
This calibrated approach fulfils the model's role as a tool for economic sanitization, ensuring that capital is preserved for firms capable of resource multiplication rather than being lost to those in terminal decay. This leads us directly to the most critical hurdle facing the Italian SME sector: the "Refinancing Cliff." As the following section will demonstrate, the model's 0.40 threshold is the only setting capable of distinguishing between those firms that can bridge the jump to normalized market rates and those that will inevitably shatter upon impact.

5.8 Synthesis, Limitations, and Theoretical Constraints

The convergence of the model's performance metrics, its forensic drivers, and the results of the stress simulations demand a final synthesis within the broader context of turnaround theory. While the empirical evidence confirms the model's efficacy as a predictive "Gatekeeper," it is essential to map the theoretical boundaries and data-driven constraints that frame these findings. This critical appraisal ensures that the model's outputs are not interpreted as binary certainties, but as high-fidelity probabilistic maps. They are assessments contingent upon specific structural parameters and, most importantly, the shifting sands of the Italian macroeconomic environment. In short, the model provides the "financial physics" of the situation, but the outcome remains a dialogue between these structural laws and the unpredictable interventions of the market.

5.8.1 The Solvency vs. Strategic Success Gap: Retrenchment and Recovery

A fundamental distinction must be maintained between the model's primary objective, the prediction of Technical Solvency Restoration, and the broader, more complex concept of long-term strategic success. To contextualize this, we must look at the classic two-stage turnaround framework proposed by Robbins and Pearce (1993), which characterizes corporate recovery as a sequential progression.

The first stage, Retrenchment, is a period of high-stakes survival focused on liquidity preservation, cost reduction, and the stabilization of a hemorrhaging balance sheet. The subsequent stage, Recovery, is a more creative endeavour, requiring strategic repositioning, product innovation, and the hard-won re-establishment of a competitive advantage.

The XGBoost model developed in this study operates effectively as a specialized auditor of the Retrenchment Phase. By identifying firms capable of reinstating positive shareholders' equity, the algorithm detects the necessary financial conditions for survival. It essentially verifies that the firm possesses the structural "integrity" required to escape the gravity of terminal insolvency. However, the model does not, and by its nature, cannot guarantee sufficient conditions for Stage 2 competitive success.

A firm classified as "Safe" has proven it can repair its hull, but its long-term voyage remains contingent upon qualitative factors, leadership quality, organizational culture, and market agility, that standardized accounting data is simply too blunt to capture. Consequently, the model should be viewed as a surgical tool for "Financial Sanitization": it clears the wreckage of the system, identifying those entities that are technically fit to continue, but it leaves the work of industrial excellence to the province of human management. In short, the model tells us who *can* survive; the market will eventually decide who *deserves* to thrive.

5.8.2 Data Constraints and Temporal Biases: The 2019–2024 Regime Shift

The validity of these findings is inherently bound by the idiosyncratic nature of the observation window. The period spanning 2019 to 2024 represents one of the most significant "regime shifts" in modern economic history, introducing specific biases and information asymmetries that frame the model's predictive limits. To use the model effectively, one must acknowledge that it was forged in an era of unprecedented distortion.

The "Market Hibernation" Distortion

As explored in the stress-testing phase, the massive influx of public support measures between 2020 and 2022 created an artificial environment where default signals were suppressively masked. This "Liquidity Artifact" presents a significant challenge: it may lead to an inherent underestimation of risk for firms whose survival was purely a byproduct of policy intervention. While the "Refinancing Cliff" simulation was designed to strip away this veneer, the underlying training data reflects a protected economic regime that may not recur. This creates a "Data-Regime Gap," suggesting that the model is a map of a world under a temporary safety net, one that must be recalibrated as that net is permanently withdrawn.

Reporting Lags and the "Recognition Gap"

Accounting data is, by definition, an ex-post reflection of firm performance. In the Italian context, the delay between the close of a fiscal year and the official filing of a balance sheet often spans four to six months, creating a dangerous "Recognition Gap."

In a distressed environment where liquidity can evaporate in a matter of weeks, the model provides a structural diagnosis, a "biopsy" of the firm's health at a specific point in time, rather

than a real-time pulse. It is a powerful tool for identifying long-term structural rotting, but it cannot anticipate "black-swan" events or sudden, out-of-court restructuring agreements that occur outside the official reporting cycle. The model sees the architecture of the building, but it cannot see the fire that starts after the inspection is filed.

The Survivor Bias of the "Red Zone"

Finally, the decision to focus exclusively on a distressed cohort with ex-ante negative equity provides high granularity but introduces a natural survival bias.

Firms that entered immediate liquidation prior to filing their 2022 or 2023 financial statements are naturally excluded from the dataset. This means the algorithm is effectively learned from the "resilient" survivors of the initial shock. While this focus allows the model to become an expert at identifying the "needles in the haystack," it also means the model is calibrated on firms that had enough latent strength to survive at least the first wave of the crisis. Essentially, the model is trained in the firms who stayed in the market; it does not see those who were knocked out in the preliminary phases.

5.8.3 Methodological Synthesis: Why Machine Learning Requires Stress Testing

The most significant methodological conclusion of this study is a sobering one: the inherent limitation of pure Machine Learning approaches when confronted with a seismic regime shift.

Beyond the Black Box: The Regime-Shift Wall

The existence of the "Rate-Sensitive Zombie" cluster proves that an algorithm, no matter how sophisticated, remains a prisoner of its training history. An XGBoost model trained on the 2019–2022 window, a period of artificial, subsidized calm, cannot "hallucinate" the structural impact of the 2023/2024 rate hikes if those dynamics were never present in its data. To rely solely on the algorithm would be to fly into a storm using a map of the clear sky.

Furthermore, while this study employs SHAP and ALE plots to bridge the interpretability gap, Crosato note that such post-hoc methods impose a 'regularization on the interpretations' and restrict the ability to fully address biases inherent in the training data. Therefore, while model-agnostic tools allow for the ranking of variable importance (e.g., identifying Profit Margin or Solvency Ratio as key drivers), they remain a reconstruction of the decision process rather than a direct view of the model's internal parameters, maintaining a degree of separation between the computational logic and the human explanation

Rudin's warning on Machine Learning Interpretation

While the implementation of SHAP in this study provides granular visibility into feature contribution, a fundamental theoretical constraint persists regarding the distinction between post-hoc explainability and models that are interpretable from the ground up. As rigorously argued by Rudin (2019), there exists a critical chasm between designing an inherently interpretable model, such as a sparse decision tree, and layering a secondary explanation over a "black box" like XGBoost.
Rudin posits that for "high-stakes decisions," a category that clearly covers credit default prediction and the classification of "Zombie Firms," relying on post-hoc explanations risks being unfaithful to the model's actual logic. In plain terms, an explanation that is easy for a human to grasp might not actually reflect how the algorithm is crunching the numbers. If an explanation perfectly matches the underlying model, the complex model is redundant; conversely, if the explanation simplifies the logic, it inevitably misses the mark when it comes to the actual computational mechanism. Consequently, the risk profiles generated by the Gradient Boosting algorithm, while highly accurate in capturing non-linear distress signals, end up as opaque approximations rather than transparent

causal mechanisms.
Therefore, the interpretability results in this chapter must be understood as summaries of prediction sensitivities rather than a direct mapping of the algorithm's internal logic. While Moscatelli argues that such opacity is a necessary trade-off to catch complex relationships, Rudin warns that it can hide confounding variables. Thus, the "Silent Killers" and "Miraculous Survivors" identified in the error analysis may partially reflect the "fidelity gap" where the SHAP tool drifts away from the ensemble model's actual decision boundary.

The Hybrid Solution: Forensic Normalization

To bridge this gap, this study advocates for a Hybrid Approach, a fusion of probabilistic machine learning and deterministic stress testing. This represents a transition from a passive "Black Box" to an active Forensic Normalization.
In this framework, the two methods perform distinct, yet complementary roles:

Component	Function	Diagnostic Focus
XGBoost Engine	Identifies Operational Distress	Detects "Terminal Burners" who are destroying credit availability
Deterministic Stress Test	Identifies Financial Fragility	Identifies "Rate-Sensitive Zombies" whose solvency is a byproduct of artificially low interest rates.

Table 6: Dual-Component Diagnostic Framework

Regime Adaptation and the "Solvency Illusion"

By integrating these two methods, we successfully pierce the "Solvency Illusion." We move beyond a model that simply predicts who *is* alive to one that evaluates who *can* survive. This dual-layered filter prevents the misclassification of firms that are technically "safe" in a subsidized hibernation but macro economically non-viable in a normalized market.
This hybrid integration represents the true state-of-the-art in predictive credit risk management for distressed portfolios. It transforms the model from a historical auditor into a dynamic strategic navigator, capable of adapting to the shifting sands of European monetary policy. We are no longer just counting the survivors; we are vetting the resilient.

5.8.4 The Economic Imperative of the Gatekeeper

Despite these inherent constraints, this study provides a robust empirical and theoretical justification for the adoption of high-performance machine learning frameworks within the distressed debt landscape. The red thread connecting these chapters reveals that corporate recovery in distressed SMEs is not a random occurrence, but a measurable outcome of structural resilience, one that can be caught in the net of a properly calibrated algorithm.
The model fulfils a critical dual forensic purpose, acting as both a shield and a scalpel for the Italian financial system:

Systemic Sanitization

By surgically identifying the "Terminal Burners", the model provides the objective evidence required to cease the misallocation of credit. This effectively reduces the "Congestion Tax" on the real economy, clearing up the industrial landscape of "Zombies" to make room for new growth.

Active Credit Selection

Conversely, by isolating the "Resilient Core," the model prevents the "Systemic Credit Halt" that typically follows a macroeconomic shock. It ensures that liquidity is not indiscriminately rationed but is instead funnelled toward enterprises that have proven they can thrive even within the gravity of a normalized, 6.5% interest rate environment.

6 - Conclusions

6.1 Conceptual Summary and Results

6.1.1 Topology of Crisis

The empirical reconstruction of the Distressed Cohort calls into question the foundational premise of traditional credit scoring, which is the idea that financial risk is a purely additive phenomenon. By analysing the higher-order interactions captured by the XGBoost framework, this study concludes that corporate crisis is not a cumulative sum of independent weaknesses, but rather a system of destructive interactions. To put it simply, a company does not fail just because it has many small problems, but because those problems combine to create a much larger, more complex failure.
The identification of the Operational Trap in manufacturing serves as the definitive proof of this topological complexity. The results demonstrate that Asset Heaviness in Manufacturing, which is traditionally viewed by linear models such as Altman's Z-Score as a static variable representing collateral value, functions as a context-sensitive variable. XGBoost is able to identify this relationship, changing interpretation as the value tips below zero. Essentially, the same assets that once protected the firm begin to accelerate its downfall once profits disappear.
This phenomenon is empirically captured by the distinction between Financial Scale and Organizational Complexity. While the model assigns a 'Stability Premium' to large asset bases (confirming the 'Too Big to Fail' bias), it identifies a critical trap in the Human Capital dimension. The dependence plot for Number of Employees reveals that for distressed firms, scaling the workforce toward the 50-employee threshold without corresponding profitability triggers a 'Complexity Trap'. Here, the rigidity of fixed labour costs acts as a multiplier for operational losses, confirming that while financial size acts as a shield, organizational weight acts as an anchor. It turns out that complexity is only an accelerator: when profitable it can give a bonus, but becomes a burden when margins are tight.
This polarity shift represents the conceptual core of the findings in this study. The model has decoded a phase transition where, below the zero-profitability threshold, tangible assets cease to be a guarantee and transform into structural rigidity. In this specific topology, high capital intensity exacerbates the speed of default by increasing fixed maintenance costs and reducing the agility required for retrenchment. The extraction of these insights was only possible through Methodological Validation using Accumulated Local Effects and Shapley Interaction Values.
Consequently, the crisis mechanism is not driven by the simple sum of high assets and low profitability, but by the multiplicative toxicity of their combination. This confirms that for the SME sector, the risk function is governed by thresholds where variables do not merely change in magnitude, but reverse their economic polarity. It all boils down to the fact that risk is not just about what you have, but how those different factors play out against one another in a crisis.

6.1.2 Mechanics of Cliff Effects

The empirical reconstruction of the Distressed Cohort contests the conventional depiction of corporate decline as a smooth, linear progression. The results indicate that the transition toward insolvency does not occur through a steady, predictable deterioration of financial ratios, but rather through structural jumps and abrupt disruptions. Essentially, a company does not just slowly fade away; it reaches a specific breaking point where the entire financial structure gives way all at once.
The model has identified specific critical thresholds, or cliffs, where the operational resilience of the firm effectively ceases to exist. This is most visibly manifested in the Scissor Effect detected by the ALE analysis, where the convergence of debt acceleration and revenue stagnation triggers a vertical spike in default probability rather than a proportional increase. In this scenario, the risk does not just grow; it explodes. Similarly, the Event Horizon map reveals a terminal zone in the

Solvency Ratio, specifically where the equity deficit exceeds five times the assets, where the risk function becomes inelastic to positive operational signals. Beyond this point, the firm enters a state of systemic collapse where conventional corrective measures no longer function. Once a company crosses this line, even good news or minor improvements in sales cannot pull it back from the brink of failure.
The identification of these 'Cliff Effects' validates the hypothesis that corporate failure is a discontinuous event based on critical thresholds, rather than a gradual linear decline. The model demonstrates that once specific breaking points are breached (e.g., Equity Deficit > 5x Assets), the firm enters a state of gravitational collapse where standard recovery levers become inelastic. The model recognizes that in the real world, there is a massive difference between being on the edge of a cliff and having already fallen off it, whereas older models tend to blur that distinction together.

