DYNAMIC WEIGHT-ADJUSTED RANDOM FOREST BOOST: A NOVEL ENSEMBLE FRAMEWORK FOR FINANCIAL DISTRESS PREDICTION IN HIGHLY IMBALANCED DATA

By

HOU GUODONG

A thesis submitted to the Faculty of Information and Communication Technology,
Universiti Tunku Abdul Rahman,
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computer Science
March 2026

ABSTRACT

DYNAMIC WEIGHT-ADJUSTED RANDOM FOREST BOOST: A NOVEL ENSEMBLE FRAMEWORK FOR FINANCIAL DISTRESS PREDICTION IN HIGHLY IMBALANCED DATA

HOU GUODONG

Financial distress prediction is essential for identifying early warning signs of bankruptcy and financial instability. This study aims to improve financial distress prediction by addressing class imbalance, enhancing the learning of persistently misclassified distressed samples, and dynamically adapting ensemble weights. To achieve this, a Dynamic Weight-Adjusted Random Forest Boost (DWARFB) framework is proposed. DWARFB integrates multi-stage sampling, history-aware misclassified samples reweighting, and a dynamic performance-driven ensemble strategy. It improves the detection of distressed samples while preserving the original data distribution through gradual class-ratio adjustment without synthetic data generation. As a result, this approach maintains stability and interpretability in financial applications.

An accumulated misclassification-based mechanism is introduced to track persistently misclassified samples across iterations. This mechanism emphasizes borderline but informative observations and mitigating noisy minority effects. Combined with a reward-based sliding-iteration strategy, DWARFB stabilizes training and correctly identifies 85% of the distressed samples. This adaptive ensemble component further stratifies iterative models

based on real-time F1 performance, dynamically adjusting weights to optimize generalization in highly imbalanced settings.

Empirical findings suggest that the DWARFB framework provides more accurate and reliable predictions than conventional boosting methods and Random Forest models. DWARFB achieves superior balance between precision and recall, higher sensitivity to minority-class distress samples, and robust overall discriminative power. The result findings also confirm the importance of cumulative profitability, retained earnings, equity strength, and earnings stability, aligning with established financial distress indicators. Sensitivity analysis shows that model parameters such as tree depth, leaf size, and Alpha–Gamma weighting jointly govern learning capacity and the precision–recall trade-off, while the number of trees primarily affects efficiency.

Overall, DWARFB provides a generalizable, interpretable, and adaptive framework for enhancing tree-based classifiers under severe class imbalance. It offers a practical tool for investors and decision-makers to improve early-warning prediction of financial distress in dynamic corporate environments.

Keywords: random forest; dynamic sample-weighting; adaptive sampling; real-time performance strategy; financial distress prediction; imbalanced data

Subject Area: T58.5-58.64 Information technology

ACKNOWLEDGEMENT

It is often said that by the time one reaches this point in writing, there are always some heartfelt reflections to share. Indeed, during this doctoral journey of diligent effort and dedicated learning, there are many thoughts and emotions I wish to express.

Submitting this dissertation, and perhaps one day being granted the doctoral degree, serves as a moment of reflection. It represents a period in which, alone in a foreign country and at the age of forty, I experienced and made up for earlier regrets in my academic life. While youthful inexperience may have allowed some laxity, at this stage of life greater effort is required. Preparing for the IELTS exam and sitting alongside young candidates in their teens and twenties required courage. Yet standing at that starting point, I reminded myself that every effort is worthwhile for a dream I had not pursued wholeheartedly in my youth.

Before embarking on this journey, I anticipated various difficulties. However, for someone who has weathered storms, these challenges proved relatively minor. It is natural to hesitate when facing unfamiliar fields or knowledge. What truly sustained me, however, was the enduring belief in my dreams, along with the support of a group of kind and encouraging people who accompanied me along the way.

Throughout my doctoral research, I owe my deepest gratitude first to my three advisors. Dr. Tong is a rigorous and responsible scholar. Under her guidance, I learned many academic principles I had never encountered before. She once told me, “Most of the doctorates we pursue are philosophical in nature; many philosophical truths of life are embedded within our research, waiting for

us to discover them." This insight greatly influenced me, prompting me to constantly evaluate the validity of experimental results and interpret them through the lens of everyday life. From topic formulation to experimental design and results analysis, Dr. Tong provided invaluable guidance, teaching me the essence of doctoral research, scholarly writing, and how to develop research ideas. She has been instrumental in my holistic academic growth.

Prof. Liew offered crucial guidance whenever I encountered research bottlenecks. He taught me to identify unresolved gaps by synthesizing existing literature and encouraged me to design experiments beyond conventional thinking frameworks. Dr. Choo, with his extensive interdisciplinary perspective, helped bridge the gap between theory and engineering practice, often using real-world business cases to validate my theoretical models. Though distinct in style, all three advisors shared a common goal, shaping me into a researcher who embodies both critical thinking and pragmatic rigor.

I am also deeply grateful to my family. Without their support, it would have been impossible to focus on my studies in Malaysia for such an extended period. The sentiment of missing them strengthened my determination to achieve every goal and overcome each challenge.

There are many others to thank, and countless reflections to share. This doctoral experience is undoubtedly a vivid and defining chapter in my life, an unprecedented journey that will be cherished and remembered for a lifetime.

TABLE OF CONTENTS

ABSTRACT I
ACKNOWLEDGEMENT III
LIST OF TABLES VIII
LIST OF FIGURES X
LIST OF ABBREVIATIONS XI
CHAPTER 1 INTRODUCTION 1
1.1 Motivation 4
1.2 Research Problem Statement 7
1.2.1 Fragmented and Static Handling of Class Imbalance 8
1.2.2 High-Dimensional Feature Redundancy and Weak Representation 8
1.2.3 Inadequate Treatment of Persistently Hard-to-Classify Distress Cases 9
1.2.4 Rigid and Fragmented Ensemble Learning Frameworks 10
1.3 DWARFB: Prediction Framework 10
1.4 Research Questions and Hypotheses 14
1.5 Contributions 15
1.6 Structure of Thesis 16
CHAPTER 2 BACKGROUND AND LITERATURE REVIEW 19
2.1 Trends in Financial Distress Prediction Modelling 19
2.1.1 Traditional Statistical Models 19
2.1.2 Emergence of Machine Learning Approaches 21
2.1.3 Transition to Ensemble Learning 23
2.1.4 Summary of Model Development Trends 25
2.2 Class Imbalance in Financial Distress Prediction: Challenges and Research Progress 25
2.3 Data-Level Methods for Class Imbalance Mitigation 27
2.3.1 Oversampling Techniques 28
2.3.2 Under-sampling Techniques 29
2.3.3 Hybrid Resampling Techniques 30
2.3.4 Summary of data-level techniques 30
2.4 Cost-Sensitive Learning as an Algorithm-Level Technique 31
2.4.1 Cost-Sensitive Learning: Principles and Overview 32
2.4.2 Class-Weight Adjustments in Trees/Boosting 34
2.4.3 Threshold adjustment to deal with Class Imbalance 35
2.4.4 Hybrid Cost-Sensitive and Data-Level Approaches 36
2.4.5 Summary of Cost-Sensitive Learning Approaches 37
2.5 Iteration-Driven Learning Mechanisms for Financial Distress Prediction 38
2.5.1 Iterative Mechanisms in Boosting and Bagging 38
2.5.2 Hard Example Mining 42
2.5.3 Adaptive Sample Weighting 43
2.5.4 Emerging Flexible Iteration Approaches 46
2.6 Summary and Research Gaps 48

CHAPTER 3 METHODOLOGY 50
3.1 Empirical Data Acquisition and Integrating 51
3.2 Data Stratification and Preprocessing 54
3.2.1 Data stratification 54
3.2.2 Data preprocessing 55
3.3 Designing DWARFB 59
3.3.1 Sampling Adjustment Strategy 61
3.3.2 Misclassified Samples Reweighting Mechanism 62
3.3.3 Dynamic Adaptive Performance-Driven Ensemble Strategy 64
3.3.4 Optimal Classification Threshold Selection and Final Decision 66
3.4 Evaluation Metrics 67
3.4.1 Preliminary Evaluation 67
3.4.2 Classification Evaluation Metrics 69
3.4.3 The statistical test metrics 72
3.5 Summary 73
CHAPTER 4 DWARFB IMPLEMENTATION AND EXPERIMENTAL STUDY 75
4.1 Experimental Tools 75
4.1.1 Software Environment 76
4.1.2 Hardware Configuration 77
4.1.3 Development & Debugging Tools 77
4.2 DWARFB Implementation Architecture 78
4.2.1 Parameters Setting 78
4.2.2 Positive-to-Negative Sampling Strategy 81
4.2.3 Misclassification Tracking with Reward Mechanism 82
4.2.4 Misclassified Sample Reweighting 83
4.2.5 Performance-Driven Dynamic Ensemble Strategy 84
4.3 Comparative Study 85
4.3.1 Baseline Machine Learning Models 85
4.3.2 RF Variants 88
4.3.3 Resampling Methods 90
4.3.4 Misclassification Penalty Methods 92
4.3.5 Statistical Significance Testing 94
4.4 Parameter Sensitivity and Robustness 95
4.4.1 Parameter Scope Definition 95
4.4.2 Grid Search Optimization 98
4.5 Experimental Study 99
4.5.1 Experimental Objectives 100
4.5.2 Experimental Data Set 101
4.5.3 Experimental Design 101
4.5.4 Evaluation Metrics 104
4.6 Summary 105
CHAPTER 5 EXPERIMENTAL RESULTS AND DISCUSSION 107

5.1 Model Evaluation and Feature Analysis 107
5.1.1 Evaluation of DWARFB's Predictive Performance 108
5.1.2 Persistent Misclassified Patterns Analysis 110
5.1.3 Feature Importance Contribution 112
5.1.4 Discussion 114
5.2 The Comparison with Variety Models 118
5.2.1 Overall Performance Analysis 118
5.2.2 Confusion Matrix Comparative Analysis 124
5.2.3 The Statistical Analysis with Other Models 125
5.2.4 Discussion 128
5.3 The Comparison with RF Variants and Baseline Resampling Methods 131
5.3.1 Overall Performance Analysis 131
5.3.2 Confusion Matrix Comparative Analysis 135
5.3.3 Statistical Significance Analysis with Other Sampling Methods 136
5.3.4 Discussion 138
5.4 Parameter Sensitivity Analysis 141
5.4.1 Misclassified sample strategy Parameter Analysis 142
5.4.2 Base RF Parameters Analysis 146
5.4.4 Discussion 148
5.5 Summary 150
CHAPTER 6 CONCLUSIONS AND FUTURE WORKS 152
6.1 Conclusions of This Thesis 152
6.2 Summary of Contributions 154
6.2.1 Addressing Class imbalance using multi-stage sampling 154
6.2.2 History-aware misclassified samples reweighting 155
6.2.3 Dynamic adaptive performance-driven ensemble strategy 155
6.3 Conclusions in Relation to the Research Questions 157
6.4 Limitations of the Study 159
6.4.1 Data-related limitations 159
6.4.2 Methodological and computational limitations 159
6.4.3 Training Process Adaptability Limitations 160
6.5 Future Works 161
6.5.1 Data and Feature Expansion 161
6.5.2 Integration with Deep Learning for Temporal Dynamics 161
6.5.3 Industry-Specific Customization 162
6.5.4 Scalability and Efficiency 162
6.5.5 Enhancing Training Process Adaptability 163
6.6 Summary 163
Reference 165
Appendix A Dataset Related Information 177
Appendix B Experimental Results 183

LIST OF TABLES

Table **Page**

Chapter 2 Background and literature review

Table 2.1 Z-Score Values and Corresponding Financial Status 20

Chapter 3 Methodology

Table 3.1 Dataset Composition 51

Table 3.2 Dataset Basic Information 53

Table 3.3 Benchmark Concepts in Misclassification Handling 63

Table 3.4 Cohen's Kappa Interpretation 68

Chapter 4 DWARFB Implementation and Experimental Study

Table 4.1 Detailed Computer Specifications Configuration 77

Table 4.2 Summary of DWARFB Parameters 79

Table 4.3 Benchmark Concepts in DWARFB Misclassification Weighting 93

Table 4.4 Weight Formulas for Comparison 93

Table 4.5 Fixed DWARFB Parameters for Sensitivity Analysis 96

Chapter 5 Experimental Results and Discussion

Table 5.1 Comparative Results Across Benchmark Weight Types 110

Table 5.2 Top 20 Features Before and After DWARFB Training 113

Table 5.3 Description of Top 20 Features Contributing to DWARFB 116

Table 5.4 Overall Performance Comparison Across Models 118

Table 5.5 Confusion Matrices of Seven Models 124

Table 5.6 Overall Comparative Results of Seven Models 125

Table 5.7 Pairwise Wilcoxon Test: DWARFB vs. Other Models 128

Table 5.8 Parameter Ranges for Baseline Model Optimization 132

Table 5.9 Optimized Parameters for Comparative Methods 132

Table 5.10 Overall Results Compared with Baseline Methods 133

Table 5.11 Confusion Matrices Compared to Other Methods 136

Table 5.12 Statistical Performance of DWARFB and Baseline Models 136

Table 5.13 Pairwise Wilcoxon Test: DWARFB vs. Benchmarks 137

Table 5.14 Parameter Performance Summary Matrix 144

Table 5.15 Parameter Sensitivity Ranking 144

Table 5.16 RF Performance by Single-Parameter Grouping 147

Appendix A Data Related Information

Table A.1 Feature Names and Corresponding Indicators 177

Appendix B Experimental Results

Table B.1 Results after Winsorization (Outlier Treatment) 184

Table B.2 (a) Feature Importance Ranking (Rank 1–126) 210

Table B.2 (b) Feature Importance Ranking (Rank 127–252) 211

Table B.2 (c) Feature Importance Ranking (Rank 253–378) 212

Table B.2 (d) Feature Importance Ranking (Rank 379–504) 213
Table B.2 (e) Feature Importance Ranking (Rank 505–630) 214
Table B.2 (f) Feature Importance Ranking (Rank 631–756) 215
Table B.2 (g) Feature Importance Ranking (Rank 757–882) 216
Table B.2 (h) Feature Importance Ranking (Rank 883–1008) 217
Table B.2 (i) Feature Importance Ranking (Rank 1009–1066) 218
Table B.3 Base Random Forest Parameter Sensitivity Analysis 219
Table B.4 Overall Performance Across Parameter Combinations 221

LIST OF FIGURES

Figures **Page**

Chapter 1 Introduction

Figure 1.1 Progressive Multi-Stage Sampling Strategy 12

Figure 1.2 Cumulative Misclassification-Aware Weighting Mechanism 13

Chapter 3 Methodology

Figure 3.1 Data Integration Process 52

Figure 3.2 Structure of the Proposed Framework 55

Figure 3.3 Workflow of the Proposed Algorithm Framework 59

Chapter 4 DWARFB Implementation and Experimental Study

Figure 4.1 Iterative Sampling Process of DWARFB 81

Figure 4.2 Misclassification Tracking with Reward Mechanism 82

Figure 4.3 Misclassified Sample Reweighting Workflow 84

Figure 4.4 Experimental Workflow 102

Chapter 5 Experimental Results and Discussion

Figure 5.1 Performance Radar of DWARFB 108

Figure 5.2 Confusion Matrix of DWARFB 109

Figure 5.3 Misclassified Distress Samples During Training 111

Figure 5.4 ROC Curve Comparison of Seven Models 121

Figure 5.5 PR-AUC Curve Comparison of Seven Models 122

Figure 5.6 AUC Distribution Comparison: DWARFB vs. Other Models 126

Figure 5.7 PR-AUC Distribution Comparison: DWARFB vs. Other Models 126

Figure 5.8 ROC Comparison: DWARFB, RF Variants, and Resampling 134

Figure 5.9 PR-AUC Comparison: DWARFB, RF Variants, and Resampling 135

Figure 5.10 Sensitivity Analysis: Alpha Parameter 142

Figure 5.11 Sensitivity Analysis: Gamma Parameter 143

Figure 5.12 Interaction Effects on F1, Precision, Recall, and MCC 145

Appendix A Data Related Information

Figure A.0.1 DWARFB Parameters 182

Appendix B Experimental Results

Figure B.0.1 Correlation Heatmap of Highly Correlated Features 183

LIST OF ABBREVIATIONS

Abbreviations	Descriptions
AdaBoost	Adaptive Boosting
ADASYN	Adaptive Synthetic Sampling
CatBoost	Categorical Boosting
CSMAR	China Stock Market & Accounting Research
CV	Cross Validation
DT	Decision Tree
DWARFB	Dynamic Weight-Adjusted Random Forest Boost
ENN	Edited Nearest Neighbors
LR	Logistic Regression
MCC	Matthews Correlation Coefficient
MLP	Multilayer Perceptron
OHEM	Online Hard Example Mining
PR-AUC	Precision–Recall – Area Under the Curve
ROC-AUC	Receiver Operating Characteristic – Area Under the Curve
RUS	Random Under-Sampling
RF	Random Forests
SMOTE	Synthetic Minority Over-sampling Technique
ST	Special Treatment
SVM	Support Vector Machine
XGBoost	eXtreme Gradient Boosting

CHAPTER 1 INTRODUCTION

Financial distress occurs when a firm's financial position weakens to the extent that servicing its obligations becomes increasingly challenging. This situation is frequently associated with internal issues, such as ineffective managerial decisions and an overreliance on borrowed capital, as well as adverse external conditions [1]. The ability to identify firms facing financial deterioration at an early stage equips investors, creditors, auditors, regulators, and corporate managers with valuable information for risk assessment, strategic planning, and timely corrective action [2]. Early detection helps organizations implement strategic interventions to prevent insolvency and other adverse outcomes.

Fakhar et al. [1] c mapped the scientific literature on banking-sector financial distress through bibliometric evaluation, revealing valuable insights into key knowledge clusters and trends, underscoring the importance of this area of study. Atanasovski and Tocev [3] investigated the extent to which financial distress serves as an indicator of subsequent corporate bankruptcy and discussed the broader economic consequences of financial crises.

Many studies have examined financial distress across different industries. These include hospitality [4], [5], [6], manufacturing [7], [8], [9], [10], [11], and banking [12], [13], [14], [15], [16]. Furthermore, the COVID-19 pandemic amplified the need for robust predictive models to anticipate distress and prevent widespread economic fallout [17], [18], [19], [20], [21].

Historically, financial distress prediction relied heavily on simple statistical models such as financial ratio analysis, and trend analysis and expert

judgment. [22], [23] to distinguish financially healthy firms and distressed firms. Although these approaches provided valuable insights, their effectiveness is limited by an inability to fully characterize the complex and nonlinear relationships that commonly exist among predictive variables. These inherent limitations make it difficult for such methods to extract the rich structural information and complex variable dependencies characteristic of high-dimensional financial datasets [24]. To address these issues, many researchers adopting machine learning techniques in predicting financially distressed firms.

Machine learning enables computational systems to build predictive knowledge directly from data, allowing them to adapt to previously unseen observations without requiring handcrafted decision rules. Driven by ongoing advances in machine learning and data analytics, more sophisticated predictive models have been introduced to handle complex financial data [25]. Recent advances in machine learning have led to the widespread use of methods such as Support Vector Machines, Neural Networks, and tree-based learning algorithms, has substantially enhanced financial prediction by enabling models to learn complex data representations and multidimensional feature dependencies that are difficult for conventional approaches to capture. Among these approaches, ensemble learning methods have gained considerable attention as they synthesize information from diverse predictive models, producing more dependable classification outcomes and greater resilience to variations in the data [26].

Despite these advances, financial distress prediction remains challenging. Real-world financial datasets are typically characterized by several inherent issues. These issues include severe class imbalance, high feature

redundancy, noise contamination, and the presence of difficult-to-classify observations. These issues significantly affect the reliability and performance of predictive models.

The rarity of financially distressed firms within corporate datasets produces a highly skewed class distribution, posing a significant challenge for predictive model development. In practical financial datasets, observations of distressed firms often account for only a small proportion of the overall sample [27]. Consequently, these models often struggle to learn the distinctive patterns of distressed firms, leading to reduced predictive performance for the minority class.

In addition to class imbalance, the diverse characteristics and complex dependencies within financial data create further obstacles for effective predictive modelling, as they frequently contain redundant or highly correlated features. High-dimensional financial indicators increase computational complexity and may cause the model to capture noise and dataset-specific characteristics instead of underlying patterns, leading to weaker predictive performance when evaluated on unseen data.

Furthermore, some firms are persistently difficult to classify because their financial characteristics lie near decision boundaries or exhibit unusual patterns. Traditional machine learning algorithms typically treat all samples equally or only consider misclassification errors within a single iteration. Consequently, they fail to systematically identify and reinforce learning on persistently misclassified but informative samples.

Recognizing the shortcomings of existing financial distress prediction methods, this study develops a Dynamic Weight-Adjusted Random Forest

Boost (DWARFB) framework. The proposed approach integrates the strengths of Random Forest and boosting strategies while incorporating adaptive sampling and misclassification-aware weighting mechanisms. By dynamically emphasizing persistently misclassified samples and adjusting learning strategies across iterations, the framework is intended to strengthen learning from underrepresented distressed-firm observations, thereby improving their identification in financial datasets with markedly uneven class distributions.

1.1 Motivation

In real-world financial scenarios, distressed firms typically constitute only a small proportion of observations [28]. This causes conventional models to derive most of their predictive behaviour from the abundant healthy-firm samples while failing to identify rare but high-impact distress events. For the prediction model, accurate identification of distressed firms is particularly critical, as failing to detect a distressed company can lead to disproportionately large economic and social consequences [29].

Conventional methods for handling class imbalance, including over-sampling, under-sampling, and hybrid resampling techniques, only partially mitigate this problem [30], [31], [32], [33]. Over-sampling risks introducing noise or overfitting to duplicated cases [34], [35], [36], while under-sampling discards potentially useful information from the majority class [37], [38], [39].

To address the limitations of independently applied resampling techniques, some studies have explored embedding resampling techniques directly into ensemble learning frameworks, demonstrating their effectiveness in mitigating severe class imbalance in financial distress prediction [40], [41]. In particular, resampling techniques have been tightly integrated into boosting

[42], [43] and bagging [44], [45] ensembles, enabling data-level balancing to interact with algorithm-level learning and thereby improving minority-class detection. However, existing financial distress prediction studies often tackle resampling, feature selection, or ensemble tuning independently [46], [47]. Such combinations of resampling techniques with ensemble algorithms rarely achieve true integration into the model training process [48], [49]. In most cases, resampling remains an independent preprocessing step within the ensemble framework. This case leaves many of the inherent limitations of resampling, such as noise introduction, information distortion, destabilizing model explainability, and limited adaptivity, largely unresolved [50], [51].

In contrast, an increasing number of studies highlight the contribution of cost-sensitive learning to learning financial distress patterns from imbalanced datasets [52], [53]. Through differentiated misclassification costs that place greater learning emphasis on underrepresented observations, these methods avoid the need to generate synthetic samples while achieving higher distress case detection rates and improved geometric mean [54], [55]. Rather than altering the data distribution, cost-sensitive approaches embed asymmetric error costs directly into the optimization process. They embed the mechanisms such as weighted loss functions, penalty-adjusted boosting, meta-cost wrappers, multi-objective ensemble selection, or hybrid designs that jointly consider sample hardness, feature correlation, and misclassification costs [56], [57], [58], [59], [60]. Cost-sensitive learning has become an effective strategy for tackling class imbalance by imposing higher misclassification costs on minority classes, thereby biasing the learning process toward improved minority recognition.

However, cost-sensitive learning still suffers from several practical limitations:

Cost-sensitive learning lies in the requirement to carefully tune the relative misclassification costs between majority and minority classes [46], [61]. In practice, these costs are rarely known a priori and must be manually selected or heuristically estimated. If the assigned cost for the minority class is excessively high, the classifier may be overfit to minority samples, leading to a substantial degradation in overall accuracy and poor generalization to unseen data [62], [63]. Conversely, if the cost is set too low, the learning bias may be insufficient to counteract class imbalance, resulting in continued underprediction of minority instances.

Conventional cost-sensitive learning relies on static, class-level misclassification costs [64]. Most methods assign a single fixed weight to all samples belonging to a given class, implicitly assuming that all samples from the same class are equally informative and equally difficult to classify [65]. However, real datasets often contain heterogeneous samples, including borderline cases, outliers, and hard-to-classify instances that may require more nuanced treatment [66], [67]. By ignoring sample-level difficulty, static cost schemes fail to allocate learning focus dynamically where it is most needed. As a result, persistently misclassified or ambiguous samples may remain poorly modeled, despite their crucial role in shaping the decision boundary.

Cost-sensitive learning is particularly susceptible to label and feature noise, both of which are prevalent in real-world imbalanced datasets [64], [68]. Particularly, financial datasets often comprise hundreds of correlated or redundant variables [69], [70], which may lead the model to learn noise and

incidental patterns from the training data instead of generalizable relationships, thereby reducing its effectiveness on unseen samples. Under such conditions, increasing misclassification costs for minority classes may unintentionally magnify the influence of mislabelled instances or noisy features. Consequently, the learning process can become biased toward corrupted samples, leading to distorted decision boundaries and degraded generalization performance.

To conclude, this research is motivated by persistent challenges in financial distress prediction that remain insufficiently addressed by existing imbalance-handling strategies. Although resampling-based and cost-sensitive approaches have demonstrated partial effectiveness, current solutions are often fragmented, rely on static and undynamic tuned mechanisms, and lack deep integration within ensemble learning frameworks. Existing methods often struggle to simultaneously tackle class imbalance, persistent hard-to-classify samples, feature redundancy, and noise sensitivity coherently. The reliance on fixed resampling or cost schemes further limits adaptability, amplifies noisy minority samples, and fails to improve the recognition of consistently hard-to-classify distressed samples. To address these gaps is crucial to develop practical, robust, and interpretable early-warning systems capable of reliable financial distress prediction under real-world conditions.

1.2 Research Problem Statement

As discussed in the motivation section, the financial distress prediction methods face several unresolved challenges. These difficulties arise not only from methodological limitations identified in the literature but also from the practical demands of real-world financial risk management, where early, accurate, and stable detection of distress cases is critical.

In response to these limitations, this thesis focuses on addressing the following four core problems.

1.2.1 Fragmented and Static Handling of Class Imbalance

Financial distress events are inherently rare, resulting in datasets dominated by non-distressed firms. Existing imbalance-handling techniques, particularly generic balancing the dataset through sample addition or removal may modify the intrinsic characteristics of the original data, affecting how underlying patterns are represented during training [36] or discard valuable information [39], leading to unstable and poorly generalizing models [71]. While ensemble-based variants of resampling have been proposed, imbalance mitigation is still frequently treated as an external preprocessing step rather than being integrated into the learning process. Consequently, there remains an unresolved challenge in developing imbalance-handling mechanisms that preserve the statistical structure of financial data while consistently improving generalization performance for rare but critical distress cases.

1.2.2 High-Dimensional Feature Redundancy and Weak Representation

The models used for predicting financial distress typically rely on large sets of accounting, market, and macroeconomic variables, many of which are highly correlated or redundant. This redundancy increases computational complexity and exacerbates overfitting, particularly in imbalanced settings where minority-class samples are scarce. Conventional feature selection methods, such as correlation filtering or variance-based thresholds, often ignore the temporal dependencies and domain-specific characteristics inherent in financial data. Moreover, raw financial indicators frequently lack

transformation into higher-level representations that capture latent trends or early warning signals.

Although feature redundancy and data quality issues can influence model effectiveness, they are not the primary focus of this research. Instead, they are addressed through supporting data preparation procedures, including Winsorization and Random Forest-based feature importance screening, to improve data robustness and reduce the influence of less informative variables. The primary research focus remains on developing adaptive ensemble learning mechanisms for handling class imbalance and improving minority-class detection in financial distress prediction.

1.2.3 Inadequate Treatment of Persistently Hard-to-Classify Distress Cases

Some distressed firms consistently lie near classification boundaries due to their complex financial conditions. These cases contain important early-warning information but are often misclassified by conventional algorithms. Most existing ensemble and cost-sensitive learning methods assume uniform importance for samples within the same class or rely on single-round misclassification signals. Such assumptions fail to capture heterogeneous sample difficulty, particularly for borderline distressed samples near decision boundaries. Without mechanisms to identify persistently hard but informative samples and distinguish them from noisy minority instances, models may allocate learning effort inefficiently, leading to unstable minority-class performance and distorted decision boundaries.

1.2.4 Rigid and Fragmented Ensemble Learning Frameworks

Existing ensemble-based financial distress prediction models are constrained by rigid learning dynamics and fragmented system design. Most ensembles employ fixed iteration, sampling, and weight update mechanisms that do not accumulate historical knowledge of model weaknesses across training rounds. This rigidity prevents progressive refinement of learning focus and limits the model's capacity to adapt as training evolves. Furthermore, imbalance handling, feature processing, and ensemble optimization are typically addressed as separate components, resulting in disjointed pipelines that hinder reproducibility and complicate real-world deployment. The absence of a unified, misclassification-history-driven, and dynamically adaptive ensemble framework, therefore, represents a distinct and unresolved research problem.

In the next section, we will discuss the solution to these problems in our proposed ensemble framework.

1.3 DWARFB: Prediction Framework

This research seeks to improve the reliable identification of financially distressed samples in highly imbalanced, redundant, and noisy datasets.

The current body of research on financial distress prediction suggests that the relative performance of different prediction paradigms often depends on dataset characteristics and modelling settings, and no single approach consistently demonstrates superior performance across all scenarios. While several resampling-based, ensemble-based, and cost-sensitive methods have emerged, and some integrate sampling with ensemble learning, existing approaches still face limitations such as insufficient emphasis on persistently

hard-to-classify samples, limited adaptability to evolving data distributions, and reduced robustness in highly imbalanced and noisy financial datasets. These challenges are exacerbated by the heterogeneous characteristics of financial data, including extreme class imbalance, feature redundancy, and noise contamination, which hinder the reliable identification of rare, distressed firms. To address these issues, this study sets Random Forest (RF) as the base learner and embed it within a boosting framework to further enhance its discriminative ability and robustness in detecting financially distressed samples.

The proposed framework is built upon RF, whose characteristics make it well suited to the learning task because its robustness and high accuracy in complex, high-dimensional classification tasks [72], [73], [74], [75]. As a bagging-based ensemble approach, RF creates a diverse set of Decision Trees (DTs) using randomly sampled training data and subsets of input variables. The collective decision of these trees provides a more reliable prediction than any single tree, allowing RF to handle complex financial data containing redundant variables, correlated predictors, and noisy signals effectively [76], [77], [78], [79]. In addition, RF provides feature importance measures, which help identify the most informative financial indicators for distress prediction [80], [81], [82].

However, although RF effectively reduces variance through ensemble averaging, it does not explicitly emphasize difficult or minority-class observations during training. This limitation is particularly important in financial distress prediction, where distressed samples typically represent a small minority of the dataset. Consequently, standard RF models may still exhibit limited sensitivity to rare but critical distress signals.

To address this limitation, a boosting framework is introduced to complement RF. Boosting sequentially adjusts the learning process by increasing the emphasis on misclassified samples, thereby reducing model bias and improving the detection of difficult observations [83], [84], [85]. By integrating RF into a boosting framework, each iteration focuses more strongly on previously misclassified or hard-to-predict samples, allowing the model to progressively improve its ability to identify distress cases in imbalanced datasets.

Therefore, the proposed approach combines the complementary strengths of the two techniques: RF provides stable and low-variance base learners capable of handling high-dimensional financial data, while boosting introduces an adaptive reweighting mechanism that emphasizes difficult samples and reduces bias. The integration of these two mechanisms enables the model to achieve improved robustness and enhanced sensitivity to minority-class financial distress observations. Thus, the integration of RF and boosting forms a complementary learning mechanism that simultaneously addresses variance, bias, and class imbalance challenges in financial distress prediction.

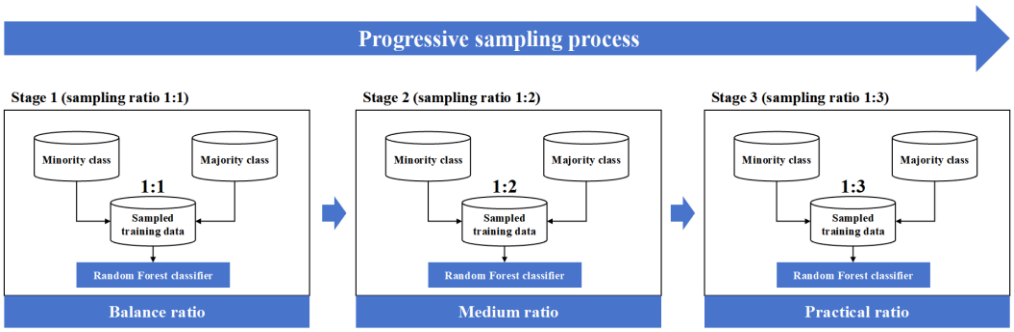

Figure 1.1 Progressive Multi-Stage Sampling Strategy

In the proposed ensemble algorithm, instead of handling class imbalance as an external preprocessing step in the ensemble framework, we keep the

original statistical structure of the data and mitigate class imbalance through progressive sampling strategies that adapt class ratios across iterations. Figure 1.1 outlines the progressive multi-stage sampling strategy.

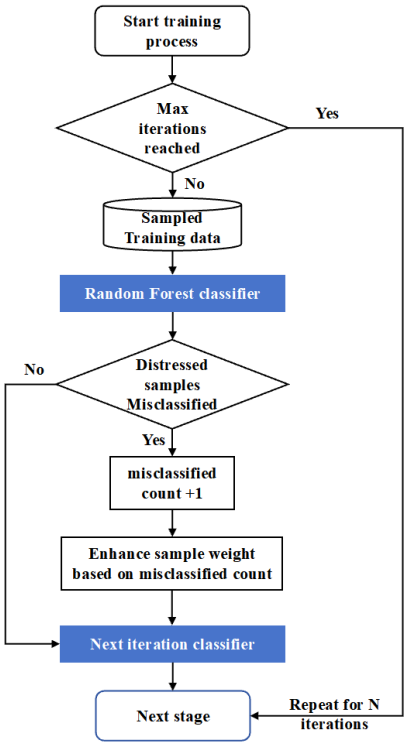

Figure 1.2 Cumulative Misclassification-Aware Weighting Mechanism

Figure 1.2 shows that DWARFB also develops a cumulative misclassification-aware weighting mechanism to emphasize persistently difficult distressed cases. In the iterative RF training process, each distressed sample that is misclassified by the base model is tracked by incrementing its misclassification count. This cumulative information is then used to adaptively increase the sample's weight in the next loop to guide the learning process toward distressed instances that remain difficult to classify across successive iterations.

1.4 Research Questions and Hypotheses

To address the major limitations and unresolved challenges identified through a systematic review of previous research, this study formulates the following research questions. The two research questions are shown as follows:

RQ1: Can we address severe class imbalance in financial distress prediction in a manner that preserves the original data distribution while adaptively improving the identification of rare but economically critical distressed firms?

In this context, addressing class imbalance refers to embedding adaptive imbalance-handling mechanisms directly within the learning process rather than relying on static or preprocessing-based resampling strategies.

RQ2: Can financial distress prediction models explicitly identify and emphasize persistently hard-to-classify yet informative distressed samples, while avoiding disproportionate influence from noisy or mislabelled minority samples?

Persistently hard-to-classify samples denote borderline observations that are repeatedly misclassified across training iterations and contain meaningful early-warning signals rather than random noise.

These research questions collectively define the overarching objectives of this research, which is to establish a dynamic learning framework for financial distress under severe class imbalance that integrates a persistent misclassification-aware reinforcement strategy with a history-aware dynamic weighting mechanism. Therefore, the hypotheses evaluate the expected outcomes of the proposed methodological contributions. The hypotheses are as follows:

H1: Adaptive imbalance-handling mechanisms that are embedded within the learning process will achieve significantly better minority-class generalization performance than static, preprocessing-based resampling methods, without distorting the statistical structure of financial distress datasets.

H2: Modelling cumulative sample-level difficulty via persistent misclassification-aware weighting improves the identification of distressed samples near decision boundaries and reduces the influence of noisy minority instances, outperforming uniform and iteration-local weighting strategies.

H3: Ensemble frameworks that incorporate history-aware learning dynamics and unified integration of imbalance handling, feature processing, and ensemble optimization will produce more stable, interpretable, and robust financial distress predictions than fragmented and rigid ensemble pipelines.

1.5 Contributions

Based on the problems identified in Section 1.2, the purpose of this study is to establish a multi-stage, adaptive ensemble learning framework for addressing class imbalance and hard-to-classify samples in financial distress prediction. This aim gives rise to the following contributions:

- The recognition of severe class imbalance and the presence of persistently hard-to-classify distressed cases in distress predicting highlight the limitations of conventional ensemble learning models that rely on fixed sampling ratios and iteration-local weighting strategies. In real financial datasets, distressed firms are rare, heterogeneous, and often located near decision boundaries, making them difficult to identify accurately using static resampling or uniform reweighting schemes. Moreover, misclassifications are not random but tend to recur for

specific firms exhibiting ambiguous or noisy financial patterns. This realization motivates the development of a multi-stage, adaptive ensemble learning framework that explicitly models both dynamic class imbalance and cumulative sample-level difficulty, enabling the learning process to progressively concentrate on truly informative and persistently challenging distressed cases, rather than being misled by noise or transient misclassification errors.

- The design and implementation of adaptive multi-stage sampling strategies that progressively adjust sampling ratios during training provide flexible control over class distributions and directly address the class imbalance problems identified in Section 1.2.
- The development of an iterative sampling adjustment mechanism that dynamically adapts class ratios across boosting iterations. This mechanism enhances minority-class (distressed samples) representation while preserving sufficient majority-class information, thereby improving discrimination capability in highly imbalanced scenarios.
- The formulation of a dynamic, misclassification-aware weighting function with tunable parameters, including an amplification factor and a non-linear adjustment exponent, to adaptively enhance the importance of persistently misclassified distressed samples.

1.6 Structure of Thesis

The thesis comprises six chapters and two appendices. Chapter 2 provides a comprehensive review of the theoretical background and existing research on financial distress prediction, with a focus on the evolution from statistical models to machine learning and ensemble-based approaches. It also

discusses the challenges of class imbalance and reviews strategies for addressing them, including data-level methods, algorithm-level techniques, and iterative learning mechanisms. The chapter concludes by consolidating the limitations identified in existing research and demonstrating how they guide the design of the proposed framework.

Chapter 3 shows the methodological foundation of this research. It describes empirical datasets, preprocessing procedures, and feature engineering strategies used to construct robust financial indicators. A detailed explanation of the proposed Dynamic Weight-Adjusted Random Forest Boost (DWARFB) framework is provided, including its sampling strategy adjustments, misclassified-sample weighting mechanism, adaptive ensemble strategy, and classification threshold optimization. The evaluation metrics employed to assess performance are also introduced.

Chapter 4 presents the DWARFB implementation and the experimental design underpinning this research. It includes a comparative study against multiple baselines, RF variants, and benchmark misclassification-penalty methods. The chapter also describes parameter sensitivity analyses and optimization procedures to assess the consistency and reliability of DWARFB.

Chapter 5 reports the empirical findings, including comparative evaluations of benchmark models, analysis of persistent misclassified patterns, and feature contribution. It also provides comparative analyses across competing models and sampling strategies. Additionally, the chapter examines parameter sensitivity, focusing on misclassified-sample strategies and RF configurations.

In Chapter 6, we conclude our research work, including the study's main contributions, limitations and opportunities for further investigation.

CHAPTER 2 BACKGROUND AND LITERATURE REVIEW

This chapter provides a comprehensive review of the literature related to these problems and focuses on the corresponding methodological developments in ensemble learning models that have been proposed to address them.

2.1 Trends in Financial Distress Prediction Modelling

The development of methods for predicting corporate financial distress has been a long-standing focus of academic research. Altman's Z-score is the earliest and most influential model, which was developed to predict corporate bankruptcy [86]. Since then, the models used for predicting bankruptcy have progressed from conventional statistical techniques to approaches using machine learning that more effectively capture the underlying complexity of financial data beyond linear assumptions [87]. This evolution has further progressed toward ensemble learning techniques, which integrate multiple base learners to improve robustness, accuracy, and stability, particularly when handling noisy, high-dimensional, and highly imbalanced financial data.

2.1.1 Traditional Statistical Models

Traditionally, scholars and practitioners have relied on statistical models to assess the likelihood of corporate insolvency or bankruptcy. The first and most influential model was Altman's Z-score [88], which used linear discriminant analysis to combine various financial indicators into a composite score. This provided a simple yet effective early warning indicator of corporate bankruptcy. The model incorporated five categories of financial indicators. The equation is expressed as follows:

$$\text{Z-score} = 1.2(\frac{\text{WC}}{\text{TA}}) + 1.4(\frac{\text{RE}}{\text{TA}}) + 3.3(\frac{\text{GE}}{\text{TA}}) + 0.6(\frac{\text{ME}}{\text{TL}}) + 1.0(\frac{\text{SA}}{\text{TA}}) \quad \text{Eq. 2.1}$$

Equation 2.1 incorporates seven financial variables: WC for working capital, TA for total assets, RE for retained earnings, GE for earnings before interest and tax, ME for the market value of equity, TL for total liabilities, and SA for sales. These ratios capture a firm's liquidity (WC/TA), profitability (RE/TA and GE/TA), leverage/solvency (ME/TL), and operational activity (SA/TA) [89]. Table 2.1 presents the Z-score values and their corresponding financial statues of distress.

Table 2.1 Z-Score Values and Corresponding Financial Status

Z-score's value	**Financial status**
Below 1.8	High risk of bankruptcy
1.8 -3.0	Financially health
Above 3.0	Excellent financial health

Z-score has guided the practice of corporate bankruptcy prediction for several decades, shaping how analysts and researchers identify and interpret key financial indicators related to financial risk. By integrating multiple accounting indicators into a composite score, the Z-score and its alternatives, such as the O-score [90] and X-score[91], inspired extensive research into refining and expanding indicator-based models for early warning purposes.

LR is another widely adopted method due to its interpretable nature and its appropriateness for modelling binary outcomes [92], such as distinguishing between healthy and distressed firms. Other methods, including multivariate discriminant analysis and probit models [93], have also been explored to improve prediction accuracy. However, these models assume linear relationships, overlook non-financial factors (e.g., market sentiment), and

struggle to adapt to dynamic economic environments, restricting their capacity to catch the complex interactions and nonlinear dependencies among financial variables.

The reliance of conventional statistical models on the linear nature of these approaches and their limited capacity for handling complex and non-linear interactions in data exposed significant shortcomings, particularly in dynamic and high-dimensional settings. These limitations spurred the adoption of machine learning techniques, which offered greater flexibility in modelling non-linear relationships and capturing intricate patterns, marking a significant shift in financial distress prediction research.

2.1.2 Emergence of Machine Learning Approaches

Since 1990s, research on predicting financial distress began to shift toward the application of machine learning techniques. These approaches offered greater flexibility than traditional statistical models and were better suited to capturing the highly complex and nonlinear nature of financial variables. As a result, researchers have increasingly employed standalone machine learning models, such as Decision Trees (DTs), K-nearest Neighbours [94], Support Vector Machines (SVMs) [95], [96], and Artificial Neural Networks [97], [98], emerged as effective alternatives for bankruptcy prediction.

Among these methods, DTs construct a hierarchical decision structure by repeatedly choosing the feature and split value that most effectively discriminate between the target classes. Tree construction is generally based on impurity measures, including Gini impurity [99], where each split is chosen to achieve the largest decrease in the weighted impurity of the resulting partitions. The core idea of SVM is to learn a classification boundary that maximizes the

distance between the closest training samples from opposing classes, thereby improving classification robustness in high-dimensional feature spaces [100]. Artificial Neural Networks are feedforward neural networks capable of modelling complex nonlinear relationships between financial indicators and distress outcomes. Their parameters are optimized through backpropagation by minimizing a loss function such as cross-entropy [101].

Empirical studies demonstrate that these machine learning approaches can achieve strong predictive performance in financial distress prediction tasks. SVMs excel in high-dimensional settings, achieving approx. 83% accuracy by navigating complex feature spaces that overwhelm statistical models, while demonstrating strong generalization to new data [102]. Artificial Neural Networks achieve up to 98% accuracy by modelling intricate, non-linear relationships between financial variables, such as the compounding effects of liquidity and leverage on distress, without the linear constraints of statistical methods [38]. DTs offer intuitive IF-ELSE rules for non-linear thresholds, enhancing accessibility [36]. Despite these advances, single models are vulnerable to data imperfections, including outliers, multicollinearity (e.g., correlated profit ratios), and class imbalance, which degrade performance in financial contexts.

While single machine learning models overcome the linear limitations of traditional statistical approaches and provide strong predictive capabilities, their sensitivity to data imperfections, such as class imbalance, multicollinearity, and outliers, remains a significant challenge in financial distress prediction. These issues can lead to unstable predictions and reduced generalizability, particularly in high-dimensional and highly imbalanced financial datasets. To

mitigate the aforementioned limitations, researchers have increasingly adopted ensemble learning strategies, which combine multiple classifiers to leverage their complementary strengths.

2.1.3 Transition to Ensemble Learning

A growing body of research leverages ensemble learning to leverage the complementary strengths of multiple learners, thereby reducing the weaknesses of individual models and delivering superior predictive capability in complex financial environments. Ensemble techniques combine multiple base learners to exploit their complementary strengths, thereby improving prediction accuracy, robustness, and generalization. Common ensemble strategies include bagging [103], [104], [105], [106], boosting [107], [108], [109], stacking [110], [111], [112], [113], and voting [114], each integrating predictions from multiple models to capture diverse patterns within data.

Among these approaches, several ensemble algorithms have been widely adopted in financial distress prediction. RF, developed by Breiman [115], generates an ensemble of decision trees by repeatedly sampling the training data with replacement and introducing randomness into feature selection at each split. Each decision tree contributes one vote, and the class supported by the majority of trees is assigned as the ensemble's final prediction, resulting in lower model variance and improved generalization performance relative to a single decision tree. Boosting algorithms provide another important ensemble strategy. AdaBoost [116] iteratively combines weak learners by increasing the weights of misclassified samples, thereby directing the learning process toward difficult observations. Building on the boosting principle, XGBoost introduces regularization and parallel computation to improve model scalability and

predictive performance [117]. By sequentially adding decision trees that fit the negative gradient of the loss function, XGBoost achieves strong predictive accuracy and computational efficiency.

Numerous studies have applied ensemble learning techniques to financial distress prediction. Sun et al. [46] proposed dynamic financial distress prediction models addressing concept drift and class imbalance, leveraging Adaptive Boosting (Adaboost)-SVM ensembles to assess financial health. Liang et. al. [110] employed stacking ensembles to develop a bankruptcy prediction model that outperformed baseline models, particularly when misclassification costs were high. Du et. al. [40] proposed a heterogeneous ensemble approach combining clustering-based techniques with under-sampling, which showed superior performance in comparative testing. In particular, boosting has gained traction for its ability to iteratively correct the previous models' errors, making it effective in financial distress prediction [118].

In practical applications, RFs [103], [119], along with some boosting models [107], have been widely used to analyse complex financial data and extract meaningful patterns. Yang et. al. [120] highlighted that Gradient Boosting enhanced prediction accuracy and provided interpretable results compared to traditional models, which are essential for detecting early warning signals of financial trouble. XGBoost, combined with techniques such as bagging and RF, often outperforms other methods in predicting company health and financial stability [105].

Rather than relying on a single learning algorithm, these approaches integrate multiple predictive models to produce more consistent, accurate, and

generalizable classification outcomes. By utilizing Ensemble learning approaches, researchers have demonstrated their ability to better capture the multifaceted patterns and uncertainties present in financial problems, thereby enabling more reliable predictions in complex financial scenarios.

2.1.4 Summary of Model Development Trends

The evolution of prediction methods illustrates a steady transition from fixed, assumption-driven statistical methods to more flexible machine learning and ensemble frameworks. While ensemble models have significantly improved predictive accuracy and robustness, these advances have not fully resolved the fundamental data characteristics that limit model performance. Among them, the imbalance issue remains the most influential problem in financial distress prediction persistently.

Regardless of whether traditional statistical models, machine learning methods, or ensemble frameworks are employed, the extreme scarcity of distressed firms relative to healthy ones continues to bias learning and evaluation. Therefore, before reviewing specific methodological solutions, it is necessary to systematically examine the nature, underlying causes, and practical implications of the imbalance issue in datasets, which constitutes the focus of the following section.

2.2 Class Imbalance in Financial Distress Prediction: Challenges and Research Progress

Corporate distress prediction often struggles with the imbalance issue. It refers to the data that typically exhibits a significant disparity between the normal class and the distress class.