6.1.3 Inertia as an Opportunity Cost

The Hybrid Target Variable allows for a fundamental redefinition of corporate failure, demonstrating that economic expiration often predates formal legal adjudication. By classifying Technical Failure, the persistence of negative equity without liquidation, as a terminal event, the model challenges the binary assumption that an active tax identity equates to operational viability. This perspective shifts the focus from courtroom outcomes to the firm's actual economic pulse, accounting for the institutional latency and evergreening that typically mask entities that are already *de facto* failed.

The empirical evidence dismantles the notion that a distressed firm can exist in a benign state of dormancy. The results from the Distress History Score analysis reveal that inertia is not a static condition but a kinetic process of accelerated erosion. For firms classified as Terminal Burners, time acts exclusively as a compounding factor of decay. In this state, the firm does not merely survive; it actively consumes working capital to service non-productive liabilities. In other words, insolvency functions as a metabolic deficit. An entity can sustain operations within a brief period of distress, but only if it has a rapid operational retrenchment. If the firm fails to pivot, it will exhaust the structural reserves necessary for credit viability restoration.

This leads to a counter-intuitive logical consequence on risk management, where the survival of a Terminal Burner does not represent a success of crisis management, but rather a failure of capital allocation. When a financial institution extends forbearance to such an entity, allowing it to persist as a Miraculous Survivor or False Positive, it is not preserving economic value; instead, it is formalizing a structural opportunity cost. The Schivardi Swap simulation demonstrates that the resources trapped in these inertial entities, which are estimated at over twelve billion euros in the sample, generate a negative return, whereas their reallocation to resilient firms would yield positive EBITDA. Put simply, keeping a failing company on life support takes away the oxygen that healthy companies need to grow. Therefore, the perceived success of keeping a zombie alive is, in macroeconomic terms, a net loss.

Finally, this framework provides the granulated technical apparatus necessary to measure the congestion effects theorized by Caballero, Hoshi, and Kashyap. By quantifying the exact volume of credit absorbed by Rate-Sensitive Zombies and Burners, the model moves beyond abstract theory to offer a precise accounting of the congestion tax imposed on the Italian SME ecosystem. It proves that the congestion is not a vague market friction, but a measurable sum of misallocated credit that actively suppresses the recovery of the resilient core. Bottom line is a clear trade-off: every euro tied up in a terminal case is a euro that cannot be used to drive an actual economic recovery.

6.2 Policy and Managerial Applications

Under the light of the extended temporal scope and its aggregate macroeconomic consequences, the prevailing policy interventions were structurally suboptimal from a purely banking and financial perspective. What were previously localized phenomena in credit risk, characterized by isolated and infrequent failures, suddenly reached a systemic proportion of the corporate population. This shift transformed credit risk from a niche concern into a broad market crisis. Therefore, drawing on the lessons learned from this experience, this study advocates for a sympathetic paradigm shift in the management of financial distress.
This new approach does not only apply to future crises but aims to overhaul the overarching management of financial distress as a whole. The financial system must set up a framework that moves beyond the reactive, blanket support seen in recent years to favor a more surgical, data-driven methodology. It all boils down to the fact that we can no longer afford to treat systemic distress with the same tools used for isolated defaults. By incorporating the non-linear insights provided by machine learning, regulators can rule out the broad inefficiencies of the past and build a more resilient path for credit allocation. This strategy ensures that when the next shock hits, the system is already prepared to distinguish between firms that are worth saving and those that are simply consuming capital without a future.

6.2.1 Towards a Protocol for Sanitization

The identification of close to thirteen billion euros in trapped capital, necessitates a shift from passive risk monitoring to proactive allocative intervention. With the counterfactual analysis of the Schivardi Swap, we elevate its use from a theoretical exercise, to a pragmatic protocol to dismantle the solvency illusion prevalent in the post-pandemic economic environment. Essentially, the goal is to stop claiming that dead capital is still alive and to start moving it to where it can actually grow.

The core simulation result demonstrates that banking inertia incurs a substantial economic penalty. Capital locked in terminal entities generates negative returns, effectively eroding 1.5 billion euros in annual EBITDA. This misallocation suggests that while a managed write-off is technically demanding, it remains a mathematically imperative strategy to liberate recoverable capital. If reallocated to the resilient cohort, this capital would generate a net economic delta of approximately two billion euros per year. Consequently, the write-off serves not as a destruction of value, but as the necessary entry fee to catalyse a GDP multiplier, converting stagnant stocks into productive flows. It comes down to the fact that absorbing a loss today is the only viable path to funding the growth of tomorrow.

Simulation results confirm that banking inertia erodes €1.5 billion in annual EBITDA. A managed write-off is mathematically essential to liberate this capital. Reallocating these funds to XGBoost-prioritized, high-recovery firms yields a €2 billion annual net delta, converting stagnant stocks into a productive GDP multiplier. Ultimately, absorbing immediate losses is the only viable path to funding future growth.

This restructuring operationalizes IFRS 9 by forcing the immediate recognition of Expected Credit Losses (ECL). While crystallizing an initial P&L impact, the Schivardi Swap eliminates the "capital inefficiency paradox" of holding Stage 2 assets with negative real returns. Reducing portfolio asset density liberates Tier 1 capital, restoring the monetary policy transmission channel and enabling banks to shift lending toward resilient, growth-oriented borrowers.

6.2.2 The Refinancing Cliff

The deterministic stress test at a normalized rate of 6.5%, reveals a fundamental structural disparity in the resilience of the Italian small-to-medium enterprise ecosystem. The simulation results confirm that the normalization of monetary policy does not impact the population uniformly.

Instead, it acts as a regressive filter that disproportionately targets young enterprises. In short, the same competitive interest rate hike that merely inconveniences a veteran company can completely dismantle a newer one.

Stress test results confirm the Liability of Newness: firms aged 0–5 years are the first to enter the rate-sensitive "zombie" cluster. Lacking the organizational slack and accumulated retained earnings of legacy firms (20+ years), these younger entities operate with thinner capital buffers. While mature firms benefit from the Lindy Effect, the solvency of younger firms remains structurally dependent on the low-interest environments of their inception.

Simulation results confirm that banking inertia erodes € 1.5 billion in annual EBITDA, rendering managed write-offs a mathematical imperative for capital liberation. By operationalizing IFRS 9 through the Schivardi Swap, banks can purge toxic Stage 2 assets to liberate Tier 1 capital, generating a € 2 billion annual net delta through reallocation to high-potential firms. This transition must, however, navigate the liability of newness; while legacy firms leverage the Lindy Effect, younger firms lack the organizational slack to withstand structural debt repricing. Within the ICAAP framework, failing to distinguish between operational viability and financial exposure risks a Refinancing Cliff that aborts the economy's replacement mechanism, sacrificing the growth of tomorrow to sustain the terminal burners of today.

Consequently, policy must transition from blanket guarantees toward maturity-calibrated interventions that distinguish "terminal burners" from rate-sensitive startups. Credit guarantees, such as the MCC Fund, should be recalibrated as strategic bridges to mitigate the liability of newness without subsidizing structural inertia. Avoiding a wholesale retreat into safety, which favours legacy reserves while calcifying market entry, is essential to prevent "zombie" preservation at the expense of future champions. Ultimately, the regulatory choice lies between sustaining a stagnant status quo and capitalizing the next cycle of industrial growth.

6.2.3 On Cost-Sensitivity

The final implication of this study is the proposal to move beyond the static classification paradigm that currently dominates credit risk. The empirical comparison between the Academic Baseline, the Risk-Averse Baseline, and the Turnaround Target demonstrates that a fixed decision boundary is dangerously inefficient in a volatile macroeconomic environment. In other words, a rule that works when money is cheap might be disastrous when interest rates rise. Setting a single, unmoving target for risk is like trying to sail a ship without ever adjusting the sails for the wind.

The application of a Point-in-Time (PiT) XGBoost architecture reveals the inherent lag of traditional Through-The-Cycle (TTC) rating models, which often smooth out volatility at the cost of risk accuracy. This disconnect becomes critical in a high-rate environment (above 4%), where a false positive, financing a "zombie" firm, represents a direct destruction of expensive Tier 1 capital rather than a mere accounting error. Conversely, in low-rate regimes, the opportunity cost of a false negative increases as banks struggle for yield. Ultimately, the cost of predictive error is dynamic and regime-dependent, necessitating a transition toward adaptive, machine-learning-driven models that can calibrate risk in real-time.

Therefore, we propose replacing static acceptance cut-offs with a dynamic macro-prudential threshold, effectively operationalizing the Risk Appetite Framework (RAF) as an algorithmic lever. Rather than adhering to a rigid probability of default (PD) mark, risk managers should utilize a threshold that minimizes total expected loss relative to the current cost of funds and Loss Given Default (LGD). This approach allows the algorithm to recalibrate the acceptance boundary: tolerating higher false discovery rates to capture marginal growth during expansion, while automatically tightening to prioritize capital preservation during contractionary, high-rate phases. By shifting

from static to adaptive boundaries, the bank ensures its posture is synchronized with the prevailing macroeconomic cycle.

These results advocate for a shift from point-estimate precision toward the minimization of total expected loss. As demonstrated, a model may be globally less accurate in statistical terms yet economically superior by aggressively filtering "silent killers". For risk managers, this validates that a model's worth lies in its ability to preserve capital during downturns rather than its fit to a mathematical curve. Adopting cost-sensitive learning serves as a dynamic response protocol, ensuring that credit allocation logic adapts to the fluctuating price of risk. Ultimately, the optimal model is defined not by its statistical score, but by its capacity to protect the institution's solvency.

Efficiency in a rehabilitative framework depends on the model's ability to distinguish between salvageable entities and those beyond structural recovery. By deliberately excluding firms where operational retrenchment is no longer feasible, resources can be concentrated on firms with genuine turnaround potential, recognizing that support for those who can grow is impossible if consumed by those already lost. This represents a paradigm shift from the broad-based support of the "economic dormancy" era toward an accelerating philosophy designed to cultivate productive, efficacious firms.

The objective is to move beyond "life support" toward a robust market structure capable of withstanding global competitive pressures. By prioritizing the resilient core over the artificial preservation of terminal inertia, the Italian SME ecosystem can transition from mere survival to international competitiveness. Ultimately, protecting inefficiency weakens the collective; a market's strength is defined by the health of its most efficient players.

6.3 Research Limitations

Despite the breadth of the dataset and the robust predictive power of the model, several critical constraints warrant a thorough examination. Even though the model has proven its value in identifying failure patterns, it is essential to point out where its boundaries lie. This section aims to peel back the layers of the analysis to reveal the factors that could influence the results in different contexts, ensuring that the findings are not taken as a universal truth without considering their specific environment.

The Regime Shift

The most significant limitation acknowledged by this study is that the XGBoost model is effectively constrained by its temporal training distribution. The model was developed using data from the 2019–2022 period, a timeframe characterized by extensive institutional intervention. Because the model learned during an abnormal time, its baseline for what a healthy company looks like is somewhat skewed. This internal constraint led to liquidity-induced distortions, often referred to as a solvency illusion. Many firms classified as safe within the training set were surviving solely due to subsidized credit costs, which masked their underlying operational inefficiency. Therefore, the model risks undervaluing the risk profile of interest-rate-contingent entities because the training data did not contain the high-interest-rate dynamics that materialized in the years immediately following. This limitation reflects the weaning-off effect. While this study attempts to mitigate this gap through the Refinancing Cliff stress test, it must be admitted that the core predictive engine cannot extrapolate structural shifts that are absent from its training distribution. This creates a ceiling on the model's predictive accuracy when faced with a regime change where the cost of capital suddenly becomes a primary driver of default.

The Qualitative Blind Spot

A significant limitation of this study is its dependence on annualized fiscal reporting, which inherently creates a recognition gap due to significant reporting lags. While the internal work successfully diagnoses structural insolvency or zombie stagnation, it lacks the high-frequency telemetry typically available to commercial banks, such as the credit behavioural triggers found in the Central Credit Register or real-time payment delays for VAT and social security. Essentially, by the time a problem shows up in the annual report, the company might have already been struggling for months. The model can see that the ship is sinking, but it cannot see the exact moment the leak started. Evidence from Moscatelli et al. (2020) show that integrating these behavioural indicators enhances the model's discriminative capacity, potentially increasing the AUC-ROC by ten percentage points. These signals provide a window into the firm's immediate financial flexibility, which is something annual statements simply cannot capture. This hypothesis posits that detecting high-frequency irregularities is the only way to anticipate the cliff, which is the abrupt liquidity exhaustion that occurs between reporting periods. The current model is excellent for identifying long-term structural decay, but it cannot yet act as a daily alarm system for sudden cash-flow failures.

A Fidelity Gap

While this study employs SHAP to provide transparency, it acknowledges a significant theoretical constraint regarding the accuracy of these explanations. The analysis admits that post-hoc explanations, such as those generated by Shapley values, are essentially approximative reconstructions rather than direct windows into the internal logic of the model. In other words, just because the model gives us an explanation for why it made a choice does not mean that explanation perfectly captures every nuance of its complex internal calculations. This limitation is grounded in the work of Rudin, who argues that for high-stakes financial decisions, post-hoc explainability is functionally secondary to intrinsic interpretability. Consequently, there is always a risk that the explanation simplifies the reality of the risk too much, potentially leading a risk manager to overlook a subtle but critical combination of factors that the XGBoost model actually detected. The adoption of a shallow learner protocol aids in mitigating the risk of redundancy. It all boils down to the fact that while we can see the general reasons behind a model's decision, we must remain aware that the map provided by the explanation is not the same thing as the actual terrain of the model's logic.

Survivor Bias of the Red Zone

The specific design of this study, which filters exclusively for firms with negative equity, incorporates an inherent survivorship bias into the analysis. The dataset includes only those firms that demonstrated sufficient operational endurance to complete the fiscal reporting cycle and file 2022 financial statements. This means that any business that entered immediate liquidation or collapsed during the peak of the pandemic before filing is naturally excluded. Consequently, the algorithm is constrained to a cohort of survivor-biased entities, which could lead to upwardly biasing the projected recovery potential of the general distressed population. This limitation aligns with the warnings of Balcaen and Ooghe (2006), who argue that sample selection criteria in bankruptcy prediction often create artificial populations. These filtered groups do not perfectly mirror the stochastic flux of corporate demography, where companies enter and exit the market in unpredictable ways. By focusing only on the red zone of negative equity survivors, the study provides a high-resolution view of a specific type of distress, but it must be acknowledged that this view is not a universal map of all business failures. It effectively tells us why some survivors might eventually fail, but it cannot tell us about the ones that were gone before the data could even be recorded.

The empirical evidence within this research suggests that while the model functions as a resilient diagnostic instrument for structural triage, its application requires the explicit awareness that it was developed within an artificially stabilized economic environment. The model is effectively an

expert in the rules of a game that was played under high levels of state protection and liquidity. Consequently, its predictions should be interpreted as evaluations of structural financial mechanics, specifically focusing on the long-term interaction between solvency and asset heaviness, rather than as predictive indicators of near-term liquidity shocks.

6.4 Future Developments

The preceding analysis has established a robust framework for identifying the structural mechanisms of corporate distress through the lens of non-linear machine learning. However, as with any empirical inquiry into complex systems, the validity of these findings is intrinsically tied to the temporal and structural boundaries of the data utilized. Recognizing these boundaries is not a mere academic formality; it is a critical step in transitioning the "Schivardi Swap" and the "Refinancing Cliff" from theoretical simulations into operational protocols.
As the economic landscape shifts from the subsidized dormancy of the pandemic era toward a more volatile, high-interest-rate environment, the tools used to measure risk must undergo a parallel evolution. This final chapter outlines the critical constraints of the current research, while simultaneously charting the trajectories for the next generation of credit models. By acknowledging where the current map ends, we can define the technical and methodological innovations required to navigate the future of the Italian SME ecosystem with greater precision and resilience.

6.4.1 Beyond the Pandemic Timeframe

The primary direction of development of the framework would be the overcoming of the temporal dependency. Effectively, the architecture applied behaves well within its specific time-period, but there's no concrete guarantee of its performance with other timeframes.

Validating the Refinancing Cliff

Future developments will focus on integrating 2023–2024 fiscal data to transition the model from deterministic simulation to empirical verification. By replacing the theoretical "Refinancing Cliff" stress test with realized default data, the XGBoost algorithm can treat interest rate sensitivity as an endogenous feature rather than an external factor. This allows the model to bridge the data-regime gap by learning the actual non-linear boundaries where debt servicing costs trigger default.

Different Temporal Window

A more radical approach would be the execution of cross-temporal validation across heterogeneous economic cycles. By transposing the architecture onto distinct temporal windows, such as the 2005–2011 global financial crisis or the 2014–2019 expansionary period, the model undergoes a rigorous robustness benchmarking. This stress test is designed to verify the invariance of predictive signals, ensuring that the internal logic remains adaptive regardless of the prevailing macro-regime. Such transposition will demonstrate the model's unique ability to decouple firm-specific failure patterns from broader market volatility, further refining the proprietary parameters that define its predictive stability.

Macro-Economic Normalization

To mitigate temporal overfitting, future iterations will integrate exogenous macroeconomic covariates directly into the feature space, replacing indirect "scissor effect" proxies. Incorporating variables such as the Euribor 3M spot rate, sectoral inflation indices, and energy cost fluctuations will transition the model toward a regime-aware architecture. By training on a dataset spanning both zero-interest and tightening cycles, the algorithm can dynamically weight solvency ratios and EBITDA margins relative to the prevailing cost of capital. This evolves the architecture from a point-

in-time classifier into a through-the-cycle system capable of autonomously calibrating acceptance boundaries during monetary contractions.

Event-centric Panel Dataset

A significant evolution of this research involves transitioning from static binary classification to a multidimensional longitudinal architecture. By extending the predictive horizon to five years, the system shifts focus from immediate insolvency to latent failure detection, specifically identifying "zombie" entities that remain operational despite terminal economic non-viability. This approach redefines corporate decline as a progressive trajectory rather than an isolated event. To facilitate this, the data architecture will utilize an event-based panel. Data for insolvent firms is aligned relative to the insolvency year (t_0) using a t_{-5} to t_{-1} window. For healthy firms, the framework generates a synthetic t_0 pseudo-event to serve as a comparative anchor. This temporal alignment enables the algorithm to effectively distinguish standard operational fluctuations from the early-stage signals of terminal distress.

6.4.2 Qualitative Data

Further direction of development is over the data included in the dataset.

High Frequency Behavioural Data

To substantiate the marginal input hypothesis, future iterations will integrate centralized credit-registry telemetry. Following the findings of Moscatelli et al. (2020), augmenting static financial ratios with monthly behavioural indicators can increase discriminative capacity by approximately ten percentage points. This transition effectively bridges the gap between structural insolvency and effective default. The model will ingest specific covariates indicative of instantaneous liquidity strain. This involves monitoring overdraft utilization as a primary metric for liquidity depletion, specifically looking for credit line usage beyond granted limits. Additionally, the system will track past-due flow dynamics by monitoring the monthly transition from performing to past-due status, which serves as a leading indicator of impending solvency crises. Finally, the architecture will account for banking multiplicity, as an increase in bank relationships often signals credit shopping behaviour where distressed entities seek emergency funding to cover existing gaps. By prioritizing high-frequency behavioural data over lagging fiscal reports, the system can identify the early warning signs of distress significantly before they appear in statutory filings.

Temporal Aware Architecture

To manage the temporal dependence of longitudinal data streams, future research will explore augmenting the XGBoost framework through ensemble integration with Temporal-Balanced Random Forest (T-BRF) methodologies. This structural advancement incorporates a temporal decay function that prioritizes data recency, assigning significantly higher weights to proximal behavioural triggers relative to distal, static structural data. By shifting the model's focus toward the most recent fiscal actions, the system can identify urgent signals of distress that are frequently obscured by lagging balance sheet indicators. Adopting this temporal-aware logic ensures the system reacts dynamically to immediate financial mechanics rather than being constrained by the inertia of historical data. This capability is vital in volatile economic contexts where a firm's historical solvency may appear stable while its current cash-flow patterns indicate a rapid approach to a liquidity cliff. Integrating T-BRF allows the model to filter out the noise of outdated performance and focus on the current operational reality, transforming the past into a reference point rather than a predictive anchor.

6.5 On Interpretability

Rashmon Set

A critical trajectory for model evolution involves challenging the perceived trade-off between predictive power and transparency. Grounded in Rudin's hypothesis, this research will investigate the existence of a Rashomon Set, a collection of functionally equivalent models where diverse architectures achieve comparable accuracy on the same structured dataset. The proposed development involves rigorously testing whether a fully interpretable glass-box model, such as a Generalized Additive Model (GAM) or a sparse decision tree, exists within the Rashomon Set of the Distressed Cohort. By benchmarking these architectures against the current XGBoost model, the study aims to quantify the fidelity gap.

Knowledge Distillation

If the XGBoost architecture is empirically necessitated by the non-linearities of the "Event Horizon," development will shift toward knowledge distillation. This approach addresses the inherent limitations of SHAP, which provides local feature attribution rather than a comprehensive functional mapping of the model's logic. The proposed strategy involves constructing a global surrogate model, such as a single deep decision tree or a sparse rule set, trained specifically on the probability predictions generated by the teacher XGBoost model. This technique maps the high-dimensional decision boundary into a human-readable format, providing risk managers with a simplified rule-based logic that can be audited for consistency against SHAP outputs.

6.5.1 Mitigating the Survivorship Bias

The final and perhaps most critical research trajectory involves a systematic effort to counteract the inherent optimism bias present in the current survivor-biased sample. Future iterations must incorporate the graveyard dataset, to provide a balanced view of corporate distress.

Pre-Filing Defaults

To mitigate the current temporal anchor, future research should transition from a static observation point at T_0 to a rolling window approach spanning 2019–2024. This methodological shift allows the model to observe firms that entered the red zone of negative equity in 2020 but liquidated in 2021, thus never appearing in the 2022 survivor-biased dataset. Reintroducing this fragile population into the training set would likely lower the model's overall predicted probability of recovery. Essentially, by looking at the companies that did not make it through the first shock, the model can establish a more realistic and grounded baseline for distress turnaround. It turns out that a doctor cannot learn the full extent of a disease if they only study patients who survived the intensive care unit.

Sudden Death Variables and Two-Stage Filtering

To capture acute failure alongside chronic stagnation, future iterations will integrate covariates associated with rapid corporate collapse, such as high-frequency director turnover and intra-year legal filings. Following Wilson and Altanlar (2014), these behavioural indicators serve as precursors to immediate liquidation, capturing distress signals that occur between annual reporting cycles. To formalize this, the research proposes a two-stage filtering framework. The first stage utilizes a selection model to predict the probability of a firm surviving long enough to file a balance sheet, thereby quantifying latent survivor bias. In the second stage, the XGBoost model is applied using weights derived from the inverse probability of the first stage, functioning similarly to a Heckman correction. This statistical adjustment mathematically recalibrates the model by accounting for the "invisible" cohort of firms that exit the market before statutory deadlines. By correcting for this bias,

the framework ensures a more conservative and realistic assessment of insolvency risk, particularly in high-volatility sectors.

From Financial Physics to Ecological Modeling

Finally, future studies must move beyond individual balance sheet analysis toward ecological modelling that accounts for the stochastic flux of corporate demography. By including sectoral churn rates as an exogenous feature, the model can learn that the environment is just as important as the individual firm's health. For instance, a firm in a high-churn sector like hospitality has a much lower survival probability than a firm with identical financial ratios in a low-churn sector like manufacturing. Adding this environmental hazard layer ensures that the diagnostic process is not just a calculation of structural solvency, but a holistic view of the company's ecosystem. It all boils down to the fact that survival is not just about how strong you are, but how dangerous the neighbourhood is where you operate.

6.6 In Conclusion

The journey from the static, linear frameworks of the past to the dynamic, non-linear architectures of the future represent more than just a technical upgrade; it is a fundamental shift in how we conceptualize economic stability. This research has demonstrated that corporate failure is not a predictable slide down a smooth curve, but a series of catastrophic discontinuities and threshold effects that only high-capacity models like XGBoost can adequately map. By identifying the specific "cliffs" where financial physics break down, we have moved beyond simple prediction toward a genuine diagnostic understanding of the SME ecosystem.

The "Schivardi Swap" and the "Refinancing Cliff" are not merely academic simulations; they are urgent calls for a proactive approach to capital allocation. In an environment where the "Zombie" state is no longer a localized anomaly but a systemic risk, the cost of inaction is a direct tax on future growth. The findings suggest that by leveraging explainable AI, financial institutions can finally bridge the gap between regulatory caution and economic efficiency, ensuring that credit flows not toward those with the most "historical fat," but toward those with the highest potential for revitalization.

Ultimately, this study serves as a blueprint for the "Metabolic" management of financial distress. By recognizing the data-regime gaps and the inherent biases of our current models, we have defined the exact coordinates for the next stage of risk engineering. As we integrate real-time behavioural telemetry and transition toward glass-box architectures, the goal remains the same: to create a financial system that is robust enough to withstand global shocks and precise enough to foster the next generation of industrial champions. The future of credit risk lies in our ability to adapt our logic as fast as the market reverses its polarity, ensuring that resilience is no longer a matter of chance, but a product of design.