In many contemporary financial applications, particularly corporate bankruptcy prediction and credit risk assessment, the distressed class represents only a small part of observations, typically constituting between 2% and 10% of the samples. Safi et al. [121] analyse the dataset of Spanish companies for using Neural Network to predict financial distress, there are 2860 samples in all, and only 62 correspond to insolvency, the insolvency cases only formed 2%. Similarly, in the research of bankruptcy forecasting in enterprises, Gaurav et al. [122] use the dataset contains 6599 records labelled with 'Not Bankrupt' (96.8 % of records) and only 220 (3.2 %) marked as 'Bankrupt'.

The literature consistently highlights the prevalence of class imbalance in corporate financial datasets, with imbalance ratios often ranging from 10:1 to 10,000:1, depending on the dataset and context. A comprehensive survey about big data defines the class imbalance ratio of the majority and minority, which is particularly relevant for predicting corporate distress due to the rarity of default events. [123]. The analysis of datasets with over 100,000 instances underscores the severity of imbalance in financial applications, where non-distressed companies dominate. Kim et al. analyzed corporate default prediction models and noted that datasets often exhibit ratios around 30:1, particularly in markets with low default rates, such as developed economies [124]. Yang et al. analyze an imbalance ratio of approximately 200:1 in Chinese bond default data, highlighting the challenge of modelling rare default events in emerging markets [125]. These ratios are often quantified using statistical measures such as the Gini coefficient or entropy, with high values (e.g., Gini > 0.8) indicating severe skewness that complicates predictive modelling [123].

Distressed events, though rare, often lead to substantial financial losses for stakeholders, including creditors, investors, and employees. Zhao et al. [126] indicate that while distress occurrences represent a small proportion of corporate outcomes, they are associated with significant negative impacts on firm value and stakeholder wealth, necessitating accurate early detection. Kim et al. [124] emphasize that misclassifying a distressed company as non-distress (false negative) incurs significantly higher costs, often 10 to 20 times greater, than misclassifying a non-distressed company (false positive), due to the severe consequences of undetected distress. Jan [127] highlights that because financial distress often emerges without transparent warning, highly accurate prediction models are essential to protect investors, creditors, and employees from unanticipated losses.

To tackle class imbalance, scholars have proposed methods at both the data and algorithmic levels, designed to rebalance datasets and enhance the identification of minority-class instances. These strategies represent a pivotal focus for improving model performance in imbalanced settings. The subsequent sections offer a detailed review of the principal techniques and models applied in this domain.

2.3 Data-Level Methods for Class Imbalance Mitigation

Data-level strategies seek to mitigate the poor detection of minority-class instances in imbalanced datasets by applying resampling techniques to rebalance the data. Resampling techniques can generally be categorized into three major groups: oversampling, under-sampling, and hybrid resampling [128]. Each approach seeks to address class imbalance by adjusting the size of

the majority or minority class to enhance the model's prediction power in detecting minority class instances.

2.3.1 Oversampling Techniques

Oversampling techniques address class imbalance by generating synthetic samples for the minority class, thereby improving the performance of machine learning models in detecting rare but critical events.

The commonly used oversampling technique is the Synthetic Minority Oversampling Technique (SMOTE), which generates synthetic samples for the minority class by interpolating new synthetic instances between minority class instances and their nearest neighbors [129]. It has been widely applied in biomedical domains to generate synthetic samples of the minority class in protein-protein interactions [130], breast cancer prediction [131], and medical diagnosis [132]. In some practice studies, Alex and Nayahi [133] reported that SMOTE introduced noise during the process of creating synthetic samples. This could be due to instances that are sparsely distributed, and synthetic samples do not accurately reflect the true characteristics of the minority class.

To alleviate the limitations of SMOTE, Han, Wang, and Mao [134] proposed using Borderline-SMOTE, which focuses on generating samples near the class boundary rather than the entire minority class. It operates on the premise that minority samples far from the class boundary contribute minimally to classification improvement, while those closer to the boundary play a critical role in distinguishing between classes. The effectiveness of Borderline-SMOTE has been reported in fault detection systems [135] and text classification [136], where careful parameter tuning is crucial to prevent misclassification at the class boundaries.

Adaptive Synthetic Sampling (ADASYN) is an advanced oversampling technique that extends SMOTE by adaptively focusing on minority class instances that are harder to learn [137]. The method increases the sampling rate for minority instances that are misclassified or located near the decision boundary, thereby enhancing the classifier's focus on complex regions of the input space. It has been reported to outperform other methods in multiclass imbalanced scenarios when combined with stacking algorithms [138]. Analogous to other oversampling techniques, ADASYN can easily overfit noisy regions if not combined with any other denoising techniques.

2.3.2 Under-sampling Techniques

Unlike oversampling techniques that increase the number of instances for the minority class, under-sampling techniques tackle class imbalance by reducing the number of instances in the majority class.

Random Under Sampling (RUS) is an under-sampling technique that reduces the size of the majority class by randomly removing instances. Due to its random nature, RUS carries the risk of discarding important information, which is why it is often combined with cleaning strategies. Arifin et al. [139] showed that RUS with Decision Tree (DT) and RF classifiers performed better than other resampling techniques.

Tomek Links is another widely used under-sampling technique, which identifies minority-majority pairs where each is the nearest neighbor of the other, and removing these pairs improves class separation [140]. This technique has been applied in credit card fraud detection [141] and abnormal traffic detection for filtering noise samples in the data [142]. However, it may excessively trim borderline instances to cause an overly aggressive reduction in the dataset's

complexity [143], potentially losing valuable, nuanced patterns that could contribute to a more comprehensive understanding of the predictive model's behavior and performance [144].

2.3.3 Hybrid Resampling Techniques

While oversampling and under-sampling techniques can effectively balance datasets, relying on a single resampling method faces inherent challenges: oversampling may introduce noise, whereas under-sampling can discard some useful information. To address these shortcomings, hybrid techniques that combine multiple resampling approaches have been proposed.

SMOTE-Tomek, which combines oversampling and under-sampling techniques, is widely used in enhancing classifier performance for cancer prediction datasets [145]. It combines SMOTE with Tomek Links to remove noisy majority samples near minority instances.

SMOTE-ENN integrates SMOTE with Edited Nearest Neighbors (ENN) to delete misclassified majority samples post-oversampling, refining decision boundaries. It has been used in medical datasets such as missed abortion diagnosis to enhance the model's classification accuracy [146]. Similar to most of the under-sampling techniques, ENN may remove useful samples in its aggressive pruning process.

2.3.4 Summary of data-level techniques

In summary, oversampling techniques such as SMOTE and ADASYN have proven effective in improving minority class detection, particularly in dense regions of the minority class [147]. However, in financial distress prediction, where distressed companies may be highly heterogeneous and sparsely distributed, these methods risk introducing synthetic noise or

generating unrealistic samples [148], potentially distorting the true risk profile of companies. Under-sampling techniques such as Tomek Links can enhance class separation and reduce overlap, but they may also discard valuable information from the majority class, which is especially problematic in financial datasets where each observation may carry unique, informative signals about corporate health.

Hybrid techniques, such as SMOTE-Tomek and SMOTE-ENN, often outperform individual methods by combining data synthesis with noise reduction [149]. Nevertheless, these approaches can lead to over-pruning, inadvertently removing borderline or informative samples along with noise, which may reduce the model's ability to capture subtle patterns critical for early warning in financial distress scenarios. Moreover, the effectiveness and side effects of these resampling techniques are highly dependent on dataset characteristics, such as the degree of imbalance, feature distribution, and the presence of outliers, as well as on computational resources, an important consideration for large-scale financial applications.

Given these limitations, increasing attention has turned to algorithm-level strategies, which address imbalance without destroy the structure of the original data.

2.4 Cost-Sensitive Learning as an Algorithm-Level Technique

The strategies in algorithm-level tackle class imbalance by adjusting the training focus, instead of modifying the underlying data distribution. Cost-sensitive learning has emerged as the leading algorithm-level approach for addressing class imbalance in financial distress prediction, as evidenced by a substantial body of literature [64], [107], [121], [150]. By embedding imbalance

awareness directly into the training process, cost-sensitive methods guide models to place greater emphasis on minority-class observations, making them a principled and flexible framework for managing asymmetric classification risks.

2.4.1 Cost-Sensitive Learning: Principles and Overview

Cost-sensitive learning addresses class imbalance by adjusting the loss function or incorporating misclassification costs directly into the training process. Incorporating class-specific costs into the loss function ensures that the model reduces false negatives for the minority class, which is critical in financial distress prediction [151].

A variety of models have benefited from cost-sensitive strategies:

In ensemble models, such as RF with class weight balancing, also known as cost-sensitive RF [152], adjust the contribution of each sample during tree construction. Minority-class observations exert greater influence on the split criteria, affecting impurity measures such as Gini or entropy. The overall tree structure and feature randomness remain identical to standard RF, but the weighted impurity calculation biases splits toward improving minority-class performance. This approach avoids noise introduced by resampling while effectively counteracting class imbalance.

Cost-sensitive strategies have also been applied to neural network architectures with notable success. Zubair and Yoon [153] propose a cost-sensitive loss function for convolutional neural networks, showing improved minority-class detection in imbalanced medical datasets, a concept transferable to financial distress scenarios. Priatna et al. [154] focus on optimizing Multilayer Perceptron (MLP) with cost-sensitive learning to enhance credit card

fraud detection. Their findings show that incorporating class-dependent costs into the training process significantly enhances detection performance, thereby confirming the effectiveness of cost-sensitive strategies in handling class imbalance in financial applications.

Boosting frameworks extend cost-sensitive learning to gradient boosting methods. Zou et al. [107] develop a Weighted extreme Gradient Boosting (XGBoost-W) for business-failure prediction. They embed a weighted cross-entropy loss into the boosting objective, where the weights reflect the relative misclassification costs of distressed vs. non-distressed firms. Similarly, Yotsawat et al. [54] proposed Cost-Sensitive Extreme Gradient Boosting, which assigns different misclassification penalties to each class to address bankruptcy prediction without resampling, resulting in a higher Bankruptcy Detection Rate and improved Geometric Mean.

Beyond single-model modifications, several studies extend cost-sensitive learning into ensemble and optimization frameworks. De Bock et al. [56] demonstrate that cost-sensitive ensemble selection is crucial when misclassification costs are uncertain, offering a multi-objective NSGA-II–based approach that identifies Pareto-optimal sets of classifiers aligned with different cost scenarios. Papíková and Papík [57] similarly review the effects of cost-sensitive, feature selection, and resampling methods in bankruptcy prediction, pointing out that cost-sensitive algorithms alongside thresholding remain among the most effective. Chan et al. [58] employ the Meta-Cost algorithm combined with attribute selection to address class imbalance in financial distress prediction, illustrating how cost-sensitive classification can improve model robustness. Safi et al. [59] propose a cost-sensitive neural network ensemble

optimized by Particle Swarm Optimizer and its cost-sensitive variant Competitive Swarm Optimizer, where misclassification costs are embedded in the fitness function. This design prioritizes reducing false negatives and achieves superior detection of distressed firms compared with cost-agnostic models across multiple datasets.

In summary, cost-sensitive learning provides a principled and flexible framework for handling class imbalance, enhancing minority-class detection across tree-based models, neural networks, boosting methods, and ensemble frameworks. These approaches are particularly valuable in financial distress prediction, where accurately identifying rare but high-impact events is essential.

2.4.2 Class-Weight Adjustments in Trees/Boosting

Among various cost-sensitive implementations, class-weight adjustment is widely adopted due to its simplicity and effectiveness. By modifying the loss function, these adjustments impose greater penalties on misclassified minority-class observations, increasing their influence during training. This strategy has received considerable attention in financial distress prediction, where class imbalance can severely degrade model performance.

In ensemble approaches, including RF and boosting methods, class-weight adjustments play a key role in guiding the model to focus on difficult cases. For instance, the Balanced RF classifier [155], [156] extends standard RF by constructing balanced bootstrap samples for each decision tree. Unlike traditional RF, it integrates imbalance handling directly into the training process rather than relying on oversampling or under-sampling, reducing the risk of overfitting minority samples while ensuring that each tree contributes meaningfully to minority-class detection. This method is particularly effective

for highly imbalanced datasets, where standard ensembles often produce skewed decision boundaries and low sensitivity to the minority class.

Similarly, in boosting frameworks such as XGBoost, class-weight adjustments influence the gradient calculations and loss evaluation at each iteration. Liu et al. [157] demonstrate that Bayesian optimization of class weights in XGBoost can enhance feature selection and improve predictive accuracy for financial distress. Nevasalmi [158] further analyzes the effect of class weights on gradient boosting with J-terminal node regression trees, showing how they help the algorithm handle skewed class distributions more effectively. In ensemble classifier frameworks, integrating class-weight adjustments alongside feature selection, as reported in [159], improves model balance and sensitivity to the minority class.

These studies collectively highlight that class-weight adjustments are essential for enhancing the performance of both tree-based and boosting models. By modifying loss and gradient calculations, these techniques improve prediction accuracy and robustness in high-dimensional, imbalanced financial datasets.

2.4.3 Threshold adjustment to deal with Class Imbalance

Beyond modifying the loss function, threshold adjustment offers a complementary mechanism by managing the trade-off between false positives and false negatives during the decision-making stage.

This technique involves dynamically setting decision boundaries to account for class imbalance, enabling the model to more effectively identify minority-class instances. Unlike traditional classifiers that use a fixed threshold (e.g., 0.5 for binary classification), threshold adjustment optimizes the decision

boundary based on performance metrics that are more suitable for imbalanced data. Recent research highlights that adaptive thresholding can improve minority class recall in highly skewed datasets, offering a flexible approach for financial distress prediction where minority class detection is critical [160].

The literature suggests that threshold adjustment is not only about the precision-recall trade-off but also about optimizing the rate of true positives in the presence of class imbalance. Techniques such as sampling-based class distribution adjustments [161] and kernel-based parameter tuning [162] also contribute to the broader strategy of threshold calibration, aiming to improve classifier robustness against imbalance.

The threshold adjustment is a critical and versatile tool for addressing class imbalance in financial and other data domains. Whether through direct threshold tuning, algorithmic integration, or domain-specific standards, these strategies intend to improve the fairness, accuracy, and reliability of classifiers that tackle imbalanced data distributions.

2.4.4 Hybrid Cost-Sensitive and Data-Level Approaches

Recent studies increasingly combine cost-sensitive learning with resampling strategies, integrating complementary data-level and algorithm-level strategies to improve predictive performance [163]. Peykani et al. [60] propose a two-phase approach for business-failure prediction. First, correlation-based oversampling generates minority-class samples in meaningful subspaces, reducing the risk of overfitting. Second, cost-sensitive ensemble learning, using AdaBoost, Categorical Boosting (CatBoost), or other base learners, optimizes misclassification costs to strongly penalize false negatives. Experiments on an Iranian capital-market dataset show that the method achieves very high

sensitivity, particularly when CatBoost is used as the base learner. Meanwhile, Building upon the Single Self-Paced Ensemble, Gao et al. (2025) [164] developed the Multi-Hybrid Self-Paced Ensemble to improve financial distress prediction under severe class imbalance. SPE applies a self-paced factor to emphasize difficult and borderline instances while performing controlled under-sampling, preserving each bin's overall hardness contribution. These selected majority samples are combined with all minority-class instances to form a balanced dataset, reducing noise amplification and sample overlap. The model extends this approach by employing multiple heterogeneous base classifiers, improving predictive stability, generalizability, and overall accuracy in complex financial distress prediction.

2.4.5 Summary of Cost-Sensitive Learning Approaches

Overall, cost-sensitive learning has become a central strategy for financial distress prediction, offering a principled alternative to resampling. Rather than altering the dataset, these methods embed asymmetric error costs directly into the optimization process, through weighted loss functions, penalty-adjusted boosting, meta-cost wrappers, multi-objective ensemble selection, or hybrid designs combining hardness-, correlation-, and cost-awareness.

Through diverse studies, cost-sensitive methods consistently enhance minority-class detection, especially by reducing false negatives. Implementations such as class-weight adjustment, threshold adjustment and hybrid cost-sensitive frameworks offer flexible and effective tools for aligning model behaviour with real-world risk preferences. Cost-sensitive learning therefore establishes a robust foundation for handling class imbalance and

bridges traditional optimization objectives with adaptive learning mechanisms discussed in the following sections.

2.5 Iteration-Driven Learning Mechanisms for Financial Distress Prediction

While cost-sensitive learning defines what the model should emphasize, iteration-driven mechanisms focus on how learning should evolve, dynamically reallocating attention to difficult and informative samples over successive training rounds.

Although cost-sensitive learning effectively modifies the loss function to emphasize minority-class errors, it assigns static importance to samples throughout training. In contrast, iteration-driven learning mechanisms dynamically adjust sample importance based on misclassification behaviour and learning progress, enabling models to focus increasingly on hard and informative cases [165]. This section therefore reviews iterative strategies that further enhance minority-class detection by exploiting training dynamics.

2.5.1 Iterative Mechanisms in Boosting and Bagging

Boosting and bagging are foundational ensemble learning techniques that utilize iterative processes to construct predictive models.

Boosting, originally proposed by Freund and Schapire [166], iteratively trains weak learners (e.g., DTs) by assigning higher weights to misclassified instances, thereby focusing on difficult cases in each iteration. TUNIO et al. [167] evaluated AdaBoost as a meta-learner for multiple classifiers, including LR, Neural Network, DT, and SVM, using Pakistan Stock Exchange data in a financial distress prediction setting. Across all models, AdaBoost improved predictive performance, with the AdaBoost-SVM combination achieving 89.07%

accuracy and a 91.9% ROC area, outperforming the baseline SVM. The study highlights AdaBoost's iterative emphasis on misclassified observations, which is particularly beneficial in distress prediction where distressed firms constitute the minority class. This aligns with findings from Sun et al. [46], who showed that AdaBoost enhances minority-class recall by progressively focusing on high-risk misclassified cases. Pavlicko et al. [168] extended boosting research by applying RobustBoost, a noise-tolerant boosting variant, to financial distress prediction in Visegrad Group economies. Unlike AdaBoost, which is highly sensitive to mislabeled or noisy data, RobustBoost incorporates a non-convex loss that adapts during training, improving resilience to data imperfections common in large corporate datasets. The ensemble achieved strong performance (AUC 0.964; accuracy 94.25%), outperforming classical statistical models and individual boosting variants, thereby highlighting the potential of noise-robust boosting approaches for financial distress prediction in emerging European markets.

Bagging, introduced by Breiman [115], constructs multiple weak learners independently on bootstrap samples and aggregates their outputs via majority voting or averaging. In financial distress prediction, bagging methods such as RFs have been shown to reduce variance and enhance robustness against noise. Liu rt al. [169] show that Exactly Balanced Bagging consistently outperforms other machine learning algorithms, including LR, XGBoost, RUSBoost, C4.5 DT, and Roughly Balanced Bagging, in predicting fiscal distress of local governments, regardless of the under-sampling method used. Its superior F1-scores (53.11–55.33%) highlight Bagging's effectiveness in handling imbalanced datasets, suggesting that balanced ensemble sampling is

particularly advantageous for improving prediction accuracy in minority-class instances. Khalil et al. [170] prove that ensemble learning techniques, especially CatBoost and Bagging, are more effective for predicting insurance company insolvency in the Egyptian market than classic machine learning methods. However, their independence across iterations prevents them from adaptively emphasizing the minority class. A recent survey by Le et al. [63] highlight that while bagging-based ensembles remain competitive, they are generally less adaptive than boosting approaches, particularly under severe class imbalance.

The fundamental difference between the two methods lies in their iterative mechanisms: boosting employs sequential dependency, where each learner corrects the errors of its predecessors, strengthening the model's capability to identify distressed firms in imbalanced datasets across iterations. Conversely, bagging treats each learner independently, which helps mitigate overfitting but fails to adaptively focus on persistently misclassified minority-class samples. As emphasized by Sun et al. [46] and Le et al. [63], this rigidity, sequential dependency in boosting and independence in bagging, limits their capacity to address dynamic class distributions in financial distress prediction, where economic conditions often induce temporal shifts in distress patterns.

Some researchers try to explore the combination of boosting and bagging, combining the complementary learning mechanisms of bagging and boosting within a unified framework to improve predictive consistency, increase resilience to noisy and heterogeneous data, and strengthen financial distress prediction. Zou et al. [171] proposed a unified bagging–cascading–boosting framework for corporate financial risk prediction. The model combines Light Gradient Boosting Machine for bias reduction, bagging for

variance control, and a depth-adaptive cascading structure for iterative feature refinement. Using 46 indicators from 2,826 Chinese listed firms, the model achieved strong performance with an AUC of 0.9502 and an F1 score of 0.4274 and still maintained an AUC of 94.09% when only half of the training data were used. Wang et al. [172] propose an adaptive weighted XGBoost-Bagging framework that combines XGBoost with Bagging and an adaptive weighting mechanism. Using stratified non-replacement under-sampling guided by K-Means clustering, all minority class samples are retained while the majority class is under-sampled to balance the data. Multiple balanced datasets train individual XGBoost classifiers, whose minority class predictions are adaptively weighted based on local neighbourhood information and aggregated via weighted soft voting, enhancing the detection of minority class instances in imbalanced datasets. Liu et al. [173] compared Easy Ensemble and SMOTE with benchmark models, including XGBoost, AdaBoost, Gradient Boosting Decision Trees, RF, SVM, LR, and Deep Neural Networks on imbalanced financial data. Easy Ensemble, an under-sampling method combining Bagging and AdaBoost, balances the dataset by sampling the majority class to match the minority class. The results show that Easy Ensemble improves the ability of benchmark models to predict distressed firms compared to SMOTE, with slight decreases in overall accuracy but increases in AUC, H-measure, and Kolmogorov–Smirnov statistics.

These findings demonstrate that combining bagging's stability, boosting's accuracy, and cascading's structural flexibility can enhance robustness and balance the bias-variance trade-off in financial risk early-warning systems.

2.5.2 Hard Example Mining

Predicting financial distress models are inherently challenging due to severe class imbalance, subtle minority-class patterns, and high noise levels in financial indicators. A persistent issue is that certain firms, often those near distress thresholds, are consistently misclassified during training. Standard ensemble methods typically rely on static sample weights, which fail to adapt to these evolving classification challenges [174]. Hard example mining seeks to overcome this limitation by systematically identifying and emphasizing samples that are repeatedly misclassified, thereby improving both minority-class recall and model robustness.

Hard example mining refers to the targeted reinforcement of samples that are most informative for improving decision boundaries. In the financial domain, several strategies have been adopted:

Cost-sensitive learning: Sun et al. [175] integrated misclassification cost matrices into bankruptcy prediction models, penalizing distress-class errors more heavily to counterbalance class imbalance. Similarly, Bock et al. [56] applied asymmetric misclassification costs in business failure prediction, improving sensitivity to rare distress events.

Adaptive boosting: Tanha et al. [176] reviewed boosting methods for imbalanced data, highlighting how algorithms like AdaBoost improve minority-class prediction by iteratively increasing the weights of misclassified samples, thereby enhancing recall in challenging bankruptcy or failure prediction tasks.

Focal loss and margin-based penalties: Yi [177] incorporated focal loss into credit risk scoring models, dynamically scaling the influence of easy versus hard samples and thereby mitigating the dominance of majority-class patterns.

Hybrid sampling with difficulty prioritization: Wang et al. [178] enhanced SMOTE by integrating adaptive SVM-based weighting, generating synthetic minority samples preferentially in harder-to-classify regions of the feature space to reduce persistent misclassification.

While these methods contribute significantly to improving model performance on difficult samples, they also exhibit limitations. Most approaches emphasize hard samples only in the current iteration without explicitly accounting for their cumulative difficulty across training rounds. For example, a sample repeatedly misclassified may not receive additional reinforcement beyond its current weight. Moreover, hard example mining techniques often rely on static designs, such as fixed loss penalties, pre-defined oversampling strategies, which restrict their ability to adapt dynamically to evolving misclassification patterns in financial datasets. This leaves systematic weaknesses uncorrected, particularly in high-noise environments.

2.5.3 Adaptive Sample Weighting

Adaptive sample weighting builds on the principle of hard example mining by dynamically reallocating importance to samples throughout training. In classification tasks, traditional static ensemble methods often struggle to handle class imbalance and evolving data distributions, motivating the use of adaptive weighting to emphasize minority or difficult-to-classify instances and improve overall model performance. Xu et al. [179] note that while static ensembles can provide baseline performance, they are outperformed by selective ensemble learning approaches that dynamically adapt by rejecting uncertain models, thereby improving predictive accuracy. Li et al. [180] develop a dynamic feature weighting strategy that updates feature importance according

to the coefficient of variation and Pearson correlation coefficients, enabling the model to emphasize more informative and stable financial indicators during training. This strategy allows the deep neural network to better capture heterogeneous financial patterns across industries and time, improving robustness and predictive accuracy. Wand et al. [181] concentrate on enhancing classifier performance when the distribution of training samples is heavily dominated by one class by introducing an adaptive class weight adjustment strategy. The weighting strategy continuously reallocates learning emphasis toward underrepresented observations, allowing the classifier to capture minority-class patterns more effectively. Huang et al. [182] design a weighting mechanism that continuously adjusts the contribution of training samples. The adjustment is guided by class prevalence, distributional density, and classification complexity, enabling the model to devote more learning capacity to infrequent and difficult-to-identify instances.

Among the benchmark strategies that have informed adaptive sample weighting, AdaBoost introduces a progressive misclassification-based weighting mechanism. In each iteration, misclassified samples are assigned higher weights, directing the model's attention toward difficult cases. Such a progressive learning strategy is well suited to financial distress prediction, where minority-class samples carry critical early-warning information [116]. Building on this idea, Focal Loss adds an exponent term to scale the influence of each sample based on its predicted probability. Well-classified samples are down-weighted, while hard-to-classify samples receive greater emphasis, which helps models focus on challenging cases in high-dimensional imbalanced datasets [183]. Complementing these approaches, Online Hard Example Mining

(OHEM) prioritizes samples that exceed a difficulty threshold or exhibit high loss values during training, directing learning toward the most challenging observations in the current training stage [184]. By focusing learning on the most challenging observations, OHEM can improve decision boundary refinement and minority-class recognition.

Despite these advances, many weighting strategies remain relatively static. They rely on one-time adjustments such as oversampling or cost assignment and often fail to iteratively incorporate feedback from misclassification history across training rounds [185], [186]. Even iterative methods such as boosting and OHEM primarily focus on the current learning state. Boosting algorithms increase the weights of samples misclassified in the current iteration, while OHEM prioritizes samples according to their current loss values or prediction difficulty. Although these mechanisms improve attention to hard samples, they do not explicitly maintain a cumulative record of sample-level difficulty across multiple training rounds. Consequently, persistently difficult samples and temporarily misclassified samples may receive similar treatment, limiting the model's ability to identify observations that consistently challenge the learning process [187], [188].

These shortcomings underscore the need for misclassification-history-driven sample weighting frameworks that not only respond to immediate errors but also accumulate knowledge of misclassification difficulty across iterations. By incorporating cumulative misclassification information, such frameworks can dynamically distinguish persistently difficult samples from transient errors and allocate learning resources more effectively. This capability is particularly important in financial distress prediction, where distressed firms often exhibit

overlapping characteristics with healthy firms and remain difficult to classify throughout training. Therefore, a cumulative difficulty-tracking mechanism offers a promising direction for improving robustness and minority-class detection in highly imbalanced financial datasets.

2.5.4 Emerging Flexible Iteration Approaches

Merging flexible iteration approaches aim to overcome the limitations of rigid ensemble frameworks by introducing greater adaptability into the learning process. Variants of Adaptive Boosting enhance flexibility by dynamically adjusting weighting schemes to emphasize difficult instances, and studies in bankruptcy prediction confirm that combining boosting with resampling strategies such as SMOTE significantly improves minority-class detection compared to traditional boosting [189].

Although gradient boosting frameworks such as LightGBM [190] and CatBoost [191] have demonstrated excellent predictive performance in financial distress prediction, they primarily rely on sequential gradient optimization and fixed boosting strategies. These models generally focus on minimizing overall prediction error rather than explicitly improving the learning process for persistently difficult minority-class samples. In contrast, Random Forest provides stronger model diversity through bootstrap aggregation and randomized feature selection, making it a robust foundation for incorporating adaptive sample weighting and dynamic iteration mechanisms [192]. The independence of individual trees also enables greater flexibility in modifying the training process without affecting the stability of the underlying ensemble.

More recently, research has explored hybrid frameworks that integrate boosting's error-focused weighting with bagging's model diversity to handle

complex and imbalanced tasks more effectively. In the financial distress prediction domain, Aljawazneh et al. [193] reported that although boosting generally outperforms bagging under severe class imbalance, hybrid ensemble methods can further enhance predictive performance. Similarly, Malek et al. [194] showed that in imbalanced water quality classification, bagging combined with resampling yielded superior sensitivity, highlighting the complementary strengths of bagging and boosting.

Building on this line of work, our earlier study conducted a comparative analysis of resampling techniques for financial distress prediction using XGBoost, demonstrating that Bagging–SMOTE and boosting hybrids achieved consistently higher recall rates for distressed firms under evolving imbalance scenarios [195]. Recent studies have further advanced flexible ensemble learning through self-paced training strategies. For example, Gao et al. [196] proposed a Multi-Heterogeneous Self-Paced Ensemble (MHSPE) framework that combines multiple self-paced ensembles with different base classifiers through weighted aggregation to improve financial distress prediction under class imbalance. Although MHSPE enhances minority-class learning through self-paced training, it does not explicitly exploit cumulative misclassification history to guide adaptive weighting across iterations.

Consequently, most flexible iteration approaches remain limited in their ability to account for persistent misclassification patterns or dynamically reallocate learning resources based on cumulative sample difficulty. Addressing this gap requires learning-history-informed adaptive iterative frameworks that can capture persistent weaknesses and improve generalization in highly imbalanced financial distress prediction.

2.6 Summary and Research Gaps

Despite significant advances in financial distress prediction, several limitations remain in existing studies. Traditional statistical models have provided foundational insights but often rely on linear assumptions that may not hold in complex financial environments. While machine learning models have improved predictive performance and can capture nonlinear patterns, their effectiveness is frequently compromised by severe class imbalance, which can lead to biased predictions.

Existing approaches typically address class imbalance using static resampling methods. These one-time adjustments, however, may introduce noise or discard valuable information from the dataset, limiting the model's ability to fully exploit minority-class patterns. In addition, few studies have explored dynamic learning mechanisms that adaptively emphasize persistently misclassified minority instances throughout the training process. Without such mechanisms, models struggle to consistently identify high-risk firms and may fail to adapt to evolving financial conditions. Similarly, many hybrid ensemble frameworks combine bagging, boosting, and resampling techniques to improve robustness, yet most rely on fixed iteration structures and static integration rules, limiting their flexibility to respond to temporal shifts in economic conditions or changing class distributions.

Current models exhibit three main limitations:

- Most ensembles struggle to effectively handle severe class imbalance, as fixed iteration strategies and static integration rules prevent them from adequately emphasizing minority-class instances.

- Persistently hard-to-class minority-class samples are not systematically emphasized across iterations, limiting the model's ability to learn from critical high-risk firms.
- Sample weighting and learning resources are rarely dynamically reallocated based on cumulative misclassification trends, reducing sensitivity to evolving patterns in the minority class.

Addressing these gaps requires the development of a history-aware, dynamically adaptive iterative ensemble framework that integrates long-term hard sample reinforcement, continuous adaptive weighting, and flexible iteration strategies. Such a framework would improve minority-class detection, enhance robustness to evolving financial data distribution, and provide a more reliable tool for early-warning of financial distress.

CHAPTER 3 METHODOLOGY

In this chapter, a RF–based framework enhanced with a boosting is proposed to address the challenges outlined in Section 1.2. Existing sample reweighting techniques typically ignore misclassification history and therefore fail to progressively emphasize samples that are persistently misclassified. Moreover, the high redundancy inherent in financial data increases computational complexity and intensifies the risk of overfitting, especially under class-imbalanced conditions where minority samples are limited. In addition, many ensemble methods adopt fixed weight adjustment schemes, overlooking the fact that the performance of base learners varies with the underlying data structure.

The key innovations include the integration of lagged feature construction to capture temporal dynamics, a dynamic sample-weighting mechanism, adaptive class ratio adjustment, and real-time performance-based model weighting. Together, these components enable the framework to more effectively emphasize persistently misclassified samples and to more accurately model the evolving patterns associated with financial distress.

This chapter contains six sections. Section 3.1 outlines the procedures for acquiring and preparing the empirical data. Section 3.2 describes the steps taken to process data and engineering features to enhance model performance. Section 3.3 presents the design of the DWARFB framework, detailing its mechanisms for adaptive sampling, weight adjustment, and ensemble optimization. Section 3.5 introduces the metrics and statistical tests used to assess model performance. Finally, Section 3.6 provides a summary and connects the chapter to the subsequent work.

3.1 Empirical Data Acquisition and Integrating

The analysis is conducted using financial records obtained from the China Stock Market & Accounting Research (CSMAR) database, a leading research database that provides broad coverage of publicly listed companies. In this study, financial distress is identified based on the Special Treatment (ST) status of Chinese A-share listed firms. The ST designation is selected because it provides an observable and standardized classification mechanism widely used in empirical studies of Chinese corporate distress prediction.

Table 3.1 Dataset Composition

Category	Dataset	Description	Number of Indicators	Range
Financial Indicators	Solvency	Ability to repay short- and long-term debt using enterprise assets.	37	1990-2024
	Disclosed Index	Indicators directly reported to reflect company operations.	16	2007-2024
	Ratio Structure	Financial structure based on the proportions of key financial indicators.	44	1990-2024
	Operating Capacity	Efficiency in utilizing assets for business operations.	76	1990-2024
	Earning Capacity	Ability to generate profits.	77	1990-2024
	Cash Flow	Cash-related indicators derived from financial statement ratios.	41	1991-2024
	Risk Level	Risk of financial instability due to weak structure or poor financing practices.	12	1991-2024
	Growth Capability	Potential for future expansion and performance improvement.	66	1990-2024
	Index per Share	Financial condition measured on a per-share basis.	97	1990-2024
	Relative Value Index	Derived from comparisons among related indicators.	36	1990-2024
	Dividend Distribution	Metrics derived from comparisons among related financial indicators.	19	1990-2024
Bankruptcy Reorganization	Business Risk	Risk profile of bankrupt or restructured listed firms.	33	2000-2024
Total			554	2007-2024*

* The overall period range is based on the overlapping period covered by all datasets.

Twelve datasets cover quarterly financial information for 1,174 collected from Chinese A-share listed firms throughout the 1990–2024 observation period were downloaded from the CSMAR database and Table 3.1 summarizes these datasets.

This dataset comprises 554 distinct indicators, which are organized into two major categories, i.e. Financial Indicator and Bankruptcy Reorganization.

The Financial Indicator describes seven perspectives on the financial performance and condition of a firm. These perspectives are the ability to debt obligations (solvency, firm's operational (disclosed indices), asset utilization (financial ratio structures, operating capacity), profit generation (earning capacity), financial stability (cash flow indicators, risk level), firm's expansion (growth capability), comparative metrics (per-share financial measures, relative value indices), and firm's dividend (dividend distribution metrics). The Bankruptcy Reorganization provides variables that characterize the business risk profiles of firms undergoing bankruptcy or restructuring, supporting more detailed analyses of corporate distress and recovery.

The downloaded twelve datasets are integrated using STATA, a statistical software widely used for data analysis. Figure 3.1, illustrates the integration process of these data.

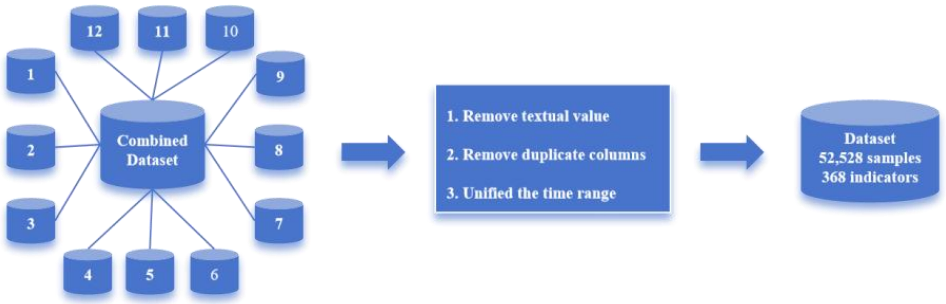

Figure 3.1 Data Integration Process

To construct a unified analytical dataset, all datasets are systematically merged based on firm identifiers and reporting periods. Duplicate columns arising from overlapping variables across sub-datasets are identified and removed to prevent redundancy and ensure data integrity. Furthermore, indicators containing non-numeric or textual entries are excluded, as they are unsuitable for quantitative modelling and statistical analysis. To address the inconsistencies in statistical periods and report frequencies in the merged dataset, a temporal harmonization procedure is applied. All records are converted to a quarterly frequency, covering the period from 2007 to 2024 to ensure temporal alignment across firms and facilitating rigorous panel data analysis. Following the integration and cleaning processes, the consolidated dataset comprises 368 features and 52,528 firm-quarter observations. This unified dataset provides a robust foundation for subsequent stages of feature extraction, variable selection, and model development. Table 3.2 summarizes the unified dataset.

Table 3.2 Dataset Basic Information

Items	**Information**
Total samples	52,528
Total features	370 (include company ID and dates)
Target variable	Company's financial status (non-distressed and distressed)
Memory usage	148.68 MB
Class 0 (non-distressed)	49,555 (94.340%)
Class 1 (Distressed)	2,973 (5.660%)
Imbalance ratio	16.67:1

After completing the data preparation process, the analytical dataset consists of 52,528 firm-quarter observations and 370 variables, including

company identifiers, observation dates, and a binary target variable indicating each company's financial health status. It occupies 148.68 MB of memory and exhibits a highly imbalanced class distribution, with 49,555 non-distressed instances (94.340%) and 2,973 distressed instances (5.660%), corresponding to an imbalance ratio of approximately 16.67:1.

3.2 Data Stratification and Preprocessing

3.2.1 Data stratification

The merged dataset is partitioned into training and testing sets using a company-level stratified split with a 7:3 ratio. Companies are first grouped according to whether they contained at least one distressed observation, and distressed and non-distressed companies are then randomly allocated to the training (70%) and testing (30%) sets while preserving similar class proportions. To evaluate the robustness of the proposed method in handling class imbalance, both training and test sets maintain similar proportions of distressed and non-distressed firms. As a result, the training set contains 820 companies with 36,911 firm-quarter observations, including 2,056 distressed samples, while the test set includes 354 companies with 15,617 observations, including 917 distressed cases. This splitting design establishes a reliable and unbiased basis for model training and out-of-sample evaluation.

To prevent data leakage and maintain temporal integrity, all observations from a given firm are placed entirely within a single subset, either training or testing. This approach ensures that the model is evaluated on completely unseen firms, offering a more realistic measure of its generalization performance.

3.2.2 Data preprocessing

Preprocessing tasks are then performed on the training data, and these tasks, include lagged features construction, winsorization to reduce the impact of outliers, and the removal of duplicate and highly correlated variables. The top 200 features are then selected for training. The cleaned training data is putting into DWARFB to train. Finally, the test data is used to generate the final predictions. The structure of the preprocessing step is shown in Figure 3.2.

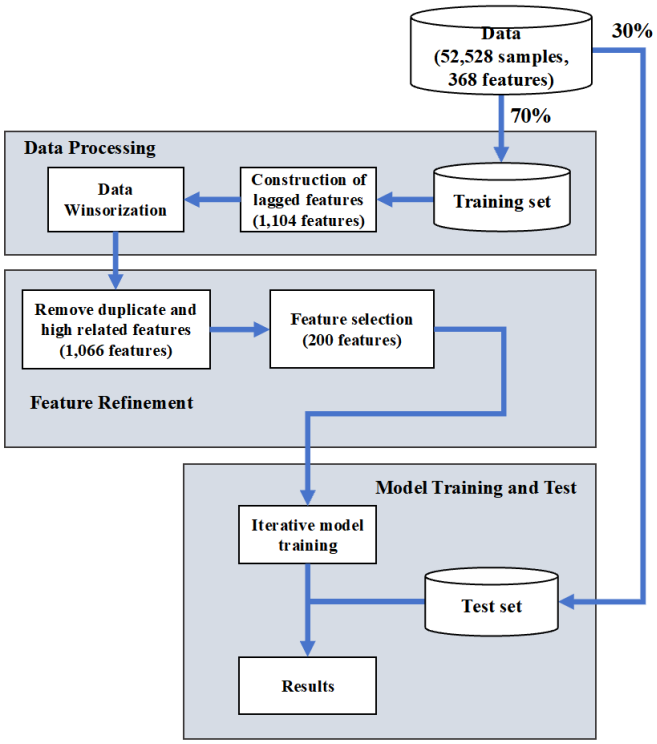

Figure 3.2 Structure of the Proposed Framework

(1) Outlier Treatment

To mitigate the potential distortion caused by extreme observations, Winsorization is applied to all continuous variables in the dataset. Winsorization is a robust statistical transformation technique designed to reduce the influence of outliers without discarding data points. Instead of removing extreme values, this method replaces them with the nearest value within a predefined percentile range, typically between the 1st and 99th percentiles. This approach effectively

limits the impact of extreme deviations while maintaining the overall data structure and full sample size, thereby improving the stability and reliability of subsequent model estimation and inference [197].

Mathematically, for a variable X:

$$X_i^{\text{Winsorization}} = \begin{cases} Q_L = \text{if } X_i < Q_L \\ Q_U = \text{if } X_i > Q_U \\ X_i, \text{otherwise} \end{cases} \quad \text{Eq. 3.1}$$

Where,

Q_L = lower quantile (e.g., 1st percentile)

Q_U = upper quantile (e.g., 99th percentile)

X_i = original value

By capping extreme values within this percentile range, Winsorization effectively minimizes the leverage of outliers on the model's learning process, ensuring that the resulting financial distress prediction models are both robust and generalizable across firms with varying financial characteristics.

(2) Construction of Lagged Financial Indicators

Following the initial dataset partitioning and outlier treatment, lagged financial indicators are systematically constructed for the training and test sets independently to capture firms' historical dynamics while maintaining temporal integrity. For each company, the original set of financial and accounting indicators is shifted by predetermined lag periods, specifically two quarters (Q2) and four quarters (Q4), to generate past values that reflect prior financial conditions. These lagged variables are then appended to the dataset as additional features, enriching the temporal dimension of the model input space.

The inclusion of lagged features enables the model to incorporate historical dependencies and persistence effects commonly observed in financial

performance indicators, which are crucial for identifying early signals of financial distress. Importantly, the lagging process are performed separately within each dataset split to ensure that future information from subsequent quarters was not inadvertently introduced during training, thereby eliminating look-ahead bias and preserving the chronological order of observations.

As a result of this transformation, the total number of features expand from 368 to 1,104, reflecting the addition of both Q2 and Q4 lagged variables. This extended feature set provides the foundation for subsequent redundancy analysis and feature selection procedures aim at optimizing model efficiency and interpretability.

(3) Detection and Elimination of Redundant Features

To address potential multicollinearity and reduce feature redundancy, a two-stage feature screening procedure is implemented to identify and remove duplicate or highly correlated variables. The objective of this process is to enhance model stability, improve computational efficiency, and retain only the most informative indicators for subsequent modelling.

In the first stage, exact duplicates are identified through elementwise comparison of feature vectors. Two features X_i and X_j are considered identical if $X_i = X_j$ for all observations. This step ensures the elimination of redundant variables that may result from different naming conventions or overlapping data sources representing identical financial measures.

In the second stage, a graph-based correlation grouping method is applied to detect and cluster highly correlated features. An undirected graph $G = (V, E)$ is constructed, where each vertex $v \in V$ represent a feature, and an edge $e_{ij} \in E$ is added between two vertices $v_i$and v_j if the absolute value of

their Pearson correlation coefficient satisfies $|p_{ij}| \geq \tau$, where τ is a predefined threshold (default: 0.999). The connected components of this graph correspond to sets of features that exhibit near-perfect correlation and thus potential redundancy.

For each group of duplicate or highly correlated features $G_k = \{X_1, X_2, \ldots, X_m\}$, only the most informative feature is retained based on its absolute Spearman correlation with the target variable Y:

$$X_{keep} = \arg\max_{X_i \in G} |\rho X_i, Y| \quad \text{Eq. 3.2}$$

This selection criterion ensures that the most informative feature within each redundant group is preserved, following the principle of maximizing feature relevance while minimizing redundancy.

By combining elementwise duplicate detection with graph-theoretical correlation clustering, this approach efficiently reduces feature redundancy, alleviates multicollinearity, and preserves the most informative indicators for subsequent model training.

(4) Feature Selection and Dimensionality Reduction

Feature selection is conducted using a RF-based importance ranking [198]. In this method, a preliminary RF model is trained on the full feature set, and each predictor is evaluated based on its contribution to reducing prediction error across the ensemble of DTs. For classification tasks, RF computes feature importance by measuring the reduction in Gini impurity at each split [115]. The formula for the importance of a feature j is shown in Eq. 3.3:

$$\text{Feature\_importance}_j = \frac{1}{T}\sum_{t=1}^{T} \sum_{s \in S_{j,t}} \Delta I(s,t) \qquad \text{Eq. 3.3}$$

where T represents the total number of trees, $S_{j,t}$ denotes the set of nodes in the tree t where feature j is used for splitting, and $\Delta I(s,t)$ is the decrease in impurity (Gini or entropy) from the split at the node s.

A top N strategy is employed to retain the most informative features, balancing predictive power with computational efficiency and interpretability. To prevent data leakage, feature importance is computed solely on the training set, and the resulting feature subset is applied to both validation and test sets.

3.3 Designing DWARFB

The primary contribution of this research is the development of the DWARFB framework for financial distress prediction.

The proposed framework integrates three key mechanisms:

(1) Adaptive sampling strategy

(2) Misclassification-aware sample weighting

(3) Dynamic ensemble learning

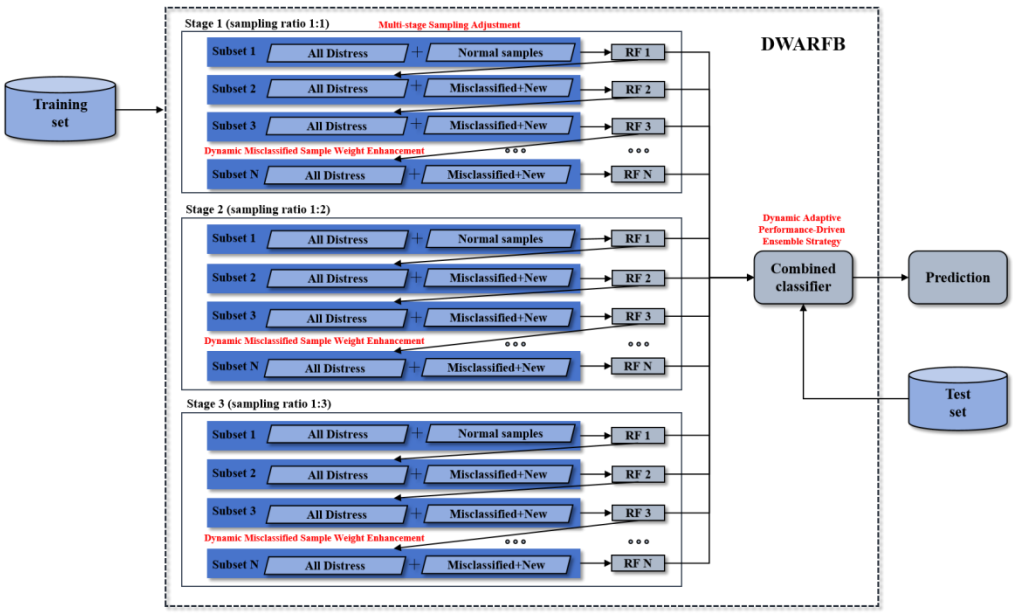

Figure 3.3 Workflow of the Proposed Algorithm Framework

These components work together to enhance the model's effectiveness in identifying financially distressed minority-class observations while retaining strong overall predictive capability. An overview of the proposed DWARFB framework is provided in Figure 3.3.

Algorithm 1: Pseudocode for DWARFB

Input:

- Training dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$
- Number of iterations T
- Weight parameters α and γ

Output:

- Final ensemble classifier $H(x)$

Procedure:

1. **Initialize sample weights:**
   $w_i = \frac{1}{N}$ for all samples
2. **For** $t = 1$ **to** T **do:**
   (1) Apply adaptive sampling
   (2) Train Random Forest classifier h_t
   (3) Predict labels for training samples
   (4) Identify misclassified samples
   (5) Update sample weights
   (6) Normalize weights
3. **End For**
4. **Aggregate classifiers using weighted voting:**

$$H(x) = \text{sign}(\sum_{t} w_t \, h_t(x))$$

The training procedure of the proposed DWARFB framework is summarized in Algorithm 1. The training set is used to build and iteratively refine the ensemble, while the test set is reserved for final evaluation. During each iteration, sample weights are updated based on cumulative misclassification information, allowing the model to focus more on difficult cases. The algorithm iteratively trains multiple Random Forest classifiers while dynamically adjusting sample weights. Finally, the trained models are combined using a weighted ensemble based on validation performance, and adaptive thresholding is applied to determine the optimal classification boundary. The finalized DWARFB model is evaluated on the test set to measure

its accuracy, robustness, and effectiveness in handling imbalanced financial data.