Appendix A: Appendix

A.1 Target Definition

The following code illustrates the logic used to construct the binary dependent variable (**default_label**). The classification employs a hybrid approach that integrates formal legal status with the firm's economic substance, following three hierarchical criteria: - **Noise Filter:** Preliminary exclusion of non-relevant administrative events (e.g., registered office transfers or duplications). - **Legal Mortality:** Identification of firms in insolvency or liquidation proceedings through keyword matching (assigned **Target = 1**). - **Economic Verification ("Zombie" Detection):** For firms legally classified as "Active" or "Under Restructuring," the latest available Total Equity is verified. - If $Equity > 0$: The turnaround is considered effective (**Target = 0**). - If $Equity \leq 0$: The firm is classified as an economic failure or "Zombie" (**Target = 1**).

```
# STATUS DEFINITION
# Failure Cases (Legal Mortality)
status_legal_failure <- c(
  "Fallimento", "Liquidazione giudiziale", "Liquidazione giudiziaria",
  "Liquidazione coatta amministrativa", "Concordato fallimentare",
  "Stato di insolvenza", "Liquidazione controllata",
  "Chiusura del fallimento", "Chiusura per fallimento o liquidazione",
  "Chiusura per liquidazione", "Chiusura della liquidazione",
  "Scioglimento per atto dell'autorità", "Decreto cancellazione tribunale",
  "Cancellazione dal registro delle imprese",
  "Cancellata d'ufficio ai sensi art. 2490 c.c.",
  "Liquidazione volontaria", "Scioglimento e liquidazione",
  "Scioglimento", "Liquidazione", "In liquidazione"
)

# Success Candidates (Legal Life / Mergers)
status_active_candidate <- c(
  "Attiva", "Fusione mediante incorporazione in altra società",
  "Fusione mediante costituzione di nuova società", "Scissione",
  "Trasformazione di natura giuridica", "Revoca del fallimento",
  "Accordi di ristrutturazione dei debiti", "Piano di ristrutturazione omologato
(CCI)",
  "Concordato preventivo", "Concordato preventivo omologato",
  "Concordato semplificato", "Concordato minore",
  "Misure cautelari e protettive", "Procedimento unitario",
  "Amministrazione straordinaria"
)

# Noise Removal
status_noise <- c("Trasferimento in altra provincia", "Trasferimento sede all'estero",
"Duplicazione", "Chiusura dell'unità locale")

# TARGET CONSTRUCTION LOGIC
target_var <- function(df) {
  n_start <- nrow(df)
  # Use both status and procedure fields
  df[, combined_status := paste(Legal_Status, ifelse(is.na(Cessation_Procedure), "",
Cessation_Procedure))]
  # Filter Noise
  noise_pattern <- paste(status_noise, collapse = "|")
  df <- df[!grepl(noise_pattern, combined_status, ignore.case = TRUE)]
```

```r
  print(paste("Removed Noise Rows:", n_start - nrow(df)))

  # Financial Reality Check
  # Retrieve latest available Equity (2024 preferred, fallback 2023)
  col_24 <- "Total_Net_Equity_kEUR_2024"
  col_23 <- "Total_Net_Equity_kEUR_2023"

  if (col_24 %in% names(df) & col_23 %in% names(df)) {
    df[, last_available_equity := fcoalesce(get(col_24), get(col_23))]
  } else if (col_23 %in% names(df)) {
    df[, last_available_equity := get(col_23)]
  } else {
    df[, last_available_equity := NA_real_]
  }

  # Final Labeling Logic
  # 0 = Turnaround Success ( Acive/Meged + Positive Equity)
  # 1 = Turnaround failure (Legal Bankruptcy OR Economic Insolvency)
  fail_pattern <- paste(status_technical_failure, collapse = "|")
  active_pattern <- paste(status_active_candidate, collapse = "|")

  df[, default_label := fcase(
    # Priority 1: Verified Success (Positive Equity + Active/Merged)
    grepl(active_pattern, combined_status, ignore.case = TRUE) & last_available_equity
> 0, 0,
    # Priority 2: Legal Failure
    grepl(fail_pattern, combined_status, ignore.case = TRUE), 1,
    # Priority 3: Technical Failure (Zombies)
    grepl(active_pattern, combined_status, ignore.case = TRUE) & last_available_equity
<= 0, 1,
    # Fallback
    is.na(combined_status) & last_available_equity <= 0, 1,
    is.na(combined_status) & last_available_equity > 0, 0,
    default = NA_real_
  )]

  # CLEANUP
  # Removing observations with insufficient data to determine label
  n_drop <- nrow(df[is.na(default_label)])
  df <- df[!is.na(default_label)]
  print(paste("Dropped rows with insufficient data:", n_drop))

  df[, default_label := as.integer(default_label)]
  df[, `:=`(combined_status = NULL, last_available_equity = NULL)]

  return(df)
}

# Operationalization
DF <- target_var(DF)

sample_tracker$Stage_3_Clean <- nrow(DF)
```

A.2 Dynamic Sector Specific Imputation

The following algorithm handles missing financial data through a **sector-specific ratio imputation** method designed to preserve the firm's economic structure. Variables are first classified into **Stock** (Balance Sheet) or **Flow** (Income Statement) items. Missing values are then estimated by anchoring them to a relevant driver (*Total Assets* for Stock, *Sales Revenues* for Flow) based on the sector's median ratio:

$$\hat{X}_{i,s} = Driver_i \times \text{Median} \left(\frac{X}{Driver} \right)_s \tag{20}$$

Where s represents the 2-digit NACE/ATECO sector. If a sector has insufficient observations ($N < 10$), the global median is used as a fallback. Finally, variables with a post-imputation missingness rate exceeding 20% are discarded, and the dataset is filtered to retain only the **Distressed Cohort** (firms with Negative Equity in the target year, $t = 2022$).

```
# MISSING DATA IMPUTATION AND COHORT SELECTION

# Define driver variables for ratio calculation
# - Assets drive Stock variables (e.g., Debt, Equity)
# - Sales drive Flow variables (e.g., EBITDA, Costs)
drivers <- list(
  stock = "Total_Assets_kEUR",
  flow  = "Sales_Revenues_kEUR"
)

# Identify Patrimonial (Stock) keywords
keywords_stock <- c(
  "assets", "equity", "debt", "liabilities", "inventory",
  "receivables", "payables", "cash", "funds", "reserve", "nfp"
)

# Identify P&L (Flow) keywords
keywords_flow <- c( "sales", "revenues", "turnover", "profit", "loss", "income",
"ebitda", "ebit", "costs", "expenses", "amortization", "depreciation", "tax",
"interests", "production_value", "added_value"
)

impute_via_sector_ratio <- function(df, target_col, driver_col, group_col, year) {

  # DEFINE TRAINING SET (COMPLETE CASES)
  # Must have valid target, valid driver (>0), and exist in the reference year
  valid_rows <- !is.na(df[[target_col]]) &
                !is.na(df[[driver_col]]) &
                df[[driver_col]] != 0 &
                df$incorporation_year <= year

  complete_data <- df[valid_rows, ]

  # Minimum threshold for statistical reliability
  if (nrow(complete_data) < 10) return(df)

  # CALCULATE RATIOS (TARGET / DRIVER)
  # Global Median (Fallback)
  global_ratio <- median(complete_data[[target_col]] / complete_data[[driver_col]],
na.rm = TRUE)
```

```
  # Sector-Specific Medians
  sector_stats <- complete_data[, .(
    Median_Ratio = median(get(target_col) / get(driver_col), na.rm = TRUE),
    N = .N
  ), by = group_col]

  # Filter sectors with insufficient data points
  sector_stats <- sector_stats[N >= 10]

  # APPLY IMPUTATION
  # Target rows: Missing target, valid driver, active company
  to_impute <- is.na(df[[target_col]]) &
               df$incorporation_year <= year &
               !is.na(df[[driver_col]])

  if (sum(to_impute) == 0) return(df)

  df[to_impute, (target_col) := {
    # Match sector median
    idx <- match(get(group_col), sector_stats[[group_col]])
    ratio <- sector_stats$Median_Ratio[idx]

    # Apply Fallback if sector median is missing
    ratio[is.na(ratio)] <- global_ratio

    # Calculate imputed value
    val <- get(driver_col) * ratio

    # Integer rounding for employees
    if (grepl("employees", target_col, ignore.case = TRUE)) val <- pmax(1, round(val,
0))
    val
  }]

  return(df)
}

# IMPUTATION ROUTINE
years_to_process <- c(2019, 2020, 2021, 2022)

for (yr in years_to_process) {

  # Define dynamic column names for the year
  col_driver_flow  <- paste0(drivers$flow, "_", yr)
  col_driver_stock <- paste0(drivers$stock, "_", yr)

  if (!col_driver_flow %in% names(DF) || !col_driver_stock %in% names(DF)) next

  # Identify variables to impute
  cols_in_year <- grep(paste0("_", yr, "$"), names(DF), value = TRUE)
  exclude_cols <- c(col_driver_flow, col_driver_stock, "tax_id", "ateco_code",
                    "incorporation_year", "sector_group")

  cols_numeric <- setdiff(cols_in_year, exclude_cols)
  cols_numeric <- cols_numeric[sapply(DF[, cols_numeric, with = FALSE], is.numeric)]
```

```
  # Split into Stock vs Flow
  cols_stock <- grep(paste(keywords_stock, collapse = "|"), cols_numeric, value = TRUE,
ignore.case = TRUE)
  cols_flow <- grep(paste(keywords_flow, collapse = "|"), cols_numeric, value = TRUE,
ignore.case = TRUE)

  # Safety check: Report any variable that was NOT classified
  cols_unclassified <- setdiff(cols_numeric, c(cols_stock, cols_flow))
  if(length(cols_unclassified) > 0) {
      warning(paste("Unclassified numeric variables found (skipped):",
      paste(cols_unclassified, collapse=", ")))
  }


  # Apply Imputation
  # Stock variables driven by Total Assets
  for (col in cols_stock) {
    DF <- impute_via_sector_ratio(DF, col, col_driver_stock, "sector_group", yr)
  }

  # Flow variables driven by Sales Revenues
  for (col in cols_flow) {
    DF <- impute_via_sector_ratio(DF, col, col_driver_flow, "sector_group", yr)
  }
}
```

A.3 Feature Engineering

The predictor set is enriched through the construction of synthetic ratios covering *Operational Momentum*, *Liquidity*, and *Financial Structure*. Key methodological adjustments include: 1. **Robust Arithmetic:** Implementation of `safe_ratio` to prevent division-by-zero errors in financial ratios. 2. **Ohlson Correction:** For firms with negative equity (*OENEG=1*), the Leverage Ratio is capped at a maximum risk value (50x) to maintain ordinal consistency, while a normalized *Solvency Ratio* captures the severity of the deficit. 3. **Dimensional Controls:** Raw scale variables (`Sales_Volume`, `Num_Employees`, `Total_Size`) are explicitly retained to control for firm size effects in the multivariate analysis.

```
# FEATURE ENGINEERING AND RATIO CONSTRUCTION

# HELPER FUNCTIONS AND ROBUST ARITHMETICS
# These functions prevent division-by-zero errors and handle infinite values.

safe_ratio <- function(num, den) {
  res <- num / den
  res[is.na(res) | is.infinite(res)] <- 0
  return(res)
}

safe_growth <- function(curr, prev) {
  res <- (curr - prev) / abs(prev)
  res[is.na(res) | is.infinite(res)] <- 0
  return(res)
}
```

```r
# OHLSON CORRECTION AND DISTRESSED FLAG

# OENEG (Negative Equity Dummy)
# Binary indicator: 1 if Shareholders' Equity is negative (Insolvency), else 0.
DF[, OENEG := ifelse(!is.na(Shareholders_Equity_kEUR_2022) &
                     Shareholders_Equity_kEUR_2022 < 0, 1, 0)]

# Leverage Adjustment for Distressed Firms
# When Equity < 0, standard Leverage (Debt/Equity) loses meaning.
# We impute a ceiling value (50x) to represent "Maximum Financial Risk".
if ("Leverage_Ratio_2022" %in% names(DF)) {
  DF[OENEG == 1, Leverage_Ratio_2022 := 50]
}

# Solvency Ratio (Severity of Deficit)
# Ratio: Equity / Total Assets.
# Capped at -5 to limit the influence of extreme outliers (deficit > 5x assets).
DF[, Solvency_Ratio := safe_ratio(Shareholders_Equity_kEUR_2022,
Total_Assets_kEUR_2022)]
DF[Solvency_Ratio < -5, Solvency_Ratio := -5]

# Historical Distress Score (Gradient)
# Sum of years with Negative Equity (2019-2022). Range: 0 (Healthy) to 4 (Chronic).
DF[, distress_history_score := rowSums(cbind(
    DF$Shareholders_Equity_kEUR_2019 < 0,
    DF$Shareholders_Equity_kEUR_2020 < 0,
    DF$Shareholders_Equity_kEUR_2021 < 0,
    DF$Shareholders_Equity_kEUR_2022 < 0
  ), na.rm = TRUE)]


# 3. ADVANCED FEATURE GENERATION

DF[, `:=`(
  # MOMENTUM & PERFORMANCE INDICATORS
  sales_growth_1y      = safe_growth(Sales_Revenues_kEUR_2022,
Sales_Revenues_kEUR_2021),
  sales_growth_2y      = safe_growth(Sales_Revenues_kEUR_2022,
Sales_Revenues_kEUR_2020),

  delta_ebitda_margin  = safe_ratio(EBITDA_kEUR_2022, Sales_Revenues_kEUR_2022) -
                         safe_ratio(EBITDA_kEUR_2021, Sales_Revenues_kEUR_2021),

  delta_asset_turnover = safe_ratio(Sales_Revenues_kEUR_2022, Total_Assets_kEUR_2022) -
                         safe_ratio(Sales_Revenues_kEUR_2021, Total_Assets_kEUR_2021),

  # LIQUIDITY & WORKING CAPITAL
  # "Scissor Effect": Growth of Payables vs. Growth of Sales
  scissor_debt_sales   = safe_growth(Total_Payables_kEUR_2022,
Total_Payables_kEUR_2021) -
                         safe_growth(Sales_Revenues_kEUR_2022,
Sales_Revenues_kEUR_2021),

  # Operating Cash Flow Proxy (normalized by Assets)
  cash_flow_proxy      = safe_ratio(
                           (Cash_Equivalents_kEUR_2022 - Cash_Equivalents_kEUR_2021),
```

```
                          (Total_Assets_kEUR_2022 + 1)
                        ),

  # Trade Financing Stress: Change in Payables / Revenues
  trade_financing_stress = safe_ratio(
                            (Total_Payables_kEUR_2022 - Total_Payables_kEUR_2021),
                            Sales_Revenues_kEUR_2022
                          ),

  # STRUCTURAL & PROFITABILITY DRIVERS
  # Cash Flow Generation: Net Income + D&A (Proxy: EBITDA - EBIT)
  CF_Generative        = Net_Income_kEUR_2022 + (EBITDA_kEUR_2022 - EBIT_kEUR_2022),

  # Gearing: Total Debt / Total Assets
  Gearing_Ratio        = safe_ratio(Total_Payables_kEUR_2022, Total_Assets_kEUR_2022),

  # Return on Equity (ROE)
  ROE                  = safe_ratio(Net_Income_kEUR_2022,
Shareholders_Equity_kEUR_2022),

  # Return on Capital Employed (ROCE) Proxy
  # Denominator: Total Assets - Short Term Debt (~ Capital Employed)
  ROCE                 = safe_ratio(EBIT_kEUR_2022, (Total_Assets_kEUR_2022 -
Short_Term_Debts_kEUR_2022)),

  # Profit Margin
  Profit_Margin        = safe_ratio(Net_Income_kEUR_2022, Sales_Revenues_kEUR_2022),

  # SCALE & DIMENSIONAL CONTROLS
  # Explicitly retained for model specifications
  Num_Employees        = Employees_2022,
  Sales_Volume         = Sales_Revenues_kEUR_2022,
  Total_Size           = Total_Assets_kEUR_2022
)]


# INTEREST COVERAGE RATIO ESTIMATION
# Calculate absolute interest expenses from turnover ratio, then compute ICR.
DF[, Interest_Expenses_Est := Sales_Revenues_kEUR_2022 *
(Financial_Charges_to_Turnover_2022 / 100)]

DF[, ICR_2022 := ifelse(Interest_Expenses_Est > 0,
                        EBITDA_kEUR_2022 / Interest_Expenses_Est,
                        100)] # Cap at 100 for firms with zero interest expenses


# FINAL SANITIZATION
# Replace any remaining Infinite or NA values with 0 to ensure model stability.
cols_to_sanitize <- c(
  "sales_growth_1y", "sales_growth_2y", "delta_ebitda_margin",
  "scissor_debt_sales", "Gearing_Ratio", "ROE", "ROCE", "Solvency_Ratio",
  "Num_Employees", "Sales_Volume", "Total_Size"
)

for (col in cols_to_sanitize) {
  if (col %in% names(DF)) {
    set(DF, i = which(!is.finite(DF[[col]])), j = col, value = 0)
```

```
  }
}
```

A.4 Firewall

The feature selection process applies a strict exclusion protocol to ensure model validity: 1. **Temporal Leakage Filter:** Removal of all variables containing future information (post-2022) or procedural status labels that would reveal the target outcome. 2. **Identifier Removal:** Exclusion of non-predictive metadata (Tax IDs, Company Names) to prevent overfitting. 3. **Multicollinearity Control:** Raw *Shareholders' Equity* and standard *Leverage* are removed to avoid collinearity with the engineered *Solvency Ratio* and *Ohlson Correction* terms.

```
# FEATURE SELECTION AND LEAKAGE PREVENTION

# DEFINE VARIABLE CATEGORIES
# ENGINEERED PREDICTORS
engineered_features <- c(
  # Momentum & Performance
  "sales_growth_1y", "sales_growth_2y", "delta_ebitda_margin",
  "delta_asset_turnover", "recovery_sales_vs_2019",

  # Liquidity & Working Capital
  "scissor_debt_sales", "cash_flow_proxy", "gap_receivables_sales",
  "trade_financing_stress", "delta_cost_weight", "delta_debt_maturity",

  # Structural & Distress Indicators
  "OENEG", "Solvency_Ratio", "distress_history_score",
  "CF_Generative", "Gearing_Ratio", "Profit_Margin", "ROCE", "ROE",

  # Dimensional Controls
  "Num_Employees", "Sales_Volume", "Total_Size"
)

# EXCLUSION LIST 1: TEMPORAL LEAKAGE & METADATA
# Keywords that imply future knowledge or non-predictive IDs
patterns_leakage <- c(
  "2023", "2024",             # Future financial years
  "Procedure", "Cessation",   # Target-related info
  "Legal_Status",             # Target-related info
  "Tax_ID", "Company_Name",   # Identifiers
  "ATECO_Code"                # Granular Code (Sector Group used instead)
)

# EXCLUSION LIST 2: REDUNDANT / BIASED VARIABLES
# Variables removed to prevent multicollinearity with engineered features
vars_redundant <- c(
  "Shareholders_Equity_kEUR_2022",  # Replaced by Solvency_Ratio + OENEG
  "Total_Net_Equity_kEUR_2022",     # Duplicate
  "Leverage_Ratio_2022"             # Replaced by Capped Leverage / Solvency
)


# EXECUTION OF SELECTION LOGIC
```

```
# Start with all available columns
all_cols <- names(DF)

# Step A: Initial Selection (Demographics + Base 2022 Financials + Engineered)
# We select 2022 columns generally, then refine.
candidates <- grep("_2022$|Company_Age|sector_|Macro.Area|Legal_Form", all_cols, value
= TRUE)
candidates <- unique(c(candidates, engineered_features))

# Step B: Apply Exclusion Filters (Set Operations)
# 1. Remove Leakage Patterns
final_features <- setdiff(candidates, grep(paste(patterns_leakage, collapse = "|"),
candidates, value = TRUE))

# 2. Remove Redundant Variables
final_features <- setdiff(final_features, vars_redundant)

# 3. Remove Target Variable (if present in list)
final_features <- setdiff(final_features, "default_label")


# FINAL CHECK
# Ensure no intersection between final features and banned variables
if (length(intersect(final_features, vars_redundant)) > 0) {
  warning("Critical Error: Redundant variables still present in feature set.")
}

# Subset the dataframe for the modeling stage
X <- DF[, ..final_features]
y <- DF$default_label
```

A.5 XGBoost Optimization & Training

The classification model is an Extreme Gradient Boosting (XGBoost) algorithm. The training process implements two key regularization strategies to handle the specific characteristics of the distressed debt dataset: 1. **Cost-Sensitive Learning:** To address the severe class imbalance, positive instances (distressed firms) are weighted inversely to their frequency using the `scale_pos_weight` parameter (N_{neg}/N_{pos}). This penalizes false negatives more heavily than false positives. 2. **Stratified Cross-Validation with Early Stopping:** The optimal number of boosting rounds (T) is determined via 5-fold stratified cross-validation. Training is halted if the validation AUC does not improve for 100 consecutive rounds ("Early Stopping") to prevent overfitting on the noise.

```
# MODEL TRAINING AND HYPERPARAMETER TUNING

# COST-SENSITIVE WEIGHING
# Calculate the ratio of negative to positive instances.
# This value is passed to 'scale_pos_weight' to balance the loss function.

n_pos <- sum(y_train == 1)
n_neg <- sum(y_train == 0)
imbalance_ratio <- n_neg / n_pos
```

```r
# HYPERPARAMETER CONFIGURATION

xgb_params <- list(
  # General Booster Settings
  booster          = "gbtree",
  tree_method      = "hist",             # Optimized histogram-based algorithm
  objective        = "binary:logistic",  # Output probability for binary classification
  eval_metric      = "auc",              # Metric to maximize

  # Learning Rate & Complexity
  eta              = 0.02,               # Learning rate (shrinkage)
  max_depth        = 4,                  # Interaction depth (prevent overfitting)

  # Stochastic Subsampling (Noise Robustness)
  subsample        = 0.7,                # Row subsampling
  colsample_bytree = 0.7,                # Column subsampling per tree

  # Imbalance Handling
  scale_pos_weight = imbalance_ratio
)


# 3. CROSS-VALIDATION
# Run 5-fold Stratified CV to find the optimal number of trees (nrounds).

set.seed(123) # Ensure reproducibility of folds

cv_model <- xgb.cv(
  params                = xgb_params,
  data                  = dtrain,        # xgboost.DMatrix object
  nrounds               = 3000,          # Max potential rounds
  nfold                 = 5,
  stratified            = TRUE,
  early_stopping_rounds = 100,           # Stop if no improvement for 100 rounds
  maximize              = TRUE,
  verbose               = 0              # Silent execution
)

# Extract optimal iteration
best_iter <- cv_model$best_iteration
best_auc  <- max(cv_model$evaluation_log$test_auc_mean)


# LEARNING CURVE VISUALIZATION

eval_log <- cv_model$evaluation_log
eval_log$iter <- 1:nrow(eval_log)

plot_learning_curve <- ggplot(eval_log, aes(x = iter)) +
  geom_line(aes(y = train_auc_mean, color = "Train AUC"), linewidth = 0.8) +
  geom_line(aes(y = test_auc_mean,  color = "CV Test AUC"), linewidth = 0.8) +
  geom_vline(xintercept = best_iter, linetype = "dashed", color = "black", alpha = 0.5)
+
  scale_color_manual(values = c("Train AUC" = "#999999", "CV Test AUC" = "#E41A1C")) +
  labs(
    title = "XGBoost Cross-Validation Learning Curve",
    subtitle = paste("Optimal Complexity:", best_iter, "Trees | Peak AUC:",
```

```
round(best_auc, 4)),
    x = "Boosting Rounds (Trees)",
    y = "AUC Score",
    color = "Metric"
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")


# FINAL MODEL TRAINING
# Retrain on the full training set using the optimal number of rounds found via CV.

set.seed(123) # Seed for final model initialization

xgb_final <- xgb.train(
  params    = xgb_params,
  data      = dtrain,
  nrounds   = best_iter,
  watchlist = list(train = dtrain, test = dtest),
  verbose   = 0
)
```

A.6 Economic Thresholds - Gatekeeper

The transformation of continuous probabilities into binary classes requires the selection of a decision threshold (T). We evaluate three distinct strategies, each corresponding to a different risk appetite: 1. **Balanced Approach (Youden's J):** Maximizes the sum of Sensitivity and Specificity ($J = Sens + Spec - 1$). This is the standard statistical benchmark. 2. **Prudential/Economic Approach (Basel III Proxy):** Minimizes the total economic loss of the portfolio. We assign a Loss Given Default ($LGD = 45\%$) to False Negatives and an Opportunity Cost ($NIM = 4.5\%$) to False Positives.