3.3.1 Sampling Adjustment Strategy

DWARFB adopts a multi-stage progressive sampling strategy to address the inherent class imbalance in financial distress prediction. The positive (distressed minority) and negative (healthy majority) sample ratio changes gradually across iterative training rounds. This progressive ratio setting allows the model to first build a stable fitting baseline on balanced data, then adapt stepwise to real-world imbalanced data distributions without incurring excessive computational cost or overfitting.

Within each sampling stage, all positive samples are included, and a subset of negative samples is selected to meet the target class ratio. The number of negative samples required for the current iteration is computed as:

$$\mathrm{n_{required}} = \min(n_{neg_{pool}}, n_{pos} \times \frac{neg_{ratio}}{pos_{ratio}}) \qquad \text{Eq. 3.4}$$

where $\mathrm{n_{pos}}$ is the number of positive samples, $\frac{\mathrm{neg\_ratio}}{\mathrm{pos\_ratio}}$ is the target negative-to-positive ratio, and $\mathrm{n_{required}}$ is the number of negative samples to be drawn.

To prioritize difficult cases, negative samples are selected according to their misclassification history using the following rule:

$$\mathrm{NegSamples} = \begin{cases} \mathrm{RandomSubset}(\mathrm{MisNeg}, \mathrm{n_{required}}), & \text{if } \mathrm{n_{mis}} \geq \mathrm{n_{required}} \\ \mathrm{MisNeg} \cup \mathrm{RandomSubset}\left(\mathrm{NeverMisNeg}, \mathrm{n_{required}} - \mathrm{n_{mis}}\right), & \text{if } \mathrm{n_{mis}} < \mathrm{n_{required}} \end{cases} \qquad \text{Eq. 3.5}$$

Here, MisNeg denotes the set of previously misclassified negative samples, NeverMisNeg denotes negative samples that have never been misclassified, $\mathrm{n_{mis}}$ is the number of misclassified negatives, and $\mathrm{RandomSubset}(\mathrm{S}, \mathrm{n})$ represents a random selection of n samples from set S. This approach ensures that previously misclassified negatives are prioritized

while maintaining the target negative-to-positive ratio, enabling the model to focus on difficult or borderline cases that carry high predictive significance.

3.3.2 Misclassified Samples Reweighting Mechanism

The misclassification weight adjustment mechanism increases the importance of samples that are frequently misclassified during iterative training. By giving more attention to difficult instances, the model can better learn from challenging cases, improving the overall performance of the ensemble.

This study proposes a novel weighting function that integrates concepts from existing machine learning methods. The weight assigns to sample i after mis_i misclassifications is defined as:

$$w_i^{(\mathrm{mis})} = 1 + \alpha \left(\frac{\mathrm{mis}_i}{\max(\mathrm{mis}_{\max}, 1)} \right)^{\gamma} + \alpha \cdot \mathrm{I}(\mathrm{mis}_i \geq \text{threshold}) \quad \text{Eq. 3.6}$$

Key inspirations include:

Adaptive Boosting [116]: Increasing the importance of misclassified samples progressively motivates using the cumulative misclassification count mis_i.

Focal Loss [199]: The exponent γ introduces nonlinear scaling. Larger γ emphasizes heavily misclassified samples, while smaller γ prevents excessive weighting of outliers.

Online Hard Example Mining [184]: The indicator function $\mathrm{I}(\mathrm{mis}_i \geq \text{threshold})$ boosts samples that are repeatedly misclassified beyond a defined limit.

As shown in Table 3.3, this design achieves continuous scaling (via normalized misclassification ratio and γ) and dynamic correction (via reward adjustment), allowing the model to focus on hard-to-classify instances without destabilizing training.

Table 3.3 Benchmark Concepts in Misclassification Handling

Benchmark Concept	Original formula	Role in the formula
Adaptive Boosting (AdaBoost)	Provides the progressive weighting idea based on misclassification count $w_i^{t+1} = w_i^t \times exp(\alpha_t \times I[h_t(x_i) \neq y_i])$	The weighting mechanism assigns higher weights to misclassified samples, where α and mis_i form the basis of the calculation.
Focal Loss	Inspires the exponent term $(1 - p_i)^\gamma$ controlling emphasis on hard samples	$\left(\frac{mis_i}{\max(mis_{max}, 1)}\right)^\gamma$
Online Hard Example Mining (OHEM)	Forms the indicator term $I(mis_i \geq threshold)$	$\alpha \times I(mis_i \geq threshold)$

The weighting strategy uses the misclassification history as a proxy for sample difficulty. Difficult samples are identified through a percentile-based threshold, which adapts to the characteristics of each dataset. The procedure consists of four stages:

(1) Normalized Misclassification Count

$$\text{mis\_count}_{\text{normalized}} = \frac{\text{mis\_count}}{\max(\text{mis\_count})} \quad \text{Eq. 3.7}$$

Where mis_count is the cumulative number of misclassifications for each sample.

This formulation is inspired by three research streams:

(2) Primary Weight Calculation

$$\text{mis\_count\_weight} = 1.0 + \alpha(\text{mis\_count}_{\text{normalized}})^\gamma \quad \text{Eq. 3.8}$$

- α controls the overall intensity of weight adjustment.
- γ shapes the nonlinear growth of weights.
- Base weight 1.0 ensures all samples participate in training.

(3) Enhanced Weight for Difficult Samples:

$$\text{if } \text{mis}_{\text{count}} >= \text{ threshold then } \text{ mis\_count\_weight} += \alpha \quad \text{Eq. 3.9}$$

Where the threshold is determined by:

$$\text{threshold } = \text{ percentile}(\text{mis\_count}, \text{threshold\_percentile}) \quad \text{Eq. 3.10}$$

(4) Reward-based Weight Adjustment:

$$w_{\text{final}} = \text{mis\_count\_weight} \times \text{reward\_factor} \quad \text{Eq. 3.11}$$

The reward_factor (typically, 0.5) is the reward factor used to reduce historical misclassification counts for correctly predicted samples.

And if the misclassified history will be reset to 0 when it reaches consecutive threshold (typically, 2).

This three-stage process ensures that frequently misclassified samples receive greater attention while avoiding extreme weights that could destabilize learning.

3.3.3 Dynamic Adaptive Performance-Driven Ensemble Strategy

We employ the Dynamic Adaptive Performance-Driven Ensemble Strategy to address the limitations of traditional static integration methods. This strategy continuously monitors cross-validation performance metrics and adaptively adjusts model weights based on real-time performance feedback, enabling fine-grained control over the ensemble composition and enhancing the model's capacity to maintain optimal integration performance.

(1) Performance-Based Weight Calculation

At the core of the Dynamic Adaptive Performance-Driven Ensemble Strategy, we employ a logit transformation of cross-validation F1 scores to compute adaptive model weights. For each base learner m_t trained at iteration t, the weight α_t is calculated as:

$$\alpha_t = \log\left(\frac{F1_t}{1-CV_{F1_t}+\varepsilon}\right) \quad \text{Eq. 3.12}$$

where CV_{F1_t} represents the cross-validation F1 score of model m_t, and ε is a small constant (1e-8) to prevent division by zero. This transformation ensures that higher-performing models receive exponentially larger weights while maintaining numerical stability.

(2) Dynamic Weight Adjustment Framework

Using the computed weight α_t, a new weight $w_i^{dynamic}$ is computed as follows:

$$w_i^{dynamic} = \alpha_t \times M_{performance}\left(CV_{F1_i}\right) \quad \text{Eq. 3.13}$$

where $M_{performance}\left(CV_{F1_i}\right)$ is the performance multiplier function:

$$M_{performance}\left(CV_{F1_i}\right) = \begin{cases} 0.5, \text{if } CV_{F1_i} \geq 0.85 \text{ (Very high performance)} \\ 0.8, \text{ if } 0.80 \leq CV_{F1_i} < 0.85 \text{ (High performance)} \\ 1.2, \text{ if } 0.75 \leq CV_{F1_i} < 0.80 \text{ (Satisfactory performance)} \\ 2.0, \text{if } CV_{F1_i} < 0.75 \text{ (Low performance)} \end{cases} \quad \text{Eq. 3.14}$$

(3) Adaptive Learning Rate Mechanism

This ensemble strategy incorporates an adaptive learning rate mechanism that adjusts based on performance trends:

$$\eta_t = \begin{cases} \max(0.1, \eta_{t-1} \times 0.9), \text{if } CV_{F1_t} - CV_{F1_{t-1}} < \delta \\ \min(2.0, \eta_{t-1} \times 1.05), \text{otherwise} \end{cases} \quad \text{Eq. 3.15}$$

where $\delta = 0.001$ is the performance change threshold, and learning rates are constrained between 0.1 and 2.

(4) Ensemble Prediction with Dynamic Weights

For an ensemble $M = \{m_1, m_2, ..., m_T\}$ with corresponding dynamic weights $W = \{w_1, w_2, ..., w_T\}$, the final prediction probability for class c is computed as:

$$P(c|X) = \frac{\sum_{i=1}^{T}(w_i \cdot P_i(c|X))}{\sum_{i=1}^{T} w_i} \quad \text{Eq. 3.16}$$

where $P_i(c|X)$ denotes the probability assigned to class c by model m_i for input X. The final classification decision is obtained through:

$$y_{pred} = \arg\max_{c} P(c|X) \quad \text{Eq. 3.17}$$

(5) Weight Normalization and Stability

To ensure numerical stability and prevent dominance by extreme weights, the ensemble strategy applies weight normalization:

$$w_i^{normalized} = \frac{w_i^{dynamic}}{\sum_{j=1}^{T} w_j^{dynamic}} \quad \text{Eq. 3.18}$$

This normalization guarantees that all weights sum to unity while preserving the relative performance-based weight ratios.

3.3.4 Optimal Classification Threshold Selection and Final Decision

To maximize classification performance, an optimal threshold selection strategy is employed. The threshold is determined by optimizing the F1 score across a range of probability thresholds:

Step 1: Threshold Optimization

Test various threshold values τ from 0 to 1.

For each threshold, convert probabilities to binary predictions:

$$y_{pred(\tau)} = 1 \text{ if } P(y = 1|x) \geq \tau, \text{else } 0 \quad \text{Eq. 3.19}$$

Calculate F1 score for each threshold.

Select the threshold τ^* that gives the highest F1 score.

Step 2: Final Classification

Apply the optimal threshold τ^* to the ensemble probability $P(y = 1|x)$.

Make final prediction:

$$y_{final} = 1 \text{ if } P(y = 1|x) \geq \tau, \text{else } y_{final} = 0 \qquad \text{Eq. 3.20}$$

This two-step process ensures that the classification threshold is optimized for the specific dataset and model performance, rather than using a fixed 0.5 threshold.

3.4 Evaluation Metrics

A rigorous evaluation framework is essential to ensure that the proposed DWARFB model is assessed comprehensively across multiple aspects of classification performance. This section outlines the primary and secondary performance metrics adopted for the study, as well as the role of confusion matrix analysis in interpreting classification behaviour. These metrics are particularly important in the context of imbalanced datasets, where conventional measures such as accuracy can be misleading.

3.4.1 Preliminary Evaluation

(1) Cohen's Kappa

Cohen's Kappa coefficient [200] is a statistical measure of inter-rater agreement that has been widely adopted in machine learning to assess the reliability of classification models. Unlike accuracy, which only measures the proportion of correct predictions, Kappa adjusts for the level of agreement that might occur purely by chance, thereby providing a more robust and interpretable evaluation metric, especially in imbalanced datasets.

The formula for Cohen's Kappa is defined as:

$$\kappa = \frac{P_o - P_e}{1 - P_e} \qquad \text{Eq. 3.21}$$

where:

- P_o= Observed agreement, i.e., the proportion of instances for which the predicted and actual classes match (equivalent to overall accuracy).

- P_e = Expected agreement by chance, estimated from the marginal distributions of the confusion matrix.

  For a binary classification problem, P_e can be computed as:

$$P_e = \frac{(TP+FP)(TP+FN)+(TN+FN)(TN+FP)}{N^2} \quad \text{Eq. 3.22}$$

The interpretation of Cohen's Kappa values is generally categorized as follows:

Table 3.4 Cohen's Kappa Interpretation

The value of κ	Explanations
$\kappa < 0$	Agreement worse than chance
$0 \leq \kappa < 0.2$	Minimal or slight agreement
$0.2 \leq \kappa < 0.4$	Fair level of agreement
$0.4 \leq \kappa < 0.6$	Moderate agreement
$0.6 \leq \kappa < 0.8$	Strong or substantial agreement
$K \geq 0.8$	Near-perfect agreement

Cohen's Kappa is particularly valuable when evaluating models in imbalanced classification scenarios such as financial distress prediction, where accuracy may overestimate performance due to the dominance of the majority class. By considering both the observed and chance agreement, Kappa provides a more meaningful reflection of the classifier's effectiveness.

(2) Feature Importance Metric

We employ a loss-based feature contribution metric to measure how much each feature reduces the weighted misclassification loss during the DWARFB training process. A feature is regarded as important if it consistently contributes to a significant reduction in classification loss throughout the boosting iterations. The loss-based feature importance for the feature f is defined as:

$$FI_f = \frac{\sum_{t=1}^{T} \alpha_t (L^{(t-1)} - L_f^{(t)})}{\sum_k \sum_{t=1}^{T} \alpha_t (L^{(t-1)} - L_k^{(t)})} \quad \text{Eq. 3.23}$$

Where:

T denotes the total number of boosting iterations.

f denotes a specific feature, and k indexes all features.

$L^{(t-1)}$ is the weighted classification loss before the t-th iteration.

$L_f^{(t)}$ presents the loss after the splits involve the feature f at iteration t.

α_t is the boosting weight of the t-th base learner

This metric provides a performance-driven and interpretable measure of feature importance for predicting distress.

3.4.2 Classification Evaluation Metrics

The following metrics form the core of the evaluation due to their robustness in imbalanced classification scenarios:

The F1-score measures classification performance by combining precision and recall using their harmonic mean. This metric assigns equal importance to false-positive and false-negative errors, making it an appropriate evaluation criterion for financial distress prediction, where both forms of misclassification are critical.

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{Eq. 3.24}$$

The F1-score accounts for the trade-off between false-positive and false-negative errors, providing a balanced indicator of classification effectiveness. This makes it an appropriate metric for tasks in which minimizing both error types is important.

The Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) assesses the overall discrimination performance of a classifier by

evaluating its ability to differentiate between classes across every possible classification threshold. Higher values indicate stronger separability between distressed and non-distressed classes.

$$\mathrm{AUC} = \int_0^1 \mathrm{TPR}(\mathrm{FPR})\,\mathrm{d}(\mathrm{FPR}) \qquad \text{Eq. 3.25}$$

Where,

TPR (True Positive Rate) is also known as sensitivity or recall.

FPR (False Positive Rate) is defined as $\frac{\mathrm{FP}}{\mathrm{FP+TN}}$.

The Precision–Recall Area Under the Curve (PR-AUC) quantifies classifier performance by aggregating precision–recall pairs over the full range of decision thresholds. Since it focuses on the positive class rather than overall class separation, it is especially appropriate for evaluating models on imbalanced datasets. It is more sensitive to class imbalance than ROC-AUC and is therefore critical in this study:

$$\mathrm{PR\_AUC} = \int_0^1 \mathrm{P}(\mathrm{R})\,\mathrm{dR} \qquad \text{Eq. 3.26}$$

where $\mathrm{P}(\mathrm{R})$ is precision as a function of recall R.

In discrete form (trapezoidal approximation):

$$\mathrm{PR_{AUC}} \approx \sum_{\mathrm{i}=1}^{\mathrm{n}-1} (\mathrm{R_{i+1}} - \mathrm{R_i}) \cdot \frac{\mathrm{P_{i+1}} + \mathrm{P_i}}{2} \qquad \text{Eq. 3.27}$$

Precision is the proportion of predicted distressed companies that are truly distressed, reflecting the model's ability to avoid false positives.

$$\mathrm{Precision} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \qquad \text{Eq. 3.28}$$

Recall is the proportion of actual distressed companies correctly identified, indicating the model's sensitivity and its ability to minimize false negatives.

$$\text{Recall} = \frac{TP}{TP+FN} \quad \text{Eq. 3.29}$$

The overall predictive performance of a classifier can be assessed using accuracy, which is computed as the ratio of correctly classified observations to the total number of observations. While easy to interpret, accuracy is less reliable in highly imbalanced datasets and is therefore used here only as a supplementary indicator.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{Eq. 3.30}$$

The Matthews Correlation Coefficient (MCC) serves as a balanced indicator of binary classification performance by measuring the association between actual and predicted class labels. The coefficient ranges from −1, indicating total disagreement, to +1, indicating complete agreement, while a value of 0 suggests performance comparable to random classification. By accounting for all four confusion matrix components, MCC provides a robust and balanced performance metric, particularly for imbalanced datasets. It is calculated as:

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad \text{Eq. 3.31}$$

Where,

TP = True Positives, distressed companies correctly identified as distressed.

TN= True Negatives, non-distressed companies correctly classified, indicating accurate identification of stable entities.

FP = False Positives, non-distressed companies incorrectly classified as distressed, potentially leading to unnecessary interventions or investigations.

FN = False Negatives, distressed companies incorrectly predicted as non-distressed, which can be particularly costly in financial contexts where early detection is critical.

3.4.3 The statistical test metrics

The Wilcoxon signed-rank test [201] is a distribution-free statistical technique designed to assess whether the median difference between two paired sets of observations is significantly different from zero. Unlike the paired t-test, which assumes normality of the differences, the Wilcoxon signed-rank test makes no such assumption, making it particularly suitable for performance comparison across cross-validation folds where metric distributions may not be Gaussian.

Formally, let x_i and y_i denote the performance scores (e.g., F1-score, AUC, PR-AUC) of two models on the $i-th$ fold. The differences are computed as $d_i = x_i - y_i$. Observations with no difference are first removed from the paired sample. The remaining absolute differences, $|d_i|$, are ordered and assigned ranks, after which separate rank totals are computed for positive and negative differences, yielding the statistics W^+ and W^-, and the Wilcoxon statistic is defined as:

$$W = \min(W^+, W^-) \qquad \text{Eq. 3.32}$$

For sufficiently large samples, the statistic can be approximated by a normal distribution using:

$$Z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \qquad \text{Eq. 3.33}$$

where n is the number of non-zero differences. The corresponding p-value is then derived from the standard normal distribution.

In this study, the Wilcoxon signed-rank test is applied to paired cross-validation results to determine whether the performance improvements of DWARFB over baseline approaches are statistically significant. In this study, statistical significance is determined at the 5% level. Accordingly, a p-value below 0.05 indicates that the null hypothesis should be rejected, implying that the observed difference in performance is unlikely to be explained by random sampling variability alone.

The Wilcoxon signed-rank test, anon-parametric statistical method, is employed instead of parametric alternatives (such as paired t-tests) to ensure robust statistical inference. This choice is motivated by several considerations: (1) the non-parametric nature of the test eliminates the need for normality assumptions, which may not hold for machine learning performance metrics; (2) the test is robust to outliers and skewed distributions that commonly arise in cross-validation results; (3) it provides a conservative assessment of statistical significance, making our conclusions more reliable. The consistent statistical significance across all performance metrics (F1-score, AUC, PR-AUC, precision, recall, and MCC) demonstrates that the observed improvements are not due to random chance but represent genuine gains in predictive performance.

3.5 Summary

This chapter presents the research methodology used in this study. It describes the dataset for financial distress prediction, the data preprocessing procedures, and the feature engineering techniques applied to prepare the financial data.

The chapter also introduces the proposed DWARFB framework, including its architecture, algorithmic structure, and mathematical formulation.

The dynamic weighting mechanism and adaptive sampling strategy enable the model to effectively address class imbalance and improve learning from difficult samples.

The next chapter presents the experimental setup and evaluation procedures for assessing the performance of the proposed model.

CHAPTER 4 DWARFB IMPLEMENTATION AND EXPERIMENTAL STUDY

Chapter 3 introduced the conceptual framework of our predictive model. This chapter, we introduce the implementation of DWARFB, which emphasizes a positive-to-negative sampling strategy, enhances the weights of misclassified positive instances, tracks persistent misclassification patterns and loss-based feature contributions, and employs a performance-driven ensemble strategy.

The chapter also describes the experimental study designed to evaluate the DWARFB framework. This includes sensitivity analyses of iterative learning parameters and baseline RF hyperparameters, benchmarking against state-of-the-art models and resampling strategies, and comparisons with misclassification penalty methods. Statistical significance testing and repeated independent runs ensure that the results are robust, reliable, and reproducible.

This chapter is organized into six sections. Section 4.1 outlines the tools utilized in this study. Section 4.2 presents the architectural design of the DWARFB implementation. Section 4.3 describes the comparative study, including baseline machine learning models, benchmark misclassification penalty methods, and statistical significance testing. Section 4.4 explains the parameter sensitivity and robustness analysis, including parameter scope definition and grid search optimization. Section 4.5 presents the experimental study, detailing the objectives, datasets, and experimental design. Finally, Section 4.6 provides the chapter's conclusion.

4.1 Experimental Tools

The implementation and evaluation of the proposed DWARFB framework require an integrated experimental environment capable of handling large-scale financial datasets and iterative machine learning training. This

section describes the software environment, hardware configuration, and development tools employed in the study.

The experiment for developing and validating the financial distress prediction framework relied on a structured, end-to-end toolchain, covering data integration, preprocessing, model implementation, and statistical validation. This toolchain was designed to handle the uniqueness of financial data (e.g., class imbalance of distressed firms) while ensuring experimental reproducibility, robustness, and efficiency. The detailed configuration of hardware and software is as follows:

4.1.1 Software Environment

STATA: Employed for large-scale database integration (merging firm-quarter financial indicators from multiple sources) and initial data standardization. It handled critical preprocessing steps such as aligning time-series records, resolving duplicate entries, and filtering invalid observations.

Python: Although several programming languages and platforms can be used for machine learning experiments, The proposed framework was developed using Python, which was chosen for its robust support for scientific computing, flexible programming environment, and broad range of machine learning libraries that facilitate efficient model implementation and experimentation. Python served as a multi-functional platform throughout the entire experimental process.

It complemented the workflow by handling advanced data manipulation and transforming the raw datasets into a format compatible with model training. Additionally, Python was employed as the core platform for implementing the entire program and for computing evaluation metric scores.

4.1.2 Hardware Configuration

The hardware setup was selected to handle the computational demands of large-scale financial data processing (52k+ observations) and iterative model training (18+ epochs per experiment). Detailed specifications are as shown in Table 4.1.

Table 4.1 Detailed Computer Specifications Configuration

Hardware Component	**Configuration Details**
Processor (CPU)	Intel(R) Core (TM) i7-10700 CPU @ 2.90 GHz (8 cores, 16 threads, max turbo 4.80 GHz)
Installed RAM	64.0 GB (63.9 GB usable, DDR4-2933 MHz)
Storage	1. 932 GB HDD (ST1000DM010-2EP102) 2. 477 GB SSD (HS-SSD-E100 512G)
Graphics Card (GPU)	NVIDIA GeForce GTX 750 Ti (2 GB VRAM)
Operating System (OS)	Windows 10 Education, Version 22H2, OS Build 19045.6093 (64-bit, x64-based processor)

4.1.3 Development & Debugging Tools

To ensure code reliability, version control, and efficient experimentation, the following auxiliary tools were used:

Jupyter Notebook: Enabled interactive code development and step-by-step testing of preprocessing and model iterations.

PyCharm (Community Edition): Primary IDE for project management, code debugging, Git integration, and virtual environment support.

This integrated hardware-software toolchain supported end-to-end experimentation, from raw financial data integration to statistical validation,

laying the foundation for reliable conclusions about the proposed framework's superiority in predicting financial distress.

4.2 DWARFB Implementation Architecture

Referring to the theoretical design presented in Section 3.3, this section illustrates the implementation architectural structure of the DWARFB framework. The DWARFB framework consists of three core modules that are tightly integrated into the ensemble model described in Chapter 3, namely sampling ratio adjustment, misclassified sample reweighting, and performance-driven ensemble strategies. These three components are unified into a bespoke and coherent framework, which is implemented as a complete program using Python. The pseudocode of the DWARFB implementation is provided in Algorithm 2.

The DWARFB framework integrates several mechanisms designed to address the limitations of conventional ensemble learning methods in imbalanced financial distress prediction. The following subsections describe each component in detail.

4.2.1 Parameters Setting

Prior to the implementation of the DWARFB, the model parameters must be specified. These settings are specific to the proposed framework and are directly incorporated into the Python implementation, as detailed in Appendix A. Table 4.2 summarizes the parameter configuration used in DWARFB.

Each parameter plays a distinct role in guiding the training process. For example, MAX_ITERATIONS controls the number of training rounds,

MISCLASSIFIED_RATIO emphasizes difficult samples, and CV_SPLITS and VAL_SPLIT govern cross-validation and validation monitoring.

Table 4.2 Summary of DWARFB Parameters

Parameter	Description	Default Value
MAX_ITERATIONS	Maximum number of iterations in the training process.	50
MISCLASSIFIED_RATIO	Proportion of misclassified samples emphasized in each iteration.	0.5
CV_SPLITS	Number of folds for cross-validation during subset training and evaluation.	5
VAL_SPLIT	Fraction of data reserved for validation to monitor model performance.	0.2
ALPHA	Controls the intensity of sample weight adjustment.	15
GAMMA	Governs the non-linear scaling of sample weights.	1
THRESHOLD_PERCENTILE	Percentile cutoff for identifying hard-to-classify samples.	80
USE_ENHANCED_TRACKING	Enables advanced tracking of misclassified samples.	True
REWARD_FACTOR	Reduces historical misclassification counts for correctly predicted positive-class samples.	0.5
CONSECUTIVE_THRESHOLD	Number of consecutive correct predictions required to reset a sample's misclassification history.	2
MISCLASSIFICATION_WINDOW_SIZE	Number of recent iterations used to track misclassification counts.	5
USE_SLIDING_WINDOW	Enables or disables the sliding iteration mechanism for misclassification tracking.	True
FEATURE_CORRELATION_ANALYSIS	Correlation threshold for detecting highly correlated feature pairs.	0.9
ADD_LAG_FEATURES	Lag periods used for constructing temporal features.	[2, 4]
TRAIN_RATIO	Proportion of the dataset allocated to the training set.	0.7
TEST_RATIO	Proportion of the dataset allocated to the test set.	0.3

Sample weighting is adjusted via ALPHA and GAMMA, while the treatment of hard-to-classify samples is managed through THRESHOLD_PERCENTILE, USE_ENHANCED_TRACKING, REWARD_FACTOR, CONSECUTIVE_THRESHOLD, and USE_SLIDING_WINDOW.

Feature engineering and data partitioning are controlled by FEATURE_CORRELATION_ANALYSIS, ADD_LAG_FEATURES, and TRAIN_RATIO, TEST_RATIO for dataset allocation. Collectively, these parameters ensure that DWARFB effectively addresses class imbalance and captures critical risk patterns in financial distress prediction.

The detailed implementation process is shown in Algorithm 2.

Algorithm 2: The pseudocode implementation of DWARFB

Require: Training set (X, y); sampling ratios (r_{pos}, r_{neg}); maximum iterations; early stopping rounds; model parameters, learning rate η_0; RF hyperparameters.
Ensure: An ensemble of trained models $\{M_t\}$ and corresponding weights $\{w_t\}$.
Initialize:
Set misclassification counters, learning rate, and tracking variables for all samples.
for each sampling ratio (r_{pos}, r_{neg}) **do**
    Reset misclassification counters for this stage.
    Initialize best cross-validation F1 score = 0 and early stopping counter = 0.
    for t=1 to max_iterations **do**
        // **Sampling according to sampling Ratio**
        - Select all positive samples (ST = 1).
        - Select required number of negative samples (ST = 0), prioritizing misclassified negatives.
        // **Enhance misclassified positive samples' weights**
        - If there are misclassified positive samples:
            - Create misclassification tracking based on misclassified count
            - Apply enhanced sample weighting
            - identify "extremely difficult" samples for additional enhancement
        // **Enhanced misclassification tracking with reward mechanism**
        - For correctly predicted positive samples with misclassification history:
            - Reduce their misclassification count by reward factor (default: 0.5)
            - If a sample is correctly predicted consecutively for 2 rounds
                - Clear its misclassification history completely
        - Use sliding iteration mechanism to prevent indefinite accumulation of misclassification counts
        // **Model Training**
        - Train model on sampled subset with weights.
        // **Update Statistics**
        - Update misclassification counters using cross-validation and predictions on the full pool.
        // **Model Weight & Learning Rate**
        - Compute model weight (e.g., logit of CV F1); adjust learning rate based on F1 improvement (increase if improved, decrease if not).
        - Adjust learning rate based on F1 improvement.
        // **Early Stopping**
        - If CV F1 does not improve for early stopping rounds, break.
    end for (iterations)

33. **end for (sampling ratios)**
34. **Ensemble Prediction:**
35. Combine predictions from all trained models $\{M_t\}$ using their weights $\{w_t\}$.
36. **Evaluation & Analysis:**
37. Evaluate on test set, plot curves, analyse misclassified samples, and save results.

4.2.2 Positive-to-Negative Sampling Strategy

At each stage of the sampling ratio, the sampling logic is a targeted stratified sampling strategy designed for imbalanced classification corporate distress prediction (positive samples (1) are the distress samples and negative samples (0) are the non-distress). Its core objective is to retain all positive samples and prioritize "hard-to-learn" negative samples, while maintaining the desired positive-to-negative ratio in the training subset.

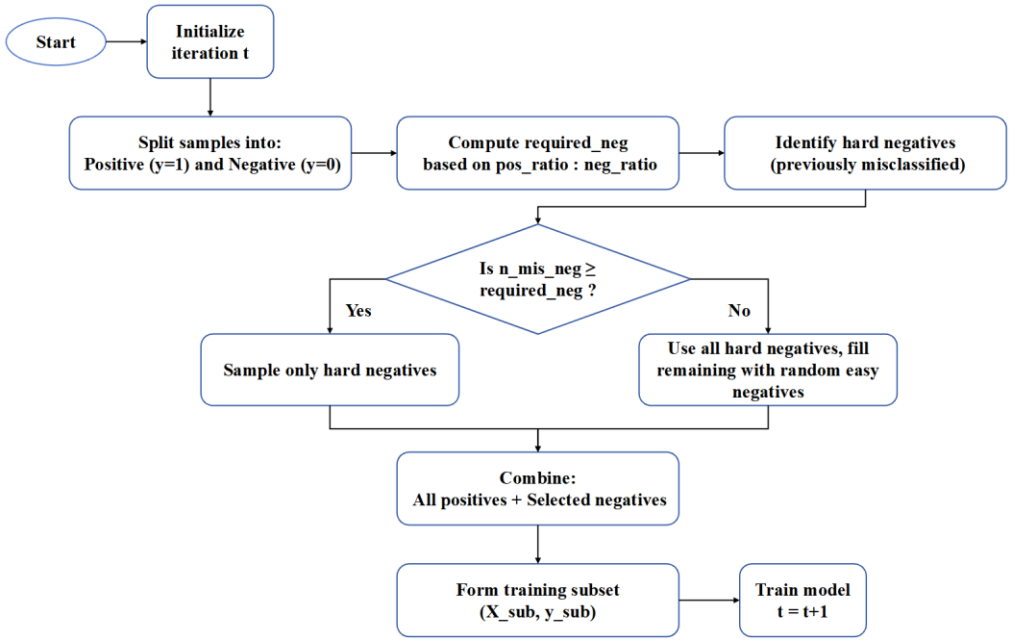

Figure 4.1 Iterative Sampling Process of DWARFB

Figure 4.1 illustrates the iterative sampling process of DWARFB. At each iteration, the training data are initially separated into positive and negative samples, with the number of negative samples selected based on a predefined positive-to-negative ratio. Previously misclassified negative samples are identified as hard instances and are given priority in the sampling process. If the

number of hard negatives is sufficient, only these samples are selected; otherwise, all hard negatives are retained, and the remaining slots are filled with randomly chosen easy negatives. Finally, all positive samples and the selected negative samples are combined to form the training subset for the current iteration.

4.2.3 Misclassification Tracking with Reward Mechanism

During each iteration, a misclassification tracking system is introduced to monitor misclassified samples and record their misclassification frequencies. To prevent samples from being over-penalized, a fixed sliding iteration is adopted to store misclassification information from the most recent iterations; when the window size is exceeded, older records are discarded. In addition, a reward mechanism is incorporated to reduce the penalty for samples that are predicted correctly in consecutive iterations, thereby encouraging stable and consistent correct predictions.

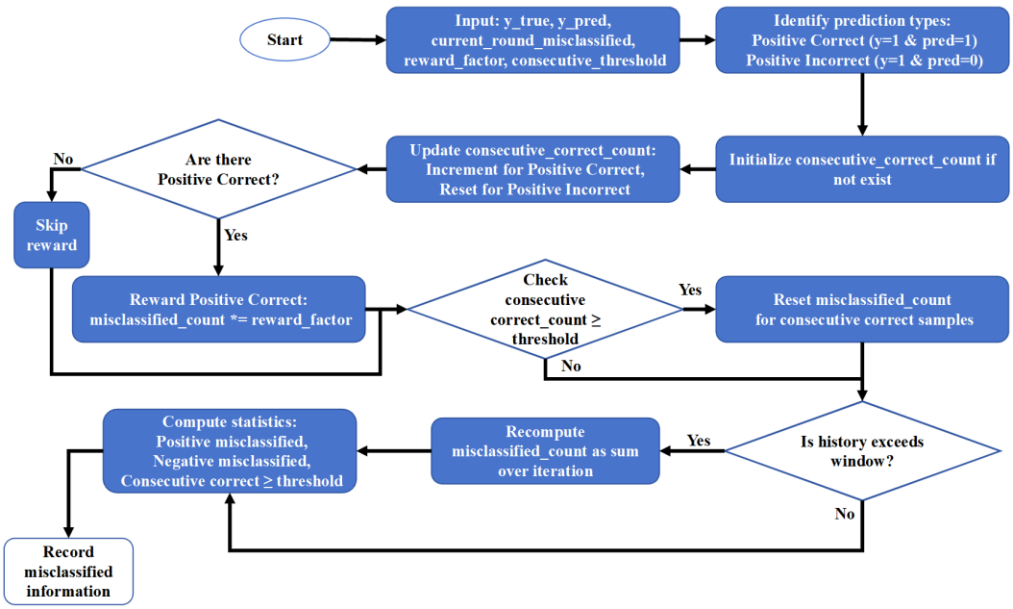

Figure 4.2 Misclassification Tracking with Reward Mechanism

Figure 4.2 shows the workflow of this tracking mechanism, after receiving the input, the system first identifies whether the current prediction is

a true positive or a false negative. A counter tracks consecutive correct positive predictions, incrementing with each true positive and resetting with any misclassification. For each correct positive prediction, a reward is applied by adjusting the misclassification count; otherwise, the reward step is skipped. Then, the mechanism examines whether the number of consecutive correct predictions for each sample has reached the predefined threshold. If the threshold is satisfied, the historical misclassification count of the corresponding sample is reset to zero.

When the history length exceeds the predefined window size, older records are discarded to maintain a fixed window, and the misclassified count is recomputed as the sum of misclassifications within the window; otherwise, all current records are retained. Finally, aggregated statistics are computed, including the numbers of false negatives, false positives, and whether the consecutive correct prediction threshold has been reached, after which these results are recorded. This process is repeated iteratively, ensuring that misclassified samples are not over-penalized while consistently correctly predicted samples are appropriately rewarded.

4.2.4 Misclassified Sample Reweighting

The misclassified sample reweighting adjusts sample weights based on misclassification counts, with a focus on hard-to-learn positive samples (distress samples). It aims to assign higher weights to positive samples that the model repeatedly misclassifies, forcing the model to prioritize learning these “hard positive samples” during training and addressing class imbalance bias.

Figure 4.3 shows how the sample reweighting mechanism adjusts sample importance according to their misclassification difficulty. All samples

are first assigned a uniform baseline weight. Misclassification counts obtained from the sliding iteration are then normalized and used to exponentially scale sample weights, where α controls the strength of adjustment and γ determines the degree of nonlinearity. To further emphasize the hardest samples, those in the top percentile of misclassification frequency receive an additional weight boost. This two-stage strategy ensures that more learning attention is allocated to frequently misclassified samples, improving model performance on difficult and minority-class instances while maintaining stable training behaviour.

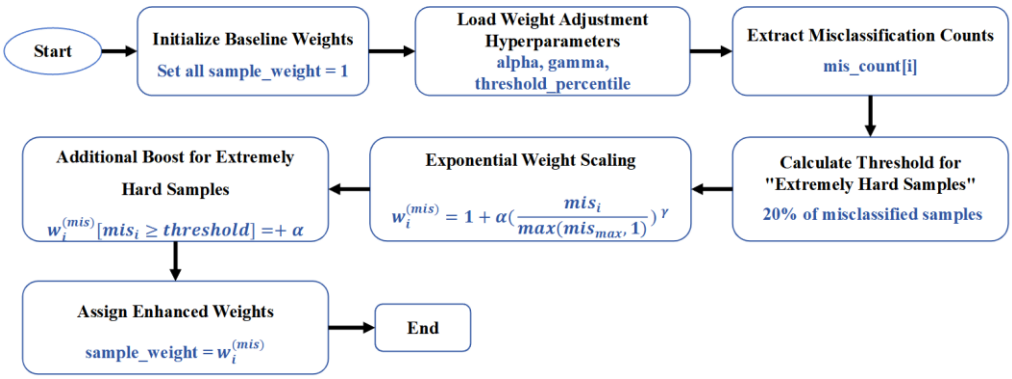

Figure 4.3 Misclassified Sample Reweighting Workflow

4.2.5 Performance-Driven Dynamic Ensemble Strategy

The dynamic ensemble strategy in this program follows an iterative model accumulation + adaptive weighting paradigm, intended to dynamically combine multiple iterative models for improved performance. The ensemble does not use equal weighting; instead, it assigns weights to each model based on iteration cross-validation performance.

Algorithm 3 illustrates the process of the dynamic ensemble strategy. The performance-driven dynamic ensemble strategy begins by initializing all models with equal weights. During training, each model's performance, such as cross-validation F1-scores, is tracked and recorded. For prediction, the strategy calculates dynamic model weights based on past performance, and a weighted

average of predicted probabilities is computed using an optimal threshold to generate the final output. This process is repeated iteratively, continuously updating the ensemble and weights to enhance predictive accuracy over successive iterations.

Algorithm 3: DWARFB Iterative Ensemble Training and Prediction

1. Initialize Iterative Model by setting each model equal weight.
2. For each training iteration:
   2.1 Train a model using adaptive sample weights.
   2.2 Save the trained model to SAVED_MODELS.
   2.3 Track performance metrics (such as Cross-validation F1-score) in results.
3. When prediction is needed:
   3.1 Calculate dynamic weights for each model based on past performance.
   3.2 Generate weighted average of predicted probabilities (soft voting).
   3.3 Output final prediction using the optimal threshold.
4. Repeat: Update the ensemble and dynamic weights in subsequent iterations.

4.3 Comparative Study

To assess the performance of the proposed DWARFB ensemble framework in financial distress prediction, this study undertakes a thorough comparative evaluation. The experiments compare DWARFB with various state-of-the-art machine learning models, as well as with models incorporating different resampling techniques under class-imbalanced scenarios, to evaluate its predictive performance and robustness. In addition, comparisons with benchmark misclassification penalty approaches are performed to demonstrate the superiority of the proposed misclassified sample reweighting mechanism.

4.3.1 Baseline Machine Learning Models

Six benchmark models are selected to compare. Since DWARFB is a tree-based framework, the DT is chosen as the fundamental baseline to assess its performance relative to a single-tree classifier. The RF, widely recognized in prior research as an enhanced ensemble extension of the DT, is also included for comparison. The SVM, a classical machine learning algorithm frequently

employed in financial distress prediction studies, is selected to represent margin-based classifiers. Considering the increasing success of deep learning techniques in predictive tasks, the MLP is included as a representative neural network model. Finally, two widely used boosting frameworks, AdaBoost and XGBoost, are incorporated to benchmark the proposed boosting-based RF design.

(1) DT

The DT is defined as:

$$\text{Gini} = 1 - \sum_{k=1}^{K} {p_k}^2 \qquad \text{Eq. 4.1}$$

where p_k is the proportion of samples belonging to class k in a node.

(2) RF

The final classification prediction of RF is obtained by majority voting:

$$\hat{y} = \text{model}\{h_1(x), h_2(x), \ldots., h_B(x)\} \qquad \text{Eq. 4.2}$$

(3) SVM

SVM aims to find the optimal hyperplane $w^T x + b = 0$ that maximizes the margin between two classes:

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \qquad \text{Eq. 4.3}$$

Subject to

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \qquad \text{Eq. 4.4}$$

where C controls the trade-off between maximizing the margin and minimizing classification errors ξ_i.

(4) MLP

The mapping function of MLP is defined as:

$$\hat{y} = f(x;\theta) = \sigma(W^{(L)}\sigma(W^{(L-1)}\cdots\sigma(W^{(1)}x + b^{(1)})\cdots) + b^{(L)}) \quad \text{Eq. 4.5}$$

where $W^{(L)}$ and $b^{(L)}$ denote the weight matrix and bias vector of layer L, and $\sigma(\cdot)$ is an activation function (e.g., ReLU, sigmoid). Model parameters θ are optimized using backpropagation to minimize the loss function, such as cross-entropy.

(5) AdaBoost

The final ensemble of AdaBoost is defined as:

$$H(x) = \text{sign}(\sum_{t=1}^{T}\alpha_t h_t(x)) \quad \text{Eq. 4.6}$$

where $h_t(x)$ is the weak learner at iteration t, and its corresponding weight α_t depends on its classification error ε_t:

$$\alpha_t = \frac{1}{2}\ln(\frac{1-\varepsilon_t}{\varepsilon_t}) \quad \text{Eq. 4.7}$$

(6) XGBoost

The objective function of XGBoost at iteration t is defined as:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad \text{Eq. 4.8}$$

where l is the differentiable loss function and $\Omega(f_t)$ is the regularization term:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda\sum_{j=1}^{T} {w_j}^2 \quad \text{Eq. 4.9}$$

Here, T is the number of leaves, and w_j represents the leaf weight. The model incrementally adds trees f_t to fit the negative gradient of the loss function, achieving improved generalization and computational efficiency.

These six baseline models represent diverse learning paradigms, including decision-based, ensemble-based, margin-based, neural, and boosting algorithms. Their inclusion provides a comprehensive benchmark for assessing

the proposed DWARFB framework's performance, robustness, and adaptability in financial distress prediction tasks.

4.3.2 RF Variants

As the DWARFB framework is a boosting-based extension of the RF algorithm, this comparative experiment involves benchmarking it against RF models integrated with different resampling techniques and their balanced variants under class imbalance scenarios. The baseline methods are classified into three categories: standard ensemble models, balanced variants, and sampling-based approaches.

(1) Standard RF

The prediction of standard RF classifier for an input x is given by:

$$\mathrm{RF(x)} = \mathrm{majority\_vote}(\{\mathrm{h}_1(\mathrm{x}), \mathrm{h}_2(\mathrm{x}), \ldots, \mathrm{h}_n(\mathrm{x})\}) \quad \text{Eq. 4.10}$$

To encourage diversity among the ensemble members, each Decision Tree, h_i, is trained using a bootstrap-resampled version of the training dataset, and only a randomly selected subset of predictors is evaluated at each split. The trees are induced using the Classification and Regression Tree (CART) algorithm, where the optimal partition is determined by minimizing impurity according to criteria such as the Gini index or entropy, and the bootstrap sampling preserves the original class distribution. No specific mechanism is applied to address class imbalance; consequently, the optimization process may be dominated by the majority class in highly imbalanced datasets, potentially reducing its sensitivity to destress samples. The Standard RF implementation uses the standard configuration without any special handling for class imbalance.

(2) Balanced RF Variants

Two balanced variants are implemented to address class imbalance:

RF with Class Weight Balancing, the sample weight for observation i is computed as:

$$\text{sample\_weight}[i] = \frac{\text{class\_weight}[y_i]}{n_{samples}} \quad \text{Eq. 4.11}$$

where y_i is the class label of sample i, and $n_{samples}$ is the total number of training observations. The class weight itself is calculated as:

$$\text{class\_weight}[c] = \frac{n_{samples}}{n_{class} \times n_c} \quad \text{Eq. 4.12}$$

where n_{class} is the total number of unique classes, and n_c is the number of samples belonging to class c. This formula ensures that the total weight assigned to each class is balanced, with rarer classes receiving proportionally higher weights.

Balanced RF is another variant of RF. Given a training dataset $D = \{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in R^p$ are feature vectors and $y_i \in \{0,1\}$ denote class labels, let C_{min} and C_{maj} represent the minority and majority class sets, respectively, and N_{min} the number of minority samples. For each tree $t \in \{1, \dots, T\}$, Balanced RF generates a balanced bootstrap dataset D_t by sampling N_{min} instances from each class:

$$D_t = \text{Sample}(C_{min}, N_{min}) \cup \text{Sample}(C_{maj}, N_{min}) \quad \text{Eq. 4.13}$$

This ensures that each tree is trained on an equal number of minority and majority examples, reducing the bias toward the majority class.

After sampling, a DT h_t is trained on D_t using standard CART splitting criteria, such as Gini impurity or entropy, with random feature selection at each split to enhance diversity. The ensemble prediction for a new sample xxx is obtained via majority voting:

$$\hat{y} = \arg\max_{c \in \{0,1\}} \sum_{t=1}^{T} I(h_t(x) = c) \quad \text{Eq. 4.14}$$

where $I(\cdot)$ is the indicator function. This balanced sampling occurs independently for each tree, meaning that different trees see different balanced subsets of the data, which improves generalization.

4.3.3 Resampling Methods

The comparative study evaluates multiple resampling strategies to address class imbalance.

(1) SMOTE

The default number of k neighbours in the standard SMOTE algorithm is 5. The process involves selecting a minority class sample x_i, identifying one of its nearest neighbours x_j, and generating a synthetic sample x_{synth} as follows:

$$x_{synth} = x_i + \lambda \cdot (x_j - x_i) \quad \text{Eq. 4.15}$$

where x_i, and x_j are minority class neighbours, and λ is a random weight $\in [0, 1]$.

(2) Borderline-SMOTE

Borderline-SMOTE starts by analysing the k-nearest neighbours of each minority class instance. Based on the distribution of class labels within this neighbourhood, each instance is categorized as safe, noisy, or dangerous. Only those identified as dangerous, which are surrounded by a mix of majority and minority class neighbours, are selected for oversampling.

For a given dangerous minority instance x_i and its k-nearest minority neighbours $\{k_1, k_2, \ldots, k_j\}$, the SMOTE interpolation process generates synthetic samples by combining x_i with a selected neighbour k_j. The

synthetic instance is created using a linear interpolation between the two points based on their distance. The mathematical representation is given as follows:

$$x_{synth} = x_i + \lambda \cdot (k_j - x_i) \quad \text{Eq. 4.16}$$

where λ is a random weight $\in [0, 1]$.

(3) ADASYN

ADASYN assigns a higher weight to such samples when generating synthetic data. Formally, for each minority sample x_i, the number of synthetic samples to be generated is proportional to its local density measure. The followed equation is the mathematical expression of this method.

$$g_i = \frac{r_i}{\sum_{j=1}^{n} r_j} \quad \text{Eq. 4.17}$$

where r_i is the ratio of the majority class examples to x_i, and n is the total number of observations in the minority class.

(4) RUS

The core assumption of RUS is that the majority class contains redundant or less informative instances that can be discarded without significantly compromising model performance. However, due to the random nature of instance removal, there is a potential risk of discarding important majority examples and thus losing valuable information.

(5) SMOTE-Tomek

SMOTE-Tomek is a two-stage hybrid technique that integrates SMOTE oversampling with Tomek Links under-sampling. Firstly, SMOTE generates synthetic minority class samples using interpolation between existing instances and their k-nearest neighbours (see Equation of SMOTE). In the latter stage, Tomek Links are identified within the augmented dataset to remove the nearest-neighbour instances from the majority class:

$$NN(x_i) = x_j \ \wedge \ NN(x_j) = x_i \ \wedge \ y_i \neq y_j \qquad \text{Eq. 4.18}$$

Two instances (x_i,x_j) form a Tomek Link if they are nearest neighbours ($NN(x)$) from opposite classes (y_i,y_j).