$$Loss(T) = FN(T) \cdot LGD + FP(T) \cdot NIM \tag{21}$$

3. **High-Recall Approach (Turnaround Strategy):** Fixes a minimum Sensitivity target (80%) to ensure the majority of distressed firms are detected, accepting a higher False Positive rate. This is suitable for early-warning systems where missing a crisis is more costly than a false alarm. The final operational threshold is selected based on this comparative analysis.

```
# THRESHOLD DEFINITION AND PERFORMANCE EVALUATION

# GENERATE PREDICTIONS
probs_test <- predict(xgb_final, dtest)
labels_test <- as.numeric(as.character(y_test)) # Ensure numeric 0/1

# Initialize ROC Object for analysis
roc_curve <- roc(response = labels_test, predictor = probs_test,
                 levels = c(0, 1), direction = "<")

# ACADEMIC APPROACH (Youden's Index)
# Objective: Maximize (Sensitivity + Specificity - 1)
coords_youden <- coords(roc_curve, "best", best.method = "youden",
                        ret = c("threshold", "sensitivity", "specificity"))

thresh_balanced <- coords_youden$threshold
```

```r
# PRUDENTIAL CUT-OFF (Basel III Proxy)
# Objective: Minimize Total Portfolio Loss
# Parameters:
# - LGD (Loss Given Default): 45% (Standard unsecured recovery rate)
# - NIM (Net Interest Margin): 4.5% (Proxy for opportunity cost of rejecting a good
firm)

LGD <- 0.45
NIM <- 0.045

# Retrieve all coordinates to find the minimum
all_coords <- coords(roc_curve, "all", ret = c("threshold", "sensitivity",
"specificity"))

# Calculate False Negatives (FN) and False Positives (FP) at each threshold
n_pos <- sum(labels_test == 1)
n_neg <- sum(labels_test == 0)

all_coords$FN_Count <- (1 - all_coords$sensitivity) * n_pos
all_coords$FP_Count <- (1 - all_coords$specificity) * n_neg

# Calculate Total Loss Function
all_coords$Total_Loss <- (all_coords$FN_Count * LGD) + (all_coords$FP_Count * NIM)

# Find optimal point
opt_economic <- all_coords[which.min(all_coords$Total_Loss), ]
thresh_economic <- opt_economic$threshold

# TURANROUND/HIGH SENSITIVITY FOCUS

# Objective: Guarantee at least 80% recall (catch 4 out of 5 crises)
# Logic: Select the highest threshold that satisfies Sensitivity >= 0.80

candidates <- all_coords[all_coords$sensitivity >= 0.80, ]
# Maximize specificity among valid candidates
thresh_high_recall <- candidates[which.max(candidates$specificity), "threshold"]

# COMPARISON TABLE

# Helper function to calculate full metrics for a given threshold
calc_metrics <- function(prob, actual, threshold, strategy_name) {
  pred_class <- factor(ifelse(prob > threshold, 1, 0), levels = c("0", "1"))
  act_class  <- factor(actual, levels = c("0", "1"))

  cm <- confusionMatrix(pred_class, act_class, positive = "1", mode = "prec_recall")

  data.frame(
    Strategy       = strategy_name,
    Threshold      = round(threshold, 4),
    Sensitivity    = round(cm$byClass[["Sensitivity"]], 4),
    Specificity    = round(cm$byClass[["Specificity"]], 4),
    Precision      = round(cm$byClass[["Pos Pred Value"]], 4),
    F1_Score       = round(cm$byClass[["F1"]], 4),
    Balanced_Acc   = round(cm$byClass[["Balanced Accuracy"]], 4),
    stringsAsFactors = FALSE
  )
```

```
}

# Compile results
results_table <- rbind(
  calc_metrics(probs_test, labels_test, thresh_balanced, "Statistical (Youden)"),
  calc_metrics(probs_test, labels_test, thresh_economic, "Economic (Basel III)"),
  calc_metrics(probs_test, labels_test, thresh_high_recall, "High Recall (Target 80%)")
)

# Display Comparison
print("--- THRESHOLD STRATEGY COMPARISON ---")
print(results_table)

# OPERATIONAL SELECTION
# For the purpose of this study, we adopt the Economic Threshold.
final_threshold <- thresh_economic
final_predictions <- ifelse(probs_test > final_threshold, 1, 0)
```

A.7 Counterfactual Analysis

This module quantifies the economic cost of lending to "Terminal" firms using a counterfactual reallocation framework. We define **Trapped Capital** (K_{trap}) as the total debt exposure to firms with negative EBITDA and chronic distress. We then simulate a "Liquidation & Reallocation" scenario: 1. **Liquidation:** Terminal firms are liquidated, recovering a fraction of capital based on a conservative Loss Given Default ($LGD = 60\%$).

$$K_{rec} = K_{trap} \times (1 - LGD) \tag{22}$$

2. **Reallocation:** The recovered capital K_{rec} is theoretically re-deployed to "Viable" firms (Rate-Sensitive + Resilient). 3. **Net Value Impact:** We compare the EBITDA currently destroyed by Terminal firms against the potential EBITDA generated by Viable firms using the same capital (Marginal ROI).

```
# OPPORTUNITY COST ANALYSIS: THE "RESCUE RATIO"

# CONFIGURATION AND AGGREGATION

# Parameter: LGD for Unsecured Distressed Debt (Conservative Estimate)
LGD_TERMINAL <- 0.60

# Aggregation: Group metrics by Status
misallocation_summary <- zombie_df[, .(
  Count = .N,
  Total_Debt_Exposure    = sum(Total_Fin_Debt, na.rm = TRUE),
  Total_EBITDA_Generated = sum(EBITDA_kEUR_2022, na.rm = TRUE)
), by = Status]

# EXTRACT SPECIFIC CLUSTERS (Strict Adherence to Thesis)

# THE ISSUE: "Terminal" Firms (Operational Burners)
# Capital trapped in firms that destroy value regardless of interest rates.
debt_terminal   <- misallocation_summary[Status == "Terminal (Operational Burner)",
Total_Debt_Exposure]
ebitda_terminal <- misallocation_summary[Status == "Terminal (Operational Burner)",
Total_EBITDA_Generated]
```

```r
# THE VICTIM: "Rate-Sensitive" Zombies
# Firms that are operationally viable but crushed by financial costs.
debt_sensitive   <- misallocation_summary[Status == "Rate-Sensitive Zombie (Risk >
6.5%)", Total_Debt_Exposure]
ebitda_sensitive <- misallocation_summary[Status == "Rate-Sensitive Zombie (Risk >
6.5%)", Total_EBITDA_Generated]

# OPPORTUNITY COST CALCULATION (The Swap)

# Step A: Calculate Trapped Capital
trapped_capital <- debt_terminal

# Step B: Calculate Liquidation Dividend (Recoverable Cash)
# If we liquidate Terminal firms today, we recover (1 - LGD)
capital_freed <- trapped_capital * (1 - LGD_TERMINAL)

# Step C: The "Rescue Ratio"
# What % of the Rate-Sensitive cluster's debt could be refinanced by the freed capital?
rescue_ratio_pct <- (capital_freed / debt_sensitive) * 100

# REPORTING

opportunity_report <- data.frame(
  Metric = c(
    "1. Trapped Capital (Terminal Debt)",
    "2. Recoverable Capital (Post 60% LGD)",
    "3. Funding Need (Rate-Sensitive Debt)",
    "4. RESCUE RATIO (Coverage Potential)"
  ),
  Value = c(
    format(trapped_capital, big.mark = ","),
    format(capital_freed, big.mark = ","),
    format(debt_sensitive, big.mark = ","),
    paste0(round(rescue_ratio_pct, 1), "%")
  )
)

print("--- OPPORTUNITY COST: TERMINAL VS RATE-SENSITIVE ---")
print(opportunity_report)

# VISUALIZATION: The Rescue Potential
# We visualize the specific coverage of Rate-Sensitive debt by recovered capital.

plot_data <- data.frame(
  Category = c("Rate-Sensitive Debt (Needs Funding)", "Recoverable Capital (From
Terminal Liq.)"),
  Amount   = c(debt_sensitive, capital_freed),
  Type     = c("Demand", "Supply")
)

p_rescue <- ggplot(plot_data, aes(x = Category, y = Amount, fill = Type)) +
  geom_col(width = 0.6, alpha = 0.9) +
  geom_text(aes(label = paste0(format(round(Amount / 1000, 0), big.mark = ","), "
M€")),
            vjust = -0.5, fontface = "bold") +
  scale_fill_manual(values = c("Demand" = "#E67E22", "Supply" = "#27AE60")) +
```

```
  scale_y_continuous(labels = scales::label_number(scale = 1e-3, suffix = " M€")) +
  labs(
    title = "The Rescue Potential",
    subtitle = paste0("Recovered capital from Terminal liquidation covers ",
                      round(rescue_ratio_pct, 1), "% of Rate-Sensitive debt"),
    x = NULL,
    y = "Capital Amount (EUR)"
  ) +
  theme_minimal() +
  theme(legend.position = "none")

# Note: Only p_rescue is printed to keep the appendix concise.
```

A.8 Institutional Entrenchment

This module tests the hypothesis of **Lending Bias** towards mature firms. We investigate whether "Legacy" companies (Age > 20 years) exhibit significantly higher debt loads at the point of distress compared to "Start-ups" (Age $\leq$ 5 years). A positive differential suggests that creditors may extend softer budget constraints to established incumbents. The analysis proceeds in three steps: 1. **Stratification:** Firms are categorized into four age cohorts (*Start-up, Young, Mature, Legacy*). 2. **Metric Calculation:** We compute *Debt Load* ($\frac{\text{Total Payables}}{\text{Total Assets}}$) as the proxy for financial leverage tolerance. 3. **Statistical Inference:** A non-parametric **Mann-Whitney U Test** (one-sided) assesses if the median debt load of distressed legacy firms is strictly greater than that of distressed start-ups.

```
# INSTITUTIONAL ENTRENCHMENT: AGE-DEPENDENT LEVERAGE TOLERANCE

# COHORT STRATIFICATION AND METRICS

# Create a focused subset for analysis
analysis_df <- copy(DF)

# COMPUTE DEBT LOAD (TOTAL PAYABLES/TOTAL ASSETS %)
# We cap the ratio at 150% to minimize the impact of extreme outliers (e.g., near-zero
assets).
analysis_df[, Debt_Load := (Total_Payables_kEUR_2022 / Total_Assets_kEUR_2022) * 100]
analysis_df[Debt_Load > 150, Debt_Load := 150]
analysis_df[Debt_Load < 0, Debt_Load := 0] # Sanity check

# AGE BANDS
# Categorization based on firm maturity lifecycle
analysis_df[, Age_Band := fcase(
  Company_Age <= 5,  "Start-up (0-5y)",
  Company_Age <= 10, "Young (6-10y)",
  Company_Age <= 20, "Mature (11-20y)",
  default = "Legacy (21+y)"
)]

# Set factor levels for ordinal consistency in plotting
analysis_df$Age_Band <- factor(
  analysis_df$Age_Band,
  levels = c("Start-up (0-5y)", "Young (6-10y)", "Mature (11-20y)", "Legacy (21+y)"),
  ordered = TRUE
)
```

```r
# STATUS LABEL
analysis_df[, Status := ifelse(default_label == 1, "Distressed", "Healthy")]
analysis_df$Status <- factor(analysis_df$Status, levels = c("Healthy", "Distressed"))

# Remove incomplete cases
analysis_df <- analysis_df[!is.na(Debt_Load) & !is.na(Age_Band)]


# DESCRIPTIVE STATISTICS

# Calculate median Debt Load by Age Cohort and Status
summary_stats <- analysis_df %>%
  group_by(Age_Band, Status) %>%
  summarise(
    N = n(),
    Median_Debt = round(median(Debt_Load, na.rm = TRUE), 2),
    IQR_Debt = round(IQR(Debt_Load, na.rm = TRUE), 2),
    .groups = "drop"
  )

print("--- DEBT LOAD DISTRIBUTION BY COHORT ---")
print(summary_stats)

# HYPOTHESIS TESTING (Mann-Whitney U)
# H0: Median Debt(Legacy_Distressed) <= Median Debt(Startup_Distressed)
# H1: Median Debt(Legacy_Distressed) >  Median Debt(Startup_Distressed)

# Extract vectors for the two target groups
vec_legacy  <- analysis_df[Age_Band == "Legacy (21+y)" & Status == "Distressed",
Debt_Load]
vec_startup <- analysis_df[Age_Band == "Start-up (0-5y)" & Status == "Distressed",
Debt_Load]

if (length(vec_legacy) > 20 && length(vec_startup) > 20) {

  test_result <- wilcox.test(vec_legacy, vec_startup, alternative = "greater")

  # Output Test Results
  cat("\n=== HYPOTHESIS TEST RESULTS ===\n")
  cat("Test Method: Mann-Whitney U (One-sided)\n")
  cat(paste("Legacy Median Debt:", median(vec_legacy), "%\n"))
  cat(paste("Start-up Median Debt:", median(vec_startup), "%\n"))
  cat(paste("P-value:", format.pval(test_result$p.value, digits = 4), "\n"))

  if (test_result$p.value < 0.05) {
    cat("CONCLUSION: Reject H0. Evidence supports Institutional Entrenchment.\n")
  } else {
    cat("CONCLUSION: Fail to reject H0. No significant evidence of lending bias.\n")
  }

} else {
  warning("Insufficient sample size for statistical testing.")
}

# VISUALIZATION (Boxplot)

p_entrenchment <- ggplot(analysis_df, aes(x = Age_Band, y = Debt_Load, fill = Status))
```

```
+
  geom_boxplot(outlier.shape = 1, outlier.alpha = 0.3, width = 0.6) +
  scale_fill_manual(values = c("Healthy" = "#27AE60", "Distressed" = "#C0392B")) +
  labs(
    title = "Evidence of Institutional Entrenchment",
    subtitle = "Distribution of Debt Load (Payables/Assets) at Distress point across
Age Cohorts",
    x = "Firm Maturity Stage",
    y = "Debt Load (%)"
  ) +
  theme_minimal() +
  theme(legend.position = "bottom")

# Note: Print logic is omitted for the appendix to focus on structure.
```

A.9 Post-Hoc Interpretability

To validate the economic consistency of the model, we employ a model-agnostic interpretability framework using the `DALEX` library. 1. **Global Importance:** We assess variable contribution via **Permutation Feature Importance** (Fisher et al., 2019). This measures the drop in model performance (AUC Loss) when a specific feature is randomized. 2. **Marginal Effects:** We utilize **Accumulated Local Effects (ALE)** (Apley & Zhu, 2020) to visualize the relationship between key financial ratios and the probability of distress. Unlike Partial Dependence Plots (PDP), ALE is robust to the high multicollinearity inherent in financial statements. 3. **Heterogeneity Analysis:** ALE profiles are segmented by **Macro-Area** and **Sector** to detect structural breaks in risk behavior across different sub-populations.

```
# XAI: INTERPRETABILITY AND BEHAVIORAL ANALYSIS (DALEX)

# Define custom prediction wrapper for XGBoost
# DALEX requires a function that takes the model and a data frame, returning a numeric
vector.
custom_predict <- function(model, newdata) {
  # Handle potential type mismatches or missing columns if necessary
  dmat <- xgb.DMatrix(data = data.matrix(newdata), missing = NA)
  predict(model, dmat)
}

# Prepare data subset for explanation
# We use a random sample of 2,000 observations to ensure computational efficiency
# while maintaining statistical representativeness.
set.seed(42)
valid_features <- intersect(final_features, names(test_set))
data_dalex     <- as.data.frame(test_set)[, valid_features, drop = FALSE]
y_dalex        <- as.numeric(as.character(test_set$default_label))

idx_sample     <- sample(1:nrow(data_dalex), min(2000, nrow(data_dalex)))
explainer_data <- data_dalex[idx_sample, , drop = FALSE]
explainer_y    <- y_dalex[idx_sample]

# Initialize DALEX Explainer
explainer_xgb <- DALEX::explain(
  model            = xgb_final,
  data             = explainer_data,
```

```
  y                = explainer_y,
  predict_function = custom_predict,
  label            = "XGBoost_Distressed",
  verbose          = FALSE
)

# GLOBAL FEATURE IMPORTANCE (Permutation Method)

# Calculate variable importance based on "1 - AUC" loss.
# B = 5 permutations per variable for stability.
vip_xgb <- model_parts(explainer_xgb, loss_function = loss_one_minus_auc, B = 5)

# Visualization
plot(vip_xgb, max_vars = 15, title = "Feature Importance (Permutation-based)")

# ACCUMULATED LOCAL EFFECTS (ALE) - Key Drivers Analysis

# Define the 9 key economic drivers for detailed behavioral analysis
target_vars <- c(
  "CF_Generative", "Gearing_Ratio", "Num_Employees",
  "Profit_Margin", "ROCE", "ROE",
  "Sales_Volume", "Solvency_Ratio", "Total_Size"
)
target_vars <- intersect(target_vars, colnames(explainer_data))

# A. Generate Overall ALE Profiles
ale_overall <- model_profile(explainer_xgb, variables = target_vars, type =
"accumulated")
df_overall  <- as.data.frame(ale_overall$agr_profiles)
df_overall$Label <- "Global Average"

# Center the curves at 0 for comparability
df_overall <- df_overall %>%
  group_by(`_vname_`) %>%
  mutate(`_yhat_` = `_yhat_` - mean(`_yhat_`, na.rm = TRUE)) %>%
  ungroup()

# HETEROGENEITY ANALYSIS: Macro-Area Segmentation
# We decode the One-Hot encoded Macro-Area variables back to a single factor
# to allow DALEX to group the ALE plots.

# Identify Macro-Area columns
cols_macro <- c("Macro.Area_North", "Macro.Area_Center", "Macro.Area_South_Islands")

if (all(cols_macro %in% colnames(explainer_data))) {

  # 1. Reconstruct Categorical Feature
  explainer_data_geo <- explainer_data %>%
    mutate(Region = case_when(
      `Macro.Area_North` == 1 ~ "North",
      `Macro.Area_Center` == 1 ~ "Center",
      `Macro.Area_South_Islands` == 1 ~ "South/Islands",
      TRUE ~ "Other"
    )) %>%
    filter(Region != "Other")

  # 2. Update Predict Function to ignore the new 'Region' column
```

```r
  predict_geo <- function(model, newdata) {
    if ("Region" %in% colnames(newdata)) newdata$Region <- NULL
    dmat <- xgb.DMatrix(data = data.matrix(newdata), missing = NA)
    predict(model, dmat)
  }

  # 3. New Explainer for Geography
  explainer_geo <- DALEX::explain(
    model = xgb_final, data = explainer_data_geo, y = explainer_y,
    predict_function = predict_geo, label = "Geo_Split", verbose = FALSE
  )

  # 4. Generate Regional ALE
  ale_geo <- model_profile(explainer_geo, variables = target_vars,
                           groups = "Region", type = "accumulated")

  df_geo <- as.data.frame(ale_geo$agr_profiles)
  colnames(df_geo)[colnames(df_geo) == "_groups_"] <- "Label"

  # Center curves by group
  df_geo <- df_geo %>%
    group_by(`_vname_`, Label) %>%
    mutate(`_yhat_` = `_yhat_` - mean(`_yhat_`, na.rm = TRUE)) %>%
    ungroup()

  # 5. Combined Visualization (Global + Regional)
  df_combined <- bind_rows(df_overall, df_geo)

  p_ale_geo <- ggplot(df_combined, aes(x = `_x_`, y = `_yhat_`, color = Label, linetype
= Label)) +
    geom_line(linewidth = 1) +
    facet_wrap(~`_vname_`, scales = "free", ncol = 3) +
    scale_color_manual(values = c(
      "Global Average" = "black", "North" = "#27AE60",
      "Center" = "#2980B9", "South/Islands" = "#E67E22"
    )) +
    scale_linetype_manual(values = c(
      "Global Average" = "dashed", "North" = "solid",
      "Center" = "solid", "South/Islands" = "solid"
    )) +
    labs(
      title = "Regional Heterogeneity in Default Drivers (ALE)",
      subtitle = "Accumulated Local Effects centered at mean prediction",
      x = "Feature Value", y = "Marginal Effect on Log-Odds"
    ) +
    theme_minimal() +
    theme(legend.position = "bottom")

  print(p_ale_geo)
}

# HETEROGENEITY ANALYSIS: Manufacturing Sector
# Compare "Manufacturing" vs "Rest of Economy"

col_manuf <- "sector_division_MANUFACTURING"

if (col_manuf %in% colnames(explainer_data)) {
```

```r
  # 1. Reconstruct Sector Label
  explainer_data_sec <- explainer_data %>%
    mutate(Sector = ifelse(get(col_manuf) == 1, "Manufacturing", "Rest of Economy"))

  # 2. Update Predict Function
  predict_sec <- function(model, newdata) {
    if ("Sector" %in% colnames(newdata)) newdata$Sector <- NULL
    dmat <- xgb.DMatrix(data = data.matrix(newdata), missing = NA)
    predict(model, dmat)
  }

  # 3. New Explainer
  explainer_sec <- DALEX::explain(
    model = xgb_final, data = explainer_data_sec, y = explainer_y,
    predict_function = predict_sec, label = "Sector_Split", verbose = FALSE
  )

  # 4. Generate Sector ALE
  ale_sec <- model_profile(explainer_sec, variables = target_vars,
                           groups = "Sector", type = "accumulated")
  df_sec <- as.data.frame(ale_sec$agr_profiles)

  # 5. Visualization (Uncentered to show absolute risk differences)
  p_ale_sec <- ggplot(df_sec, aes(x = `_x_`, y = `_yhat_`, color = `_groups_`)) +
    geom_line(linewidth = 1) +
    facet_wrap(~`_vname_`, scales = "free", ncol = 3) +
    scale_color_manual(values = c("Manufacturing" = "#C0392B", "Rest of Economy" =
"#7F8C8D")) +
    labs(
      title = "Sectoral Impact: Manufacturing vs. Economy",
      subtitle = "Absolute Accumulated Local Effects (Uncentered)",
      x = "Feature Value", y = "ALE Contribution"
    ) +
    theme_minimal() +
    theme(legend.position = "bottom")

  print(p_ale_sec)
}
```

A.10 Variable List

1. Identification & Registry Data - `Company_Name` - `Tax_ID` - `ATECO_Code` (NACE Rev. 2) - `ateco_2_digit` - `sector_group` - `Incorporation_Year` - `Company_Age` - `Legal_Status` - `Last_Balance_Sheet_Date` - `is_quoted`

2. Status & Target Variables

- `Cessation_Procedure`
- `Procedure_Start_Date`
- `Procedure_End_Date`
- `default_label` (Target Variable: 0=Active/Success, 1=Distressed)
- `Active_2019` (Selection Flag)

3. Categorical Classifiers (Grouped)

- `Macro.Area` (North, Center, South & Islands)

- `Legal_Form` (e.g., S.R.L., S.P.A., Consorzio, etc.)
- `Sector_Division` (e.g., Manufacturing, Construction, Trade, etc.)

4. Raw Financial Variables (Panel Data: $t_{2019} \dots t_{2024}$)

- `Sales_Revenues_kEUR`
- `Sales_Services_Revenues_kEUR`
- `Production_Value_kEUR`
- `Production_Costs_kEUR`
- `EBITDA_kEUR`
- `EBIT_kEUR`
- `EBT_kEUR` (Earnings Before Tax)
- `Net_Income_kEUR` / `Net_Profit_Loss_kEUR`
- `Total_Assets_kEUR`
- `Shareholders_Equity_kEUR` / `Total_Net_Equity_kEUR`
- `Net_Financial_Position_kEUR`
- `Total_Payables_kEUR`
- `Short_Term_Debts_kEUR`
- `Long_Term_Debts_kEUR`
- `Bank_Debt_to_Turnover` (Ratio)
- `Financial_Income_Charges_kEUR`
- `Financial_Charges_to_Turnover`
- `Extraordinary_Items_kEUR`
- `Cash_Equivalents_kEUR`
- `Current_Assets_kEUR`
- `Current_Financial_Assets_kEUR`
- `Inventory_kEUR`
- `Total_Receivables_kEUR`
- `Intangible_Fixed_Assets_kEUR`
- `Tangible_Fixed_Assets_kEUR`
- `Financial_Fixed_Assets_kEUR`
- `Accruals_Deferrals_Assets_kEUR`
- `Accruals_Deferrals_Liabilities_kEUR`
- `Total_Liabilities_Equity_kEUR`
- `Memorandum_Accounts_kEUR`
- `Employees` (Count)

5. Financial Ratios (Panel Data)

- `ROS` (Return on Sales)
- `ROA` (Return on Assets)
- `ROE` (Return on Equity)
- `ROI` (Return on Investment)
- `EBITDA_Margin`
- `Debt_to_Equity`
- `Debt_to_EBITDA`
- `Leverage_Ratio`
- `Liquidity_Ratio`
- `Current_Ratio`
- `Invested_Capital_Turnover`
- `Short_Term_Debt_Index`
- `Long_Term_Debt_Index`

- Fixed_Asset_Coverage_Equity
- Fixed_Asset_Coverage_Financial

6. Engineered Features (Model Specific)

- OENEG (Negative Equity Dummy)
- Solvency_Ratio (Equity/Assets Capped)
- distress_history_score
- sales_growth_1y
- sales_growth_2y
- recovery_sales_vs_2019
- delta_ebitda_margin
- delta_asset_turnover
- scissor_debt_sales
- delta_debt_maturity
- gap_receivables_sales
- trade_financing_stress
- delta_cost_weight
- cash_flow_proxy
- CF_Generative
- ICR (Interest Coverage Ratio)
- Gearing_Ratio
- Profit_Margin
- ROCE (Return on Capital Employed)
- Sales_Volume (Alias)
- Total_Size (Alias)
- Num_Employees (Alias)

A.11 Relevant Tables

Benchmark Comparison Table

Model Benchmarking: Comprehensive Performance Comparison

	Discrimination	Classification Metrics					
Model	AUC	Sensitivity	Specificity	F1-Score	PPV	NPV	Balanced Acc.
XGBoost	**0.7791**	**0.7947**	**0.5952**	**0.8092**	**0.8243**	**0.5482**	**0.6950**
Random Forest	0.7734	0.9005	0.3862	0.8348	0.7781	0.6189	0.6433
Logistic Regression	0.7151	0.7954	0.4437	0.7843	0.7736	0.4758	0.6196
Altman Z'-Score	0.6465	0.7943	0.3482	0.7685	0.7444	0.4146	0.5712

Figure 31: Comparison of the most common models,adopting the same predictive threshold, over the same Distressed Cohort sample.