The combination of oversampling through SMOTE and data cleaning via Tomek Links enables SMOTE-Tomek to enhance both class balance and the discriminative quality of the feature space.

4.3.4 Misclassification Penalty Methods

To evaluate the effectiveness of the proposed Misclassified Sample Reweighting Mechanism in DWARFB, we designed a comparative experiment grounded in three fundamental penalty concepts identified in prior literature. This experiment aims to isolate and assess the contribution of different penalty strategies to model performance. To ensure fairness, all experimental settings were held constant, with the only variation being the penalty mechanism applied to misclassified samples.

It is recognized that some benchmark penalty strategies were originally developed for specific learning contexts and may not be fully optimized within the DWARFB framework. Nonetheless, their inclusion serves an important purpose: to illustrate the performance gains achieved by integrating and enhancing these penalty concepts within our unified mechanism. The comparative results therefore highlight not only the effectiveness of the proposed model but also the added value generated by combining complementary misclassification penalty strategies.

The benchmark misclassification penalty concepts incorporated into the comparative analysis are summarized in Table 4.3. By integrating these established ideas into a unified structure, the comparison experiment evaluates

how the combined penalty mechanism enhances DWARFB's ability to emphasize informative misclassified samples.

Table 4.3 Benchmark Concepts in DWARFB Misclassification Weighting

Benchmark Concept	Original formula	Role in the formula
Adaptive Boosting (AdaBoost)	Provides the progressive weighting idea based on misclassification count $\mathrm{mis_i}$ $w_i^{(t+1)} = w_i^{(t)} \times \exp(\alpha_t \times I[h_t(x_i) \neq y_i])$	The weighting mechanism assigns higher weights to misclassified samples, where α and $\mathrm{mis_i}$ form the basis of the calculation.
Focal Loss	Inspires the exponent term $(1 - p_i)^{\gamma}$ controlling emphasis on hard samples	$(\frac{\mathrm{mis_i}}{\max(\mathrm{mis_{max}}, 1)})^{\gamma}$
Online Hard Example Mining (OHEM)	Forms the indicator term $I(\mathrm{mis_i} \geq \mathrm{threshold})$	$\alpha \times I(\mathrm{mis_i} \geq \mathrm{threshold})$

Table 4.4 Weight Formulas for Comparison

Weight Type	Benchmark $w_i^{(mis)}$
AdaBoost-like	$w_i^{(mis)} = \exp(\alpha \cdot \mathrm{mis_i})$
Focal-like	$w_i^{(mis)} = 1 + \alpha(\frac{\mathrm{mis_i}}{\max(\mathrm{mis_{max}}, 1)})^{\gamma}$
Proposed weight	$w_i^{(mis)} = 1 + \alpha(\frac{\mathrm{mis_i}}{\max(\mathrm{mis_{max}}, 1)})^{\gamma} + \alpha \times I(\mathrm{mis_i} \geq \mathrm{threshold})$

Based on the original formulations of the benchmark penalty concepts and their applicability within the DWARFB framework, the corresponding comparative weighting functions are constructed and summarized in Table 4.4. These adapted formulas represent the penalty mechanisms that can be directly applied within DWARFB's Misclassified Sample Reweighting structure. For clarity, the three variants are referred to as AdaBoost-like, Focal-like, and the

Proposed Weight, each reflecting a progressively richer incorporation of misclassification information.

The original AdaBoost weight update rule is driven by a binary misclassification indicator rather than a continuous loss value. Since the DWARFB framework employs the cumulative misclassification count mis_i as its core difficulty signal, the AdaBoost penalty is adapted to ensure consistency across all benchmark mechanisms. Replacing the indicator term with mis_i preserves AdaBoost's fundamental principle, progressively amplifying the weight of harder samples, while aligning the formulation with the unified misclassification-based structure required by DWARFB. This adaptation enables a methodologically fair comparison among the AdaBoost-like, Focal-like, and proposed weighting strategies.

4.3.5 Statistical Significance Testing

To verify the statistical significance of DWARFB relative to comparative approaches, we conducted multiple independent experiments. Each method was retrained and evaluated over 20 independent runs, using distinct random seeds (42 + run) to ensure independence of results. Performance differences between methods were statistically assessed using the Wilcoxon rank-sum test at a significance level of $\alpha = 0.05$. For DWARFB, hyperparameters were optimized based on parameter sensitivity analysis, with automatic detection of available parameter sources. Base model parameters for RF-based methods were set to either method-specific optimized values or consistent default settings, while other state-of-the-art models were evaluated using default parameters. This multi-run experimental design enhances the

reliability of performance estimates and strengthens the validity of statistical inferences regarding method superiority.

4.4 Parameter Sensitivity and Robustness

Parameter sensitivity analysis is essential for assessing the robustness and determining the optimal configuration of the proposed DWARFB model. This analysis systematically evaluates how variations in key parameters affect model performance, providing insights into parameter stability and guiding the selection of settings that maximize predictive accuracy and generalization capability.

4.4.1 Parameter Scope Definition

To systematically evaluate the sensitivity of the DWARFB model to key hyperparameters, we first define the scope of parameters to be analysed and their respective value ranges based on both domain knowledge and empirical validation. The sensitivity analysis encompasses two categories of parameters: the core iterative learning parameters (alpha and gamma) and the base RF model parameters, each evaluated through distinct experimental designs.

(1) Core Misclassified Sample Weighting Parameters

The misclassification-based weighting mechanism updates sample weights dynamically at each iteration. The formula is shown in the conceptual definition in the section 3.3.2. This adaptive scheme ensures that difficult samples exert greater influence on subsequent iterations, thereby improving the model's capacity to correctly classify minority and borderline instances.

Alpha (Weight Adjustment Intensity):

This parameter controls the intensity of weight adjustments for misclassified samples during the iterative training process. A larger alpha value

amplifies the emphasis on correcting misclassified instances, potentially improving recall but risking overfitting to noisy samples. Based on preliminary experiments, the value range for alpha is set to [2, 4, 6, 8, 10, 12, 15] to cover both moderate and strong adjustment intensities.

Gamma (Nonlinear Adjustment Coefficient):

This parameter regulates the nonlinearity of weight updates, influencing how the model balances between exploring new patterns and exploiting known ones. A higher gamma value introduces stronger nonlinearity, enabling more flexible adaptation to complex data distributions. The value range for gamma is defined as [1, 2, 3, 4, 5, 6], which was determined to balance computational efficiency and the ability to capture nonlinear relationships.

Other model parameters are fixed during the sensitivity analysis to isolate the impact of alpha and gamma. These fixed parameters are shown in Table 4.5.

Table 4.5 Fixed DWARFB Parameters for Sensitivity Analysis

Items	**Parameter**	**Values**
Base RF settings	N_ESTIMATORS	100
	MAX_DEPTH	10
	MAX_FEATURES	'sqrt'
	MIN_SAMPLES_SPLIT	10
	MIN_SAMPLES_LEAF	10
Iterative training configurations	MAX_ITERATIONS	30
	NOISE_SAMPLE_CONTAMINATION	0.01
Ensemble settings	THRESHOLD_PERCENTILE	80

This parameter scope ensures comprehensive coverage of potential configurations while maintaining focus on the core hyperparameters that drive the model's adaptive learning behaviour.

(2) Base RF Parameters

In addition to the misclassified sample weighting parameters, a dedicated sensitivity analysis is conducted on the foundational RF hyperparameters that determine the structure and behavior of the base learners. These parameters include:

N_ESTIMATORS: Number of DTs in the ensemble, tested across [50, 100, 150] to evaluate the impact of ensemble size on performance and computational efficiency.

MAX_DEPTH: Maximum depth of individual trees, with values [5, 10, 15] to assess the trade-off between model complexity and overfitting risk.

MIN_SAMPLES_SPLIT: Minimum samples required to split an internal node, tested at [2, 5, 10] to control tree growth and prevent over-specialization.

MIN_SAMPLES_LEAF: Minimum samples required at a leaf node, evaluated across [1, 5, 10] to regulate leaf node granularity.

For this base parameter analysis, MAX_FEATURES is fixed at ‘sqrt’ to maintain consistency, while alpha and gamma are set to their default values (4 and 3 respectively) to isolate the impact of the base model parameters.

This comprehensive parameter scope ensures that both the foundational structure (base RF) and adaptive learning mechanisms (alpha and gamma) are thoroughly evaluated, providing a complete understanding of parameter sensitivities while maintaining focus on the core hyperparameters driving the model’s performance.

4.4.2 Grid Search Optimization

Grid search optimization systematically examines the parameter space by testing every possible combination of parameter values within specified ranges. This exhaustive search strategy ensures identification of the optimal configuration under the specified bounds. While computationally intensive for large parameter spaces, grid search provides a reliable benchmark for performance tuning and ensures reproducibility in parameter selection.

(1) Misclassified Parameter Space Exploration

Alpha Parameter (α): Governs the intensity of weight adjustment for misclassified samples. Higher α values assign substantially greater weight to difficult samples, making the reweighting more aggressive.

Gamma Parameter (γ): Controls the degree of nonlinearity in the weight adjustment function. Larger γ values accentuate weight disparities between frequently misclassified and correctly classified samples.

The grid search evaluates all $7 \times 6 = 42$ parameter combinations. For each configuration, the specific α and γ values are first assigned, after which the DWARFB model is trained using the iterative boosting–RF structure. Key classification metrics are then computed. Both the performance metrics and the corresponding parameter configuration are recorded, and the current best configuration is updated whenever superior performance is achieved.

(2) Base RF parameters Space Exploration

For sensitivity analysis of the base RF parameters, a systematic grid search over 81 combinations was conducted for key hyperparameters governing tree structure and ensemble behaviour.

The parameter ranges included: N_ESTIMATORS [50, 100, 150] to assess the effect of ensemble size on predictive performance and computational cost; MAX_DEPTH [5, 10, 15] to balance model complexity and overfitting risk; MIN_SAMPLES_SPLIT [2, 5, 10] to control tree growth by setting the minimum number of samples required to split an internal node; and MIN_SAMPLES_LEAF [1, 5, 10] to regulate leaf granularity and prevent over-specialization. During this analysis, MAX_FEATURES was fixed at 'sqrt' for consistency, while α and γ were set to their default values (15 and 1, respectively) to isolate the effect of the base model parameters.

This comprehensive grid search provides a thorough understanding of how foundational RF settings influence performance while maintaining focus on the core hyperparameters of the base learners.

4.5 Experimental Study

To support the objectives of this research, an experimental study is conducted in this thesis. Experimental study is a widely adopted methodology in computer science for validating conceptual designs and evaluating system performance. In this work, the experimental study is used to examine whether the proposed DWARFB framework satisfies the hypotheses stated in Section 1.4, particularly in terms of adaptive imbalance handling, misclassification-aware learning, and robust ensemble construction under severe class imbalance.

The effectiveness of DWARFB is validated through multiple independent runs, systematic comparisons with baseline and state-of-the-art methods, and statistical significance testing using the Wilcoxon rank-sum test. In addition, sensitivity analyses and grid search procedures are performed to

investigate the impact of key hyperparameters, ensuring that the conclusions drawn are reliable, robust, and reproducible.

4.5.1 Experimental Objectives

This refers to the hypotheses outlined in Section 1.4 of the study, which emphasize:

- The superiority of adaptive learning-embedded imbalance handling over static resampling strategies.
- The effectiveness of persistent misclassification-aware weighting in identifying informative distressed firms while suppressing noisy samples.
- The advantages of incorporate accumulated misclassification information ensemble learning for achieving stable, interpretable, and robust predictions.

To test these hypotheses, the experimental study evaluates the proposed DWARFB framework through a comprehensive set of comparative and analytical experiments. Therefore, the experimental study serves the following purposes:

The first purpose is to test Hypothesis **H1** on adaptive imbalance handling. DWARFB is compared with conventional classifiers and resampling-based methods to show that embedding imbalance handling within learning improves minority-class generalization while preserving the original data distribution.

The second purpose is to test Hypothesis **H2** on persistent misclassification-aware weighting. DWARFB is evaluated against misclassification penalty methods while tracing cumulative misclassification

patterns to demonstrate that historical misclassification-guided weighting more effectively identifies borderline distressed firms and mitigates the influence of noisy minority samples.

The third purpose is to test Hypothesis **H3** on history-aware ensemble learning. Through the experiments to assess DWARFB's stability, robustness, and interpretability compared with conventional ensemble pipelines, thereby verifying that the integrated ensemble model design improves prediction reliability.

The fourth purpose is to ensure experimental reliability and validity. Multiple independent runs with different random seeds and statistical significance testing confirm that performance differences are consistent and meaningful.

The fifth purpose is to examine parameter sensitivity and model stability. Variations in iterative learning parameters and RF hyperparameters are analyzed to understand their effects on DWARFB's performance and robustness.

4.5.2 Experimental Data Set

The experimental dataset combines 12 sub-datasets from the CSMAR database and comprises 368 features with a total of 52,528 firm-quarter observations. A detailed description of the dataset is provided in Section 3.1.

4.5.3 Experimental Design

To ensure consistency with the research framework, the experimental procedures are designed to address the objectives identified in Section 4.5.1 and to validate the hypotheses proposed in this thesis, as illustrated in Figure 4.4.

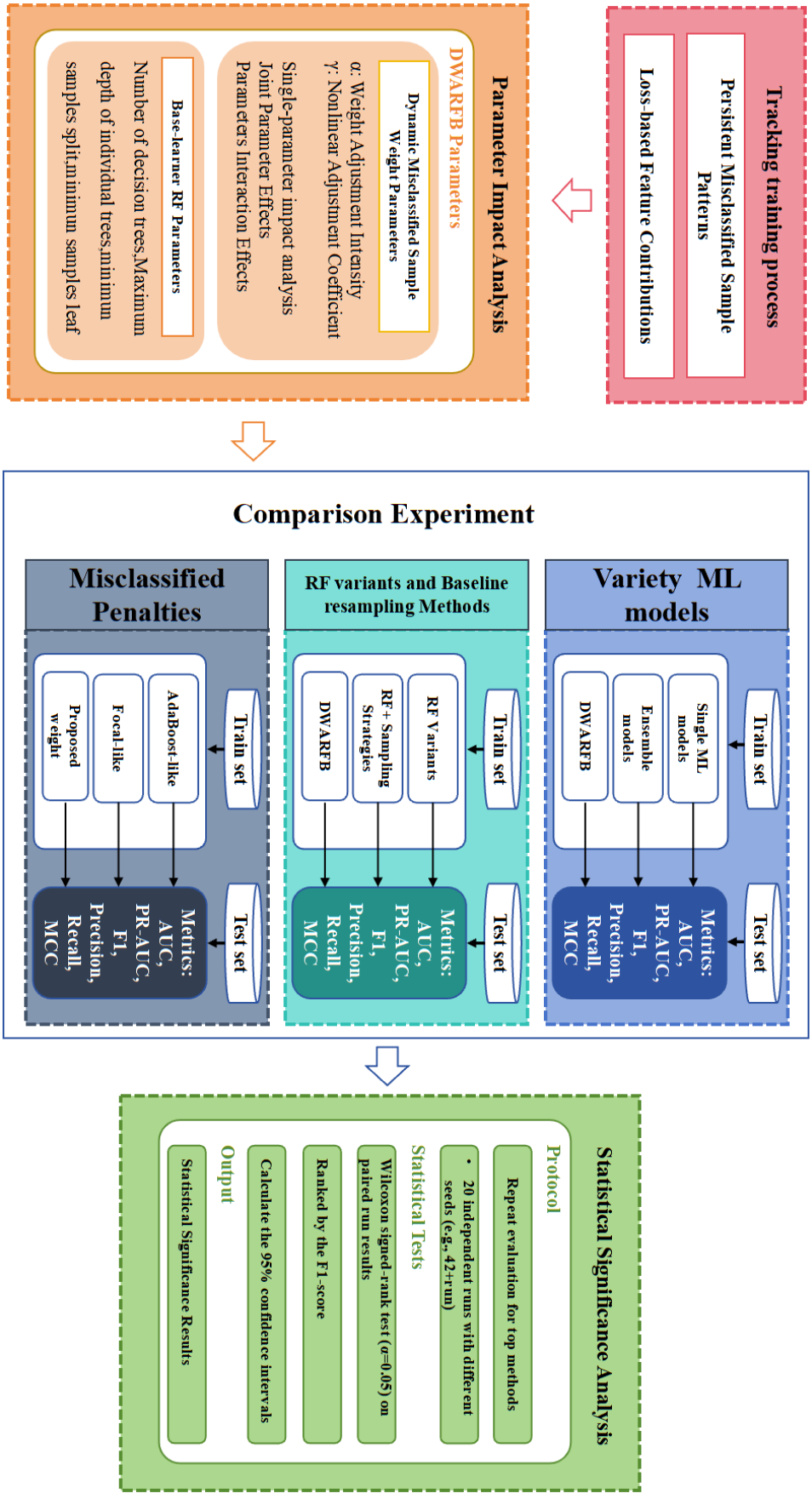

Figure 4.4 Experimental Workflow

(1) Tracking Training Process

The first experiment focuses on tracking how the DWARFB model learns from difficult samples during training. Persistent misclassified sample patterns are recorded across iterations to identify borderline distressed firms and assess how the adaptive weighting mechanism adjusts their influence. Loss-based feature contributions are analyzed to determine which features most strongly drive predictions. This stage provides insights into the model's learning dynamics, highlighting the effectiveness of sample weighting based on accumulated misclassification information and its impact on minority-class generalization.

(2) Parameter Sensitivity Analysis

The second experiment examines how variations in iterative learning parameters (weight adjustment intensity α and nonlinear adjustment coefficient γ) and baseline RF hyperparameters (number of trees, maximum depth, minimum samples per split, and minimum samples per leaf) affect DWARFB performance. Sensitivity analysis enables assessment of model stability, adaptability, and robustness under different parameter configurations.

(3) Comparative Benchmarking

The third experiment benchmarks DWARFB against existing methods using a Train/Test split. Comparisons include:

- Variety of ML models: Single machine learning models, ensemble models, and DWARFB.
- RF variants and resampling methods: RF variants, RF with resampling, and DWARFB.

- Misclassification penalty methods: AdaBoost-like, Focal-like, and DWARFB with persistent misclassification-aware weighting.

Evaluation metrics include AUC, PR-AUC, F1-score, Precision, Recall, and MCC, allowing comprehensive assessment, particularly for imbalanced scenarios. This experiment quantifies how DWARFB outperforms conventional approaches while leveraging adaptive weighting and tracking persistent misclassification.

(4) Statistical Significance Analysis

The fourth experiment is designed to investigate the robustness and reproducibility of the proposed framework. To minimize the impact of stochastic factors, each model is evaluated across 20 independent runs, with a unique random seed used for every execution. Paired results are analyzed using the Wilcoxon signed-rank test ($\alpha = 0.05$), and 95% confidence intervals for F1 scores are calculated. Ranking models by F1 score provides strong evidence of performance consistency, reliability, and superiority of DWARFB.

Through these experiments, the experimental design provides a thorough evaluation of DWARFB, demonstrating its ability to track and learn from persistently misclassified samples, its parameter adaptability, superior predictive performance, and robustness in handling imbalanced financial distress datasets.

4.5.4 Evaluation Metrics

Recognizing that no single metric can fully characterize model performance on imbalanced datasets, this study adopts a combination of complementary evaluation measures. The selected metrics include ROC-AUC, PR-AUC, F1-score, Precision, Recall, and the Matthews Correlation Coefficient

(MCC), providing a balanced assessment of classification effectiveness from multiple perspectives.

Among these metrics, PR-AUC and F1-score are particularly important for imbalanced classification tasks because they emphasize minority-class prediction performance. MCC is also included as it provides a balanced measure even when class distributions are highly skewed.

4.6 Summary

In this chapter, we introduced the tools used in this research and presented the implementation of the DWARFB framework. We also described the baseline and comparative models, as well as the parameter sensitivity analysis and the corresponding parameter ranges. An experimental study was designed to evaluate the framework's performance.

The challenges of severe class imbalance and persistently hard-to-classify distressed cases remain unresolved. Questions stated in the section 1.4, such as "Can we address severe class imbalance in financial distress prediction while preserving the original data distribution and improving detection of rare but critical distressed firms?" and "Can models identify and emphasize persistently hard-to-classify yet informative distressed firms without being misled by noisy minority samples?" guided the design of the DWARFB implementation.

To address these challenges, a misclassification tracking mechanism was embedded in the DWARFB framework. This mechanism monitors persistently misclassified samples, providing insights into sample difficulty patterns and enabling adaptive weighting to improve minority-class detection and overall predictive robustness.

In the next chapter, we conduct a comprehensive experimental study of the DWARFB framework and compare its performance with baseline and state-of-the-art methods, demonstrating how the integration of adaptive imbalance handling, persistent misclassification-aware weighting, and performance-driven ensemble strategy is effective for predicting financial distress under severe class imbalance.

CHAPTER 5 EXPERIMENTAL RESULTS AND DISCUSSION

Having established the implementation details and experimental methodology in Chapter 4, this chapter concentrates on the empirical evaluation of the proposed DWARFB framework. The experimental results are presented, interpreted, and discussed to demonstrate the effectiveness and practical value of the proposed approach.

To investigate the effectiveness of the boosting-based enhancement to the RF model, the impact of DWARFB's misclassification-aware mechanisms was evaluated. Eight RF variants combined with different resampling techniques were systematically compared to assess improvements in classification performance. Furthermore, to demonstrate the suitability and robustness of DWARFB for financial distress prediction, seven state-of-the-art models were benchmarked against the proposed method. Additionally, the chapter also presents comparative results against benchmark misclassification penalty mechanisms and provides a comprehensive parameter sensitivity analysis.

The figures and tables presented in this chapter provide empirical support for the experimental objectives and the corresponding research hypotheses formulated in this study. Supplementary details related to these results are provided in Appendix B.

5.1 Model Evaluation and Feature Analysis

In this section, we assess the overall performance of DWARFB by analysing its learning ability on misclassified samples and examining the contributions of relevant features within the misclassification tracking system. We also compare feature importance distributions before and after training.

These evaluations provide a comprehensive view of DWARFB's effectiveness in addressing class imbalance and reducing persistent misclassifications. The experimental results support the objectives outlined in Section 1.4, offering both an overview and an internal model perspective to demonstrate DWARFB's capability in handling class imbalance and hard-to-classify samples.

5.1.1 Evaluation of DWARFB's Predictive Performance

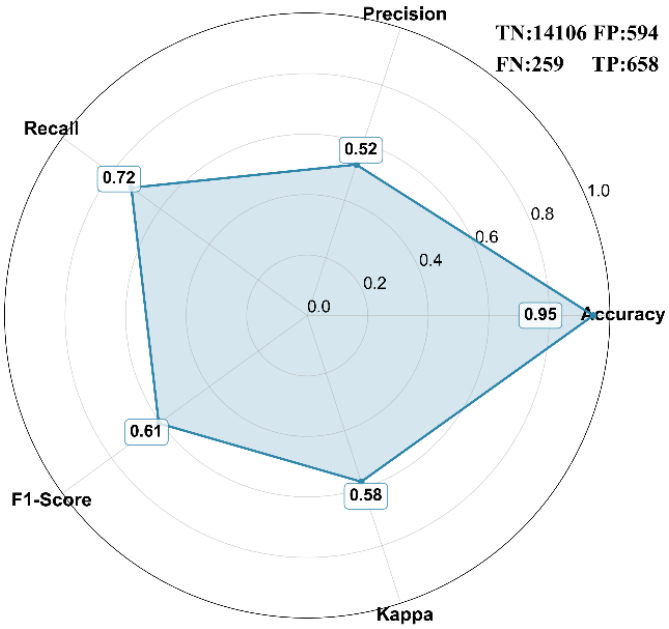

Figure 5.1 Performance Radar of DWARFB

The Figure 5.1 illustrates the model's predictive effectiveness from multiple perspectives, as reflected by Cohen's Kappa, F1-score, Recall, Precision, and Accuracy. The model achieved a very high accuracy (close to 0.95), which suggests strong overall predictive capability. However, accuracy alone may overestimate performance due to the class imbalance, as most firms belong to the non-distress category. The relatively lower values for Precision (~0.52) and Cohen's Kappa (~0.58) highlight that the model's predictive power for the distress samples is less robust, indicating that some positive predictions (financial distress) are false alarms. On the other hand, the Recall (~0.72)

suggests that the model can correctly identify a substantial proportion of distressed samples, although 259 cases remain undetected. An F1-score of around 0.61 indicates that the model maintains a reasonable balance between Precision and Recall, reflecting its overall effectiveness in identifying minority-class distress cases while limiting misclassification errors.

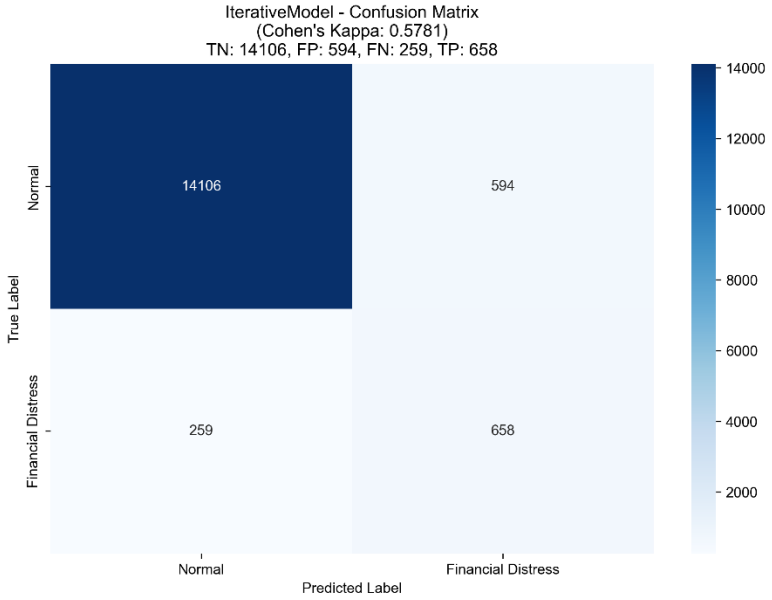

Figure 5.2 Confusion Matrix of DWARFB

The Figure 5.2 further contextualizes these findings. Evaluation on the test set of 15,617 observations showed that DWARFB correctly predicted 14,106 non-distress cases and 658 distress cases, corresponding to the true negative and true positive classifications, respectively. However, the confusion matrix also reveals 594 false-positive outcomes, indicating that 594 non-distressed observations were mistakenly assigned to the distressed class, and 259 False Negatives, where distressed observations were missed. This distribution indicates that the model is more conservative in misclassifying distressed observations (low FN), which is desirable in risk-sensitive domains such as financial distress prediction. Nonetheless, the moderate number of False Positives could increase the burden of unnecessary investigations.

5.1.2 Persistent Misclassified Patterns Analysis

Table 5.1 reports the comparative performance of three benchmark weighting strategies, AdaBoost-like, Focal-like, and the Proposed Weight, integrated into the DWARFB framework.

Table 5.1 Comparative Results Across Benchmark Weight Types

Weight type	Accuracy	Precision	Recall	F1	AUC	PR-AUC	MCC
AdaBoost-like	0.36	0.04	0.45	0.08	0.39	0.05	-0.10
Focal-like	0.95	0.60	0.56	0.58	0.95	0.64	0.55
Proposed weight	0.95	0.54	0.70	0.61	0.95	0.66	0.59

The AdaBoost-like weighting scheme performs the worst among the three approaches, with extremely low accuracy (0.36), precision (0.04), and F1-score (0.08). Only recall reaches a moderate level (0.45), indicating that the mechanism tends to overemphasize misclassified samples indiscriminately and produces excessive false positives. This outcome suggests that the direct exponential amplification of misclassification counts destabilizes the learning process under the multi-stage sampling structure used in DWARFB.

The Focal-like strategy shows substantially stronger performance across all metrics (e.g., accuracy = 0.95, F1 = 0.58, MCC = 0.55). By introducing a difficulty-aware modulation term based on normalized misclassification counts, the focal-like weighting provides a more balanced emphasis on harder samples. Although this adaptation avoids the extreme weight inflation seen in the AdaBoost-like mechanism, its performance remains slightly inferior to the proposed approach, particularly in recall and PR-AUC, two key metrics for imbalanced learning.

The Proposed Weight mechanism achieves the best overall performance, with improvements particularly evident in recall (0.70), PR-AUC (0.66), and MCC (0.59). These gains highlight the effectiveness of incorporating a step-penalty term inspired by OHEM, which selectively emphasizes severely misclassified samples without destabilizing the weight distribution. The higher recall indicates better sensitivity to minority-class instances, while the superior PR-AUC and MCC demonstrate improved robustness in imbalanced prediction scenarios.

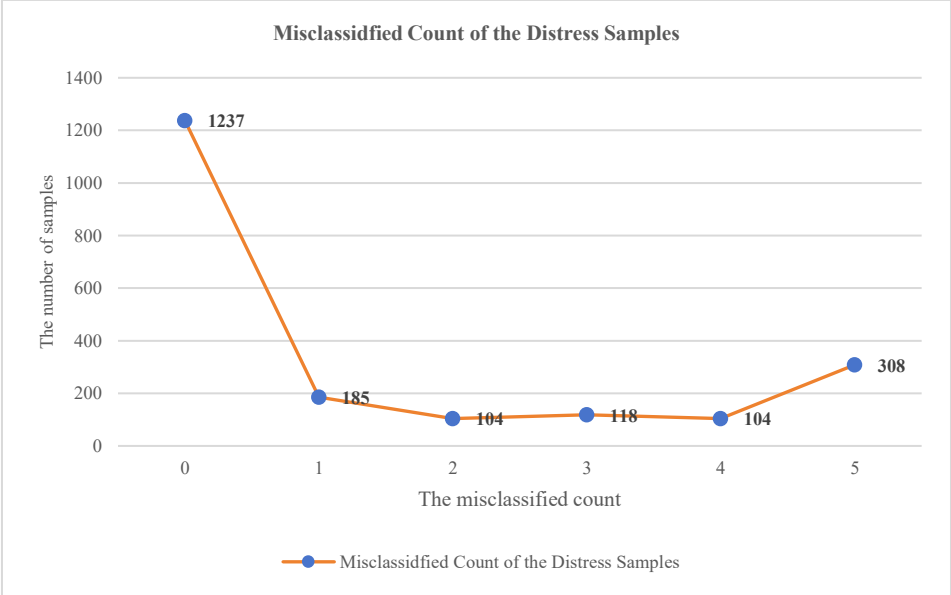

Figure 5.3 Misclassified Distress Samples During Training

Figure 5.3 reveals critical insights into how distress samples are repeatedly misclassified in DWARFB model training process. Amongst 2,056 distress samples, 1,237 (60%) were consistently classified correctly across all training iterations, suggesting that the DWARFB model effectively learned the underlying patterns of these samples. However, 308 (15%) samples were cumulatively misclassified five times, suggesting that these are particularly challenging cases, and that retaining misclassification information from the previous five training iterations is still insufficient for the DWARFB model to

effectively capture their underlying patterns. 514 (25%) samples that were persistently misclassified four times or fewer appear to have had their “hardness” patterns successfully learned by the model.

Approximately 85% of the distress samples were correctly learned by the DWARFB model, indicating that the misclassification-based reweighting mechanism effectively adjusts sample weights based on the distress samples’ pattern recognition. The process also rewards samples that are consistently predicted correctly for two consecutive training rounds. This adaptive weighting strategy strengthens the model’s focus on informative samples and enhances classification stability throughout the iterative learning process.

The DWARFB model was configured to retain misclassification information from the five most recent training iterations to prevent cumulative excessive weighting of misclassified samples. However, some samples are still repeatedly misclassified across five or more iterations. Removing this retention threshold could lead to excessive weight being assigned to misclassified samples. This imbalance may distort the model’s behavior, granting undue influence on difficult samples or even weakening the effectiveness of the penalty mechanism. Therefore, careful balancing between enhanced weighting and preventing error accumulation is essential to identify underlying causes and guide future model refinement.

5.1.3 Feature Importance Contribution

Before implementing the proposed model, 200 features were selected based on feature importance computed from a RF classifier, where importance was measured using the Gini impurity. During the training process, the iterative model tracked the contributions of features to the reduction of classification loss.

Table 5.2 presents the top 20 features before and after training, ranked according to their Gini-based importance and loss-reduction contribution, respectively.

As shown, the top features before and after training exhibit both stability and adaptive reordering driven by the iterative learning process of DWARFB. Several variables, including Var318, Var313, Var319, Var320, Q2_Var319, Var17, and Q2_Var313, remain highly ranked throughout training, indicating their consistent contribution to distinguishing between classes. Particularly, Var318 retains the first rank in both stages, demonstrating its persistent effectiveness in guiding the model toward correct predictions and its dominant role in the convergence of the training process.

Table 5.2 Top 20 Features Before and After DWARFB Training

Before training			After training		
Rank	**Feature**	**Importance**	**Rank**	**Feature**	**Contribution**
1	Var318	0.020	1	Var318	0.040
2	Var319	0.019	2	Var61	0.030
3	Var313	0.018	3	Var313	0.028
4	Var61	0.016	4	Var319	0.025
5	Var320	0.013	5	Var320	0.018
6	Q2_Var319	0.010	6	Q2_Var319	0.016
7	Var17	0.008	7	Var17	0.015
8	Q2_Var313	0.007	8	Q2_Var61	0.012
9	Q2_Var61	0.007	9	Q2_Var318	0.010
10	Var314	0.007	10	Var324	0.010
11	Var253	0.007	11	Q2_Var312	0.010
12	Var191	0.007	12	Q2_Var320	0.009
13	Var251	0.007	13	Q2_Var47	0.009
14	Q2_Var318	0.006	14	Q2_Var313	0.009
15	Q2_Var312	0.006	15	Var323	0.009
16	Var260	0.006	16	Q2_Var292	0.009
17	Var187	0.006	17	Var348	0.009
18	Q2_Var320	0.005	18	Var314	0.009
19	Var252	0.005	19	Var47	0.009
20	Var170	0.005	20	Q2_Var294	0.008

At the same time, noticeable rank shifts emerge as the iterative model progressively focuses on persistent misclassification samples. For instance, Var61 rises from rank 4 to rank 2, reflecting its increasing contribution to

correcting prediction errors during the training processing. This indicates that Var61 becomes more influential when the model concentrates on hard-to-classify samples. Similarly, features such as Var324, Var323, Var348, and Var47 newly appear in the top 20 after training, while Var253, Var191, Var251, Var260, Var187, Var252, and Var170 drop out. This replacement demonstrates that variables with greater effectiveness in reducing misclassification error attract more attention from the iterative learning mechanism, whereas those with limited contribution to loss reduction are deemphasized.

The composition of the top 20 feature set further reflects the adaptive nature of the iterative model. Before training, the ranking is dominated by a cluster around Var318–Var320 and their quarterly variants, showing that these features provide strong initial guidance for model learning. After training, while this core group remains important, additional features such as Var324, Var323, Var348, Var47, Q2_Var292, and Q2_Var294 are incorporated into the top ranks. Their emergence suggests that the model progressively identifies complementary features that become critical for refining decision boundaries and improving predictions on difficult or borderline cases.

5.1.4 Discussion

The findings of this section highlight DWARFB's overall performance under class imbalance and its ability to learn and adapt to patterns in cumulatively misclassified samples. It demonstrates strong overall predictive performance under class imbalance, achieving high accuracy (~0.95). However, accuracy overestimates effectiveness due to the dominance of non-distressed firms. More realistic metrics, such as Precision (~0.55), Recall (~0.72), F1-score (~0.61), and Cohen's Kappa (~0.58), these results show that the model correctly

identifies most distressed firms but still produces a moderate number of false alarms (594 False Positives) and misses 259 distressed cases (False Negatives). Its strength lies in reducing the risk of overlooking distressed samples, which aligns with the conservative requirements of financial risk management. Nonetheless, the relatively high false positive rate signals the need for further refinement, particularly in enhancing precision without sacrificing recall.

The iterative misclassification tracking mechanism effectively reduces persistent errors, approximately 85% of distress samples were correctly learned. Conversely, 308 samples (about 15%) remained misclassified across five iterations, highlighting inherently difficult cases that the current retention window could not fully capture. Another 514 samples (about 25%) were misclassified fewer than five times, indicating gradual learning of moderately hard patterns. The retention threshold for misclassification history serves as a necessary balance, preventing excessive weighting that could distort the model by overemphasizing difficult samples. Future refinements could explore dynamic adjustment of the retention window or feature augmentation to improve recognition of persistently misclassified cases without compromising the model's stability.

Table 5.3 presents the top 20 features' financial description identified by DWARFB. Notably, many are two-quarter lagged variables, providing early warning signals beyond static financial measures and emphasizing that financial trajectories matter more than one-time values in distress prediction. For interpretability, the features are grouped into five categories.

(1) Retained earnings and accumulated profitability: Var318 (Undistributed Profit per Share), Var61 (Retained Earnings / Total Assets),

Var319 (Retained Earnings per Share), and their Q2 counterparts dominate the top rankings, occupying 6 of the top 9 positions. Var318 contributes the most, highlighting that cumulative profitability and internal capital accumulation are decisive in reducing misclassification loss. This aligns with classical models such as Altman's Z-score, in which the retained earnings related ratio X2 indicator (Retained Earnings / Total Assets) are core explanatory variables [202], [203].

Table 5.3 Description of Top 20 Features Contributing to DWARFB

Rank	Feature	Description
1	Var318	Undistributed Profit per Share
2	Var61	Ratio of Retained Earnings to Total Assets
3	Var313	Net Assets per Share
4	Var319	Retained Earnings per Share
5	Var320	Net Assets per Share Attributable to the Parent Company
6	Q2_Var319	Q2_Retained Earnings per Share
7	Var17	Earnings Volatility
8	Q2_Var61	Q2_Ratio of Retained Earnings to Total Assets
9	Q2_Var318	Q2_Undistributed Profit per Share
10	Var324	Net Cash Flow from Investing Activities per Share - TTM
11	Q2_Var312	Q2_Operating Profit per Share - TTM
12	Q2_Var320	Q2_Net Assets per Share Attributable to the Parent Company
13	Q2_Var47	Q2_Basic Earnings per Share after Deducting Extraordinary Items
14	Q2_Var313	Q2_Net Assets per Share
15	Var323	Net Cash Flow from Investing Activities per Share
16	Q2_Var292	Q2_Earnings per Share - TTM3
17	Var348	Price to Earnings Attributable to the Parent Company
18	Var314	Tangible Assets per Share
19	Var47	Basic Earnings per Share after Deducting Extraordinary Items
20	Q2_Var294	Q2_Earnings per Share - TTM4

(2) Equity strength and asset backing: Var313, Var320, Q2_Var313, Q2_Var320, and Var314 measure shareholders' equity and tangible asset support. Their high contributions suggest that balance-sheet solidity is crucial for correcting classification errors, consistent with the notion that a strong equity base enhances resilience to financial shocks [204].

(3) Current profitability and earnings quality: Q2_Var312, Var47, Q2_Var47, Q2_Var292, and Q2_Var294 capture short-term operational performance. Their moderate importance indicates that contemporaneous earnings refine predictions but are secondary to accumulated profitability and equity strength [205].

(4) Earnings stability: Var17 (Earnings Volatility) ranks seventh (0.015), confirming that firms with unstable earnings are more prone to distress, in line with previous findings linking volatility to higher distress risk [206].

(5) Cash flow and market valuation: Var324, Var323, and Var348 play a complementary role in loss reduction, though their impact is smaller compared to accounting fundamentals.

These findings provide empirical support for Hypothesis **H2** in section 1.4, which posits that modelling cumulative sample-level difficulty through persistent misclassification-aware weighting improves the identification of distressed firms near decision boundaries. The iterative misclassification tracking mechanism allows the model to progressively focus on samples that remain difficult across multiple learning rounds, enabling more effective discrimination of borderline distressed cases. The observation that approximately 85% of distressed samples were eventually correctly learned demonstrates the effectiveness of cumulative difficulty modelling in enhancing minority-class detection. Meanwhile, the presence of a limited number of persistently misclassified observations suggests that some cases are inherently ambiguous or affected by noise, confirming the necessity of the retention threshold to prevent excessive weighting of outliers. Overall, the results indicate

that incorporating historical misclassification information helps the model allocate learning resources effectively.

5.2 The Comparison with Variety Models

In Section 5.1, we discussed the preliminary performance of DWARFB, as well as the patterns of persistently misclassified samples and the contributions of features during the training process.

A comparative analysis is conducted in this section to evaluate the performance of the proposed DWARFB framework relative to established baseline models for financial distress prediction. Since DWARFB is a tree-based boosting framework built upon RF, DT and RF are included as baseline compared tree models, while AdaBoost and XGBoost are compared to evaluating the effectiveness of the boosting framework and the hard-to-classify sample handling mechanism. In addition, SVM and deep learning model MLP are incorporated to represent alternative structural paradigms and provide a comprehensive comparison. This comparative experiment aims to verify that the integrated ensemble model design enhances prediction reliability.

5.2.1 Overall Performance Analysis

The comparison of the seven models (DT, RF, SVM, MLP, AdaBoost, XGBoost, and DWARFB) is presented in Table 5.4.

Table 5.4 Overall Performance Comparison Across Models

Models	Accuracy	Precision	Recall	F1	MCC	Training time
DT	0.92	0.37	0.58	0.45	0.42	10.34
RF	0.95	0.63	0.47	0.54	0.52	25.86
SVM	0.63	0.12	0.85	0.21	0.22	1557.08
MLP	0.94	0.59	0.19	0.28	0.31	2.96
AdaBoost	0.95	0.62	0.44	0.51	0.50	41.12
XGBoost	0.96	0.69	0.45	0.55	0.54	1.27
DWARFB	0.95	0.53	0.71	**0.61**	**0.59**	515.56

The comparative results reveal that most top-performing accuracy models, such as XGBoost with 0.96 and RF with 0.95, achieve high correctness by biasing toward the majority class. Their low-to-moderate Recall scores, 0.45 for XGBoost, 0.47 for RF, and 0.19 for MLP, indicate that they fail to identify the distress samples, which is often the critical target in imbalanced scenarios. In contrast, DWARFB avoids this trade-off. It maintains strong accuracy of 0.95, comparable with RF and AdaBoost, while achieving the highest Recall of 0.71 among all models except SVM. Unlike SVM, which exhibits a high Recall of 0.85 paired with extremely low Precision of 0.12, DWARFB balances Recall with a 0.53 Precision, ensuring that its minority-class predictions are reliable and actionable. SVM's high Recall is largely meaningless in practice, as 88% of its positive predictions are false due to extreme majority-class bias.

Most importantly, DWARFB leads in F1-score with 0.61 and MCC with 0.59, two metrics specifically designed to mitigate imbalance bias. F1 provides a balanced assessment by jointly considering Precision and Recall, showing how well the model captures the distress samples, while MCC is computed using all four elements in confusion matrix to avoid distortions caused by class imbalance. XGBoost and RF, despite their higher accuracy and Precision, fall short in this regard, achieving lower F1 scores of 0.55 and 0.54 and MCC values of 0.54 and 0.52, because they sacrifice minority-class detection to maximize majority-class correctness. MLP and DT perform even worse, as MLP's 0.94 accuracy hides a 0.19 Recall, missing 81% of minority samples, while DT's 0.92 accuracy comes with only 0.37 Precision, meaning most minority-class predictions are incorrect. AdaBoost, although reaching 0.95 accuracy, delivers only 0.44 Recall and 0.50 MCC, offering little protection against imbalance bias.

In imbalanced datasets, the objective is not simply to maximize overall correct predictions but to correctly identify minority-class samples without generating overwhelming false alarms. DWARFB is the only model in this comparison that achieves this balance with leading F1 and MCC, strong Recall, and solid Precision. Unlike other models, which let accuracy obscure imbalance issues, DWARFB's metrics reflect meaningful performance on the distress samples, making it the clear choice for imbalanced scenarios.

As shown in Table 5.4, training times vary considerably across models due to differences in algorithmic structure and computational complexity. Traditional tree-based models, including DT, RF, AdaBoost, and XGBoost, complete training within seconds, demonstrating high computational efficiency. In contrast, SVM requires the longest training time of 1557.08 seconds, primarily due to the quadratic complexity of the RBF kernel when handling large-scale datasets. The DWARFB framework also requires a relatively longer training time of 515.56 seconds compared with other ensemble models. This is mainly due to its iterative training structure, which involves multiple stages of dynamic weighting, adaptive resampling, and model updating executed sequentially to enhance recognition of difficult and minority-class samples. Each iteration retrains a base learner on the reweighted dataset, cumulatively increasing computational load.

The ROC curves presented in Figure 5.4 indicate that DWARFB maintains a consistently higher level of discriminative performance than the baseline models, regardless of the decision threshold selected. DWARFB achieves the highest ROC-AUC of 0.953, indicating the strongest class separation. XGBoost and RF follow closely with ROC-AUCs of 0.947 and

0.945, while AdaBoost performs slightly below the top three at 0.940. SVM and DT show moderate to low performance with AUCs of 0.827 and 0.720, respectively, and MLP is the weakest at 0.589, as its curve closely tracks the random guessing baseline, reflecting poor class discrimination.

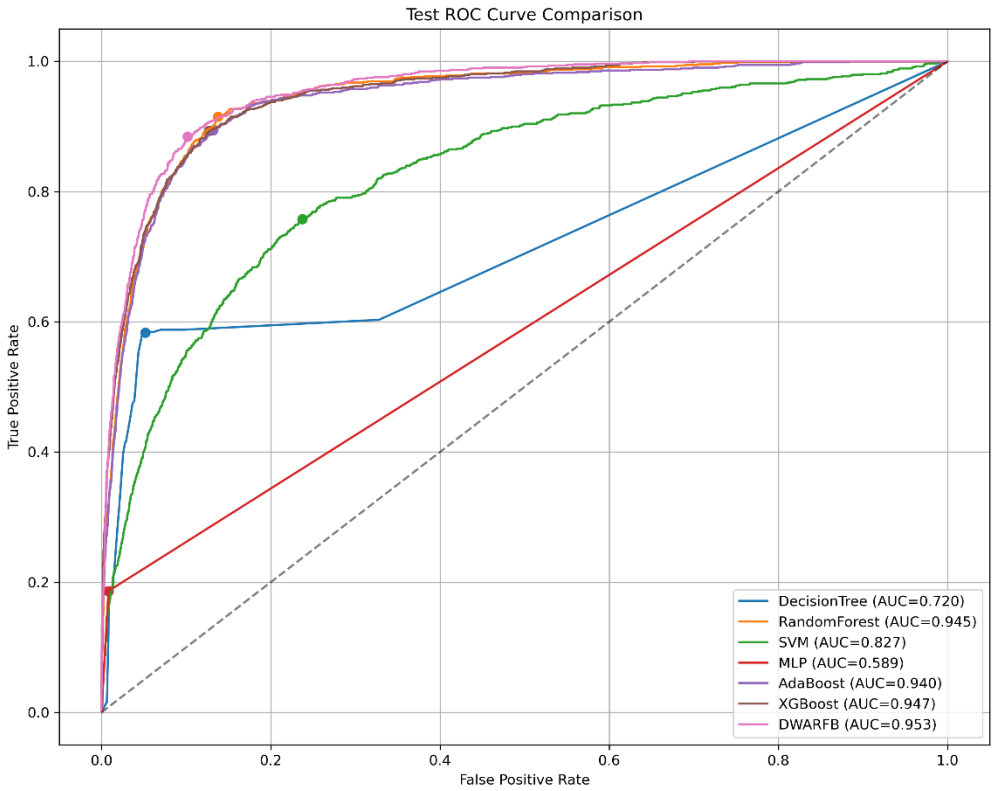

Figure 5.4 ROC Curve Comparison of Seven Models

The shapes of the ROC curves reveal practical implications for real-world applications. DWARFB, XGBoost, RF, and AdaBoost rise steeply at low False Positive Rates (FPR), achieving high True Positive Rates (TPR) without excessive false alarms, which is critical in areas such as fraud detection or distress prediction. SVM, although better than DT and MLP in AUC, requires more false positives to reach comparable TPR due to a less steep curve. DT plateaus early, showing limited ability to balance TPR and FPR, while MLP offers no meaningful discrimination, making it unsuitable for most classification tasks.

This ROC analysis reinforces previous metric evaluations. DWARFB's superior AUC aligns with its top F1 and MCC scores, reflecting balanced and reliable performance. XGBoost and RF also show strong overall performance, consistent with their high accuracy and precision. SVM's moderate AUC explains its high recall but low precision and accuracy, and MLP's poor AUC mirrors its extremely low recall and imbalanced predictions. Overall, DWARFB emerges as the most effective model for robust classification, while XGBoost and RF perform well, and SVM, DT, and MLP are less reliable. The ROC results emphasize DWARFB's consistent excellence across both threshold-specific and overall performance metrics.

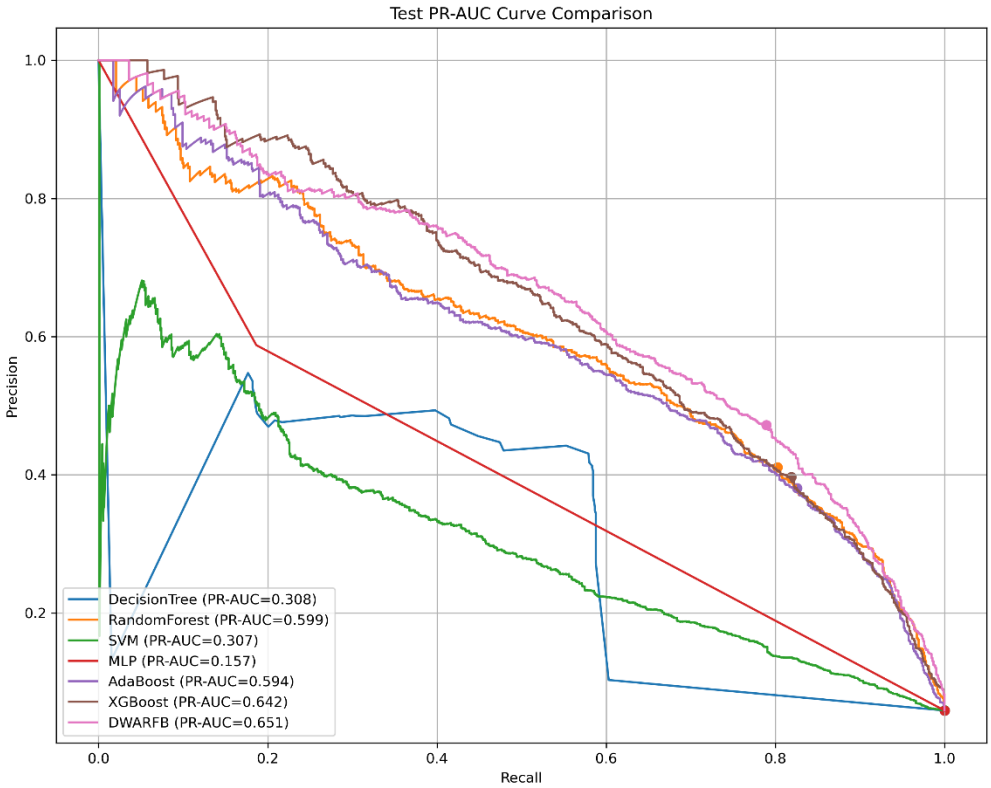

Figure 5.5 PR-AUC Curve Comparison of Seven Models

The Precision–Recall curves shown in Figure 5.5, reveal that DWARFB achieves the largest PR-AUC value (0.651), indicating the most effective overall precision–recall performance among the evaluated models across all

decision thresholds. It maintains high Precision even as Recall increases, making it reliable for identifying true positives without excessive false alarms. Closely trailing is XGBoost with a PR-AUC of 0.642, providing further evidence of the model's ability to achieve reliable minority-class detection while controlling false-positive predictions. RF with the PR-AUC of 0.599 and AdaBoost with the PR-AUC of 0.594 deliver solid but less robust performance; their curves decline more steeply as Recall rises, meaning Precision drops faster than in DWARFB and XGBoost.

SVM with PR-AUC of 0.307 and DT with the PR-AUC of 0.308 have low PR-AUC value, indicating a poor balance between Precision and Recall. SVM's curve peaks early but collapses, while DT's curve is erratic and fails to maintain meaningful Precision at higher Recall. MLP with the PR-AUC of 0.157 is the weakest, with a value close to random performance, as its curve is nearly indistinguishable from the baseline at low Recall, showing it cannot reliably distinguish positive classes.

For imbalanced datasets, DWARFB and XGBoost stand out as the top choices, as their high PR-AUC ensures they capture true positives without excessive false alarms. In contrast, SVM, DT, and MLP are unsuitable for tasks where balancing Precision and Recall is critical, especially in scenarios with rare positive classes. This PR-AUC analysis, alongside earlier ROC and metric comparisons, solidifies DWARFB and XGBoost as the most robust models, with DWARFB excelling in PR-AUC and XGBoost also delivering strong balanced performance.

5.2.2 Confusion Matrix Comparative Analysis

To further analyse the comparative performance of DWARFB with other models, Table 5.5 presents the confusion matrix comparison results across the seven models. The confusion matrix results provide a more detailed perspective on the models' classification behaviours by examining the number of TN, FP, FN, and TP.

From the confusion matrix results, it can be observed that the DWARFB model achieves a substantially higher number of true positives with the value of 654 compared with all other models, indicating a stronger capability in correctly identifying financially distressed firms. Although its true negative value of 14,128 is slightly lower than that of RF and XGBoost, DWARFB effectively reduces false negatives with the value of 263, which is critical in early warning tasks where false-negative predictions may expose investors, creditors, and other stakeholders to substantial financial risk.

Table 5.5 Confusion Matrices of Seven Models

Models	TN	FP	FN	TP
DT	13786	914	381	536
RF	14445	255	489	428
SVM	9061	5639	137	780
MLP	14580	120	746	171
AdaBoost	14453	247	515	402
XGBoost	14517	183	502	415
DWARFB	14128	572	263	654

By contrast, the RF and XGBoost models exhibit balanced performance, achieving high TN values of 14,445 and 14,517 respectively, but with relatively higher false negatives of 489 and 502. This suggests that while they are conservative in flagging distress, they may fail to detect a portion of actual distressed firms. The SVM model presents a distinct pattern with the highest

true positive count among the baseline models with the TP of 780, but at the cost of many FP of 5,639, reflecting an over-sensitive classification tendency.

The DT, MLP, and AdaBoost models demonstrate relatively weaker performance. In particular, the DT suffers from both a lower TP with value of 536 and higher FN with value of 381, suggesting limited generalization ability. The MLP shows the lowest TP with the value of 171 and highest FN with the value of 746, indicating poor sensitivity toward the distressed class. AdaBoost performs moderately, with results close to RF but slightly inferior in TP detection.

5.2.3 The Statistical Analysis with Other Models

To establish that the observed performance advantages of the proposed model over the benchmark models are unlikely to have arisen from random sampling variability, statistical analysis are conducted across 20 independent runs to evaluate the performance differences among all models.

Table 5.6 Overall Comparative Results of Seven Models

Model	F1	Precision	Recall	Accuracy	MCC
DT	0.45 ± 0.0026	0.37 ± 0.0027	0.59 ± 0.0042	0.92 ± 0.0007	0.42 ± 0.0028
RF	0.50 ± 0.0086	0.75 ± 0.0129	0.38 ± 0.0086	0.96 ± 0.0007	0.51 ± 0.0086
SVM	0.21 ± 0.0000	0.12 ± 0.0000	0.85 ± 0.0000	0.63 ± 0.0000	0.22 ± 0.0000
MLP	0.30 ± 0.0541	0.62 ± 0.0634	0.20 ± 0.0521	0.95 ± 0.0017	0.33 ± 0.0333
AdaBoost	0.48 ± 0.0000	0.61 ± 0.0000	0.39 ± 0.0000	0.95 ± 0.0000	0.46 ± 0.0000
XGBoost	0.53 ± 0.0000	0.68 ± 0.0000	0.44 ± 0.0000	0.95 ± 0.0000	0.52 ± 0.0000
Proposed Method	**0.61** ± 0.0045	0.54 ± 0.0119	**0.70** ± 0.0126	0.95 ± 0.0016	**0.59** ± 0.0043

Table 5.6 provides a comprehensive summary of the average performance achieved by DWARFB against six representative baseline models across five key evaluation metrics.

Over 20 independent runs, the proposed DWARFB model still shows outperformance compared to all benchmarks with the highest F1 (0.61 ± 0.0045)

and MCC (0.59 ± 0.0043), achieving strong recall (0.70 ± 0.0126) for distressed firms while maintaining competitive precision (0.54 ± 0.0119) and accuracy (0.95 ± 0.0016), demonstrating balanced, robust, and stable predictive performance. Along with the ROC-AUC (μ = 0.954, σ = 0.000) (see Figure 5.6) and PR-AUC (μ = 0.658, σ = 0.002) (see Figure 5.7) shows the best performance over all the comparative models.

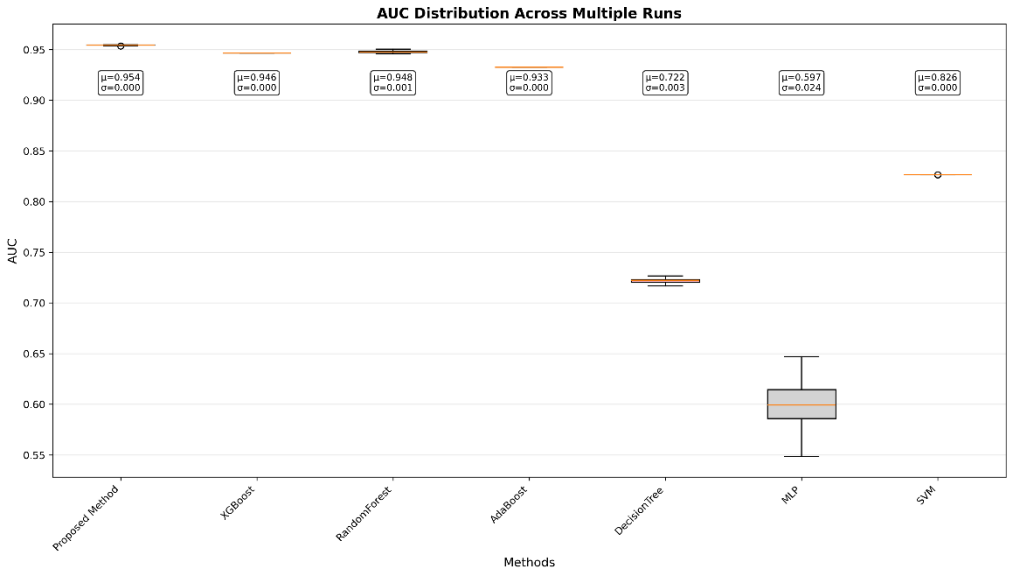

Figure 5.6 AUC Distribution Comparison: DWARFB vs. Other Models

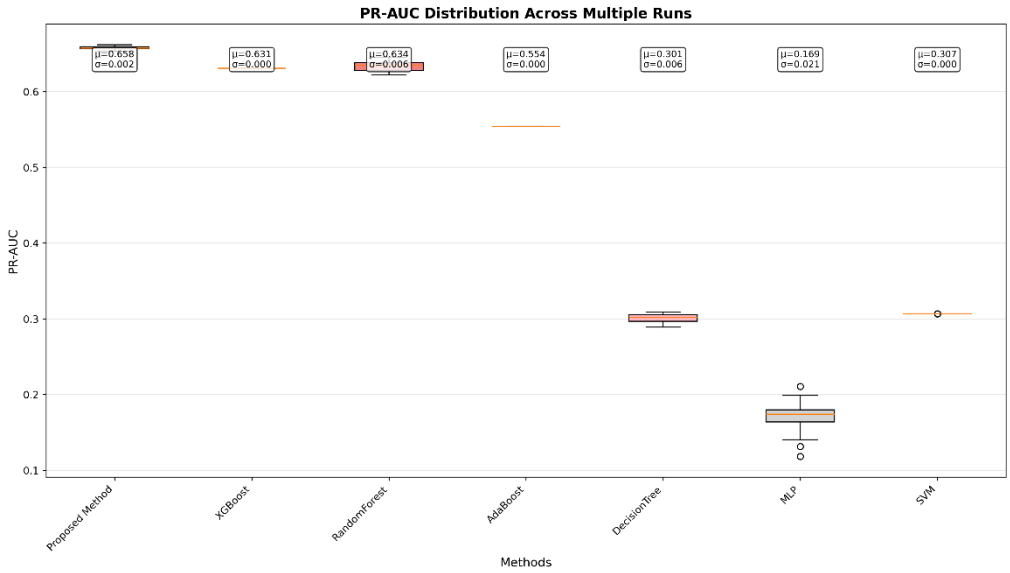

Figure 5.7 PR-AUC Distribution Comparison: DWARFB vs. Other Models

Collectively, these results establish DWARFB as the most effective and consistent model in most imbalance-sensitive metrics, delivering superior discrimination, exceptional stability, and the strongest detection of minority-class samples.

Following the descriptive comparison, the inferential statistical analyses to rigorously examine whether the performance gains achieved by the proposed DWARFB model over the benchmark algorithms are unlikely to be attributable to random variation.

Table 5.7 presents the pairwise Wilcoxon signed-rank test results comparing DWARFB with six widely used baseline models. The statistical analysis consistently produced p-values below 0.001 for all model comparisons, suggesting that the observed differences in performance are unlikely to be explained by random variation. Moreover, all effect sizes fall within the large category, demonstrating that the improvements delivered by DWARFB are not only statistically meaningful but also of substantial practical relevance. The strongest effects are observed against SVM, DT, and AdaBoost, while significant advantages are also maintained over competitive ensemble methods such as RF and XGBoost. In terms of predictive metrics, DWARFB consistently yields positive gains in both F1 and AUC. The largest improvements in F1 are achieved over SVM (+0.397) and MLP (+0.311), whereas notable AUC gains are observed over MLP (+0.357). Even for strong benchmark ensembles, DWARFB achieves measurable increases in both metrics, underscoring its superior ability to capture the characteristics of distress observations. Overall, the results confirm that DWARFB delivers significantly enhanced and robust predictive performance compared with all baseline models.

Table 5.7 Pairwise Wilcoxon Test: DWARFB vs. Other Models

Model	p-value	Significance	Effect Size	Interpretation	ΔF1	ΔAUC
DT	<0.001	Yes***	42.19	Large	0.159	0.233
RF	<0.001	Yes***	14.93	Large	0.106	0.007
SVM	<0.001	Yes***	121.46	Large	0.397	0.128
MLP	<0.001	Yes***	7.90	Large	0.311	0.357
AdaBoost	<0.001	Yes***	40.52	Large	0.133	0.022
XGBoost	<0.001	Yes***	23.96	Large	0.078	0.008

5.2.4 Discussion

(1) Comparison with Base Models: DT → RF → DWARFB

The comparison begins with DT, which serves as the simplest baseline. Its performance on the imbalanced dataset is weak, with low Recall and unstable predictions. RF improves these results through its bagging strategy. It increases sensitivity and reduces false negatives relative to DT.

DWARFB further advances performance beyond the RF. It achieves an Accuracy close to 0.95, a Recall of 0.72, and an F1-score of 0.61. The model correctly identifies 658 distressed cases, which is higher than both DT and RF. Its False Negative count is only 259. DWARFB also demonstrates stronger agreement, reflected in Cohen's Kappa of 0.5781. These improvements show that RF achieves higher performance than its base learner, and DWARFB delivers an additional performance gain over RF.

(2) Comparison with Other Single Models: SVM and MLP

DWARFB also surpasses the SVM and MLP benchmarks. SVM captures distressed firms with relatively high Recall but produces many false positives. This pattern is common in financial distress prediction when kernel-based models' over-flag the distress samples. MLP shows strong Accuracy but misses many distressed observations because it is primarily optimized to correctly classify majority-class instances.

DWARFB provides more balanced and reliable classification. Its Precision (~0.55) exceeds that of the SVM, while its Recall (0.72) exceeds that of the MLP. The confusion matrix confirms these advantages: DWARFB records 658 true positives and only 594 false positives, outperforming both SVM and MLP in minority-class discrimination. These results show that single learners struggle to maintain balanced performance, whereas DWARFB remains stable across all core metrics.

(3) Comparison Within Boosting Structures: Static Boosting vs. Adaptive Boosting

Within boosting-based methods, DWARFB also shows clear improvements. AdaBoost and XGBoost apply fixed boosting strategies and constant sample weighting. In contrast, DWARFB updates sampling ratios and misclassified sample weights at each iteration. Unlike AdaBoost, where sample weights are updated according to errors generated in the immediately preceding iteration, DWARFB incorporates cumulative misclassification information across iterations. Therefore, the model can distinguish between temporarily misclassified samples and persistently difficult cases. This historical perspective enables DWARFB to continuously emphasize distressed firms that remain challenging to classify, improving minority-class representation without relying only on iteration-specific errors.

Empirical results support this improvement. DWARFB achieves higher Recall, F1-score, and MCC than AdaBoost and XGBoost. It identifies more distressed firms while keeping false positives at manageable levels. These gains confirm that iterative sampling and adaptive boosting provide meaningful advantages over static boosting methods.

DWARFB achieves the strongest performance among all benchmark models. It records the highest number of true positives (658) and one of the lowest false negative counts (259). It also produces competitive false-positive levels (594). Its high Recall, moderate Precision, strong AUC, and stable Kappa all demonstrate its effectiveness in imbalanced financial distress prediction. From an economic perspective, the reduction of false negatives is particularly important in financial distress prediction because missed distressed firms may delay risk identification and limit opportunities for timely intervention by investors, creditors, and regulators. Meanwhile, false positives represent additional monitoring and evaluation costs caused by incorrectly warning healthy firms. DWARFB achieves a practical balance by identifying 658 distressed firms while limiting false negatives to 259 cases and maintaining a manageable false-positive level of 594 cases. This balance indicates that the proposed model improves distress detection effectiveness while avoiding excessive unnecessary warnings.

DWARFB's improvements arise from three core mechanisms: adjust sampling ratio, misclassified sample reweighting, and adaptive boosting. These elements enable the model to outperform single learners, bagging ensembles, and static boosting algorithms.

Overall, these comparative results provide empirical support for Hypothesis **H3** in section 1.4, which proposes that DWARFB incorporating integrated imbalance handling, adaptive learning dynamics, and optimized ensemble structures can deliver more robust and stable financial distress predictions than fragmented modelling approaches. By outperforming single learners (DT, SVM, MLP), bagging ensembles (RF), and conventional boosting

algorithms (AdaBoost and XGBoost), DWARFB demonstrates that the unified integration of adaptive sampling, misclassification-aware weighting, and boosting optimization leads to more balanced and reliable classification performance under severe class imbalance conditions.

5.3 The Comparison with RF Variants and Baseline Resampling Methods

Section 5.2 compared a variety of models to verify the reliability and effectiveness of DWARFB in enhancing RF, the boosting framework, and the misclassified hard-to-learn mechanism.

This section focuses on comparing DWARFB with different techniques for handling class imbalance. For a fair comparison, all methods are based on RF, combined either with different data-level strategies or with different algorithm-level imbalance handling strategies. The comparative experiments aim to answer to assess the extent to which the experimental findings address RQ1 established in Section 1.4: *"Can severe class imbalance in financial distress prediction be addressed in a way that preserves the original data distribution while adaptively improving the identification of rare but economically critical distressed firms?*"

5.3.1 Overall Performance Analysis

To ensure that the comparative evaluation was conducted under identical experimental settings, all RF-based baseline models were hyperparameter-tuned prior to evaluation to identify the best-performing versions for comparison with DWARFB.

Table 5.8 summarizes the hyperparameter ranges explored during optimization. Random search [207] was employed across four key parameters,

including N_ESTIMATORS, MAX_DEPTH, MIN_SAMPLES_SPLIT, and MIN_SAMPLES_LEAF.

Table 5.8 Parameter Ranges for Baseline Model Optimization

Parameter	Range of values
N_ESTIMATORS	50, 100, 200
MAX_DEPTH	5,10,15
MIN_SAMPLES_SPLIT	1, 2, 5
MIN_SAMPLES_LEAF	1,2,5

The optimized hyperparameter configurations obtained through this procedure, along with their corresponding cross-validation F1 scores, are presented in Table 5.9.

Table 5.9 Optimized Parameters for Comparative Methods

Method	N_ESTIMATORS	MAX_DEPTH	MIN_SAMPLES_SPLIT	MIN_SAMPLES_LEAF	CV F1 Mean
Standard RF	50	15	2	1	0.63 ± 0.0194
Cost-sensitive RF	100	15	5	5	0.67 ± 0.0044
Balanced RF	50	15	2	1	0.47 ± 0.0005
RF + RUS	50	10	5	1	0.45 ± 0.0091
RF + SMOTE	50	15	2	1	0.63 ± 0.0065
RF + Borderline SMOTE	50	15	2	1	0.64 ± 0.0043
RF + ADASYN	50	15	2	1	0.62 ± 0.0062
RF + SMOTE Tomek	50	15	2	1	0.64 ± 0.0083

This hyperparameter tuning procedure ensures that each baseline is evaluated under its most effective configuration, thereby enhancing the reliability of performance assessment. More importantly, it guarantees that subsequent comparisons with the DWARFB reflect the genuine predictive capacity of each method rather than artifacts of under-optimization.

Table 5.10 presents the comparative performance of DWARFB against several baseline approaches. Across all evaluated models, DWARFB delivers

the strongest overall classification performance, as evidenced by the highest F1-score (0.62) and MCC (0.60). The superior F1-score indicates an effective balance between Precision and Recall, while the higher MCC demonstrates robust overall classification performance by accounting for all four outcomes of the confusion matrix. While the Standard RF achieved the highest accuracy (0.96) and precision (0.74), its recall performance was relatively low (0.38), suggesting a strong bias toward the majority class and a tendency to overlook distressed firms.

Table 5.10 Overall Results Compared with Baseline Methods

Method	F1	Precision	Recall	MCC	Accuracy	Training Time
Standard RF	0.50	0.74	0.38	0.51	0.96	28.6
Cost-sensitive RF	0.58	0.61	0.55	0.55	0.95	49.0
Balanced RF	0.46	0.31	0.89	0.48	0.88	3.5
RF + RUS	0.44	0.30	0.89	0.47	0.87	1.7
RF + SMOTE	0.56	0.44	0.76	0.55	0.93	67.0
RF + Borderline SMOTE	0.56	0.44	0.74	0.54	0.93	51.0
RF + ADASYN	0.54	0.42	0.76	0.53	0.93	50.9
RF + SMOTE Tomek	0.56	0.44	0.76	0.54	0.93	60.8
DWARFB	0.62	0.56	0.69	0.60	0.95	387.9

The Cost-sensitive RF model offered a more balanced trade-off, improving recall to 0.55 and yielding an F1-score of 0.58, but it was still outperformed by DWARFB in both F1 and MCC. Similarly, resampling-based approaches such as RF + SMOTE (F1 = 0.56) and RF + Borderline SMOTE (F1 = 0.56) enhanced recall (0.74–0.76) compared to Standard RF, but their lower precision values (0.44) limited overall balance. The RF + RUS and Balanced RF methods achieved the highest recall (0.89), yet their precision dropped substantially (0.30–0.31), resulting in significantly lower F1-scores (0.44–0.46) and reduced accuracy (0.87–0.88).

Among all evaluated methods, DWARFB demonstrates the most favorable balance between correctly identifying distressed samples and limiting false-positive predictions, as reflected by its precision (0.56), recall (0.69), and the highest F1-score, while also achieving a competitive accuracy of 0.95. Importantly, the MCC score of 0.60 reinforces its robustness, as MCC is widely regarded as a balanced measure that accounts for both false positives and false negatives. The primary limitation of the proposed approach lies in its substantially higher training time (387.9 seconds) compared to all baselines, which may pose challenges in time-sensitive applications.

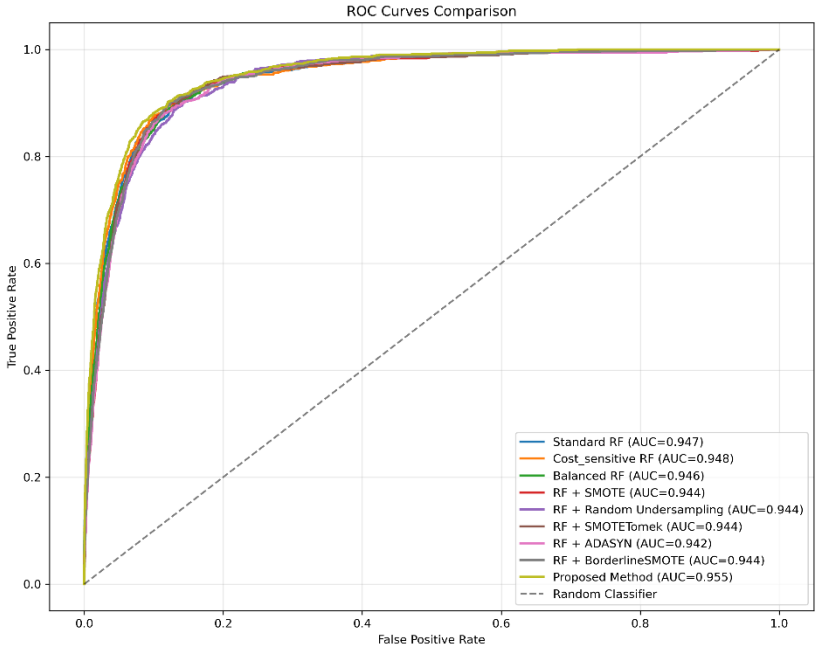

Figure 5.8 ROC Comparison: DWARFB, RF Variants, and Resampling

Figure 5.8 presents the ROC-AUC of the DWARFB compared with several baseline approaches. All models show strong discriminatory power with AUC values above 0.94, confirming their effectiveness in distinguishing between normal and financially distressed firms. Among the baselines, Cost-sensitive RF (AUC = 0.948) and Standard RF (AUC = 0.947) perform slightly better than resampling-based methods such as RF with SMOTE (AUC = 0.944), RF with

SMOTE Tomek (AUC = 0.944), and RF with ADASYN (AUC = 0.942). DWARFB achieves the highest AUC of 0.954, surpassing all baselines.

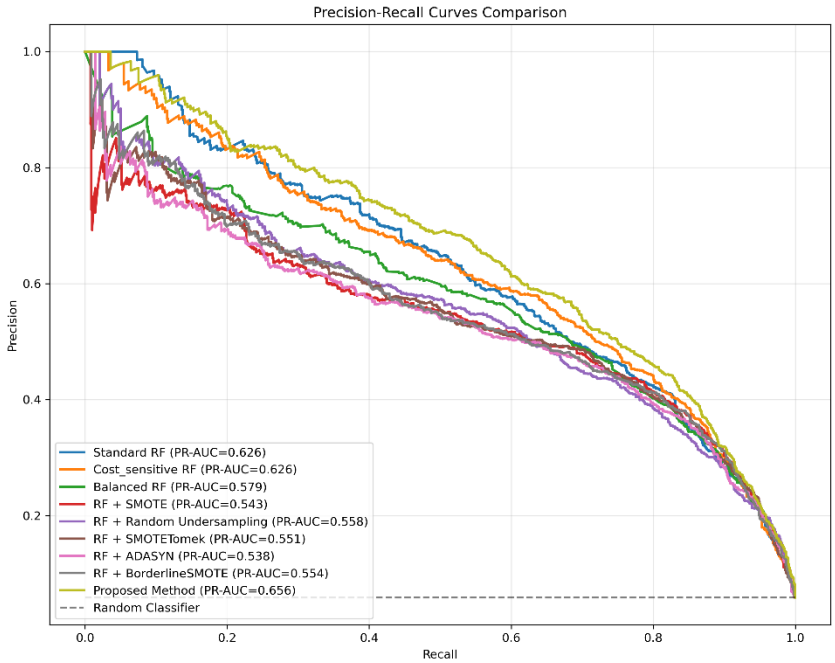

Figure 5.9 PR-AUC Comparison: DWARFB, RF Variants, and Resampling

The Precision-Recall curves in Figure 5.9 highlight DWARFB's strong performance on the imbalanced dataset, achieving the highest PR-AUC of 0.656, compared with Standard RF and Cost-sensitive RF, both at 0.626. DWARFB maintains higher precision across a wide range of recall values, demonstrating its superior balance between minimizing false alarms and identifying distressed firms. The large gap between DWARFB's PR-AUC and the random classifier baseline of approximately 0.05 further confirms the effectiveness of its dynamic weighting strategies in handling class imbalance.

5.3.2 Confusion Matrix Comparative Analysis

Table 5.11 shows the confusion matrix results compared to other methods. DWARFB generates only 502 false positives, which is substantially lower than aggressive resampling methods such as Balanced RF (1,838) and RF + RUS (1,942), and lower than SMOTE-based approaches, which range from

849 to 949. DWARFB also correctly identifies 635 out of 917 distressed samples, achieving a sensitivity of 0.69. This performance markedly reduces false negatives compared with Standard RF (568) and Cost-sensitive RF (415), demonstrating its superior capability in detecting financially distressed firms while maintaining a low false alarm rate.

Table 5.11 Confusion Matrices Compared to Other Methods

Method	TN	FP	FN	TP	Specificity	Sensitivity
Standard RF	14576	124	568	349	0.99	0.38
Cost-sensitive RF	14378	322	415	502	0.98	0.55
Balanced RF	12862	1838	100	817	0.88	0.89
RF + RUS	12758	1942	102	815	0.87	0.89
RF + SMOTE	13827	873	224	693	0.94	0.76
RF + Borderline SMOTE	13851	849	238	679	0.94	0.74
RF + ADASYN	13751	949	219	698	0.94	0.76
RF + SMOTE Tomek	13812	888	222	695	0.94	0.76
DWARFB	14198	502	282	635	0.97	0.69

5.3.3 Statistical Significance Analysis with Other Sampling Methods

Table 5.12 summarizes the comparative performance of DWARFB and various RF baseline approaches, getting from 20 independent experiments. The inclusion of multiple trials ensures that the reported performance is not the outcome of random fluctuations but reflects stable and reproducible patterns.

Table 5.12 Statistical Performance of DWARFB and Baseline Models

Method	F1-score	Precision	Recall	MCC	Accuracy
Standard RF	0.51 ± 0.02	0.75 ± 0.02	0.38 ± 0.02	0.52 ± 0.02	0.96 ± 0.00
Cost-sensitive RF	0.57 ± 0.01	0.60 ± 0.01	0.55 ± 0.02	0.55 ± 0.01	0.95 ± 0.00
Balanced RF	0.46 ± 0.01	0.31 ± 0.01	0.89 ± 0.01	0.48 ± 0.01	0.87 ± 0.00
RF + RUS	0.45 ± 0.01	0.30 ± 0.01	0.88 ± 0.01	0.47 ± 0.01	0.87 ± 0.01
RF+SMOTE	0.56 ± 0.01	0.44 ± 0.01	0.76 ± 0.01	0.54 ± 0.01	0.93 ± 0.00
RF + Borderline SMOTE	0.56 ± 0.01	0.44 ± 0.01	0.76 ± 0.01	0.54 ± 0.01	0.93 ± 0.00
RF + ADASYN	0.55 ± 0.01	0.42 ± 0.01	0.77 ± 0.01	0.54 ± 0.01	0.92 ± 0.00
RF + SMOTE Tomek	0.56 ± 0.01	0.45 ± 0.01	0.74 ± 0.01	0.55 ± 0.01	0.93 ± 0.00
DWARFB	**0.62 ± 0.01**	**0.58 ± 0.04**	**0.67 ± 0.05**	**0.60 ± 0.01**	**0.95 ± 0.00**

Across all evaluation metrics, DWARFB demonstrates superior consistency and overall predictive balance. Specifically, DWARFB achieves the highest F1-score (0.62 ± 0.01) and MCC (0.60 ± 0.01), clearly

outperforming all baselines. The observed performance ranges further indicate that the method yields robust and stable performance across repeated runs. In contrast, the Standard RF, while delivering the highest accuracy (0.96) and precision (0.75 ± 0.02), exhibited relatively larger variability in precision. More critically, its low recall (0.38 ± 0.02) reveals a persistent bias toward the majority class, leading to frequent misclassification of distressed samples. Cost-sensitive RF provided a more even trade-off, raising recall to 0.55 ± 0.02 with a moderate F1-score of 0.57 ± 0.01. However, it was consistently surpassed by DWARFB in terms of both F1-score and MCC. Similarly, resampling-based approaches (RF + SMOTE, RF + Borderline SMOTE, RF + ADASYN, and RF + SMOTE Tomek) displayed improved recall values (0.74–0.77), but their reduced precision (0.42–0.45) constrained overall performance, reflected in lower F1-scores (0.55–0.56).

Table 5.13 Pairwise Wilcoxon Test: DWARFB vs. Benchmarks

Benchmark methods	**p-value**	**Significance**	**Effect Size**	**Interpretation**	**ΔF1**	**ΔAUC**
Standard RF	<0.001	Yes***	15.59	Large	0.107	0.006
Cost-sensitive RF	<0.001	Yes***	8.63	Large	0.036	0.005
Balanced RF	<0.001	Yes***	42.47	Large	0.153	0.007
RF+SMOTE	<0.001	Yes***	11.23	Large	0.053	0.01
RF+RUS	<0.001	Yes***	32.58	Large	0.165	0.01
RF + SMOTE Tomek	<0.001	Yes***	12.89	Large	0.054	0.011
RF+ADASYN	<0.001	Yes***	14.35	Large	0.064	0.011
RF + Borderline SMOTE	<0.001	Yes***	11.24	Large	0.05	0.01

To rigorously assess whether the performance improvements of DWARFB over baseline classifiers are statistically significant, pairwise Wilcoxon signed-rank tests were conducted across 20 experimental runs. Table 5.13 summarizes the results, including p-values, effect sizes (Cohen's d), and performance differences (ΔF1 and ΔAUC). Across all comparisons, DWARFB

consistently outperforms the baselines (all $p < 0.001$) with large effect sizes ranging from 8.63 to 42.47, indicating that the improvements are both statistically significant and practically meaningful.

In terms of performance metrics, ΔF1 gains range from +0.036 versus Cost-sensitive RF to +0.165 versus RF with random under-sampling, highlighting its superior balance between precision and recall. ΔAUC improvements are smaller but consistently positive (+0.005 to +0.011), reaffirming its stronger discriminative power. The largest gains occur against Balanced RF (ΔF1 = +0.153, Cohen's d = 42.47) and RF+RUS (ΔF1 = +0.165, Cohen's d = 32.58), emphasizing its advantage over under-sampling approaches. Collectively, these results provide strong evidence that DWARFB achieves higher, robust, and statistically reliable performance across multiple metrics.

5.3.4 Discussion

(1) Comparison with static resampling techniques

Static resampling methods include over-sampling approaches such as SMOTE and its variants, under-sampling strategies such as RUS, and Balanced RF, which embeds under-sampling into each tree construction. Although these techniques substantially improve recall by increasing the representation of distressed firms, they do so at the cost of severely reduced precision. As shown in the comparative results, Balanced RF and RF + RUS achieve very high sensitivity (0.89) but generate an excessive number of false positives (1,838 and 1,942, respectively), resulting in extremely low precision (0.30–0.31) and degraded F1-scores (0.44–0.46). This error pattern is problematic in real financial applications, where an overwhelming number of false alarms can

reduce decision-makers' trust in the system and increase unnecessary investigation costs.

Similarly, SMOTE-based methods (SMOTE, Borderline-SMOTE, ADASYN, and SMOTE-Tomek) provide a moderate compromise by improving recall to around 0.74–0.76, but precision remains limited (0.42–0.45), with false positives still ranging from 849 to 949. These results indicate that static resampling techniques tend to push the classifier toward aggressive minority detection by altering the original data distribution, often introducing synthetic noise or discarding informative majority samples. Consequently, the improvement in sensitivity is achieved at the expense of reliability and interpretability in prediction outcomes.

In contrast, DWARFB achieves a substantially better balance between false negatives and false positives. Its false positives (502) are far lower than those produced by Balanced RF and SMOTE-based methods, while its true positives (635) remain competitive, yielding a sensitivity of 0.69 and specificity of 0.97. This demonstrates that DWARFB can improve minority-class detection without relying on static and potentially distortive resampling procedures, preserving the integrity of the original data distribution while maintaining a practical and controllable error profile.

(2) Comparison with Standard RF and Cost-sensitive RF

Standard RF exhibits strong precision (0.74) and accuracy (0.96) but extremely poor recall (0.38), missing more than half of the distressed firms (568 false negatives). This reflects a typical majority-class bias in imbalanced learning, where the model prioritizes overall accuracy at the expense of economically critical minority samples. DWARFB effectively corrects this bias

by raising recall to 0.69 and reducing false negatives to 282, while still maintaining high accuracy (0.95). The increase in F1-score from 0.50 to 0.62 and in MCC from 0.51 to 0.60 clearly shows that, DWARFB substantially enhances standard RF's ability to identify distressed samples through the boosting framework.

Cost-sensitive RF introduces fixed penalty weights to misclassification costs and therefore improves recall to 0.55, partially alleviating the minority-class bias. However, its learning mechanism remains static: the cost parameters are predetermined and cannot adapt to evolving misclassification patterns during training. As a result, although Cost-sensitive RF performs better than Standard RF, it still fails to achieve the optimal balance attained by DWARFB. DWARFB surpasses Cost-sensitive RF in F1-score (0.62 vs. 0.58), MCC (0.60 vs. 0.55), ROC-AUC (0.954 vs. 0.948), and PR-AUC (0.656 vs. 0.626). The confusion matrix further shows that DWARFB reduces false negatives more effectively while maintaining a comparable and manageable false positive rate.

This highlights a fundamental difference between the two approaches: Cost-sensitive RF enhances minority sensitivity through static cost assignment, whereas DWARFB performs dynamic reweighting based on actual misclassification behaviour. By repeatedly emphasizing samples that are consistently misclassified, DWARFB achieves adaptive learning focused on genuinely difficult distress cases, rather than uniformly penalizing all minority samples. This adaptive mechanism enables DWARFB to better capture the complex and heterogeneous patterns of financial distress.

These comparative findings demonstrate that DWARFB outperforms both static resampling-based methods and traditional cost-sensitive learning

approaches. Static resampling techniques improve recall by altering the class distribution, but they often generate excessive false positives and distort the statistical structure of the original dataset. Cost-sensitive RF partially alleviates minority-class bias through fixed penalty weights, yet its static cost structure limits its ability to adapt to evolving misclassification patterns during training. In contrast, DWARFB introduces a dynamic, data-driven boosting mechanism that adaptively adjusts sample importance based on misclassification behaviour. This approach improves minority-class detection while maintaining a balanced trade-off between sensitivity and precision. Overall, these results provide strong empirical support for Hypothesis **H1**, which posits that imbalance-handling mechanisms embedded within the learning process can outperform static preprocessing-based resampling methods while preserving the integrity and reliability of financial distress prediction.

5.4 Parameter Sensitivity Analysis

Following the comprehensive performance comparison in previous sections, which demonstrated DWARFB's superior ability to handle class imbalance and enhance the learning of hard-to-classify samples, it is crucial to understand how its key hyperparameters influence performance. Such understanding is essential for ensuring robust, stable, and reliable deployment of DWARFB in real-world financial distress prediction applications.

This section examines the sensitivity of the parameters controlling the misclassified samples reweighting mechanism introduced in Section 3.3.2, as well as the influence of the base learner RF parameters on the overall performance of DWARFB.

5.4.1 Misclassified sample strategy Parameter Analysis

To investigate the robustness of misclassified samples reweighting mechanism and identify its optimal hyperparameter configuration, a comprehensive parameter sensitivity analysis was conducted using Grid search and 5-fold cross-validation. The analysis focused on two key parameters, Alpha and Gamma, with performance evaluated primarily by the cross-validation metrics.

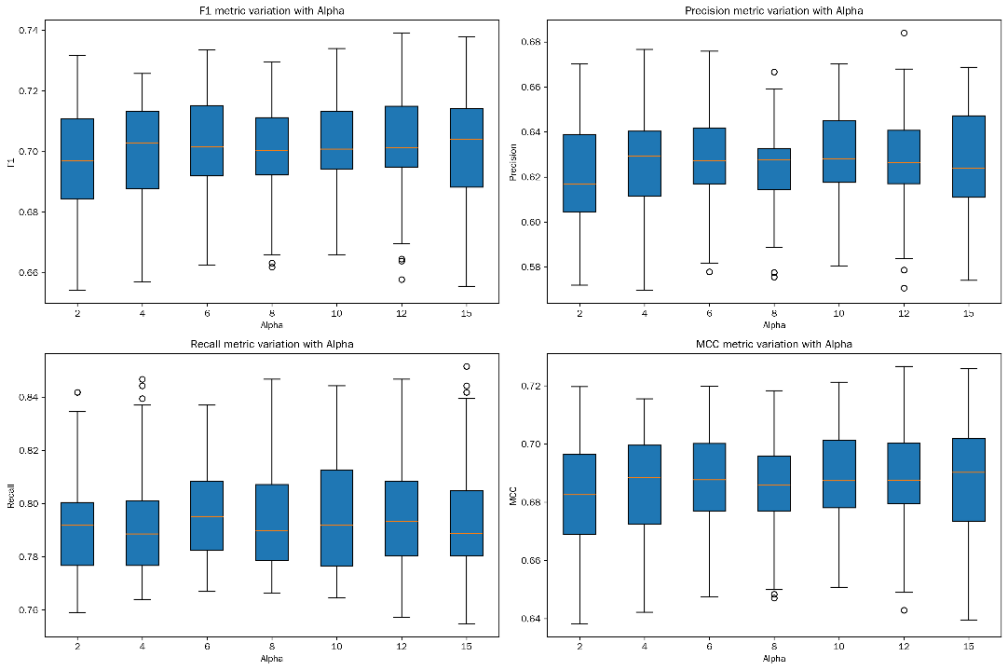

Figure 5.10 Sensitivity Analysis: Alpha Parameter

Figure 5.10 presents boxplots illustrating the variation of F1, Precision, Recall, and MCC as Alpha increases from 2 to 15. The results show that F1 and MCC remain highly stable, fluctuating narrowly between 0.69 and 0.72, and 0.68 and 0.71, respectively. Recall consistently attains relatively high values, ranging from 0.78 to above 0.80, while Precision remains lower, between 0.61 and 0.64, reflecting a stable trade-off between Precision and Recall. Increasing Alpha does not introduce substantial shifts in the central tendency or dispersion

of any metric, indicating that the model's performance is largely insensitive to changes in Alpha alone.

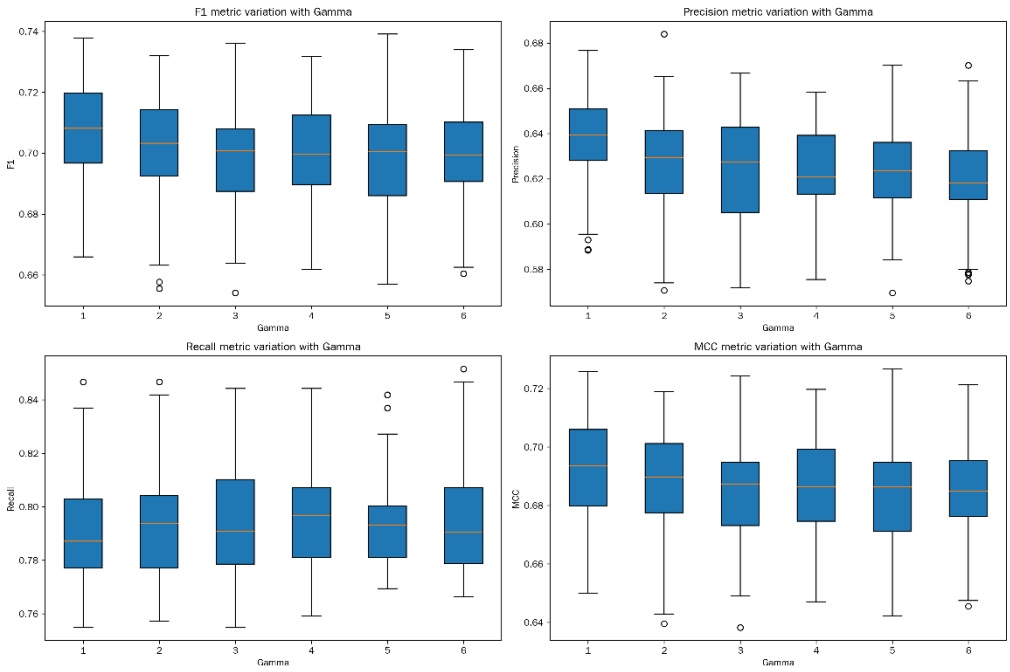

Figure 5.11 Sensitivity Analysis: Gamma Parameter

As shown in Figure 5.11, varying Gamma from 1 to 6 produces trends similar to those observed for Alpha. F1 remains stable between approximately 0.69 and 0.72, while MCC fluctuates narrowly around 0.68 to 0.70. Recall consistently exceeds Precision, ranging from roughly 0.78 to 0.81, compared with Precision values of 0.61 to 0.65, maintaining the same trade-off structure. No systematic upward or downward trend is observed for any metric, confirming that Gamma alone has a limited impact on performance. Overall, both Alpha and Gamma, when varied independently, demonstrate high stability and low sensitivity across all metrics.

The parameter performance matrix in Table 5.14 shows that the impact of Alpha and Gamma on model performance is nonlinear and metric dependent. F1 achieves its best performance when Alpha is 15 and Gamma is 1, reaching a value of 0.712, compared with the worst case at Alpha 2 and Gamma 3, which

yields 0.686, corresponding to a 3.6 percent relative improvement. This indicates that combining a high Alpha with a low Gamma consistently enhances balanced accuracy.

Table 5.14 Parameter Performance Summary Matrix

Metric	Optimal (α, γ)	Worst (α, γ)	Performance Range	Improvement
F1	(15,1) 0.7118	(2,3) 0.6862	0.0256	3.6%
Precision	(15,1) 0.6492	(2,3) 0.6043	0.0449	7.4%
Recall	(12,4) 0.8050	(2,6) 0.7875	0.0175	2.2%
MCC	(15,1) 0.6969	(2,3) 0.6723	0.0246	3.7%

Precision follows a similar pattern, with the optimal combination of Alpha 15 and Gamma 1 producing 0.649, and the lowest performance observed at Alpha 2 and Gamma 3 with 0.604, representing the largest relative improvement of 7.4 percent among all metrics. This highlights that Precision is particularly sensitive to Alpha–Gamma interactions.

Recall displays a different trend. Its peak occurs at Alpha 12 and Gamma 4 with a value of 0.805, while the minimum is observed at Alpha 2 and Gamma 6 at 0.788, corresponding to a modest 2.2 percent improvement. The smaller variation indicates that Recall remains relatively stable across parameter settings, maintaining consistent sensitivity to minority-class instances.

Table 5.15 Parameter Sensitivity Ranking

Metric	Alpha Sensitivity	Gamma Sensitivity	Dominant Parameter	Sensitivity Level
F1	0.0042	0.0037	Alpha	Medium
Precision	0.0081	0.0067	Alpha	High
Recall	0.0032	0.003	Alpha	Low
MCC	0.004	0.0036	Alpha	Medium

MCC exhibits trends similar to F1 and Precision, reaching its maximum at Alpha 15 and Gamma 1 with 0.697, and its minimum at Alpha 2 and Gamma 3 with 0.672, showing a 3.7 percent relative gain. This consistency suggests that

a combination of high Alpha and low Gamma improves the model's overall discriminative power.

Table 5.15 summarizes the sensitivity of each metric to variations in Alpha and Gamma. Precision exhibits the highest sensitivity to both Alpha, at 0.0081, and Gamma, at 0.0067, indicating it is most influenced by parameter adjustments. F1 and MCC show moderate sensitivity, with values around 0.004 for Alpha and 0.0036 to 0.0037 for Gamma, while Recall remains the most stable metric, with the lowest sensitivity of 0.0032 for Alpha and 0.0030 for Gamma. Overall, these results indicate that Precision is the most responsive metric, Recall the most robust, and that Alpha and Gamma exert comparable influence on model performance, highlighting their equal importance in shaping predictive outcomes.