Attrition Table

Attrition Table

Filtering Phases

Step	Description	Remaining_N	Removed
1. Raw Data Extraction (AIDA)	Total records imported after deduplication	319,266	-
2. Large Corporate Exclusion (EU Def.)	Exclusion of firms > 250 emp, > 50M Rev, > 43M Assets	318,458	808
3. Data Integrity & Noise Removal	Removal of transfer records and undefined status	87,311	231147
4. Healthy Firms Exclusion (Positive Equity)	Selection of Distressed Cohort (Equity < 0)	87,311	0

Figure 32: How each filtering phase removes a significant portion of the original dataset.

Descriptive Statistics Table

Variable <chr>	Mean <dbl>	Median <dbl>	SD <dbl>	Min <dbl>	Max <dbl>	Skew <dbl>	Class <chr>
Sales_Revenues_kEUR_2022	467.44923612	114	1534.9992304	0.000	3.805900e+04	10.77796991	0 (Success)
EBITDA_kEUR_2022	-27.03241452	-4	439.5014299	-29639.000	2.634700e+04	-16.38964733	0 (Success)
Leverage_Ratio_2022	50.00000000	50	0.0000000	50.000	5.000000e+01	NaN	0 (Success)
sales_growth_1y	1.21833987	0	12.9725102	-1.000	1.160000e+03	45.41413266	0 (Success)
scissor_debt_sales	1.19997819	0	48.1685274	-1159.589	5.248000e+03	62.99057437	0 (Success)
cash_flow_proxy	-0.02564054	0	0.4745784	-49.520	9.954338e-01	-55.13956471	0 (Success)
distress_history_score	2.06733490	2	1.0612353	1.000	4.000000e+00	0.46044103	0 (Success)
Sales_Revenues_kEUR_2022	289.31718583	56	1101.4505661	0.000	4.509300e+04	17.92134646	1 (Failure)
EBITDA_kEUR_2022	-48.79852643	-7	672.1965392	-44133.000	5.189900e+04	-5.28637437	1 (Failure)
Leverage_Ratio_2022	50.00000000	50	0.0000000	50.000	5.000000e+01	NaN	1 (Failure)
sales_growth_1y	0.80133829	0	16.4568935	-1.000	2.692000e+03	103.41886840	1 (Failure)
scissor_debt_sales	0.65242465	0	60.0354435	-2692.000	8.763500e+03	98.01269985	1 (Failure)
cash_flow_proxy	-0.17329175	0	3.5219859	-514.000	9.955357e-01	-89.08863135	1 (Failure)
distress_history_score	2.52193578	3	1.1619730	1.000	4.000000e+00	-0.04749063	1 (Failure)

Figure 33: Comparative Descriptive Statistics of Key Financial Drivers (Top 14) by Turnaround Outcome (Class 0 vs. Class 1)

Bibliography

Acharya, V.V., Bharath, S.T. and Srinivasan, A. (2003) "Understanding the Recovery Rates on Defaulted Securities," *SSRN Journal* [Preprint]. Available at: https://doi.org/10.2139/ssrn.442901.

Adalet McGowan, M., Andrews, D. and Millot, V. (2018) "The Walking Dead? Zombie Firms and Productivity Performance in OECD Countries," *Econ Policy*, 33(96), pp. 685–736. Available at: https://doi.org/10.1093/epolic/eiy012.

Altman, E.I. (1968) "FINANCIAL RATIOS, DISCRIMINANT ANALYSIS AND THE PREDICTION OF CORPORATE BANKRUPTCY," *The Journal of Finance*, 23(4), pp. 589–609. Available at: https://doi.org/10.1111/j.1540-6261.1968.tb00843.x.

Altman, E.I. and Hotchkiss, E. (2005) *Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt*. 1st ed. Wiley. Available at: https://doi.org/10.1002/9781118267806.

Amirshahi, B. and Lahmiri, S. (2024) "Bankruptcy Prediction Using Optimal Ensemble Models under Balanced and Imbalanced Data," *Expert Systems*, 41(8), p. e13599. Available at: https://doi.org/10.1111/exsy.13599.

Balcaen, S. and Ooghe, H. (2006) "35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems," *The British Accounting Review*, 38(1), pp. 63–93. Available at: https://doi.org/10.1016/j.bar.2005.09.001.

Barboza, F., Kimura, H. and Altman, E. (2017) "Machine Learning Models and Bankruptcy Prediction," *Expert Systems with Applications*, 83, pp. 405–417. Available at: https://doi.org/10.1016/j.eswa.2017.04.006.

Beaver, W.H. (1966) "Financial Ratios As Predictors of Failure," *Journal of Accounting Research*, 4, pp. 71–111. Available at: https://doi.org/10.2307/2490171.

Bellovary, J.L., Giacomino, D.E. and Akers, M.D. (2007) "A Review of Bankruptcy Prediction Studies: 1930 to Present," *Journal of Financial Education*, 33, pp. 1–42.

Bradley, M., Jarrell, G.A. and Kim, E.H. (1984) "On the Existence of an Optimal Capital Structure: Theory and Evidence," *The Journal of Finance*, 39(3), pp. 857–878. Available at: https://doi.org/10.1111/j.1540-6261.1984.tb03680.x.

Bragoli, D. *et al.* (2019) "Machine Learning Models for Bankruptcy Prediction in Italy: Do Industrial Variables Count?."

Braunsberger, C. and Aschauer, E. (2025) "Corporate Failure Prediction: A Literature Review of Altman Z-Score and Machine Learning Models Within a Technology Adoption Framework," *Journal of Risk and Financial Management*, 18(8), p. 465. Available at: https://doi.org/10.3390/jrfm18080465.

Caballero, R.J., Hoshi, T. and Kashyap, A.K. (2008) "Zombie Lending and Depressed Restructuring in Japan," *American Economic Review*, 98(5), pp. 1943–1977. Available at: https://doi.org/10.1257/aer.98.5.1943.

Chen, T. and Guestrin, C. (2016) "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery (KDD '16), pp. 785–794. Available at: https://doi.org/10.1145/2939672.2939785.

Crosato, L., Liberati, C. and Repetto, M. (2023) "Lost in a Black-Box? Interpretable Machine Learning for Assessing Italian SMEs Default," *Applied Stochastic Models in Business and Industry*, 39(6), pp. 829–846. Available at: https://doi.org/10.1002/asmb.2803.

Elkan, C. (2001) “The Foundations of Cost-Sensitive Learning.”

Forstner, F.F., Vassolo, R. and Sevil, A. (2025) “Different Strokes for Different Folks: A Review of SME Decline and Turnaround,” *Journal of Small Business Management*, 0(0), pp. 1–34. Available at: https://doi.org/10.1080/00472778.2025.2515072.

Grajzl, P., Cepec, J. and Mörec, B. (2023) “Weaned off Public Money: The Effect of Discontinued Reception of Public Cash on Firm Outcomes,” *Kyklos*, 76(1), pp. 41–76. Available at: https://doi.org/10.1111/kykl.12317.

Hand, D.J. and Till, R.J. (2001) “A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems,” *Machine Learning*, 45(2), pp. 171–186. Available at: https://doi.org/10.1023/A:1010920819831.

Hillegeist, S.A. *et al.* (2004) “Assessing the Probability of Bankruptcy,” *Review of Accounting Studies*, 9(1), pp. 5–34. Available at: https://doi.org/10.1023/B:RAST.0000013627.90884.b7.

Hou, G. *et al.* (2025) “Comparative Analysis of Resampling Techniques for Class Imbalance in Financial Distress Prediction Using XGBoost,” *Mathematics*, 13(13), p. 2186. Available at: https://doi.org/10.3390/math13132186.

El Kalak, I. and Hudson, R. (2016) “The Effect of Size on the Failure Probabilities of SMEs: An Empirical Study on the US Market Using Discrete Hazard Model,” *International Review of Financial Analysis*, 43, pp. 135–145. Available at: https://doi.org/10.1016/j.irfa.2015.11.009.

Kennedy, D., Lakonishok, J. and Shaw, W.H. (1992) “Accommodating Outliers and Nonlinearity in Decision Models,” *Journal of Accounting, Auditing & Finance*, 7(2), pp. 161–190. Available at: https://doi.org/10.1177/0148558X9200700205.

Kuhn, M. and Johnson, K. (2013) *Applied Predictive Modeling*. New York, NY: Springer New York. Available at: https://doi.org/10.1007/978-1-4614-6849-3.

Moscatelli, M. *et al.* (2020) “Corporate Default Forecasting with Machine Learning,” *Expert Systems with Applications*, 161, p. 113567. Available at: https://doi.org/10.1016/j.eswa.2020.113567.

Nadali, L. *et al.* (2020) “Recurrent ANNs for Failure Predictions on Large Datasets of Italian SMEs,” *Neural Approaches to Dynamics of Signal Exchanges*. Edited by A. Esposito et al. Singapore: Springer. Available at: https://doi.org/10.1007/978-981-13-8950-4_14.

Ohlson, J.A. (1980) “Financial Ratios and the Probabilistic Prediction of Bankruptcy,” *Journal of Accounting Research*, 18(1), pp. 109–131. Available at: https://doi.org/10.2307/2490395.

Pearce, J.A. and Robbins, K. (1993) “Toward Improved Theory and Research on Business Turnaround,” *Journal of Management*, 19(3), pp. 613–636. Available at: https://doi.org/10.1016/0149-2063(93)90007-A.

Peek, J. and Rosengren, E.S. (2005) “Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan,” *American Economic Review*, 95(4), pp. 1144–1166. Available at: https://doi.org/10.1257/0002828054825691.

Peykani, P. *et al.* (2025) “Evaluation of Cost-Sensitive Learning Models in Forecasting Business Failure of Capital Market Firms,” *Mathematics*, 13(3), p. 368. Available at: https://doi.org/10.3390/math13030368.

Rico, M., Pandit, N.R. and Puig, F. (2021) “SME Insolvency, Bankruptcy, and Survival: An Examination of Retrenchment Strategies,” *Small Bus Econ*, 57(1), pp. 111–126. Available at: https://doi.org/10.1007/s11187-019-00293-z.

Rudin, C. (2019) “Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead,” *Nat Mach Intell*, 1(5), pp. 206–215. Available at: https://doi.org/10.1038/s42256-019-0048-x.

Schivardi, F., Sette, E. and Tabellini, G. (2020) “Identifying the Real Effects of Zombie Lending,” *Review of Corporate Finance Studies*, 9(3), pp. 569–592. Available at: https://doi.org/10.1093/rcfs/cfaa010.

Schivardi, F., Sette, E. and Tabellini, G. (2022) “Credit Misallocation During the European Financial Crisis,” *Econ J*, 132(641), pp. 391–423. Available at: https://doi.org/10.1093/ej/ueab039.

Senyard, J. *et al.* (2014) “Bricolage as a Path to Innovativeness for Resource-Constrained New Firms,” *Journal of Product Innovation Management*, 31(2), pp. 211–230. Available at: https://doi.org/10.1111/jpim.12091.

Shleifer, A. and Vishny, R.W. (1992) “Liquidation Values and Debt Capacity: A Market Equilibrium Approach,” *The Journal of Finance*, 47(4), pp. 1343–1366. Available at: https://doi.org/10.1111/j.1540-6261.1992.tb04661.x.

Shumway, T. (2001) “Forecasting Bankruptcy More Accurately: A Simple Hazard Model,” *The Journal of Business*, 74(1), pp. 101–124. Available at: https://doi.org/10.1086/209665.

Waal, D. and Zio, D. (2008) “Recommended Practices for Editing and Imputation in Cross-Sectional Business Surveys.”

Wilson, N. and Altanlar, A. (2014) “Company Failure Prediction with Limited Information: Newly Incorporated Companies,” *Journal of the Operational Research Society*, 65(2), pp. 252–264. Available at: https://doi.org/10.1057/jors.2013.31.

Zmijewski, M.E. (1984) “Methodological Issues Related to the Estimation of Financial Distress Prediction Models,” *Journal of Accounting Research*, 22, pp. 59–82. Available at: https://doi.org/10.2307/2490859.

AI Disclosure

In accordance with the guidelines of academic integrity, it is declared the adoption of complementary AI-tools. In the preparation of this work, the author employed a hybrid methodological approach integrating Artificial Intelligence (AI) tools under strict human supervision and validation:

1. The AI-tool Elicit was utilized for the complementary identification of relevant academic literature. To ensure scientific rigor and prevent “hallucinations,” every source identified was manually cross-referenced and validated through Google Scholar to confirm its existence, authorship, and relevance to the research objectives.
2. Generative AI models (Gemini) were used exclusively as a support for stylistic enhancement, syntactic optimization, and terminological precision. The technology served as an advanced editing assistant to improve the clarity and formal quality of the manuscript.

The author declares that no original conceptual frameworks, data interpretations, or scientific conclusions were generated by AI. The final intellectual content, the critical synthesis of the literature, and the integrity of the work remain the sole responsibility of the author.

Acknowledgments

I would like to thank Professor Crosato for her guidance and availability throughout this project.
My deepest gratitude goes to my family for their encouragement over the years,
and to my friends for their endless patience.