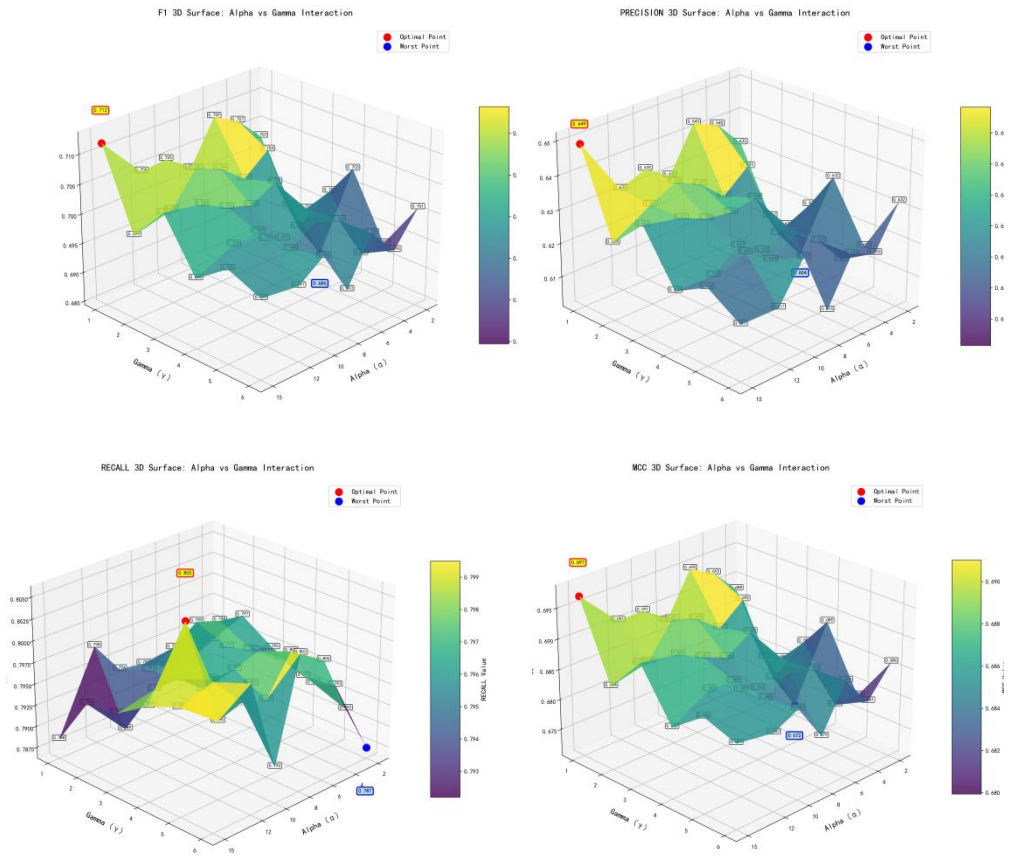

Figure 5.12 Interaction Effects on F1, Precision, Recall, and MCC

The interaction results in Figure 5.12 reveal distinct regions in the Alpha–Gamma parameter space that significantly affect model performance. Synergistic regions include High Alpha + Low Gamma (α=12–15, γ=1–2), which consistently achieve the best F1, Precision, and MCC, making this region suitable for high-precision applications. Another synergistic region, Medium Alpha + Medium Gamma (α=8–12, γ=3–4), optimizes Recall, representing a balanced Precision–Recall trade-off.

In contrast, antagonistic regions demonstrate parameter combinations that degrade performance. Low Alpha + High Gamma (α=2–6, γ=5–6) performs poorly across all metrics, while High Alpha + High Gamma (α=12–15, γ=5–6) leads to significant performance drops, suggesting conflicts between parameters and a risk of overfitting. These findings indicate that the Alpha–Gamma interaction is not purely additive but exhibits both synergistic and antagonistic effects, directly influencing the Precision–Recall balance.

5.4.2 Base RF Parameters Analysis

To evaluate the base model RF parameter effect on the iterative model performance, 81 combinations of four core parameters, N_ESTIMATORS, MAX_DEPTH, MIN_SAMPLES_SPLIT, and MIN_SAMPLES_LEAF, were evaluated. Table B.11 (shown in Appendix B) summarizes the cross-validated performance.

With different parameter configurations in the base learner RF, the iterative model's validation results obtained during iterative training demonstrate consistent predictive performance, with F1-scores varying between 0.63 and 0.74, Precision between 0.54 and 0.66, Recall between 0.75 and 0.82, and MCC between 0.61 and 0.72.

The combination of 100 estimators, a maximum depth of 15, a minimum of 2 samples required for node splitting, and a minimum of 1 sample per leaf node, achieved the highest F1-score (0.74), Precision (0.66), Recall (0.82), and MCC (0.72). Models with shallow trees, limited to a maximum depth of 5, consistently underperformed, with F1 dropping to 0.63 and MCC to 0.61. These results indicate that insufficient tree depth restricts the model's ability to capture complex decision boundaries and properly discriminate minority-class instances.

Table 5.16 RF Performance by Single-Parameter Grouping

Parameter	Value	Number of models	F1 Mean	Precision Mean	Recall Mean	MCC Mean	Train time Mean
N_ESTIMATORS	50	27	0.683	0.604	0.788	0.669	727.17
	100	27	0.686	0.610	0.787	0.672	1028.88
	150	27	0.687	0.609	0.789	0.672	2185.93
MAX_DEPTH	5	27	0.627	0.537	0.754	0.611	841.16
	10	27	0.709	0.638	0.799	0.695	1432.35
	15	27	0.720	0.647	0.812	0.707	1668.47
MIN_SAMPLES_SPLIT	2	27	0.686	0.609	0.787	0.671	1415.16
	5	27	0.686	0.607	0.790	0.671	1149.18
	10	27	0.685	0.607	0.787	0.670	1377.64
MIN_SAMPLES_LEAF	1	27	0.694	0.619	0.793	0.680	1426.89
	5	27	0.686	0.608	0.789	0.671	1285.37
	10	27	0.677	0.596	0.784	0.662	1229.72

To assess hyperparameter effects, the 81 RF configurations were grouped by the four parameters, and Table 5.16 summarizes mean F1, Precision, Recall, MCC, and training time for each parameter, highlighting trade-offs between predictive performance and computational efficiency.

The single-parameter analysis shows that MAX_DEPTH has the strongest impact on model performance. Increasing tree depth from 5 to 15 improves the mean F1-score by 14.8 percent (0.627 → 0.720) and recall by 7.7 percent (0.754 → 0.812), indicating that deeper trees better capture the non-linear, high-dimensional patterns of financial distress. MIN_SAMPLES_LEAF is also critical: a leaf size of 1 yields the highest F1 (0.694) and Precision (0.619),

whereas increasing it to 5 or 10 reduces F1 to 0.686 and 0.677, respectively, highlighting that smaller terminal nodes enhance detection of minority-class (distressed) samples.

MIN_SAMPLES_SPLIT has a minor effect, with mean F1 remaining stable between 0.685 and 0.686 for split values of 2, 5, and 10. Slightly larger splits, such as 5, reduce training time from 1415 seconds to 1149 seconds without sacrificing accuracy, suggesting moderate splits improve efficiency. In contrast, N_ESTIMATORS minimally affects predictive accuracy, with F1 increasing only from 0.683 to 0.687 when estimators rise from 50 to 150, but substantially impacts computation, with training time tripling from 727 seconds to 2186 seconds. This implies that a moderate ensemble size of approximately 100 trees balances accuracy and efficiency.

Overall, MAX_DEPTH and MIN_SAMPLES_LEAF are the most influential hyperparameters for predictive performance, while N_ESTIMATORS primarily affects computational cost, and MIN_SAMPLES_SPLIT has a relatively minor effect. Hyperparameter optimization should therefore prioritize tree depth and leaf size, while maintaining a reasonable ensemble size to balance predictive accuracy and efficiency.

5.4.4 Discussion

This sensitivity analysis provides a clear understanding of how DWARFB responds to changes in both the misclassification strategy parameters and the base RF hyperparameters.

For the misclassified-sample strategy, Alpha and Gamma have little effect when varied independently, but their interaction is crucial. High Alpha

combined with low Gamma yields the best F1, Precision, and MCC, while medium Alpha and medium Gamma maximize Recall. Precision is the most sensitive metric to parameter changes, whereas Recall remains the most stable, indicating that Alpha–Gamma tuning mainly controls the Precision–Recall trade-off.

For the base RF parameters, MAX_DEPTH has the strongest impact on performance, as deeper trees significantly improve F1 and Recall by better capturing complex financial patterns. MIN_SAMPLES_LEAF is also important, with smaller leaf sizes enhancing minority-class detection. In contrast, MIN_SAMPLES_SPLIT has only a minor effect on accuracy but can improve efficiency, and N_ESTIMATORS mainly affects computational cost rather than predictive performance.

In summary, this section shows that DWARFB is both stable and controllable. Its performance is mainly governed by:

- Alpha and Gamma (jointly), which shape the Precision–Recall trade-off,
- MAX_DEPTH and MIN_SAMPLES_LEAF, which determine the model's learning capacity,
- N_ESTIMATORS, which mainly affects computational cost.

These findings offer clear and practical tuning guidance: first adjust tree depth and leaf size, then jointly tune Alpha and Gamma according to application priorities and finally set a moderate number of trees to maintain efficiency. This makes DWARFB not only accurate but also flexible and practical for real-world financial distress early-warning systems.

5.5 Summary

In this chapter, DWARFB is systematically evaluated against a wide range of benchmark models and imbalance-handling techniques, together with an in-depth analysis of its internal misclassification patterns and parameter sensitivity. The main findings can be summarized as follows:

- DWARFB's superior performance under severe class imbalance. It consistently achieves the best balance between minority detection and overall stability, with high Recall and F1-score, strong MCC and Kappa, and controlled false-positive rates.
- Clear advantage over static resampling methods. Compared with SMOTE variants, Random Under-Sampling, and Balanced RF, DWARFB avoids distortion of the original data distribution and substantially reduces false alarms while maintaining competitive sensitivity. This demonstrates that adaptive, in-model imbalance handling is more reliable than preprocessing-based resampling.
- Improvement over standard and cost-sensitive RF. DWARFB significantly outperforms both Standard RF and Cost-sensitive RF by reducing false negatives and improving F1, MCC, ROC-AUC, and PR-AUC. Its dynamic misclassification-aware weighting is more effective than fixed penalty strategies.
- Effectiveness of persistent misclassification-aware learning. The cumulative misclassification mechanism enables DWARFB to correctly learn most distressed samples (about 85%) while preventing excessive emphasis on extremely difficult or noisy cases. This shows that

modelling sample-level difficulty improves recognition of hard-to-classify distress patterns.

- Advantage over single popular models and conventional boosting methods. DWARFB outperforms SVM, MLP, AdaBoost, and XGBoost by achieving a more stable Precision–Recall trade-off. Its history-aware and adaptive boosting strategy is superior to static and iteration-local weighting schemes.
- Interpretable financial structure captured by the model. Feature contribution analysis reveals a clear hierarchy dominated by retained earnings and accumulated profitability, followed by equity strength, current profitability, earnings stability, and cash flow/market signals. The prominence of lagged variables highlights the importance of financial dynamics for early warning.
- Stable and controllable parameter behaviour. Sensitivity analysis shows that model performance is mainly governed by the joint tuning of Alpha and Gamma (Precision–Recall trade-off) and by MAX_DEPTH and MIN_SAMPLES_LEAF (learning capacity), while N_ESTIMATORS primarily affects computational efficiency. This provides practical guidance for real-world deployment.

CHAPTER 6 CONCLUSIONS AND FUTURE WORKS

As the concluding chapter of this thesis, this chapter integrates the major findings of the research and evaluates its academic contributions. It further discusses the limitations that may influence the interpretation of the results and identifies several promising areas for future investigation.

The remainder of the chapter is organized as follows: Section 6.1 presents the conclusions, Section 6.2 summarizes the key contributions, Section 6.3 discusses the study's limitations, and Section 6.4 outlines recommendations for future research.

6.1 Conclusions of This Thesis

This thesis presents a DWARFB framework to address severe class imbalance in financial distress prediction. Our research emphasises persistently hard-to-classify yet informative distressed samples and the preservation of original data distributions while adaptively improving minority-class detection. The DWARFB is one of the very few predictive frameworks that integrates adaptive sampling, misclassification-aware weighting, and performance-driven ensemble strategies into a unified model, and it works efficiently on imbalanced financial datasets. The key conclusions of this thesis, addressing the research questions, are summarized below.

- By integrating AdaBoost-like, Focal-like, and OHEM-inspired strategies, DWARFB’s weighting mechanism provides a more effective proof-of-concept, consistently outperforming individual methods.
- Persistently misclassified positive samples can be effectively emphasized using cumulative misclassification-aware weighting,

improving model focus on borderline yet economically critical samples while mitigating the influence of noisy minority instances.

- The sliding-iteration misclassification tracking with a reward mechanism prevents over-penalization of misclassified samples, stabilizes training, and reduces persistent errors, enabling approximately 85% of distress samples to be correctly learned.
- DWARFB's loss-based feature contribution tracking shows that cumulative profitability and retained earnings are the most influential factors, followed by equity strength, current profitability, earnings stability, and cash flow/market signals, which is consistent with established findings in financial distress and corporate performance research.
- Some additional conclusions derived from the DWARFB's experimental implementation are given below.
- By embedding adaptive imbalance-handling mechanisms directly into the learning process, DWARFB achieves significantly better detection of distressed samples than static resampling strategies, without distorting the overall dataset structure.
- DWARFB's superior performance over traditional boosting ensembles demonstrates that its history-aware and adaptive boosting strategy is more effective than static and iteration-local weighting schemes.
- The parameter sensitivity analysis of DWARFB illustrates that tree depth and leaf size define learning capacity, Alpha and Gamma jointly control the Precision–Recall trade-off, and the number of trees mainly affects efficiency.

- The improvements observed in tree-based models and RF demonstrate that DWARFB can serve as an enhanced and generalizable framework for boosting the performance of both single classifiers and ensemble models.

6.2 Summary of Contributions

The main purpose of this research is to improve the accuracy of predicting financial distress, especially the prediction of distress samples in highly imbalanced dataset. To achieve this goal, this research proposes the DWARFB framework and makes three contributions. They are multi-stage sampling, history-aware ensemble learning, and dynamic adaptive performance-driven ensemble strategy.

6.2.1 Addressing Class imbalance using multi-stage sampling

The multi-stage sampling is designed for addressing class imbalance in predicting, which is an important contribution of this research. Unlike the data-level resampling techniques, it doesn't reduce the samples and synthetic new samples to balance the data, using the original samples in the dataset, through adjust the sampling ratio between majority and minority samples to make the model learn from different sample ratio and finally solve the class imbalance problem. Gradually increasing class imbalance avoids the "over-smoothing" of static oversampling and ensures the model learns from both balanced initial data (for stability) and realistic imbalance (for adaptability). This mechanism provides a new sight in dealing with class imbalance, especially in financial scenarios, because without synthetic new samples can easily explain the model's working mechanism to the thresholder and investors.

6.2.2 History-aware misclassified samples reweighting

The history-aware misclassified sample reweighting mechanism enhances DWARFB's understanding of sample difficulty by tracking misclassification patterns across multiple training iterations. Unlike traditional boosting methods, which update sample weights based only on the previous iteration, this mechanism maintains a cumulative record of consistently difficult samples, shifting the focus from single-step correction to long-term difficulty assessment. This is particularly important in financial prediction, where firms often exhibit gradual early-warning distress signals.

By leveraging accumulated difficulty information, the ensemble effectively targets structurally challenging observations while avoiding overweighting noisy instances. It promotes stable and interpretable correction behavior, improves adaptability in dynamic financial environments, and captures hard-to-classify samples more effectively. Consequently, the model produces predictions that investors and decision-makers can interpret with greater confidence.

6.2.3 Dynamic adaptive performance-driven ensemble strategy

The performance-driven ensemble strategy in DWARFB leverages the F1-score of each iterative model to stratify models into very high, high, medium, and low tiers, allowing weights to adapt dynamically based on real-time cross-validation performance. This tiered system integrates smoothly with the logit-based weight calculation and reduces reliance on fixed-weight rules, limiting the influence of unstable model outputs. Notably, models trained on special balanced sampling ratios tend to perform better during training, but when applied to practical, imbalanced data, their performance may drop. To address

this, the ensemble assigns smaller weight adjustments to strong models, while low-performing models receive larger updates. This adaptive weighting ensures that DWARFB remains robust and better suited for real-world prediction, where the training and operational data distributions differ.

Moreover, stratification enhances interpretability by mapping how performance tiers influence weight allocation, providing transparency in early-warning tasks where practitioners must justify predictions and risk assessments.

Beyond methodological improvements, DWARFB also offers practical value for financial practitioners. In a real-world financial distress early-warning system, the model can be integrated into periodic monitoring frameworks that collect financial, market, and macroeconomic information and generate distress risk scores for listed companies. Investors can use these risk assessments to support portfolio screening and risk management, while creditors, auditors, and regulators can employ them to identify firms requiring closer monitoring. Since model retraining can be conducted periodically (e.g., quarterly or annually), the generated predictions can provide timely early-warning signals without requiring continuous retraining.

It should be noted that the individual concepts underlying DWARFB are not entirely new in isolation. The originality of this research lies in the design, adaptation, and integration of these mechanisms into a unified framework for financial distress prediction. Specifically, the proposed multi-stage sampling strategy introduces a progressive adjustment of class distributions without generating synthetic samples, the history-aware misclassified samples reweighting mechanism extends conventional boosting concepts through cumulative cross-iteration difficulty tracking and a reward-based sliding-

window design, and the dynamic adaptive performance-driven ensemble strategy incorporates F1-score-based model stratification to guide weight allocation. The coordinated interaction of these three components within a single framework constitutes the primary methodological contribution of this thesis.

6.3 Conclusions in Relation to the Research Questions

The research sought to develop a reliable financial distress prediction framework that effectively addresses the challenges posed by highly imbalanced financial data. The proposed DWARFB model addresses two major challenges: (1) severe class imbalance and (2) persistently hard-to-classify minority samples. The conclusions are presented with respect to the two research questions that guided this study, as described in Section 1.4.

RQ1: Can we address severe class imbalance in financial distress prediction while preserving the original data distribution and improving minority-class detection?

Severe class imbalance significantly hinders financial distress prediction, as conventional Models often prioritize the majority class, resulting in high overall accuracy but limited capability to identify financially distressed firms. Traditional resampling methods (SMOTE, ADASYN, under-sampling) and fixed cost-sensitive approaches partially improve recall but either distort the data distribution or lack adaptability.

In contrast, DWARFB embeds adaptive imbalance-handling directly within the training process, adaptively reweighting training samples' weights and sampling ratios based on cumulative misclassification patterns. This allows the model to enhance minority-class detection while preserving the original

dataset structure. Experimental results show that DWARFB achieves a balanced trade-off between precision and recall, reduces false negatives, and maintains stable performance, providing strong support for Hypothesis **H1**.

RQ2: Can financial distress prediction models effectively identify and emphasize persistently hard-to-classify distressed samples while avoiding excessive influence from noisy or mislabelled observations?

Persistently misclassified, borderline distressed samples contain important early-warning signals that traditional models often fail to capture due to uniform or static weighting.

DWARFB addresses this by using history-aware misclassification tracking, dynamically emphasizing difficult samples while controlling the influence of noisy or anomalous observations. This mechanism progressively improves detection of hard-to-classify distressed firms. Experimental results show that DWARFB achieves higher recall, F1-score, MCC, and AUC than single learners, bagging ensembles, and static boosting models, confirming the effectiveness of cumulative difficulty modelling.

Taken together, these findings offer robust empirical evidence for Hypothesis **H2**, and by integrating adaptive misclassification weighting within an ensemble framework, DWARFB also validates Hypothesis **H3**, producing stable, interpretable, and robust predictions.

Together, the findings confirm that:

(1) Adaptive imbalance-handling mechanisms embedded within the learning process improve minority-class detection while preserving the original data distribution (RQ1 / H1).

(2) Persistent misclassification-aware weighting enables the identification of hard-to-classify but informative distressed firms, avoiding the influence of noise (RQ2 / H2 & H3).

Overall, the empirical findings suggest that DWARFB provides a dependable and transparent predictive framework that effectively addresses the challenges of financial distress prediction under highly imbalanced conditions, addressing the limitations of both static resampling and conventional cost-sensitive approaches.

6.4 Limitations of the Study

Despite its strengths, this study has several limitations that suggest directions for future research.

6.4.1 Data-related limitations

The dataset is restricted to listed Chinese A-share companies from 2007 to 2024, making it uncertain whether the observed results would be replicated in variety markets, industries, and institutional contexts. Financial distress labels based on ST classification, while practical, may not fully capture nuances such as technical defaults, covenant violations, or restructuring events. Future work could expand the scope by integrating multinational datasets, incorporating textual disclosures, or adopting alternative distress definitions to evaluate robustness.

6.4.2 Methodological and computational limitations

Although DWARFB outperforms several state-of-the-art models and resampling techniques, it remains computationally intensive, with a training time of 387.9 seconds, which is substantially longer than 28.6 seconds required by the standard Random Forest model and exceeds that of all baseline

resampling methods. The additional computational cost is primarily attributed to the multi-stage sampling process, history-aware misclassified samples reweighting mechanism, and dynamic adaptive weighting strategy, all of which introduce additional computations during model training.

This increased training time represents a trade-off between predictive performance and computational efficiency. While the higher computational burden may limit the applicability of DWARFB in highly time-sensitive environments requiring frequent model retraining, it is less problematic in many practical financial distress prediction scenarios where models are updated periodically, such as on a quarterly or annual basis. In such settings, the training process is typically performed offline, whereas prediction and risk assessment can be executed efficiently once the model has been trained. Therefore, the additional training time may be justified by the improvements achieved in enhancing the identification of financially distressed samples while maintaining robust overall classification performance.

Additionally, the current framework builds its learning process around the RF algorithm. Although the experimental results demonstrate the effectiveness of DWARFB within this setting, the generalizability of the proposed framework to other learning algorithms has not yet been fully investigated. Future studies should examine whether DWARFB can similarly enhance other models, including gradient boosting methods, neural networks, and hybrid ensemble architectures.

6.4.3 Training Process Adaptability Limitations

The current DWARFB framework applies a fixed range of sampling ratios (1:1, 1:2, 1:3) during its progressive iterative training process. Sampling

ratios outside this range, either more conservative or more aggressive, have not been explored, which may reduce the its capacity to optimally adjust to varying degrees of sample difficulty during learning. Future research could investigate a broader spectrum of sampling ratios and develop more flexible, dynamically adaptive adjustment strategies, allowing DWARFB to more effectively emphasize hard-to-classify minority samples while maintaining model stability and predictive reliability.

6.5 Future Works

While this research advances financial distress prediction, several avenues remain for further improvement and expansion.

6.5.1 Data and Feature Expansion

Future research could extend DWARFB to other markets, such as the U.S. S&P 500 or European STOXX 600, to evaluate its cross-market generalization. Incorporating non-financial features, such as textual data (annual report tone, news sentiment), ESG indicators, or market microstructure data (e.g., trading volume volatility), could uncover latent distress signals that traditional financial ratios may overlook.

While the current framework relies on quarterly observations, integrating higher-frequency data, such as monthly or daily stock price movements, and implementing real-time model updating would improve its ability to detect sudden distress triggers, including regulatory shocks or liquidity crises.

6.5.2 Integration with Deep Learning for Temporal Dynamics

Although DWARFB currently leverages lag features to capture time dependencies, combining it with sequence models such as LSTM or

Transformer networks could more effectively model the gradual, non-linear evolution of financial distress. This enhancement would be especially valuable for identifying long-term distress trajectories, such as multi-year declines in profitability.

6.5.3 Industry-Specific Customization

Financial distress patterns differ across sectors. For example, high leverage is common in real estate, while liquidity risk dominates in hospitality. Future research could develop industry-tailored versions of the DWARFB framework, incorporating sector-specific sampling ratios and relevant feature sets, such as occupancy rates for hospitality or inventory turnover for manufacturing. Moreover, integrating macroeconomic indicators like GDP growth and interest rates would help capture systemic risks that influence distress probabilities across firms.

6.5.4 Scalability and Efficiency

The proposed framework, while effective, has higher training time (387.9 seconds vs. 28.6 seconds for standard RF) due to iterative sampling, dynamic weighting, and adaptive ensemble mechanisms. Future work could optimize computational efficiency via potential approaches include algorithmic acceleration techniques, parallelized or distributed training frameworks, GPU-based implementations, adaptive early stopping mechanisms, model pruning, and lightweight ensemble variants. These strategies could help reduce training costs while maintaining predictive effectiveness, thereby improving the scalability and practicality of the framework for large-scale financial monitoring systems and facilitating its deployment in real-time or near-real-time risk monitoring applications.

6.5.5 Enhancing Training Process Adaptability

Future work could explore a broader spectrum of sampling ratios in DWARFB, including both more conservative and more aggressive options, to better tailor the learning focus to varying degrees of sample difficulty. In addition, developing dynamically adaptive adjustment strategies, where sampling ratios and misclassification weights evolve in response to cumulative learning signals, could allow the model to more effectively emphasize persistently hard-to-classify minority samples. Such enhancements would improve the model's sensitivity to borderline distressed firms while maintaining stability, reliability, and interpretability, ultimately strengthening DWARFB's practical utility in financial distress prediction under highly imbalanced conditions.

In summary, this thesis develops a robust iterative adaptive ensemble framework for financial distress prediction. By overcoming key methodological limitations and validating the approach on large-scale real-world data, it offers a foundation for more accurate, reliable, and practical early-warning systems, supporting economic stability and informed risk management for financial stakeholders.

6.6 Summary

This research demonstrates that financial distress prediction can be substantially enhanced through the proposed DWARFB framework, which integrates adaptive imbalance-handling mechanisms directly within the learning process. Empirical results show that DWARFB effectively identifies distressed samples in highly imbalanced datasets, achieving a balanced trade-off between recall, precision, and overall stability, while providing interpretable insights into

key financial predictors such as accumulated profitability, equity strength, and earnings stability.

The study highlights that conventional preprocessing-based resampling methods and static cost-sensitive approaches are limited by data distortion or non-adaptive weighting, whereas DWARFB dynamically adjusts sample importance during training, emphasizing persistently hard-to-classify minority instances without being dominated by noisy observations. These results provide strong empirical support for the research hypotheses regarding adaptive imbalance handling, cumulative difficulty modelling, and misclassification-history-driven ensemble learning.

Despite its strengths, the framework has limitations, including restricted dataset coverage, computational intensity, only reliance on RF as the base learner, and a fixed range of sampling ratios during training. Future research could address these limitations by expanding data and feature sources, integrating deep learning for temporal dynamics, customizing the framework for industry-specific patterns, improving computational efficiency, and enhancing training adaptability through more flexible and dynamically adjustable sampling strategies.

Reference

[1] S. Fakhar, F. Mohd Khan, M. I. Tabash, G. Ahmad, J. Akhter, and M. S. M. Al-Absy, 'Financial distress in the banking industry: A bibliometric synthesis and exploration', *Cogent Economics & Finance*, vol. 11, no. 2, p. 2253076, Oct. 2023, doi: 10.1080/23322039.2023.2253076.

[2] F. Zhou, L. Fu, Z. Li, and J. Xu, 'The recurrence of financial distress: A survival analysis', *International Journal of Forecasting*, vol. 38, no. 3, pp. 1100–1115, Jul. 2022, doi: 10.1016/j.ijforecast.2021.12.005.

[3] Cyril and Methodius University in Skopje, North Macedonia, A. Atanasovski, and T. Tocev, 'Research trends in disruptive technologies for accounting of the future – A bibliometric analysis', *JAMIS*, vol. 21, no. 2, pp. 270–288, Jun. 2022, doi: 10.24818/jamis.2022.02006.

[4] R. Crespí-Cladera, A. Martín-Oliver, and B. Pascual-Fuster, 'Financial distress in the hospitality industry during the Covid-19 disaster', *Tourism Management*, vol. 85, p. 104301, Aug. 2021, doi: 10.1016/j.tourman.2021.104301.

[5] S. J. Enumah and D. C. Chang, 'Predictors of Financial Distress Among Private U.S. Hospitals', *Journal of Surgical Research*, vol. 267, pp. 251–259, Nov. 2021, doi: 10.1016/j.jss.2021.05.025.

[6] L. (Don) A. N. Dioko and J. (Frank) Guo, 'Signs of imminent collapse: Can hotel bankruptcy or failure be predicted from guest reviews?', *International Journal of Hospitality Management*, vol. 119, p. 103711, May 2024, doi: 10.1016/j.ijhm.2024.103711.

[7] M. Papík and L. Papíková, 'Automated Machine Learning in Bankruptcy Prediction of Manufacturing Companies', *Procedia Computer Science*, vol. 232, pp. 1428–1436, 2024, doi: 10.1016/j.procs.2024.01.141.

[8] S. Kaur, A. Arora, and A. K. Goyal, 'A Comparative Analysis of Machine Learning Methods to Predict Bankruptcy among Manufacturing Companies in India Post-IBC', in *2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, Chennai, India: IEEE, Nov. 2023, pp. 1–6. doi: 10.1109/RMKMATE59243.2023.10369836.

[9] F. Dube, N. Nzimande, and P.-F. Muzindutsi, 'Application of artificial neural networks in predicting financial distress in the JSE financial services and manufacturing companies', *Journal of Sustainable Finance & Investment*, vol. 13, no. 1, pp. 723–743, Jan. 2023, doi: 10.1080/20430795.2021.2017257.

[10] K. A. Fachrudin, 'The Relationship between Financial Distress and Financial Health Prediction Model: A Study in Public Manufacturing Companies Listed on Indonesia Stock Exchange (IDX)', *jak*, vol. 22, no. 1, pp. 18–27, May 2020, doi: 10.9744/jak.22.1.18-27.

[11] Y. Du, F. Wang, Y. Wang, and J. Jia, 'The Research on Prediction for Financial Distress in Car Company Listed Combining Financial Indicators and Text Data', in *Artificial Intelligence in China*, vol. 871, Q. Liang, W. Wang, J. Mu, X. Liu, and Z. Na, Eds, in Lecture Notes in Electrical Engineering, vol. 871. , Singapore: Springer Nature Singapore, 2023, pp. 203–210. doi: 10.1007/978-981-99-1256-8_24.

[12] G. Manthoulis, M. Doumpos, C. Zopounidis, and E. Galariotis, 'An ordinal classification framework for bank failure prediction: Methodology and empirical evidence for US banks', *European Journal of Operational Research*, vol. 282, no. 2, pp. 786–801, Apr. 2020, doi: 10.1016/j.ejor.2019.09.040.

[13] D. P. De Jesus and C. D. N. Besarria, 'Machine learning and sentiment analysis: Projecting bank insolvency risk', *Research in Economics*, vol. 77, no. 2, pp. 226–238, Jun. 2023, doi: 10.1016/j.rie.2023.03.001.

[14] H. Keshavarz Hadadha, H. Alipour, and S. Kheradyar, 'Analysis of Challenges and Strategies to Develop More Effective Early Warning System of Bank Bankruptcy', *Environmental Energy and Economic Research*, vol. 8, no. 1, Feb. 2024, doi: 10.22097/eeer.2024.417489.1298.

[15] Z. K. Zhailybayevich and M. A. Hamada, 'Development of a Predictive Intellectual Model for Predicting the Financial Crisis in Banks', in *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India: IEEE, May 2023, pp. 661–665. doi: 10.1109/ICAAIC56838.2023.10140628.

[16] M. M. Rahman, Z. Sultana, M. Jahan, and R. Fariha, 'Prediction of Financial Distress in Bangladesh's Banking Sector Using Data Mining and Machine-Learning Technique', in *Proceedings of International Joint Conference on Computational Intelligence*, M. S. Uddin and

J. C. Bansal, Eds, in Algorithms for Intelligent Systems. Singapore: Springer Singapore, 2020, pp. 151–162. doi: 10.1007/978-981-15-3607-6_12.
[17] M. Papík and L. Papíková, 'Impacts of crisis on SME bankruptcy prediction models' performance', *Expert Systems with Applications*, vol. 214, p. 119072, Mar. 2023, doi: 10.1016/j.eswa.2022.119072.
[18] H. Elsegai, 'Improving the Process of Early-Warning Detection and Identifying the Most Affected Markets: Evidence from Subprime Mortgage Crisis and COVID-19 Outbreak—Application to American Stock Markets', *Entropy*, vol. 25, no. 1, p. 70, Dec. 2022, doi: 10.3390/e25010070.
[19] K. Halteh and H. Sharari, 'Employing Artificial Neural Networks and Multiple Discriminant Analysis to Evaluate the Impact of the COVID-19 Pandemic on the Financial Status of Jordanian Companies', *IJIKM*, vol. 18, pp. 251–267, 2023, doi: 10.28945/5112.
[20] A. A. Abdul-kareem, Z. Taha Fayed, S. Rady, S. A. El-Regaily, and B. M. Nema, 'An Intelligent Decision Support System for Forecasting Financially Distressed Businesses', in *2023 Eleventh International Conference on Intelligent Computing and Information Systems (ICICIS)*, Cairo, Egypt: IEEE, Nov. 2023, pp. 353–359. doi: 10.1109/ICICIS58388.2023.10391150.
[21] W.-T. Pan *et al.*, 'Model Construction of Enterprise Financial Early Warning Based on Quantum FOA-SVR', *Scientific Programming*, vol. 2021, pp. 1–8, Sep. 2021, doi: 10.1155/2021/5018917.
[22] E. I. Altman, 'FINANCIAL RATIOS, DISCRIMINANT ANALYSIS AND THE PREDICTION OF CORPORATE BANKRUPTCY', *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, Sep. 1968, doi: 10.1111/j.1540-6261.1968.tb00843.x.
[23] J. Sun, H. Li, Q.-H. Huang, and K.-Y. He, 'Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches', *Knowledge-Based Systems*, vol. 57, pp. 41–56, Feb. 2014, doi: 10.1016/j.knosys.2013.12.006.
[24] A. Citterio, 'Bank failure prediction models: Review and outlook', *Socio-Economic Planning Sciences*, vol. 92, p. 101818, Apr. 2024, doi: 10.1016/j.seps.2024.101818.
[25] R. Verma and D. K. Pandiya, 'The role of AI and machine learning in U.S. financial market predictions: Progress, obstacles, and consequences', *International Journal of Global Innovations and Solutions (IJGIS)*, Aug. 2024, doi: 10.21428/e90189c8.d3f08d5f.
[26] Odeyemi Olubusola, Noluthando Zamanjomane Mhlongo, Donald Obinna Daraojimba, Adeola Olusola Ajayi-Nifise, and Titilola Falaiye, 'Machine learning in financial forecasting: A U.S. review: exploring the advancements, challenges, and implications of AI-driven predictions in financial markets', *World J. Adv. Res. Rev.*, vol. 21, no. 2, pp. 1969–1984, Feb. 2024, doi: 10.30574/wjarr.2024.21.2.0444.
[27] L. Dube and T. Verster, 'Enhancing classification performance in imbalanced datasets: A comparative analysis of machine learning models', *DSFE*, vol. 3, no. 4, pp. 354–379, 2023, doi: 10.3934/DSFE.2023021.
[28] C. Jiang, X. Lyu, Y. Yuan, Z. Wang, and Y. Ding, 'Mining semantic features in current reports for financial distress prediction: Empirical evidence from unlisted public firms in China', *International Journal of Forecasting*, vol. 38, no. 3, pp. 1086–1099, Jul. 2022, doi: 10.1016/j.ijforecast.2021.06.011.
[29] J. Radovanovic and C. Haas, 'The evaluation of bankruptcy prediction models based on socio-economic costs', *Expert Systems with Applications*, vol. 227, p. 120275, Oct. 2023, doi: 10.1016/j.eswa.2023.120275.
[30] M. Altalhan, A. Algarni, and M. Turki-Hadj Alouane, 'Imbalanced data problem in machine learning: A review', *IEEE Access*, vol. 13, pp. 13686–13699, 2025, doi: 10.1109/ACCESS.2025.3531662.
[31] E. V. Lashkevich, 'Enhancing bankruptcy prediction efficiency using synthetic data', *JBI*, vol. 19, no. 3, pp. 22–47, Sep. 2025, doi: 10.17323/2587-814X.2025.3.22.47.
[32] Y.-H. Hu, C.-F. Tsai, and P.-T. Wang, 'Combining multiple data resampling methods and classifier ensembles for better financial distress prediction: Homogeneous and heterogeneous approaches', *Ann Oper Res*, vol. 353, no. 2, pp. 793–814, Oct. 2025, doi: 10.1007/s10479-025-06706-5.
[33] M. Carvalho, A. J. Pinho, and S. Brás, 'Resampling approaches to handle class imbalance: A review from a data perspective', *J Big Data*, vol. 12, no. 1, p. 71, Mar. 2025, doi: 10.1186/s40537-025-01119-4.

[34] M. Akouhar, A. Abarda, M. El Fatini, and M. Ouhssini, ‘Enhancing credit card fraud detection: The impact of oversampling rates and ensemble methods with diverse feature selection’, *reks*, vol. 2025, no. 1, pp. 85–101, Feb. 2025, doi: 10.32620/reks.2025.1.06.
[35] H. Wei, ‘SMOTE algorithm optimization and application in corporate credit risk prediction with diversification strategy consideration’, *Sci Rep*, vol. 15, no. 1, p. 23598, Jul. 2025, doi: 10.1038/s41598-025-09173-x.
[36] K. Kim, ‘Noise avoidance SMOTE in ensemble learning for imbalanced data’, *IEEE Access*, vol. 9, pp. 143250–143265, 2021, doi: 10.1109/ACCESS.2021.3120738.
[37] S.-W. Ke, C.-F. Tsai, Y.-Y. Pan, and W.-C. Lin, ‘Majority re-sampling via sub-class clustering for imbalanced datasets’, *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 36, no. 8, pp. 1581–1596, Nov. 2024, doi: 10.1080/0952813X.2023.2165715.
[38] Y. Bhargava, S. K. Shetty, and V. Baths, ‘Subjective cognitive decline prediction on imbalanced data using data-resampling and cost-sensitive training methods’, *Procedia Computer Science*, vol. 235, pp. 1964–1979, 2024, doi: 10.1016/j.procs.2024.04.186.
[39] T. Wongvorachan, S. He, and O. Bulut, ‘A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining’, *Information*, vol. 14, no. 1, p. 54, Jan. 2023, doi: 10.3390/info14010054.
[40] X. Du, W. Li, S. Ruan, and L. Li, ‘CUS-heterogeneous ensemble-based financial distress prediction for imbalanced dataset with ensemble feature selection’, *Appl. Soft Comput.*, vol. 97, p. 106758, Dec. 2020, doi: 10.1016/j.asoc.2020.106758.
[41] N. Syam and R. Kaul, ‘Random forest, bagging, and boosting of decision trees’, in *Machine Learning and Artificial Intelligence in Marketing and Sales*, Emerald Publishing Limited, 2021, pp. 139–182. doi: 10.1108/978-1-80043-880-420211006.
[42] Y. Chen, X. Kuang, and J. Guo, ‘LiFoL: An Efficient Framework for Financial Distress Prediction in High-Dimensional Unbalanced Scenario’, *IEEE Trans. Comput. Soc. Syst.*, pp. 1–12, 2024, doi: 10.1109/TCSS.2023.3276059.
[43] J. Sun, M. Zhao, and C. Lei, ‘Class-imbalanced dynamic financial distress prediction based on random forest from the perspective of concept drift’, *Risk Manag*, vol. 26, no. 4, p. 19, Dec. 2024, doi: 10.1057/s41283-024-00150-8.
[44] M. Muniappan and N. D. Paruvachi Subramanian, ‘A majority voting mechanism-based ensemble learning approach for financial distress prediction in Indian automobile industry’, *JRFM*, vol. 18, no. 4, p. 197, Apr. 2025, doi: 10.3390/jrfm18040197.
[45] B. Siswoyo, Z. Abal Abas, A. N. Che Pee, R. Komalasari, and N. Suryana, ‘Ensemble machine learning algorithm optimization of bankruptcy prediction of bank’, *IJ-AI*, vol. 11, no. 2, p. 679, Jun. 2022, doi: 10.11591/ijai.v11.i2.pp679-686.
[46] J. Sun, H. Li, H. Fujita, B. Fu, and W. Ai, ‘Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting’, *Inform. Fusion*, vol. 54, pp. 128–144, Feb. 2020, doi: 10.1016/j.inffus.2019.07.006.
[47] I. Ghosh, T. D. Chaudhuri, L. Isskandarani, and M. Z. Abedin, ‘Predicting financial cycles with dynamic ensemble selection frameworks using leading, coincident and lagging indicators’, *Research in International Business and Finance*, vol. 80, p. 103114, Aug. 2025, doi: 10.1016/j.ribaf.2025.103114.
[48] K. Fujiwara, ‘Knowledge distillation with resampling for imbalanced data classification: Enhancing predictive performance and explainability stability’, *Results in Engineering*, vol. 24, p. 103406, Dec. 2024, doi: 10.1016/j.rineng.2024.103406.
[49] S. Matharaarachchi, M. Domaratzki, and S. Muthukumarana, ‘Enhancing SMOTE for imbalanced data with abnormal minority instances’, *Machine Learning with Applications*, vol. 18, p. 100597, Dec. 2024, doi: 10.1016/j.mlwa.2024.100597.
[50] I. E. Eteng, U. L. Chinedu, and A. E. Ibor, ‘A stacked ensemble approach with resampling techniques for highly effective fraud detection in imbalanced datasets’, *J. Nig. Soc. Phys. Sci.*, p. 2066, Feb. 2025, doi: 10.46481/jnsps.2025.2066.
[51] G. Ranjbaran, D. Reforgiato Recupero, G. Lombardo, and S. Consoli, ‘Leveraging augmentation techniques for tasks with unbalancedness within the financial domain: A two-level ensemble approach’, *EPJ Data Sci.*, vol. 12, no. 1, p. 24, Jul. 2023, doi: 10.1140/epjds/s13688-023-00402-9.
[52] T. Ren, T. Lu, and Y. Yang, ‘Improved Data Mining Method for Class-Imbalanced Financial Distress Prediction’, in *2021 7th International Conference on Computing and Artificial Intelligence*, Tianjin China: ACM, Apr. 2021, pp. 308–313. doi: 10.1145/3467707.3467754.

[53] Y. Zou, C. Gao, and H. Gao, 'Business failure prediction based on a cost-sensitive extreme gradient boosting machine', *IEEE Access*, vol. 10, pp. 42623–42639, 2022, doi: 10.1109/ACCESS.2022.3168857.
[54] W. Yotsawat, K. Phodong, T. Promrat, and P. Wattuya, 'Bankruptcy prediction model using cost-sensitive extreme gradient boosting in the context of imbalanced datasets', *IJECE*, vol. 13, no. 4, p. 4683, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4683-4691.
[55] W. Liu, H. Fan, M. Xia, and C. Pang, 'Predicting and interpreting financial distress using a weighted boosted tree-based tree', *Engineering Applications of Artificial Intelligence*, vol. 116, p. 105466, Nov. 2022, doi: 10.1016/j.engappai.2022.105466.
[56] K. W. De Bock, K. Coussement, and S. Lessmann, 'Cost-sensitive business failure prediction when misclassification costs are uncertain: A heterogeneous ensemble selection approach', *European Journal of Operational Research*, vol. 285, no. 2, pp. 612–630, Sep. 2020, doi: 10.1016/j.ejor.2020.01.052.
[57] L. Papíková and M. Papík, 'Effects of classification, feature selection, and resampling methods on bankruptcy prediction of small and medium-sized enterprises', *Intelligent Sys in Account*, vol. 29, no. 4, pp. 254–281, Oct. 2022, doi: 10.1002/isaf.1521.
[58] C. P. Chan, J. H. Yang, and W.-H. Chang, 'Entropy-based time-series financial distress model based on attribute selection and MetaCost methods for imbalance class', in *Proceedings of the 2023 3rd International Conference on Artificial Intelligence, Automation and Algorithms*, Beijing China: ACM, Jul. 2023, pp. 140–151. doi: 10.1145/3611450.3611471.
[59] S. A.-D. Safi, P. A. Castillo, and H. Faris, 'Cost-Sensitive Metaheuristic Optimization-Based Neural Network with Ensemble Learning for Financial Distress Prediction', *Applied Sciences*, vol. 12, no. 14, p. 6918, Jul. 2022, doi: 10.3390/app12146918.
[60] P. Peykani, M. Peymany Foroushany, C. Tanasescu, M. Sargolzaei, and H. Kamyabfar, 'Evaluation of cost-sensitive learning models in forecasting business failure of capital market firms', *Mathematics*, vol. 13, no. 3, p. 368, Jan. 2025, doi: 10.3390/math13030368.
[61] A. Ekinci and S. Sen, 'Forecasting bank failure in the U.S.: A cost-sensitive approach', *Comput Econ*, vol. 64, no. 6, pp. 3161–3179, Dec. 2024, doi: 10.1007/s10614-023-10537-6.
[62] J. C-Rella, G. Claeskens, R. Cao, and J. M. Vilar, 'Instance-dependent cost-sensitive parametric learning', *Neurocomputing*, vol. 615, p. 128875, Jan. 2025, doi: 10.1016/j.neucom.2024.128875.
[63] T. Le, 'A comprehensive survey of imbalanced learning methods for bankruptcy prediction', *IET Communications*, vol. 16, no. 5, pp. 433–441, Mar. 2022, doi: 10.1049/cmu2.12268.
[64] K. W. De Bock, K. Coussement, and S. Lessmann, 'Cost-sensitive business failure prediction when misclassification costs are uncertain: A heterogeneous ensemble selection approach', *European Journal of Operational Research*, vol. 285, no. 2, pp. 612–630, Sep. 2020, doi: 10.1016/j.ejor.2020.01.052.
[65] J. Mao, K. Huang, and J. Liu, 'MLAWSMOTE: Oversampling in imbalanced multi-label classification with missing labels by learning label correlation matrix', *Int J Comput Intell Syst*, vol. 17, no. 1, p. 205, Aug. 2024, doi: 10.1007/s44196-024-00607-4.
[66] A. A. Khan, O. Chaudhari, and R. Chandra, 'A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation', *Expert Systems with Applications*, vol. 244, p. 122778, Jun. 2024, doi: 10.1016/j.eswa.2023.122778.
[67] M. Almeida, Y. Zhuang, W. Ding, S. E. Crouter, and P. Chen, 'Mitigating class-boundary label uncertainty to reduce both model bias and variance', *ACM Trans. Knowl. Discov. Data*, vol. 15, no. 2, pp. 1–18, Apr. 2021, doi: 10.1145/3429447.
[68] Z. Tao and J. Chao, 'Financial fraud and anomaly detection techniques: A literature review', in *Proceedings of the 2023 4th International Conference on Computer Science and Management Technology*, Xi'an China: ACM, Oct. 2023, pp. 634–641. doi: 10.1145/3644523.3644639.
[69] K. Miałkowska, K. Kaczmarczyk, M. Hernesa, and M. Dyvakb, 'Feature selection for financial data – comparison', *Procedia Computer Science*, vol. 207, pp. 3047–3056, 2022, doi: 10.1016/j.procs.2022.09.362.
[70] F. Dakalbab, M. A. Talib, Q. Nasir, and T. Saroufil, 'Artificial intelligence techniques in financial trading: A systematic literature review', *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 3, p. 102015, Mar. 2024, doi: 10.1016/j.jksuci.2024.102015.

[71] M. Fachrie, A. Musdholifah, and R. Pulungan, ‘Effectiveness of data resampling and ensemble learning in multiclass imbalance learning’, *Artif Intell Rev*, vol. 58, no. 12, p. 368, Oct. 2025, doi: 10.1007/s10462-025-11357-w.
[72] B. Pes, ‘Learning from high-dimensional and class-imbalanced datasets using random forests’, *Information*, vol. 12, no. 8, p. 286, Jul. 2021, doi: 10.3390/info12080286.
[73] J. Quist, L. Taylor, J. Staaf, and A. Grigoriadis, ‘Random forest modelling of high-dimensional mixed-type data for breast cancer classification’, *Cancers*, vol. 13, no. 5, p. 991, Feb. 2021, doi: 10.3390/cancers13050991.
[74] D. Ghosh and J. Cabrera, ‘Enriched random forest for high dimensional genomic data’, *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, vol. 19, no. 5, pp. 2817–2828, Sep. 2022, doi: 10.1109/TCBB.2021.3089417.
[75] M. Moutaib, M. Fattah, Y. Farhaoui, B. Aghoutane, and M. El Bekkali, ‘Fetal electrocardiogram prediction using machine learning: A random forest-based approach’, *IJEECS*, vol. 33, no. 2, p. 1076, Feb. 2024, doi: 10.11591/ijeecs.v33.i2.pp1076-1083.
[76] E. Halabaku and E. Bytyçi, ‘Overfitting in machine learning: A comparative analysis of decision trees and random forests’, *IASC*, vol. 39, no. 6, pp. 987–1006, 2024, doi: 10.32604/iasc.2024.059429.
[77] F. T. Kristanti, D. F. Salim, and A. F. Ihsan, ‘Financial distress prediction in emerging markets by using a machine learning approach for financial sustainability’, *Discov Sustain*, vol. 6, no. 1, p. 1296, Nov. 2025, doi: 10.1007/s43621-025-02199-1.
[78] Q. Zhang, ‘Financial data anomaly detection method based on decision tree and random forest algorithm’, *Journal of Mathematics*, vol. 2022, no. 1, p. 9135117, Jan. 2022, doi: 10.1155/2022/9135117.
[79] F. O. Aghware *et al.*, ‘Enhancing the random forest model via synthetic minority oversampling technique for credit-card fraud detection’, *J. Comput. Theor. Appl.*, vol. 1, no. 4, pp. 407–420, Mar. 2024, doi: 10.62411/jcta.10323.
[80] Z. Wang, ‘A study on early warning of financial indicators of listed companies based on random forest’, *Discrete Dynamics in Nature and Society*, vol. 2022, no. 1, p. 1314798, Jan. 2022, doi: 10.1155/2022/1314798.
[81] J. Zhang, ‘Impact of an improved random forest-based financial management model on the effectiveness of corporate sustainability decisions’, *Systems and Soft Computing*, vol. 6, p. 200102, Dec. 2024, doi: 10.1016/j.sasc.2024.200102.
[82] Y. Shi, V. Charles, and J. Zhu, ‘Bank financial sustainability evaluation: Data envelopment analysis with random forest and shapley additive explanations’, *European Journal of Operational Research*, vol. 321, no. 2, pp. 614–630, Mar. 2025, doi: 10.1016/j.ejor.2024.09.030.
[83] M. Shah, K. Gandhi, K. A. Patel, H. Kantawala, R. Patel, and A. Kothari, ‘Theoretical evaluation of ensemble machine learning techniques’, in *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, Tirunelveli, India: IEEE, Jan. 2023, pp. 829–837. doi: 10.1109/ICSSIT55814.2023.10061139.
[84] U. Ahmed *et al.*, ‘Hybrid bagging and boosting with SHAP based feature selection for enhanced predictive modeling in intrusion detection systems’, *Sci Rep*, vol. 14, no. 1, p. 30532, Dec. 2024, doi: 10.1038/s41598-024-81151-1.
[85] Y. Zhao, D. Goulias, and S. Saremi, ‘Enhancing prediction accuracy of concrete compressive strength using stacking ensemble machine learning’, *Computers and Concrete*, vol. 32, no. 3, pp. 233–246, Sep. 2023, doi: 10.12989/CAC.2023.32.3.233.
[86] E. I. Altman, M. Iwanicz-Drozdowska, E. K. Laitinen, and A. Suvas, ‘Distressed firm and bankruptcy prediction in an international context: A review and empirical analysis of altman’s Z-score model’, *SSRN Journal*, 2014, doi: 10.2139/ssrn.2536340.
[87] K. El Madou, S. Marso, M. El Kharrim, and M. El Merouani, ‘Evolutions in machine learning technology for financial distress prediction: A comprehensive review and comparative analysis’, *Expert Systems*, vol. 41, no. 2, p. e13485, Feb. 2024, doi: 10.1111/exsy.13485.
[88] F. Rashid, R. A. Khan, and I. H. Qureshi, ‘A Comprehensive Review of the Altman Z-Score Model Across Industries’, *SSRN Journal*, 2024, doi: 10.2139/ssrn.5044057.
[89] Altman, Edward I, ‘Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy’, *The Journal of Finance*, vol. 23, no. 4, pp. 589–609, Sep. 1968, doi: 10.2307/2978933.
[90] K. Seno Pamungkas, ‘Financial Distress Analysis Using the Ohlson Model in Indonesian State Owned Enterprises’, *JAFM*, vol. 3, no. 6, pp. 272–381, Feb. 2023, doi: 10.38035/jafm.v3i6.176.

[91] M. Marsenne, T. Ismail, M. Taqi, and I. A. Hanifah, 'Financial distress predictions with Altman, Springate, Zmijewski, Taffler and Grover models', *10.5267/j.dsl*, vol. 13, no. 1, pp. 181–190, 2024, doi: 10.5267/j.dsl.2023.10.002.
[92] D. M. A. S. Elkahwagy, C. J. Kiriacos, and M. Mansour, 'Logistic regression and other statistical tools in diagnostic biomarker studies', *Clin Transl Oncol*, vol. 26, no. 9, pp. 2172–2180, Mar. 2024, doi: 10.1007/s12094-024-03413-8.
[93] Y. Huo, L. H. Chan, and D. Miller, 'Bankruptcy Prediction for Restaurant Firms: A Comparative Analysis of Multiple Discriminant Analysis and Logistic Regression', *JRFM*, vol. 17, no. 9, p. 399, Sep. 2024, doi: 10.3390/jrfm17090399.
[94] A. Hassan and N. Yousaf, 'Bankruptcy Prediction using Diverse Machine Learning Algorithms', in *2022 International Conference on Frontiers of Information Technology (FIT)*, Islamabad, Pakistan: IEEE, Dec. 2022, pp. 106–111. doi: 10.1109/FIT57066.2022.00029.
[95] M. Sreedharan, A. M. Khedr, and M. El Bannany, 'A Comparative Analysis of Machine Learning Classifiers and Ensemble Techniques in Financial Distress Prediction', in *2020 17th International Multi-Conference on Systems, Signals & Devices (SSD)*, Monastir, Tunisia: IEEE, Jul. 2020, pp. 653–657. doi: 10.1109/SSD49366.2020.9364178.
[96] J. Sun, H. Fujita, Y. Zheng, and W. Ai, 'Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods', *Information Sciences*, vol. 559, pp. 153–170, Jun. 2021, doi: 10.1016/j.ins.2021.01.059.
[97] A. Sabek, 'Unveiling the diverse efficacy of artificial neural networks and logistic regression: A comparative analysis in predicting financial distress.', *Croatian Review of Economic, Business & Social Statistics*, vol. 9, no. 1, pp. 16–32, 2023, doi: 10.2478/crebss-2023-0002.
[98] F. T. Kristanti and V. Dhaniswara, 'The accuracy of artificial neural networks and logit models in predicting the companies’ financial distress', *Journal of Technology Management & Innovation*, vol. 18, no. 3, pp. 42–50, 2023, doi: 10.4067/S0718-27242023000300042.
[99] S. Tangirala, 'Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm*', *IJACSA*, vol. 11, no. 2, 2020, doi: 10.14569/IJACSA.2020.0110277.
[100] Z. Hua, Y. Wang, X. Xu, B. Zhang, and L. Liang, 'Predicting corporate financial distress based on integration of support vector machine and logistic regression', *Expert Systems with Applications*, vol. 33, no. 2, pp. 434–440, Aug. 2007, doi: 10.1016/j.eswa.2006.05.006.
[101] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, 'Learning representations by back-propagating errors', *Nature*, vol. 323, no. 6088, pp. 533–536, Oct. 1986, doi: 10.1038/323533a0.
[102] N. W. D. Ayuni, N. N. Lasmini, and A. A. Putrawan, 'Support Vector Machine (SVM) as Financial Distress Model Prediction in Property and Real Estate Companies', in *Proceedings of the International Conference on Applied Science and Technology on Social Science 2022 (iCAST-SS 2022)*, A. Azizah and E. D. Ariyani, Eds, Paris: Atlantis Press SARL, 2022, pp. 397–402. doi: 10.2991/978-2-494069-83-1_72.
[103] V. Budhidharma, R. Sembel, E. Hulu, and G. Ugut, 'Early warning signs of financial distress using random forest and logit model', *CBSR*, vol. 4, no. 4, pp. 69–88, 2023, doi: 10.22495/cbsrv4i4art8.
[104] G. A. Sandag, 'A PREDICTION MODEL OF COMPANY HEALTH USING BAGGING CLASSIFIER', *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer*, vol. 6, no. 1, Aug. 2020, doi: https://doi.org/10.33480/jitk.v6i1.1390.
[105] W. Wang and Z. Liang, 'Financial Distress Early Warning for Chinese Enterprises from a Systemic Risk Perspective: Based on the Adaptive Weighted XGBoost-Bagging Model', *Systems*, vol. 12, no. 2, p. 65, Feb. 2024, doi: 10.3390/systems12020065.
[106] A. A. Khalil, Z. Liu, A. Salah, A. Fathalla, and A. Ali, 'Predicting Insolvency of Insurance Companies in Egyptian Market Using Bagging and Boosting Ensemble Techniques', *IEEE Access*, vol. 10, pp. 117304–117314, 2022, doi: 10.1109/ACCESS.2022.3210032.
[107] Y. Zou, C. Gao, and H. Gao, 'Business Failure Prediction Based on a Cost-Sensitive Extreme Gradient Boosting Machine', *IEEE Access*, vol. 10, pp. 42623–42639, 2022, doi: 10.1109/ACCESS.2022.3168857.
[108] B. Siswoyo and N. Suryana, 'Ensemble Learning Gradient Boosting in Improving Classification and Prediction in Machine Learning', *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 8, pp. 1997–2002, 2021.
[109] Jiaming Liu, Chengzhang Li, PengOuyang, JiajiaLiu, and ChongWu, 'Interpreting the prediction results of the tree-based gradient boosting models for financial distress prediction

with an explainable machine learning approach', *Journal of Forecasting*, vol. 42, no. 5, pp. 1112–1137, Sep. 2022, doi: 10.1002/for.2931.
[110] D. Liang, C.-F. Tsai, H.-Y. (Richard) Lu, and L.-S. Chang, 'Combining corporate governance indicators with stacking ensembles for financial distress prediction', *J. Bus. Res.*, vol. 120, pp. 137–146, Nov. 2020, doi: 10.1016/j.jbusres.2020.07.052.
[111] C. Wu, X. Chen, and Y. Jiang, 'Financial distress prediction based on ensemble feature selection and improved stacking algorithm', *K*, Feb. 2024, doi: 10.1108/K-08-2023-1428.
[112] Z. Zhao, D. Li, and W. Dai, 'Machine-learning-enabled intelligence computing for crisis management in small and medium-sized enterprises (SMEs)', *Technological Forecasting and Social Change*, vol. 191, p. 122492, Jun. 2023, doi: 10.1016/j.techfore.2023.122492.
[113] M. F. Hadi, D.-R. Liang, T. K. Priyambodo, and A. Sn, 'Financial Distress Prediction with Stacking Ensemble Learning', *Indonesian J. Comput. Cybern. Syst.*, vol. 16, no. 3, p. 281, Jul. 2022, doi: 10.22146/ijccs.76575.
[114] S. Y. Kim and A. Upneja, 'Majority voting ensemble with a decision trees for business failure prediction during economic downturns', *Journal of Innovation & Knowledge*, vol. 6, no. 2, pp. 112–123, Apr. 2021, doi: 10.1016/j.jik.2021.01.001.
[115] Leo Breiman, 'Random Forests', *Machine Learning*, vol. 45, pp. 5–32, Oct. 2001, doi: 10.1023/a:1010933404324.
[116] Y. Freund and R. E. Schapire, 'A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting', *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, Aug. 1997, doi: 10.1006/jcss.1997.1504.
[117] T. Chen and C. Guestrin, 'XGBoost: A Scalable Tree Boosting System', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA: ACM, Aug. 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
[118] S. B. Jabeur, C. Gharib, S. Mefteh-Wali, and W. B. Arfi, 'CatBoost model and artificial intelligence techniques for corporate failure prediction', *Technol. Forecast. Soc.*, vol. 166, p. 120658, May 2021, doi: 10.1016/j.techfore.2021.120658.
[119] Z. Wang, 'A Study on Early Warning of Financial Indicators of Listed Companies Based on Random Forest', *Discrete Dynamics in Nature and Society*, vol. 2022, pp. 1–12, Sep. 2022, doi: 10.1155/2022/1314798.
[120] H. Yang, E. Li, Y. F. Cai, J. Li, and G. X. Yuan, 'The extraction of early warning features for predicting financial distress based on XGBoost model and shap framework', *J. Finan. Eng.*, vol. 08, no. 03, p. 2141004, Sep. 2021, doi: 10.1142/S2424786321410048.
[121] S. A.-D. Safi, P. A. Castillo, and H. Faris, 'Cost-sensitive metaheuristic optimization-based neural network with ensemble learning for financial distress prediction', *Applied Sciences*, vol. 12, no. 14, p. 6918, Jul. 2022, doi: 10.3390/app12146918.
[122] A. Gaurav, B. B. Gupta, S. Bansal, and K. E. Psannis, 'Bankruptcy forecasting in enterprises and its security using hybrid deep learning models', *Cyber Security and Applications*, vol. 3, p. 100070, Dec. 2025, doi: 10.1016/j.csa.2024.100070.
[123] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, 'A survey on addressing high-class imbalance in big data', *J Big Data*, vol. 5, no. 1, p. 42, Dec. 2018, doi: 10.1186/s40537-018-0151-6.
[124] H. Kim, H. Cho, and D. Ryu, 'Corporate Default Predictions Using Machine Learning: Literature Review', *Sustainability*, vol. 12, no. 16, p. 6325, Aug. 2020, doi: 10.3390/su12166325.
[125] M. Peng, E. R. Stern, and H. Hu, 'Forecasting China bond default with severe class-imbalanced data: A simple learning model with causal inference', *Economic Modelling*, vol. 144, p. 106985, Mar. 2025, doi: 10.1016/j.econmod.2024.106985.
[126] J. Zhao, J. Ouenniche, and J. De Smedt, 'Survey, classification and critical analysis of the literature on corporate bankruptcy and financial distress prediction', *Machine Learning with Applications*, vol. 15, p. 100527, Mar. 2024, doi: 10.1016/j.mlwa.2024.100527.
[127] C. Jan, 'Financial information asymmetry: Using deep learning algorithms to predict financial distress', *Symmetry*, vol. 13, no. 3, p. 443, Mar. 2021, doi: 10.3390/sym13030443.
[128] R. Y. Eltayeb, A. E. Karrar, W. I. Osman, and M. M. Ali, 'Handling imbalanced data through re-sampling: Systematic review', *IJEEI*, vol. 11, no. 2, pp. 503–514, Jun. 2023, doi: 10.52549/.v11i2.4471.

[129] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, 'SMOTE: Synthetic Minority Over-sampling Technique', *jair*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.
[130] B. Wang *et al.*, 'Imbalance Data Processing Strategy for Protein Interaction Sites Prediction', *IEEE/ACM Trans. Comput. Biol. and Bioinf.*, vol. 18, no. 3, pp. 985–994, May 2021, doi: 10.1109/tcbb.2019.2953908.
[131] T. Cai, 'Breast Cancer Diagnosis Using Imbalanced Learning and Ensemble Method', *ACM*, vol. 7, no. 3, p. 146, 2018, doi: 10.11648/j.acm.20180703.20.
[132] 'A Novel Distribution Analysis for SMOTE Oversampling Method in Handling Class Imbalance', in *Lecture Notes in Computer Science*, Cham: Springer International Publishing, 2019, pp. 236–248. doi: 10.1007/978-3-030-22744-9_18.
[133] S. A. Alex and J. J. V. Nayahi, 'Classification of Imbalanced Data Using SMOTE and AutoEncoder Based Deep Convolutional Neural Network', *Int. J. Unc. Fuzz. Knowl. Based Syst.*, vol. 31, no. 03, pp. 437–469, Jun. 2023, doi: 10.1142/S0218488523500228.
[134] H. Han, W.-Y. Wang, and B.-H. Mao, 'Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning', in *Advances in Intelligent Computing*, vol. 3644, D.-S. Huang, X.-P. Zhang, and G.-B. Huang, Eds, in Lecture Notes in Computer Science, vol. 3644. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 878–887. doi: 10.1007/11538059_91.
[135] C. Chen, W. Shen, C. Yang, W. Fan, X. Liu, and Y. Li, 'A New Safe-Level Enabled Borderline-SMOTE for Condition Recognition of Imbalanced Dataset', *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023, doi: 10.1109/TIM.2023.3289545.
[136] A. Glazkova, 'A Comparison of Synthetic Oversampling Methods for Multi-class Text Classification', 2020, *arXiv*. doi: 10.48550/ARXIV.2008.04636.
[137] Haibo He, Yang Bai, E. A. Garcia, and Shutao Li, 'ADASYN: Adaptive synthetic sampling approach for imbalanced learning', in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Hong Kong, China: IEEE, Jun. 2008, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969.
[138] Y. Pristyanto, A. F. Nugraha, A. Dahlan, L. A. Wirasakti, A. Ahmad Zein, and I. Pratama, 'Multiclass Imbalanced Handling using ADASYN Oversampling and Stacking Algorithm', in *2022 16th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, Seoul, Korea, Republic of: IEEE, Jan. 2022, pp. 1–5. doi: 10.1109/IMCOM53663.2022.9721632.
[139] M. A. S. Arifin *et al.*, 'Oversampling and undersampling for intrusion detection system in the supervisory control and data acquisition IEC 60870-5-104', *IET Cyber-Phy Sys Theory & Ap*, vol. 9, no. 3, pp. 282–292, Sep. 2024, doi: 10.1049/cps2.12085.
[140] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, 'A study of the behavior of several methods for balancing machine learning training data', *SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, Jun. 2004, doi: 10.1145/1007730.1007735.
[141] M. M. Kaushik, S. M. H. Mahmud, M. A. Kabir, and D. Nandi, 'The effects of class rebalancing techniques on ensemble classifiers on credit card fraud detection: An empirical study', presented at the APPLIED DATA SCIENCE AND SMART SYSTEMS, Rajpura, India, 2023, p. 030011. doi: 10.1063/5.0177524.
[142] J. Zang and H. Li, 'Abnormal Traffic Detection Based on Data Augmentation and Hybrid Neural Network', in *2024 2nd International Conference on Signal Processing and Intelligent Computing (SPIC)*, Guangzhou, China: IEEE, Sep. 2024, pp. 249–253. doi: 10.1109/SPIC62469.2024.10691452.
[143] Q. Leng, J. Guo, J. Tao, X. Meng, and C. Wang, 'OBMI: Oversampling borderline minority instances by a two-stage tomek link-finding procedure for class imbalance problem', *Complex Intell. Syst.*, vol. 10, no. 4, pp. 4775–4792, Aug. 2024, doi: 10.1007/s40747-024-01399-y.
[144] G. Tuysuzoglu *et al.*, 'Joint tomek links (JTL): An innovative approach to noise reduction for enhanced classification performance', *IEEE Access*, vol. 13, pp. 123059–123082, 2025, doi: 10.1109/ACCESS.2025.3580290.
[145] Lalu Ganda Rady Putra, Khairani Marzuki, and Hairani Hairani, 'Correlation-based feature selection and Smote-Tomek Link to improve the performance of machine learning methods on cancer disease prediction', *Engineering and Applied Science Research*, vol. 50, p. 577583, 2023, doi: 10.14456/EASR.2023.59.
[146] F. Yang, K. Wang, L. Sun, M. Zhai, J. Song, and H. Wang, 'A hybrid sampling algorithm combining synthetic minority over-sampling technique and edited nearest neighbor

for missed abortion diagnosis', *BMC Med Inform Decis Mak*, vol. 22, no. 1, p. 344, Dec. 2022, doi: 10.1186/s12911-022-02075-2.
[147] C. Díez López, D. Montiel González, A. Vidaki, and M. Kayser, 'Prediction of Smoking Habits From Class-Imbalanced Saliva Microbiome Data Using Data Augmentation and Machine Learning', *Front. Microbiol.*, vol. 13, p. 886201, Jul. 2022, doi: 10.3389/fmicb.2022.886201.
[148] S. Shaikh, S. M. Daudpota, A. S. Imran, and Z. Kastrati, 'Towards Improved Classification Accuracy on Highly Imbalanced Text Dataset Using Deep Neural Language Models', *Applied Sciences*, vol. 11, no. 2, p. 869, Jan. 2021, doi: 10.3390/app11020869.
[149] M. H. Kotb and R. Ming, 'Comparing SMOTE Family Techniques in Predicting Insurance Premium Defaulting using Machine Learning Models', *IJACSA*, vol. 12, no. 9, 2021, doi: 10.14569/IJACSA.2021.0120970.
[150] P. Peykani, M. Peymany Foroushany, C. Tanasescu, M. Sargolzaei, and H. Kamyabfar, 'Evaluation of cost-sensitive learning models in forecasting business failure of capital market firms', *Mathematics*, vol. 13, no. 3, p. 368, Jan. 2025, doi: 10.3390/math13030368.
[151] C. Elkan, 'The foundations of cost-sensitive learning', in *Proceedings of the 17th International Joint Conference on Artificial Intelligence - Volume 2*, in IJCAI'01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 973–978.
[152] M. Zhu *et al.*, 'Class Weights Random Forest Algorithm for Processing Class Imbalanced Medical Data', *IEEE Access*, vol. 6, pp. 4641–4652, 2018, doi: 10.1109/ACCESS.2018.2789428.
[153] M. Zubair and C. Yoon, 'Cost-Sensitive Learning for Anomaly Detection in Imbalanced ECG Data Using Convolutional Neural Networks', *Sensors*, vol. 22, no. 11, p. 4075, May 2022, doi: 10.3390/s22114075.
[154] W. Priatna, Hindriyanto Dwi Purnomo, Ade Iriani, Irwan Sembiring, and Theophilus Wellem, 'Optimizing Multilayer Perceptron with Cost-Sensitive Learning for Addressing Class Imbalance in Credit Card Fraud Detection', *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 8, no. 4, pp. 563–570, Aug. 2024, doi: 10.29207/resti.v8i4.5917.
[155] C. Chen, 'Using random forest to learn imbalanced data'.
[156] R. O'Brien and H. Ishwaran, 'A random forests quantile classifier for class imbalanced data', *Pattern Recognition*, vol. 90, pp. 232–249, Jun. 2019, doi: 10.1016/j.patcog.2019.01.036.
[157] Z. Liu, P. Jiang, K. W. De Bock, J. Wang, L. Zhang, and X. Niu, 'Extreme gradient boosting trees with efficient Bayesian optimization for profit-driven customer churn prediction', *Technological Forecasting and Social Change*, vol. 198, p. 122945, Jan. 2024, doi: 10.1016/j.techfore.2023.122945.
[158] L. Nevasalmi, 'Recession forecasting with high-dimensional data', *Journal of Forecasting*, vol. 41, no. 4, pp. 752–764, Jul. 2022, doi: 10.1002/for.2823.
[159] B. Van Braak, J. R. Osterrieder, and M. R. Machado, 'How can consumers without credit history benefit from the use of information processing and machine learning tools by financial institutions?', *Information Processing & Management*, vol. 62, no. 2, p. 103972, Mar. 2025, doi: 10.1016/j.ipm.2024.103972.
[160] L. Jiang, Q. Wang, Y. Chang, J. Song, H. Fu, and X. Yang, 'An Adaptive Cost-Sensitive Learning and Recursive Denoising Framework for Imbalanced SVM Classification', 2024, *arXiv*. doi: 10.48550/ARXIV.2403.08378.
[161] H. Qiu, H. Liu, and C. Zhang, 'SCD:Sampling-based Class Distribution for Imbalanced Semi-Supervised Learning', in *2023 IEEE 35th International Conference on Tools with Artificial Intelligence (ICTAI)*, Atlanta, GA, USA: IEEE, Nov. 2023, pp. 567–572. doi: 10.1109/ICTAI59109.2023.00090.
[162] C.-A. Tsai and Y.-J. Chang, 'Efficient Selection of Gaussian Kernel SVM Parameters for Imbalanced Data', *Genes*, vol. 14, no. 3, p. 583, Feb. 2023, doi: 10.3390/genes14030583.
[163] T. Le, M. T. Vo, B. Vo, M. Y. Lee, and S. W. Baik, 'A hybrid approach using oversampling technique and cost-sensitive learning for bankruptcy prediction', *Complexity*, vol. 2019, no. 1, p. 8460934, Jan. 2019, doi: 10.1155/2019/8460934.
[164] R. Gao, S. Cui, Y. Wang, and W. Xu, 'Predicting financial distress in high-dimensional imbalanced datasets: A multi-heterogeneous self-paced ensemble learning framework', *Financ Innov*, vol. 11, no. 1, p. 50, Jan. 2025, doi: 10.1186/s40854-024-00745-w.
[165] I. Araf, A. Idri, and I. Chairi, 'Cost-sensitive learning for imbalanced medical data: a review', *Artif Intell Rev*, vol. 57, no. 4, p. 80, Mar. 2024, doi: 10.1007/s10462-023-10652-8.

[166] Y. Freund and R. E. Schapire, ‘Game theory, on-line prediction and boosting’, in *Proceedings of the ninth annual conference on Computational learning theory - COLT ’96*, Desenzano del Garda, Italy: ACM Press, 1996, pp. 325–332. doi: 10.1145/238061.238163.
[167] F. H. TUNIO, Y. DING, A. N. AGHA, K. AGHA, and H. U. R. Z. PANHWAR, ‘Financial distress prediction using adaboost and bagging in pakistan stock exchange’, *The Journal of Asian Finance, Economics and Business*, vol. 8, no. 1, pp. 665–673, Jan. 2021, doi: 10.13106/JAFEB.2021.VOL8.NO1.665.
[168] M. Pavlicko, M. Durica, and J. Mazanec, ‘Ensemble model of the financial distress prediction in visegrad group countries’, *Mathematics*, vol. 9, no. 16, p. 1886, Aug. 2021, doi: 10.3390/math9161886.
[169] R. Liu, H. Li, K. Yoon, and M. A. Vasarhelyi, ‘Using machine learning algorithms to improve fiscal distress prediction models: The case of U.S. local governments’, *Journal of Information Systems*, vol. 39, no. 3, pp. 131–155, Nov. 2025, doi: 10.2308/ISYS-2023-049.
[170] A. A. Khalil, Z. Liu, A. Salah, A. Fathalla, and A. Ali, ‘Predicting insolvency of insurance companies in egyptian market using bagging and boosting ensemble techniques’, *IEEE Access*, vol. 10, pp. 117304–117314, 2022, doi: 10.1109/ACCESS.2022.3210032.
[171] Y. Zou, Y. Yuan, C. Zhu, and C. Yu, ‘Symmetric equilibrium bagging–cascading boosting ensemble for financial risk early warning’, *Symmetry*, vol. 17, no. 10, p. 1779, Oct. 2025, doi: 10.3390/sym17101779.
[172] W. Wang and Z. Liang, ‘Financial distress early warning for Chinese enterprises from a systemic risk perspective: Based on the adaptive weighted XGBoost-bagging model’, *Systems*, vol. 12, no. 2, p. 65, Feb. 2024, doi: 10.3390/systems12020065.
[173] W. Liu, Y. Suzuki, and S. Du, ‘Ensemble learning algorithms based on easyensemble sampling for financial distress prediction’, *Ann Oper Res*, vol. 346, no. 3, pp. 2141–2172, Mar. 2025, doi: 10.1007/s10479-025-06494-y.
[174] S. Vasheghani and S. Sharifi, ‘Adaptive dynamic ensemble learning with class-specific model selection for efficient and robust image classification’, *Knowledge-Based Systems*, vol. 331, p. 114842, Jan. 2026, doi: 10.1016/j.knosys.2025.114842.
[175] J. Sun, H. Li, Q.-H. Huang, and K.-Y. He, ‘Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modeling, sampling, and featuring approaches’, *Knowledge-Based Systems*, vol. 57, pp. 41–56, Feb. 2014, doi: 10.1016/j.knosys.2013.12.006.
[176] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, ‘Boosting methods for multi-class imbalanced data classification: an experimental review’, *J Big Data*, vol. 7, no. 1, p. 70, Dec. 2020, doi: 10.1186/s40537-020-00349-y.
[177] Z. Yi, ‘Credit Default Risk Measurement and Statistical Analysis Based on Improved GRU Model’, *Engineering Reports*, vol. 7, no. 2, p. e70014, Feb. 2025, doi: 10.1002/eng2.70014.
[178] J.-B. Wang, C.-A. Zou, and G.-H. Fu, ‘AWSMOTE: An SVM-Based Adaptive Weighted SMOTE for Class-Imbalance Learning’, *Scientific Programming*, vol. 2021, pp. 1–18, May 2021, doi: 10.1155/2021/9947621.
[179] H. Xu and C. Chetia, ‘An Efficient Selective Ensemble Learning with Rejection Approach for Classification’, in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, Birmingham United Kingdom: ACM, Oct. 2023, pp. 2816–2825. doi: 10.1145/3583780.3614780.
[180] X. Li, ‘Adaptive weighting and deep neural networks for automated multi-indicator financial statement analysis and risk prediction’, *IJCAI*, vol. 49, no. 6, Aug. 2025, doi: 10.31449/inf.v49i6.9008.
[181] J. Wang, ‘Markov network classification for imbalanced data with adaptive weighting’, Jan. 2025, doi: 10.5281/ZENODO.14936206.
[182] J. Huang, Y. Wang, and M. Wang, ‘Class-adaptive weighted broad learning system with hybrid memory retention for online imbalanced classification’, *Electronics*, vol. 14, no. 17, p. 3562, Sep. 2025, doi: 10.3390/electronics14173562.
[183] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, ‘Focal loss for dense object detection’, in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Oct. 2017, pp. 2999–3007. doi: 10.1109/ICCV.2017.324.
[184] A. Shrivastava, A. Gupta, and R. Girshick, ‘Training Region-Based Object Detectors with Online Hard Example Mining’, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 761–769. doi: 10.1109/CVPR.2016.89.

[185] W. Chen, K. Yang, Z. Yu, Y. Shi, and C. L. P. Chen, 'A survey on imbalanced learning: latest research, applications and future directions', *Artif Intell Rev*, vol. 57, no. 6, p. 137, May 2024, doi: 10.1007/s10462-024-10759-6.
[186] B. H. Aubaidan *et al.*, 'A review of intelligent data analysis: Machine learning approaches for addressing class imbalance in healthcare - challenges and perspectives', *Intelligent Data Analysis: An International Journal*, vol. 29, no. 3, pp. 699–719, May 2025, doi: 10.1177/1088467X241305509.
[187] Y. Liu *et al.*, 'Imbalanced data classification: Using transfer learning and active sampling', *Engineering Applications of Artificial Intelligence*, vol. 117, p. 105621, Jan. 2023, doi: 10.1016/j.engappai.2022.105621.
[188] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, 'Learning from class-imbalanced data: Review of methods and applications', *Expert Systems with Applications*, vol. 73, pp. 220–239, May 2017, doi: 10.1016/j.eswa.2016.12.035.
[189] P. Gnip, R. Kanász, M. Zoričak, and P. Drotár, 'An experimental survey of imbalanced learning algorithms for bankruptcy prediction', *Artif Intell Rev*, vol. 58, no. 4, p. 104, Jan. 2025, doi: 10.1007/s10462-025-11107-y.
[190] N.-G. B. Abrahamsen, E. Nylén-Forthun, M. Møller, P. E. De Lange, and M. Risstad, 'Financial Distress Prediction in the Nordics: Early Warnings from Machine Learning Models', *JRFM*, vol. 17, no. 10, p. 432, Sep. 2024, doi: 10.3390/jrfm17100432.
[191] S. B. Jabeur, C. Gharib, S. Mefteh-Wali, and W. B. Arfi, 'CatBoost model and artificial intelligence techniques for corporate failure prediction', *Technological Forecasting and Social Change*, vol. 166, p. 120658, May 2021, doi: 10.1016/j.techfore.2021.120658.
[192] H. A. Salman, A. Kalakech, and A. Steiti, 'Random Forest Algorithm Overview', *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, Jun. 2024, doi: 10.58496/BJML/2024/007.
[193] H. Aljawazneh, S. G. Yaseen, and Q. Al-Shayea, 'Bagging vs Boosting Ensemble Classifiers in Predicting Companies' Financial Status', in *Cutting-Edge Business Technologies in the Big Data Era*, vol. 135, S. G. Yaseen, Ed., in Studies in Big Data, vol. 135. , Cham: Springer Nature Switzerland, 2023, pp. 1–9. doi: 10.1007/978-3-031-42463-2_1.
[194] N. H. A. Malek, W. F. W. Yaacob, Y. B. Wah, S. A. Md Nasir, N. Shaadan, and S. W. Indratno, 'Comparison of ensemble hybrid sampling with bagging and boosting machine learning approach for imbalanced data', *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 1, p. 598, Jan. 2022, doi: 10.11591/ijeecs.v29.i1.pp598-608.
[195] G. Hou, D. L. Tong, S. Y. Liew, and P. Y. Choo, 'Comparative Analysis of Resampling Techniques for Class Imbalance in Financial Distress Prediction Using XGBoost', *Mathematics*, vol. 13, no. 13, p. 2186, Jul. 2025, doi: 10.3390/math13132186.
[196] R. Gao, S. Cui, Y. Wang, and W. Xu, 'Predicting financial distress in high-dimensional imbalanced datasets: a multi-heterogeneous self-paced ensemble learning framework', *Financ Innov*, vol. 11, no. 1, p. 50, Jan. 2025, doi: 10.1186/s40854-024-00745-w.
[197] C. P. Winsor, 'The Gompertz Curve as a Growth Curve', *Proc. Natl. Acad. Sci. U.S.A.*, vol. 18, no. 1, pp. 1–8, Jan. 1932, doi: 10.1073/pnas.18.1.1.
[198] K. K. Nicodemus, J. D. Malley, C. Strobl, and A. Ziegler, 'The behaviour of random forest permutation-based variable importance measures under predictor correlation', *BMC Bioinformatics*, vol. 11, no. 1, p. 110, Dec. 2010, doi: 10.1186/1471-2105-11-110.
[199] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, 'Focal Loss for Dense Object Detection', in *In Proceedings of the IEEE International Conference on Computer Vision*, Venice,Italy, Oct. 2017, pp. 2980–2988.
[200] J. L. Fleiss and J. Cohen, 'The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability', *Educational and Psychological Measurement*, vol. 33, no. 3, pp. 613–619, Oct. 1973, doi: 10.1177/001316447303300309.
[201] D. Rey and M. Neuhäuser, 'Wilcoxon-Signed-Rank Test', in *International Encyclopedia of Statistical Science*, M. Lovric, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1658–1659. doi: 10.1007/978-3-642-04898-2_616.
[202] C. Sánchez-Almeyda, J. González-Bueno, V. Koval, N. Reyes-Maldonado, H. Kryshtal, and O. Zharikova, 'Predicting financial distress in the food production sector: A dual-model approach using Z-score and O-score methods', *Discov Sustain*, vol. 6, no. 1, p. 731, Jul. 2025, doi: 10.1007/s43621-025-01348-w.
[203] N. W. Santoso, R. Kusumawardhani, and A. Maulida, 'Comparative analysis of the altman, ohlson, and zmijewski models to predict financial distress during the covid-19 pandemic', *MAKSIMUM*, vol. 14, no. 1, p. 13, Mar. 2024, doi: 10.26714/mki.14.1.2024.13-21.

[204] M. Modina and S. Zedda, ‘Diagnosing default syndromes: Early symptoms of entrepreneurial venture insolvency’, *JSBED*, vol. 30, no. 1, pp. 186–209, Feb. 2023, doi: 10.1108/JSBED-02-2022-0088.
[205] K. Mavengere and P. Gumede, ‘Financial distress prediction competence of the altman Z score and zmijewski model: Evidence from selected zimbabwe stock exchange firms’, *Pressacademia*, p. 1, Jun. 2024, doi: 10.17261/Pressacademia.2024.1892.
[206] Z. Saadaoui and S. Mokdadi, ‘Capital buffers, business models and the probability of bank distress: A dynamic panel investigation’, *JFRC*, vol. 31, no. 5, pp. 663–695, Oct. 2023, doi: 10.1108/JFRC-10-2022-0119.
[207] J. Bergstra and Y. Bengio, ‘Random search for hyper-parameter optimization’, *Journal of Machine Learning Research*, vol. 13, no. null, pp. 281–305, Feb. 2012.

Appendix A Dataset Related Information

Table A.1 Feature Names and Corresponding Indicators

var	Label name	var	Label name
var1	Accounts Receivable	var185	Operating Expense Ratio - TTM
var2	Notes Receivable	var186	Operating Profit Ratio
var3	Prepayments	var187	Operating Profit Ratio - TTM
var4	Total Assets	var188	Operating Profit Margin
var5	Accounts Payable	var189	Operating Profit Margin - TTM
var6	Notes Payable	var190	Total Operating Expenses Ratio
var7	Total Liabilities	var191	Total Operating Expenses Ratio - TTM
var8	Total Operating Revenue	var192	Selling Expense Ratio
var9	Net Profit	var193	Selling Expense Ratio - TTM
var10	Bank Loan Ratio	var194	Administrative Expense Ratio
var11	Short-term Loans Dependence	var195	Administrative Expense Ratio - TTM
var12	Debt to Assets Ratio	var196	Financial Expense Ratio
var13	Net Cash Flow from Operating Activities / Total Liabilities	var197	Financial Expense Ratio - TTM
var14	Net Cash Flow from Financing Activities of Creditors	var198	Selling Period Expense Ratio
var15	Net Cash Flow from Financing Activities of Shareholders	var199	Selling Period Expense Ratio - TTM
var16	Sustainable Growth Rate	var200	Ratio of Profit to Cost
var17	Earnings Volatility	var201	Ratio of Profit to Cost - TTM
var18	Z-score	var202	Asset Impairment Loss / Operating Revenue
var19	O-score	var203	Asset Impairment Loss / Operating Revenue - TTM
var20	Current Ratio	var204	EBITDA Margin
var21	Quick Ratio	var205	EBITDA Margin - TTM
var22	Conservative Quick Ratio	var206	Operating Margin before Interest and Taxes
var23	Cash Ratio	var207	Operating Margin before Interest and Taxes - TTM
var24	Ratio of Working Capital to Loans	var208	Ratio of Cash to Total Profit
var25	Working Capital	var209	Ratio of Cash to Total Profit - TTM
var26	Net Cash Flow from Operating Activities / Current Liabilities	var210	Return on Equity Attributable to the Parent Company - A
var27	Cash Flow-based Matured Liabilities Coverage Ratio	var211	Return on Equity Attributable to the Parent Company - B
var28	Liabilities to Assets Ratio	var212	Return on Equity Attributable to the Parent Company - C
var29	Ratio of Long-term Loans to Total Assets	var213	Return on Equity Attributable to the Parent Company - TTM
var30	Ratio of Liabilities to Tangible Assets	var214	Comprehensive Return Attributable to the Parent Company - A
var31	Ratio of Interest-bearing Liabilities to Tangible Assets	var215	Comprehensive Return Attributable to the Parent Company - B
var32	Equity Multiplier	var216	Comprehensive Return Attributable to the Parent Company - C
var33	Equity Multiplier 2	var217	Comprehensive Return Attributable to the Parent Company - TTM
var34	Ratio of Debt to Equity	var218	Return on Investment
var35	Ratio of Equity to Debt	var219	Net Profit Cash Coverage
var36	Ratio of Debt to Long-term Capital	var220	Net Profit Cash Coverage - TTM
var37	Ratio of Long-term Debt to Equity	var221	Operating Revenue Cash Coverage
var38	Ratio of Long-term Debt to Working Capital	var222	Operating Revenue Cash Coverage - TTM
var39	EBITDA / Total Liabilities	var223	Net Operating Revenue Cash Coverage

var40	Net Cash Flow from Operating Activities / Total Liabilities	var224	Net Operating Revenue Cash Coverage - TTM
var41	Net Cash Flow from Operating Activities / Interest-bearing Liabilities	var225	Net Operating Profit Cash Coverage
var42	Liabilities to Equity Market Cap Ratio	var226	Net Operating Profit Cash Coverage - TTM
var43	Ratio of Total Liabilities to Net Tangible Assets	var227	Net Cash Flow from Financing Activities Related to Creditors
var44	Extraordinary Items	var228	Net Cash Flow from Financing Activities Related to Creditors - TTM
var45	Net Profit Attributable to Shareholders of Listed Company after Deducting Extrao	var229	Net Cash Flow from Financing Activities Related to Shareholders
var46	Weighted Average Return on Equity	var230	Net Cash Flow from Financing Activities Related to Shareholders - TTM
var47	Basic Earnings per Share after Deducting Extraordinary Items	var231	Corporate Cash Flow 1
var48	Basic Earnings per Share	var232	Corporate Cash Flow 2
var49	Diluted Earnings per Share	var233	Corporate Cash Flow TTM1
var50	Ratio of Current Assets	var234	Corporate Cash Flow TTM2
var51	Ratio of Cash Assets	var235	Equity Cash Flow 1
var52	Ratio of Receivable Assets	var236	Equity Cash Flow 2
var53	Ratio of Working Capital to Current Assets	var237	Equity Cash Flow TTM1
var54	Ratio of Working Capital	var238	Equity Cash Flow TTM2
var55	Ratio of Working Capital to Net Assets	var239	Free Cash Flow to Firm (Original)
var56	Ratio of Non-current Assets	var240	Free Cash Flow to Equity (Original)
var57	Ratio of Fixed Assets	var241	Recovery Rate for All Cash
var58	Ratio of Intangible assets	var242	Operating Index
var59	Tangible Assets Ratio	var243	Cash Flow Adequacy Ratio
var60	Ratio of Owners' Equity	var244	Cash Re-investment Ratio
var61	Ratio of Retained Earnings to Total Assets	var245	Cash Flow Adequacy Ratio (Past Five Years)
var62	Confirmity Rate of Long-term Assets	var246	Free Cash Flow to Equity
var63	Ratio of Shareholders' Equity to Fixed Assets	var247	Free Cash Flow to Firm
var64	Ratio of Current Liabilities	var248	Financial Leverage
var65	Ratio of Operating liabilities	var249	Operating Leverage
var66	Ratio of financial liabilities	var250	Total Leverage
var67	Ratio of Non-current Liabilities	var251	Capital Preservation & Appreciation Rate - A
var68	Ratio of Owners' Equity Attributable to the Parent Company	var252	Capital Preservation & Appreciation Rate - B
var69	Ratio of Minority Interests	var253	Capital Preservation & Appreciation Rate of the Parent Company
var70	Ratio of Profit from Main Business	var254	Capital Accumulation Rate - A
var71	Ratio of Profit from Financial Activities	var255	Capital Accumulation Rate - B
var72	Ratio of Operating Profit	var256	Capital Accumulation Rate of the Parent Company
var73	Ratio of Non-operating Income	var257	Growth Rate of Fixed Assets - A
var74	Turnover Tax Rate	var258	Growth Rate of Fixed Assets - B
var75	Composite Tax Rate - A	var259	Growth Rate of Total Assets - A
var76	Composite Tax Rate - B	var260	Growth Rate of Total Assets - B
var77	Income Tax Rate	var261	Growth Rate of Return on Equity - A
var78	Ratio of Net Profit Attributable to the Parent Company	var262	Growth Rate of Return on Equity - B
var79	Ratio of Minority Interests to Net Profit	var263	Growth Rate of Basic Earnings per Share - A
var80	Ratio of Net Profit to Comprehensive Income	var264	Growth Rate of Basic Earnings per Share - B
var81	Ratio of Other Comprehensive Income	var265	Growth Rate of Diluted Earnings per Share - A
var82	Ratio of Comprehensive Income Attributable to the Parent Company	var266	Growth Rate of Diluted Earnings per Share - B

var83	Ratio of Comprehensive Income Attributable to Minority shareholders	var267	Growth Rate of Net Profit - A
var84	Ratio of Owners' Equity Attributable to the Parent Company to Invested Capital	var268	Growth Rate of Net Profit - B
var85	Ratio of Accounts Receivable to Income	var269	Growth Rate of Total Profit - A
var86	Accounts Receivable Turnover - A	var270	Growth Rate of Total Profit - B
var87	Accounts Receivable Turnover - B	var271	Growth Rate of Operating Profit - A
var88	Accounts Receivable Turnover - C	var272	Growth Rate of Operating Profit - B
var89	Accounts Receivable Turnover - D	var273	Growth Rate of Net Profit Attributable to the Parent Company
var90	Accounts Receivable Turnover - TTM	var274	Growth Rate of Operating Revenue - A
var91	Accounts Receivable Turnover Days - A	var275	Growth Rate of Operating Revenue - B
var92	Accounts Receivable Turnover Days - B	var276	Growth Rate of Total Operating Revenue
var93	Accounts Receivable Turnover Days - C	var277	Growth Rate of Total Operating Cost
var94	Accounts Receivable Turnover Days - TTM	var278	Growth Rate of Selling Expenses
var95	Ratio of Inventories to Income	var279	Growth Rate of Administrative Expenses
var96	Inventories Turnover - A	var280	Accrued Item
var97	Inventories Turnover - B	var281	Sustainable Growth Rate
var98	Inventories Turnover - C	var282	Growth Rate of Owners' Equity - A
var99	Inventories Turnover - D	var283	Growth Rate of Owners' Equity - B
var100	Inventories Turnover - TTM	var284	Growth Rate of Net Assets per Share - B
var101	Inventories Turnover Days - A	var285	Growth Rate of Net Assets per Share - A
var102	Inventories Turnover Days - B	var286	Sustainable Growth Rate 3
var103	Inventories Turnover Days - C	var287	Earnings per Share 1
var104	Inventories Turnover Days - TTM	var288	Earnings per Share - TTM1
var105	Business Cycle - A	var289	Earnings per Share 2
var106	Business Cycle - B	var290	Earnings per Share - TTM2
var107	Business Cycle - C	var291	Earnings per Share 3
var108	Business Cycle - TTM	var292	Earnings per Share - TTM3
var109	Accounts Payable Turnover - A	var293	Earnings per Share 4
var110	Accounts Payable Turnover - B	var294	Earnings per Share - TTM4
var111	Accounts Payable Turnover - C	var295	Comprehensive Income Per Share 1
var112	Accounts Payable Turnover - D	var296	Comprehensive Income per Share - TTM1
var113	Accounts Payable Turnover - TTM	var297	Comprehensive Income Per Share 2
var114	Working Capital Turnover - A	var298	Comprehensive Income per Share - TTM2
var115	Working Capital Turnover - B	var299	Earnings Per Share Attributable to the Parent Company
var116	Working Capital Turnover - C	var300	Earnings per Share Attributable to the Parent Company - TTM
var117	Working Capital Turnover - D	var301	Comprehensive Income per Share Attributable to the Parent Company
var118	Working Capital Turnover - TTM	var302	Comprehensive Income per Share Attributable to the Parent Company - TTM
var119	Cash and Cash Equivalents Turnover - A	var303	Total Operating Revenue per Share
var120	Cash and Cash Equivalents Turnover - B	var304	Total Operating Revenue per Share - TTM
var121	Cash and Cash Equivalents Turnover - C	var305	Operating Revenue per Share
var122	Cash and Cash Equivalents Turnover - D	var306	Operating Revenue per Share - TTM
var123	Cash and Cash Equivalents Turnover - TTM	var307	Earnings Per Share before Interest and Taxes
var124	Ratio of Current Assets to Income	var308	Earnings per Share before Interest and Taxes - TTM
var125	Current Assets Turnover - A	var309	EBITDA per Share
var126	Current Assets Turnover - B	var310	EBITDA per Share - TTM
var127	Current Assets Turnover - C	var311	Operating Profit per Share
var128	Current Assets Turnover - D	var312	Operating Profit per Share - TTM

var129	Current Assets Turnover - TTM	var313	Net Assets per Share
var130	Ratio of Fixed Assets to Income	var314	Tangible Assets per Share
var131	Fixed Assets Turnover - A	var315	Liabilities per Share
var132	Fixed Assets Turnover - B	var316	Capital Reserves per Share
var133	Fixed Assets Turnover - C	var317	Surplus Reserves per Share
var134	Fixed Assets Turnover - D	var318	Undistributed Profit per Share
var135	Fixed Assets Turnover - TTM	var319	Retained Earnings per Share
var136	Non-current Assets Turnover - A	var320	Net Assets per Share Attributable to the Parent Company
var137	Non-current Assets Turnover - B	var321	Net Cash Flow from Operating Activities per Share
var138	Non-current Assets Turnover - C	var322	Net Cash Flow from Operating Activities per Share - TTM
var139	Non-current Assets Turnover - D	var323	Net Cash Flow from Investing Activities per Share
var140	Non-current Assets Turnover - TTM	var324	Net Cash Flow from Investing Activities per Share - TTM
var141	Capital Intensity	var325	Net Cash Flow from Financing Activities per Share
var142	Total Assets Turnover - A	var326	Net Cash Flow from Financing Activities per Share - TTM
var143	Total Assets Turnover - B	var327	Free Cash Flow to Firm per Share
var144	Total Assets Turnover - C	var328	Free Cash Flow to Firm per Share - TTM
var145	Total Assets Turnover - D	var329	Free Cash Flow to Shareholder per Share
var146	Total Assets Turnover - TTM	var330	Free Cash Flow to Shareholder per Share - TTM
var147	Shareholders' Equity Turnover - A	var331	Depreciation and Amortization per Share
var148	Shareholders' Equity Turnover - B	var332	Depreciation and Amortization per Share - TTM
var149	Shareholders' Equity Turnover - C	var333	Free Cash Flow to Firm per Share (Original)
var150	Shareholders' Equity Turnover - D	var334	Free Cash Flow to Equity per Share (Original)
var151	Shareholders' Equity Turnover - TTM	var335	Net Cash Flow per Share 1
var152	Return on Total Assets - A	var336	Net Cash Flow per Share - TTM1
var153	Return on Total Assets - B	var337	Net Cash Flow per Share 2
var154	Return on Total Assets - C	var338	Net Cash Flow per Share - TTM2
var155	Return on Total Assets - TTM	var339	P/E Ratio 1
var156	Return on Assets - A	var340	P/E Ratio 2
var157	Return on Assets - B	var341	P/E Ratio - TTM
var158	Return on Assets - C	var342	P/S Ratio 1
var159	Return on Assets - TTM	var343	P/S Ratio 2
var160	Ratio of Net Profit to Current Assets - A	var344	P/S Ratio - TTM
var161	Ratio of Net Profit to Current Assets - B	var345	Price-to-Cash Flow Ratio 1
var162	Ratio of Net Profit to Current Assets - C	var346	P/B Ratio
var163	Ratio of Net Profit to Current Assets - TTM	var347	Ratio of Market Value to Tangible Assets
var164	Ratio of Net Profit to Fixed Assets - A	var348	Price to Earnings Attributable to the Parent Company 1
var165	Ratio of Net Profit to Fixed Assets - B	var349	Price to Earnings Attributable to the Parent Company 2
var166	Ratio of Net Profit to Fixed Assets - C	var350	Price to Earnings Attributable to the Parent Company TTM
var167	Ratio of Net Profit to Fixed Assets - TTM	var351	P/B Ratio Based on Owner's Equity Attributable to the Parent Company
var168	Return on Equity - A	var352	Market Value - A
var169	Return on Equity - B	var353	Market Value - B
var170	Return on Equity - C	var354	Tobin's Q - A
var171	Return on Equity - TTM	var355	Tobin's Q - B
var172	EBIT	var356	Tobin's Q - C

var173	EBIT - TTM	var357	Tobin’s Q - D
var174	EBIAT	var358	Book-to-Market Ratio - A
var175	EBITDA	var359	Book-to-Market Ratio - B
var176	EBITDA -TTM	var360	Common Stock Dividend Yield - A
var177	Ratio of Net Profit to Total Profit	var361	Common Stock Dividend Yield - B
var178	Ratio of Total Profit to EBIT	var362	EV/EBITDA
var179	Ratio of EBIT to Total Assets	var363	EV/EBITDA - TTM
var180	ROIC	var364	Dividend Distribution Ratio
var181	Return on Long-term Capital	var365	Dividend Distribution Ratio 2
var182	Gross Operating Margin	var366	Change in Dividend per Share
var183	Gross Operating Margin -TTM	var367	Dividend Distribution Ratio 3
var184	Operating Expense Ratio	var368	Earning Retention Ratio

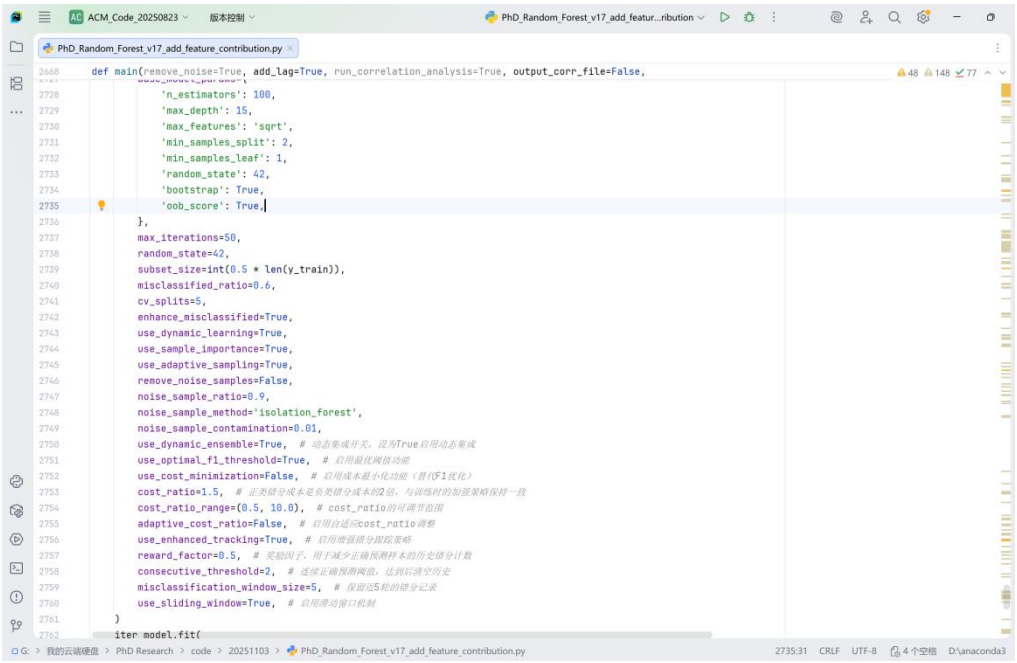

Figure A.0.1 DWARFB Parameters

Appendix B Experimental Results

This appendix contains relevant tables to be used for supporting our findings reported in Chapter 5.

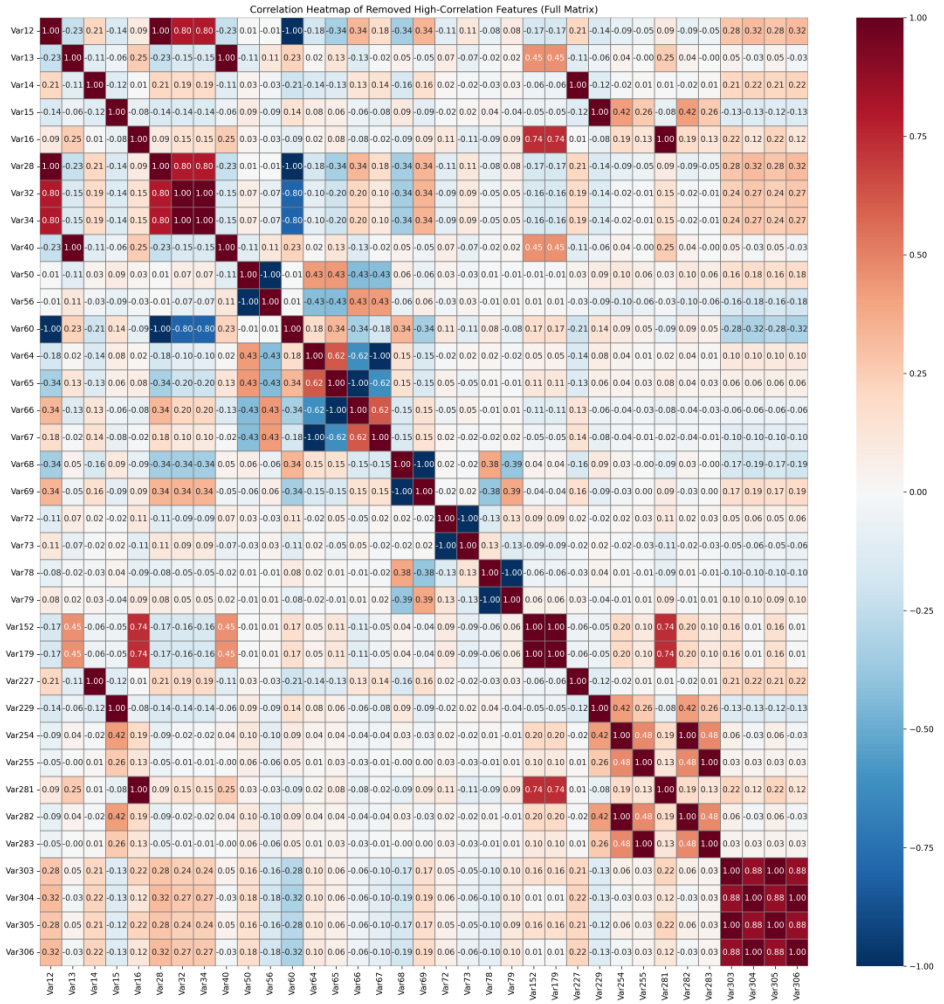

Figure B.0.1 Correlation Heatmap of Highly Correlated Features

Table B.1 Results after Winsorization (Outlier Treatment)

Variables	Num capped low	Low threshold	Num capped high	High threshold
Var1	366	1.87E+04	370	1.36E+10
Var2	1	0.00E+00	370	2.97E+09
Var3	370	4.99E+04	370	5.37E+09
Var4	370	1.96E+08	370	1.82E+11
Var5	0	0.00E+00	370	1.93E+10
Var6	0	0.00E+00	370	8.85E+09
Var7	370	3.97E+07	370	1.14E+11
Var8	370	1.79E+07	370	8.23E+10
Var9	370	4.61E+05	370	4.03E+09
Var10	0	0.00E+00	370	6.24E-01
Var11	0	0.00E+00	370	4.93E-01
Var12	370	5.99E-02	370	9.34E-01
Var13	370	-5.71E-01	370	1.07E+00
Var14	370	-2.20E+09	370	6.20E+09
Var15	370	-2.10E+09	370	2.29E+09
Var16	370	-1.77E-02	370	4.04E-01
Var17	0	0.00E+00	370	2.41E-01
Var18	370	-1.43E-01	370	3.64E+01
Var19	370	-1.23E+01	73	0.00E+00
Var20	370	3.10E-01	370	1.44E+01
Var21	370	1.91E-01	370	1.33E+01
Var22	370	8.73E-02	370	1.10E+01
Var23	370	1.28E-02	370	7.97E+00
Var24	370	-2.64E+00	370	3.73E+02
Var25	370	-9.73E+09	370	2.47E+10
Var26	370	-7.63E-01	370	1.33E+00
Var27	370	-3.42E+01	370	1.21E+02
Var28	370	5.99E-02	370	9.34E-01
Var29	0	0.00E+00	370	3.96E-01
Var30	370	6.33E-02	370	1.22E+00
Var31	5	0.00E+00	370	8.38E-01
Var32	370	1.05E+00	370	9.60E+00
Var33	370	1.05E+00	370	9.73E+00
Var34	370	4.90E-02	370	8.60E+00
Var35	370	7.05E-02	370	1.57E+01
Var36	152	0.00E+00	370	7.15E-01
Var37	199	0.00E+00	370	2.33E+00
Var38	370	-1.62E+01	370	1.47E+01
Var39	370	3.89E-03	370	1.39E+00
Var40	370	-5.71E-01	370	1.07E+00
Var41	370	-1.08E+01	370	4.19E+01
Var42	370	1.54E-02	370	8.39E-01

Variables	Num capped low	Low threshold	Num capped high	High threshold
Var43	0	0.00E+00	370	1.19E+01
Var44	370	-3.61E+07	370	7.07E+08
Var45	370	-1.85E+08	370	3.14E+09
Var46	341	0.00E+00	370	3.85E+01
Var47	370	-2.50E-01	366	1.82E+00
Var48	369	3.00E-04	370	1.96E+00
Var49	269	0.00E+00	370	1.95E+00
Var50	370	1.01E-01	370	9.66E-01
Var51	370	6.05E-03	370	6.45E-01
Var52	370	6.32E-05	370	5.05E-01
Var53	370	-2.21E+00	370	9.31E-01
Var54	370	-4.07E-01	370	7.96E-01
Var55	370	-2.16E+00	370	1.89E+00
Var56	370	3.35E-02	370	8.99E-01
Var57	370	9.22E-04	370	6.77E-01
Var58	0	0.00E+00	370	3.70E-01
Var59	370	4.92E-01	0	1.00E+00
Var60	370	6.57E-02	370	9.40E-01
Var61	370	-2.53E+00	370	5.86E-01
Var62	370	4.85E-01	370	1.35E+02
Var63	370	2.31E-01	370	4.74E+02
Var64	370	2.40E-01	37	1.00E+00
Var65	370	7.82E-02	5	1.00E+00
Var66	5	0.00E+00	370	9.22E-01
Var67	36	0.00E+00	370	7.60E-01
Var68	370	4.94E-01	370	1.02E+00
Var69	370	-2.23E-02	370	5.06E-01
Var70	370	5.64E-03	370	4.33E+01
Var71	370	-6.34E-01	370	4.57E+00
Var72	370	-2.65E+00	370	1.41E+00
Var73	370	-4.07E-01	370	3.65E+00
Var74	370	5.91E-05	370	1.50E-01
Var75	370	-1.58E-03	370	2.58E-01
Var76	370	-6.79E-02	370	2.89E+00
Var77	370	-3.48E-01	370	7.96E-01
Var78	370	1.60E-01	370	2.28E+00
Var79	370	-1.28E+00	370	8.40E-01
Var80	370	-3.99E-01	370	2.14E+00
Var81	370	-1.13E+00	370	1.39E+00
Var82	286	0.00E+00	370	2.14E+00
Var83	370	-1.14E+00	370	8.83E-01
Var84	370	8.56E-02	157	1.00E+00
Var85	370	4.25E-05	370	3.85E+00
Var86	370	1.13E-01	370	6.72E+02

Variables	Num capped low	Low threshold	Num capped high	High threshold
Var87	370	1.85E-01	370	6.03E+02
Var88	370	2.01E-01	370	5.61E+02
Var89	370	6.35E-01	370	9.55E+02
Var90	18	0.00E+00	370	8.73E+02
Var91	370	1.02E-02	370	4.74E+02
Var92	370	1.38E-01	370	4.71E+02
Var93	370	2.04E-01	370	4.78E+02
Var94	0	0.00E+00	370	3.98E+02
Var95	0	0.00E+00	370	1.45E+01
Var96	1	0.00E+00	370	1.83E+02
Var97	1	0.00E+00	370	1.97E+02
Var98	1	0.00E+00	370	1.97E+02
Var99	1	0.00E+00	370	2.90E+02
Var100	24	0.00E+00	370	2.85E+02
Var101	0	0.00E+00	370	4.47E+03
Var102	0	0.00E+00	370	4.23E+03
Var103	0	0.00E+00	370	4.21E+03
Var104	0	0.00E+00	370	3.08E+03
Var105	370	7.47E+00	370	4.52E+03
Var106	370	8.41E+00	370	4.35E+03
Var107	370	8.56E+00	370	4.33E+03
Var108	0	0.00E+00	370	3.21E+03
Var109	1	0.00E+00	370	5.95E+01
Var110	1	0.00E+00	370	5.50E+01
Var111	370	1.61E-01	370	5.89E+01
Var112	370	4.20E-01	370	8.85E+01
Var113	26	0.00E+00	370	9.47E+01
Var114	0	0.00E+00	370	5.37E+01
Var115	0	0.00E+00	370	5.73E+01
Var116	0	0.00E+00	370	5.70E+01
Var117	0	0.00E+00	370	8.99E+01
Var118	12	0.00E+00	370	8.70E+01
Var119	370	7.52E-02	370	4.47E+01
Var120	370	7.37E-02	370	3.65E+01
Var121	370	9.29E-02	370	3.62E+01
Var122	370	2.26E-01	370	5.52E+01
Var123	19	0.00E+00	370	5.80E+01
Var124	370	2.77E-01	370	3.30E+01
Var125	370	3.00E-02	370	3.60E+00
Var126	370	3.01E-02	370	3.69E+00
Var127	370	3.04E-02	370	3.75E+00
Var128	370	7.38E-02	370	4.90E+00
Var129	19	0.00E+00	370	4.93E+00
Var130	370	4.14E-03	370	8.85E+00

Variables	Num capped low	Low threshold	Num capped high	High threshold
Var131	370	1.10E-01	370	2.16E+02
Var132	370	1.10E-01	370	2.02E+02
Var133	370	1.10E-01	370	2.02E+02
Var134	370	2.55E-01	370	3.57E+02
Var135	18	0.00E+00	370	3.61E+02
Var136	370	2.88E-02	370	1.46E+01
Var137	370	2.92E-02	370	1.47E+01
Var138	370	2.92E-02	370	1.49E+01
Var139	370	7.11E-02	370	2.29E+01
Var140	19	0.00E+00	370	2.32E+01
Var141	370	5.28E-01	370	5.87E+01
Var142	370	1.68E-02	370	1.89E+00
Var143	370	1.66E-02	370	1.99E+00
Var144	370	1.66E-02	370	2.04E+00
Var145	370	4.08E-02	370	2.71E+00
Var146	19	0.00E+00	370	2.82E+00
Var147	370	1.36E-02	370	7.36E+00
Var148	370	8.55E-03	370	7.77E+00
Var149	370	8.55E-03	370	7.93E+00
Var150	370	1.93E-02	370	1.09E+01
Var151	13	0.00E+00	370	1.16E+01
Var152	369	1.08E-03	370	2.09E-01
Var153	369	1.09E-03	370	2.24E-01
Var154	369	1.09E-03	370	2.28E-01
Var155	370	-1.36E-01	370	3.32E-01
Var156	370	2.62E-04	370	1.71E-01
Var157	370	2.63E-04	370	1.85E-01
Var158	369	2.66E-04	370	1.89E-01
Var159	370	-1.52E-01	370	2.74E-01
Var160	370	4.60E-04	370	5.18E-01
Var161	370	4.67E-04	370	5.29E-01
Var162	370	4.71E-04	370	5.33E-01
Var163	370	-3.50E-01	370	7.73E-01
Var164	370	1.02E-03	370	2.79E+01
Var165	370	1.02E-03	370	2.47E+01
Var166	370	1.02E-03	370	2.43E+01
Var167	370	-1.65E+00	370	4.22E+01
Var168	370	2.31E-04	370	3.29E-01
Var169	369	9.50E-05	370	3.58E-01
Var170	370	8.96E-05	370	3.73E-01
Var171	370	-3.36E-01	370	5.37E-01
Var172	370	9.38E+05	370	5.66E+09
Var173	370	-5.11E+08	370	8.61E+09
Var174	370	3.39E+05	370	4.63E+09

Variables	Num capped low	Low threshold	Num capped high	High threshold
Var175	370	1.56E+06	370	7.06E+09
Var176	370	-4.67E+08	370	1.07E+10
Var177	370	1.60E-01	370	1.30E+00
Var178	370	1.10E-02	370	1.71E+00
Var179	369	1.08E-03	370	2.08E-01
Var180	370	9.23E-05	370	2.33E-01
Var181	203	0.00E+00	370	3.88E-01
Var182	370	1.59E-02	370	8.57E-01
Var183	255	0.00E+00	370	8.42E-01
Var184	370	1.38E-01	370	9.83E-01
Var185	10	0.00E+00	370	9.88E-01
Var186	370	-8.02E-02	370	1.09E+00
Var187	370	-5.95E-01	370	9.46E-01
Var188	369	1.15E-03	370	1.18E+00
Var189	370	-4.86E-01	370	1.01E+00
Var190	370	4.82E-01	370	1.41E+00
Var191	17	0.00E+00	370	1.68E+00
Var192	5	0.00E+00	370	4.73E-01
Var193	188	0.00E+00	370	4.82E-01
Var194	370	7.47E-03	370	5.24E-01
Var195	42	0.00E+00	370	5.20E-01
Var196	370	-5.81E-02	370	2.62E-01
Var197	370	-5.01E-02	370	2.46E-01
Var198	370	1.19E-02	370	7.87E-01
Var199	92	0.00E+00	370	8.08E-01
Var200	370	1.73E-03	370	1.59E+00
Var201	370	-4.21E-01	370	1.54E+00
Var202	370	-5.89E-02	370	6.66E-02
Var203	370	-9.14E-02	370	1.92E-01
Var204	370	7.64E-03	370	1.32E+00
Var205	370	-4.15E-01	370	1.19E+00
Var206	370	4.57E-03	370	1.47E+00
Var207	370	-4.46E-01	370	1.33E+00
Var208	370	-3.26E+01	370	2.76E+01
Var209	370	-1.15E+01	370	1.93E+01
Var210	260	0.00E+00	370	3.50E-01
Var211	255	0.00E+00	370	3.75E-01
Var212	258	0.00E+00	370	4.00E-01
Var213	370	-3.40E-01	370	5.72E-01
Var214	370	-1.40E-02	370	3.33E-01
Var215	370	-1.41E-02	370	3.61E-01
Var216	370	-1.43E-02	370	3.75E-01
Var217	370	-3.19E-01	370	5.22E-01
Var218	370	-4.23E-01	370	9.61E+00

Variables	Num capped low	Low threshold	Num capped high	High threshold
Var219	370	-5.49E+01	370	4.51E+01
Var220	370	-1.51E+01	370	2.62E+01
Var221	370	3.16E-01	370	2.47E+00
Var222	17	0.00E+00	370	1.87E+00
Var223	370	-1.70E+00	370	9.95E-01
Var224	370	-9.74E-01	370	8.44E-01
Var225	370	-3.47E+01	370	3.25E+01
Var226	370	-1.20E+01	370	2.21E+01
Var227	370	-2.20E+09	370	6.20E+09
Var228	370	-3.82E+09	370	7.32E+09
Var229	370	-2.10E+09	370	2.29E+09
Var230	370	-3.05E+09	370	2.93E+09
Var231	370	-2.79E+10	370	7.28E+09
Var232	370	-6.02E+09	370	3.50E+09
Var233	370	-2.74E+11	191	0.00E+00
Var234	370	-7.09E+09	370	5.34E+09
Var235	370	-5.22E+09	370	5.50E+09
Var236	370	-2.73E+09	370	4.31E+09
Var237	370	-3.48E+10	370	2.94E+09
Var238	370	-2.74E+09	370	5.73E+09
Var239	370	-8.08E+09	370	7.29E+09
Var240	370	-2.89E+10	370	2.64E+09
Var241	370	-1.67E-01	370	2.19E-01
Var242	370	-6.34E+01	370	5.28E+01
Var243	370	-2.90E+01	370	2.73E+01
Var244	370	-8.70E-01	370	9.93E-01
Var245	370	-5.90E+00	370	6.84E+00
Var246	370	-2.93E+10	370	1.75E+09
Var247	370	-7.03E+09	370	5.09E+09
Var248	370	7.63E-02	370	1.08E+01
Var249	215	1.00E+00	370	4.31E+00
Var250	370	3.76E-01	370	1.85E+01
Var251	370	6.84E-01	370	3.60E+00
Var252	29	0.00E+00	370	6.40E+00
Var253	32	0.00E+00	370	6.51E+00
Var254	370	-1.32E-01	370	2.60E+00
Var255	370	-3.37E-01	370	5.40E+00
Var256	370	-3.46E-01	370	5.51E+00
Var257	370	-4.48E-01	370	2.44E+00
Var258	370	-7.34E-01	370	7.34E+00
Var259	370	-2.39E-01	370	1.47E+00
Var260	370	-3.62E-01	370	4.68E+00
Var261	370	-3.50E+00	370	1.98E+01
Var262	370	-9.93E-01	370	1.28E+01

Variables	Num capped low	Low threshold	Num capped high	High threshold
Var263	369	-2.97E+00	370	1.67E+01
Var264	370	-9.64E-01	370	1.33E+01
Var265	356	-3.00E+00	368	1.60E+01
Var266	370	-9.76E-01	364	1.30E+01
Var267	370	-3.02E+00	370	2.30E+01
Var268	370	-9.57E-01	370	1.85E+01
Var269	370	-2.67E+00	370	2.39E+01
Var270	370	-9.41E-01	370	1.78E+01
Var271	370	-4.04E+00	370	2.15E+01
Var272	370	-1.33E+00	370	1.81E+01
Var273	370	-9.59E-01	370	1.87E+01
Var274	370	-7.81E-01	370	5.77E+00
Var275	370	-6.64E-01	370	8.44E+00
Var276	370	-6.62E-01	370	8.57E+00
Var277	370	-6.58E-01	370	5.99E+00
Var278	370	-8.94E-01	370	6.86E+00
Var279	370	-6.07E-01	370	2.87E+00
Var280	370	-6.40E+09	370	5.54E+09
Var281	370	-1.77E-02	370	4.04E-01
Var282	370	-1.32E-01	370	2.60E+00
Var283	370	-3.37E-01	370	5.40E+00
Var284	370	-8.03E-01	370	2.48E+00
Var285	370	-5.81E-01	370	1.60E+00
Var286	241	0.00E+00	370	3.40E-01
Var287	370	1.28E-03	370	2.25E+00
Var288	370	-7.87E-01	370	3.17E+00
Var289	370	1.28E-03	370	2.09E+00
Var290	370	-7.85E-01	370	2.79E+00
Var291	370	-1.18E-01	370	2.21E+00
Var292	370	-8.02E-01	370	3.07E+00
Var293	370	-1.18E-01	370	2.04E+00
Var294	370	-7.94E-01	370	2.72E+00
Var295	370	-3.03E-02	370	2.29E+00
Var296	370	-7.74E-01	370	3.21E+00
Var297	370	-3.03E-02	370	2.13E+00
Var298	370	-7.68E-01	370	2.85E+00
Var299	370	6.70E-04	370	2.08E+00
Var300	370	-7.64E-01	370	2.88E+00
Var301	370	-3.89E-02	370	2.12E+00
Var302	370	-7.44E-01	370	2.90E+00
Var303	370	5.09E-02	370	3.71E+01
Var304	19	0.00E+00	370	5.60E+01
Var305	370	4.49E-02	370	3.71E+01
Var306	19	0.00E+00	370	5.60E+01

Variables	Num capped low	Low threshold	Num capped high	High threshold
Var307	370	2.89E-03	370	2.83E+00
Var308	370	-6.74E-01	370	3.79E+00
Var309	370	5.22E-03	370	3.20E+00
Var310	370	-6.25E-01	370	4.24E+00
Var311	370	-9.19E-02	370	2.65E+00
Var312	370	-8.36E-01	370	3.69E+00
Var313	370	1.40E-01	370	2.03E+01
Var314	370	6.09E-01	370	4.77E+01
Var315	370	1.47E-01	370	3.62E+01
Var316	226	0.00E+00	368	9.63E+00
Var317	0	0.00E+00	367	1.29E+00
Var318	370	-3.21E+00	370	1.04E+01
Var319	370	-3.02E+00	370	1.10E+01
Var320	370	1.17E-01	370	1.87E+01
Var321	370	-2.32E+00	370	3.02E+00
Var322	370	-2.29E+00	370	4.22E+00
Var323	370	-3.71E+00	370	1.02E+00
Var324	370	-4.84E+00	370	1.21E+00
Var325	370	-2.02E+00	370	5.59E+00
Var326	370	-2.79E+00	370	6.53E+00
Var327	370	-4.00E+00	370	2.47E+00
Var328	370	-4.37E+00	370	3.39E+00
Var329	370	-2.87E+00	370	2.60E+00
Var330	370	-2.77E+00	370	3.34E+00
Var331	10	0.00E+00	370	7.70E-01
Var332	21	0.00E+00	370	9.22E-01
Var333	370	-9.19E+00	370	4.34E+00
Var334	370	-1.47E+01	370	2.55E+00
Var335	370	-2.43E+00	370	3.93E+00
Var336	370	-2.26E+00	370	3.89E+00
Var337	370	-2.31E+00	370	3.88E+00
Var338	370	-2.20E+00	370	3.76E+00
Var339	0	0.00E+00	370	1.00E+03
Var340	370	4.04E+00	370	1.83E+03
Var341	0	0.00E+00	370	1.00E+03
Var342	370	1.94E-01	370	6.83E+01
Var343	370	1.93E-01	370	6.78E+01
Var344	0	0.00E+00	370	5.66E+01
Var345	0	0.00E+00	370	9.69E+02
Var346	370	4.34E-01	370	2.76E+01
Var347	370	1.54E-01	370	1.42E+01
Var348	0	0.00E+00	370	1.03E+03
Var349	370	3.03E+00	370	1.71E+03
Var350	0	0.00E+00	370	1.00E+03

Variables	Num capped low	Low threshold	Num capped high	High threshold
Var351	370	5.15E-01	370	3.27E+01
Var352	370	8.12E+08	370	2.06E+11
Var353	370	1.09E+09	370	2.19E+11
Var354	370	8.17E-01	370	1.03E+01
Var355	370	8.45E-01	370	1.16E+01
Var356	370	7.89E-01	370	1.27E+01
Var357	370	8.17E-01	370	1.44E+01
Var358	370	9.73E-02	370	1.22E+00
Var359	370	7.90E-02	370	1.27E+00
Var360	0	0.00E+00	370	3.56E-02
Var361	370	-6.85E-01	370	2.24E+00
Var362	370	7.05E+00	370	1.14E+03
Var363	0	0.00E+00	370	3.92E+02
Var364	0	0.00E+00	370	8.80E-01
Var365	39	0.00E+00	370	8.81E-01
Var366	301	-2.00E-01	361	2.50E-01
Var367	0	0.00E+00	370	8.44E-01
Var368	370	8.80E-04	0	1.00E+00
Q2_Var1	4	0.00E+00	370	1.28E+10
Q2_Var2	1	0.00E+00	370	2.93E+09
Q2_Var3	1	0.00E+00	370	5.10E+09
Q2_Var4	0	0.00E+00	370	1.65E+11
Q2_Var5	0	0.00E+00	370	1.82E+10
Q2_Var6	0	0.00E+00	370	8.39E+09
Q2_Var7	0	0.00E+00	370	1.07E+11
Q2_Var8	0	0.00E+00	370	7.76E+10
Q2_Var9	103	0.00E+00	370	3.74E+09
Q2_Var10	0	0.00E+00	370	6.19E-01
Q2_Var11	0	0.00E+00	370	4.89E-01
Q2_Var12	0	0.00E+00	370	9.26E-01
Q2_Var13	370	-5.61E-01	370	1.05E+00
Q2_Var14	370	-2.07E+09	370	5.95E+09
Q2_Var15	370	-1.94E+09	370	2.26E+09
Q2_Var16	370	-1.43E-02	370	3.87E-01
Q2_Var17	0	0.00E+00	370	2.28E-01
Q2_Var18	370	-7.30E-02	370	3.56E+01
Q2_Var19	370	-1.22E+01	71	0.00E+00
Q2_Var20	0	0.00E+00	370	1.42E+01
Q2_Var21	0	0.00E+00	370	1.30E+01
Q2_Var22	0	0.00E+00	370	1.07E+01
Q2_Var23	50	0.00E+00	370	7.87E+00
Q2_Var24	370	-2.44E+00	370	3.58E+02
Q2_Var25	370	-9.45E+09	370	2.38E+10
Q2_Var26	370	-7.48E-01	370	1.30E+00

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q2_Var27	370	-3.13E+01	370	1.07E+02
Q2_Var28	0	0.00E+00	370	9.26E-01
Q2_Var29	0	0.00E+00	370	3.94E-01
Q2_Var30	0	0.00E+00	370	1.19E+00
Q2_Var31	5	0.00E+00	370	8.16E-01
Q2_Var32	182	0.00E+00	370	9.27E+00
Q2_Var33	0	0.00E+00	370	9.32E+00
Q2_Var34	182	0.00E+00	370	8.27E+00
Q2_Var35	182	0.00E+00	370	1.53E+01
Q2_Var36	149	0.00E+00	370	7.04E-01
Q2_Var37	193	0.00E+00	370	2.26E+00
Q2_Var38	370	-1.53E+01	370	1.38E+01
Q2_Var39	119	0.00E+00	370	1.38E+00
Q2_Var40	370	-5.61E-01	370	1.05E+00
Q2_Var41	370	-1.04E+01	370	4.14E+01
Q2_Var42	0	0.00E+00	370	8.34E-01
Q2_Var43	0	0.00E+00	370	1.11E+01
Q2_Var44	370	-3.10E+07	370	6.55E+08
Q2_Var45	370	-1.66E+08	370	2.90E+09
Q2_Var46	289	0.00E+00	370	3.77E+01
Q2_Var47	370	-2.36E-01	370	1.75E+00
Q2_Var48	230	0.00E+00	370	1.90E+00
Q2_Var49	224	0.00E+00	368	1.89E+00
Q2_Var50	0	0.00E+00	370	9.65E-01
Q2_Var51	50	0.00E+00	370	6.39E-01
Q2_Var52	4	0.00E+00	370	4.98E-01
Q2_Var53	370	-2.10E+00	370	9.30E-01
Q2_Var54	370	-3.98E-01	370	7.95E-01
Q2_Var55	370	-1.94E+00	370	1.87E+00
Q2_Var56	2	0.00E+00	370	8.95E-01
Q2_Var57	0	0.00E+00	370	6.74E-01
Q2_Var58	0	0.00E+00	370	3.58E-01
Q2_Var59	0	0.00E+00	0	1.00E+00
Q2_Var60	182	0.00E+00	370	9.39E-01
Q2_Var61	370	-2.23E+00	370	5.78E-01
Q2_Var62	136	0.00E+00	370	1.23E+02
Q2_Var63	180	0.00E+00	370	4.38E+02
Q2_Var64	0	0.00E+00	35	1.00E+00
Q2_Var65	15	0.00E+00	5	1.00E+00
Q2_Var66	5	0.00E+00	370	9.21E-01
Q2_Var67	34	0.00E+00	370	7.58E-01
Q2_Var68	8	0.00E+00	370	1.02E+00
Q2_Var69	370	-1.99E-02	370	4.99E-01
Q2_Var70	277	0.00E+00	370	4.10E+01

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q2_Var71	370	-5.52E-01	370	4.36E+00
Q2_Var72	370	-2.54E+00	370	1.35E+00
Q2_Var73	370	-3.42E-01	370	3.54E+00
Q2_Var74	66	0.00E+00	370	1.48E-01
Q2_Var75	370	-9.50E-04	370	2.54E-01
Q2_Var76	370	-3.97E-02	370	2.80E+00
Q2_Var77	370	-2.95E-01	370	7.76E-01
Q2_Var78	184	0.00E+00	370	2.19E+00
Q2_Var79	370	-1.19E+00	370	8.15E-01
Q2_Var80	370	-3.52E-01	370	2.10E+00
Q2_Var81	370	-1.08E+00	370	1.34E+00
Q2_Var82	254	0.00E+00	370	2.05E+00
Q2_Var83	370	-1.05E+00	370	8.52E-01
Q2_Var84	120	0.00E+00	146	1.00E+00
Q2_Var85	4	0.00E+00	370	3.80E+00
Q2_Var86	0	0.00E+00	370	6.44E+02
Q2_Var87	0	0.00E+00	370	5.50E+02
Q2_Var88	0	0.00E+00	370	5.38E+02
Q2_Var89	0	0.00E+00	370	9.10E+02
Q2_Var90	16	0.00E+00	370	8.07E+02
Q2_Var91	0	0.00E+00	370	4.58E+02
Q2_Var92	0	0.00E+00	370	4.52E+02
Q2_Var93	0	0.00E+00	370	4.59E+02
Q2_Var94	0	0.00E+00	370	3.85E+02
Q2_Var95	0	0.00E+00	370	1.41E+01
Q2_Var96	1	0.00E+00	370	1.60E+02
Q2_Var97	1	0.00E+00	370	1.66E+02
Q2_Var98	1	0.00E+00	370	1.66E+02
Q2_Var99	1	0.00E+00	370	2.53E+02
Q2_Var100	22	0.00E+00	370	2.51E+02
Q2_Var101	0	0.00E+00	370	4.31E+03
Q2_Var102	0	0.00E+00	370	4.12E+03
Q2_Var103	0	0.00E+00	370	4.11E+03
Q2_Var104	0	0.00E+00	370	2.94E+03
Q2_Var105	0	0.00E+00	370	4.38E+03
Q2_Var106	0	0.00E+00	370	4.23E+03
Q2_Var107	0	0.00E+00	370	4.22E+03
Q2_Var108	0	0.00E+00	370	3.08E+03
Q2_Var109	1	0.00E+00	370	5.78E+01
Q2_Var110	1	0.00E+00	370	5.34E+01
Q2_Var111	1	0.00E+00	370	5.70E+01
Q2_Var112	1	0.00E+00	370	8.56E+01
Q2_Var113	24	0.00E+00	370	9.29E+01
Q2_Var114	0	0.00E+00	370	5.02E+01

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q2_Var115	0	0.00E+00	370	5.52E+01
Q2_Var116	0	0.00E+00	370	5.57E+01
Q2_Var117	0	0.00E+00	370	8.60E+01
Q2_Var118	10	0.00E+00	370	8.44E+01
Q2_Var119	0	0.00E+00	370	4.23E+01
Q2_Var120	0	0.00E+00	370	3.55E+01
Q2_Var121	0	0.00E+00	370	3.52E+01
Q2_Var122	0	0.00E+00	370	5.34E+01
Q2_Var123	17	0.00E+00	370	5.64E+01
Q2_Var124	0	0.00E+00	370	3.16E+01
Q2_Var125	0	0.00E+00	370	3.58E+00
Q2_Var126	0	0.00E+00	370	3.67E+00
Q2_Var127	0	0.00E+00	370	3.74E+00
Q2_Var128	0	0.00E+00	370	4.85E+00
Q2_Var129	17	0.00E+00	370	4.88E+00
Q2_Var130	0	0.00E+00	370	8.77E+00
Q2_Var131	0	0.00E+00	370	1.91E+02
Q2_Var132	0	0.00E+00	370	1.81E+02
Q2_Var133	0	0.00E+00	370	1.82E+02
Q2_Var134	0	0.00E+00	370	3.07E+02
Q2_Var135	16	0.00E+00	370	2.98E+02
Q2_Var136	0	0.00E+00	370	1.40E+01
Q2_Var137	0	0.00E+00	370	1.44E+01
Q2_Var138	0	0.00E+00	370	1.46E+01
Q2_Var139	0	0.00E+00	370	2.19E+01
Q2_Var140	17	0.00E+00	370	2.21E+01
Q2_Var141	0	0.00E+00	370	5.58E+01
Q2_Var142	0	0.00E+00	370	1.88E+00
Q2_Var143	0	0.00E+00	370	1.96E+00
Q2_Var144	0	0.00E+00	370	2.02E+00
Q2_Var145	0	0.00E+00	370	2.67E+00
Q2_Var146	17	0.00E+00	370	2.79E+00
Q2_Var147	0	0.00E+00	370	7.09E+00
Q2_Var148	0	0.00E+00	370	7.52E+00
Q2_Var149	0	0.00E+00	370	7.79E+00
Q2_Var150	0	0.00E+00	370	1.06E+01
Q2_Var151	11	0.00E+00	370	1.12E+01
Q2_Var152	195	0.00E+00	370	2.06E-01
Q2_Var153	195	0.00E+00	370	2.22E-01
Q2_Var154	195	0.00E+00	370	2.26E-01
Q2_Var155	370	-1.27E-01	370	3.27E-01
Q2_Var156	103	0.00E+00	370	1.68E-01
Q2_Var157	103	0.00E+00	370	1.84E-01
Q2_Var158	103	0.00E+00	370	1.87E-01

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q2_Var159	370	-1.41E-01	370	2.69E-01
Q2_Var160	103	0.00E+00	370	5.07E-01
Q2_Var161	103	0.00E+00	370	5.12E-01
Q2_Var162	103	0.00E+00	370	5.21E-01
Q2_Var163	370	-3.12E-01	370	7.58E-01
Q2_Var164	103	0.00E+00	370	2.53E+01
Q2_Var165	103	0.00E+00	370	2.30E+01
Q2_Var166	103	0.00E+00	370	2.27E+01
Q2_Var167	370	-1.23E+00	370	3.83E+01
Q2_Var168	103	0.00E+00	370	3.19E-01
Q2_Var169	103	0.00E+00	370	3.49E-01
Q2_Var170	103	0.00E+00	370	3.61E-01
Q2_Var171	370	-3.04E-01	370	5.27E-01
Q2_Var172	195	0.00E+00	370	5.33E+09
Q2_Var173	370	-4.65E+08	370	8.07E+09
Q2_Var174	262	0.00E+00	370	4.38E+09
Q2_Var175	119	0.00E+00	370	6.70E+09
Q2_Var176	370	-4.08E+08	370	9.77E+09
Q2_Var177	22	0.00E+00	370	1.24E+00
Q2_Var178	132	0.00E+00	370	1.67E+00
Q2_Var179	195	0.00E+00	370	2.06E-01
Q2_Var180	279	0.00E+00	370	2.31E-01
Q2_Var181	193	0.00E+00	370	3.83E-01
Q2_Var182	149	0.00E+00	370	8.52E-01
Q2_Var183	225	0.00E+00	370	8.36E-01
Q2_Var184	1	0.00E+00	370	9.81E-01
Q2_Var185	10	0.00E+00	370	9.86E-01
Q2_Var186	370	-7.49E-02	370	1.01E+00
Q2_Var187	370	-5.35E-01	370	9.10E-01
Q2_Var188	103	0.00E+00	370	1.10E+00
Q2_Var189	370	-4.25E-01	370	9.64E-01
Q2_Var190	22	0.00E+00	370	1.36E+00
Q2_Var191	16	0.00E+00	370	1.65E+00
Q2_Var192	5	0.00E+00	370	4.67E-01
Q2_Var193	181	0.00E+00	370	4.79E-01
Q2_Var194	4	0.00E+00	370	4.95E-01
Q2_Var195	42	0.00E+00	370	4.91E-01
Q2_Var196	370	-5.59E-02	370	2.47E-01
Q2_Var197	370	-4.87E-02	370	2.33E-01
Q2_Var198	85	0.00E+00	370	7.48E-01
Q2_Var199	89	0.00E+00	370	7.64E-01
Q2_Var200	132	0.00E+00	370	1.55E+00
Q2_Var201	370	-3.62E-01	370	1.48E+00
Q2_Var202	370	-5.41E-02	370	6.66E-02

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q2_Var203	370	-7.11E-02	370	1.90E-01
Q2_Var204	119	0.00E+00	370	1.24E+00
Q2_Var205	370	-3.48E-01	370	1.13E+00
Q2_Var206	195	0.00E+00	370	1.40E+00
Q2_Var207	370	-3.71E-01	370	1.22E+00
Q2_Var208	370	-3.16E+01	370	2.61E+01
Q2_Var209	370	-1.11E+01	370	1.81E+01
Q2_Var210	222	0.00E+00	370	3.38E-01
Q2_Var211	217	0.00E+00	370	3.60E-01
Q2_Var212	220	0.00E+00	370	3.88E-01
Q2_Var213	370	-3.06E-01	370	5.56E-01
Q2_Var214	370	-1.20E-02	370	3.17E-01
Q2_Var215	370	-1.23E-02	370	3.51E-01
Q2_Var216	370	-1.25E-02	370	3.59E-01
Q2_Var217	370	-2.88E-01	370	5.07E-01
Q2_Var218	370	-3.90E-01	370	8.73E+00
Q2_Var219	370	-5.24E+01	370	4.13E+01
Q2_Var220	370	-1.45E+01	370	2.50E+01
Q2_Var221	1	0.00E+00	370	2.41E+00
Q2_Var222	16	0.00E+00	370	1.82E+00
Q2_Var223	370	-1.65E+00	370	9.51E-01
Q2_Var224	370	-9.47E-01	370	8.28E-01
Q2_Var225	370	-3.38E+01	370	3.02E+01
Q2_Var226	370	-1.16E+01	370	2.01E+01
Q2_Var227	370	-2.07E+09	370	5.95E+09
Q2_Var228	370	-3.59E+09	370	6.99E+09
Q2_Var229	370	-1.94E+09	370	2.26E+09
Q2_Var230	370	-2.91E+09	370	2.86E+09
Q2_Var231	370	-2.68E+10	370	6.62E+09
Q2_Var232	370	-5.70E+09	370	3.30E+09
Q2_Var233	370	-2.50E+11	178	0.00E+00
Q2_Var234	370	-6.83E+09	370	5.07E+09
Q2_Var235	370	-4.91E+09	370	5.10E+09
Q2_Var236	370	-2.62E+09	370	3.99E+09
Q2_Var237	370	-3.33E+10	370	2.86E+09
Q2_Var238	370	-2.48E+09	370	5.48E+09
Q2_Var239	370	-7.80E+09	370	6.80E+09
Q2_Var240	370	-2.77E+10	370	2.35E+09
Q2_Var241	370	-1.63E-01	370	2.17E-01
Q2_Var242	370	-6.13E+01	370	4.96E+01
Q2_Var243	370	-2.80E+01	370	2.48E+01
Q2_Var244	370	-8.13E-01	370	9.51E-01
Q2_Var245	370	-5.64E+00	370	6.35E+00
Q2_Var246	370	-2.78E+10	370	1.59E+09

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q2_Var247	370	-6.80E+09	370	4.75E+09
Q2_Var248	195	0.00E+00	370	1.03E+01
Q2_Var249	1	0.00E+00	370	4.14E+00
Q2_Var250	119	0.00E+00	370	1.75E+01
Q2_Var251	3	0.00E+00	370	3.57E+00
Q2_Var252	26	0.00E+00	370	6.31E+00
Q2_Var253	29	0.00E+00	370	6.49E+00
Q2_Var254	370	-1.29E-01	370	2.57E+00
Q2_Var255	370	-3.16E-01	370	5.31E+00
Q2_Var256	370	-3.15E-01	370	5.49E+00
Q2_Var257	370	-4.30E-01	370	2.36E+00
Q2_Var258	370	-7.19E-01	370	7.00E+00
Q2_Var259	370	-2.31E-01	370	1.46E+00
Q2_Var260	370	-3.50E-01	370	4.59E+00
Q2_Var261	370	-3.29E+00	370	1.92E+01
Q2_Var262	370	-9.85E-01	370	1.25E+01
Q2_Var263	370	-2.82E+00	367	1.60E+01
Q2_Var264	368	-9.57E-01	369	1.30E+01
Q2_Var265	370	-2.94E+00	370	1.54E+01
Q2_Var266	370	-9.68E-01	370	1.29E+01
Q2_Var267	370	-2.84E+00	370	2.24E+01
Q2_Var268	370	-9.50E-01	370	1.79E+01
Q2_Var269	370	-2.56E+00	370	2.28E+01
Q2_Var270	370	-9.35E-01	370	1.74E+01
Q2_Var271	370	-3.74E+00	370	2.08E+01
Q2_Var272	370	-1.30E+00	370	1.80E+01
Q2_Var273	370	-9.50E-01	370	1.83E+01
Q2_Var274	370	-7.76E-01	370	5.51E+00
Q2_Var275	370	-6.44E-01	370	8.19E+00
Q2_Var276	370	-6.41E-01	370	8.39E+00
Q2_Var277	370	-6.50E-01	370	5.88E+00
Q2_Var278	370	-8.81E-01	370	6.42E+00
Q2_Var279	370	-6.05E-01	370	2.84E+00
Q2_Var280	370	-6.08E+09	370	5.40E+09
Q2_Var281	370	-1.43E-02	370	3.87E-01
Q2_Var282	370	-1.29E-01	370	2.57E+00
Q2_Var283	370	-3.16E-01	370	5.31E+00
Q2_Var284	370	-7.20E-01	370	2.46E+00
Q2_Var285	370	-5.72E-01	370	1.58E+00
Q2_Var286	233	0.00E+00	370	3.31E-01
Q2_Var287	103	0.00E+00	370	2.20E+00
Q2_Var288	370	-7.44E-01	370	3.07E+00
Q2_Var289	103	0.00E+00	370	2.03E+00
Q2_Var290	370	-7.37E-01	370	2.69E+00

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q2_Var291	370	-1.17E-01	370	2.15E+00
Q2_Var292	370	-7.66E-01	370	2.98E+00
Q2_Var293	370	-1.17E-01	370	1.97E+00
Q2_Var294	370	-7.50E-01	370	2.60E+00
Q2_Var295	370	-2.68E-02	370	2.23E+00
Q2_Var296	370	-7.25E-01	370	3.11E+00
Q2_Var297	370	-2.68E-02	370	2.09E+00
Q2_Var298	370	-7.13E-01	370	2.75E+00
Q2_Var299	227	0.00E+00	370	2.04E+00
Q2_Var300	370	-7.22E-01	370	2.80E+00
Q2_Var301	370	-3.37E-02	370	2.07E+00
Q2_Var302	370	-6.84E-01	370	2.86E+00
Q2_Var303	0	0.00E+00	370	3.64E+01
Q2_Var304	17	0.00E+00	370	5.46E+01
Q2_Var305	0	0.00E+00	370	3.64E+01
Q2_Var306	17	0.00E+00	370	5.46E+01
Q2_Var307	195	0.00E+00	370	2.77E+00
Q2_Var308	370	-6.07E-01	370	3.73E+00
Q2_Var309	119	0.00E+00	370	3.15E+00
Q2_Var310	370	-5.56E-01	370	4.15E+00
Q2_Var311	370	-9.05E-02	370	2.59E+00
Q2_Var312	370	-7.82E-01	370	3.62E+00
Q2_Var313	182	0.00E+00	370	1.96E+01
Q2_Var314	0	0.00E+00	370	4.60E+01
Q2_Var315	0	0.00E+00	370	3.52E+01
Q2_Var316	212	0.00E+00	370	9.58E+00
Q2_Var317	0	0.00E+00	370	1.25E+00
Q2_Var318	370	-3.03E+00	370	9.84E+00
Q2_Var319	370	-2.85E+00	370	1.04E+01
Q2_Var320	188	0.00E+00	370	1.80E+01
Q2_Var321	370	-2.24E+00	370	2.97E+00
Q2_Var322	370	-2.21E+00	370	4.14E+00
Q2_Var323	370	-3.66E+00	370	9.69E-01
Q2_Var324	370	-4.80E+00	370	1.18E+00
Q2_Var325	370	-1.93E+00	370	5.54E+00
Q2_Var326	370	-2.68E+00	370	6.50E+00
Q2_Var327	370	-3.90E+00	370	2.40E+00
Q2_Var328	370	-4.34E+00	370	3.30E+00
Q2_Var329	370	-2.86E+00	370	2.53E+00
Q2_Var330	370	-2.75E+00	370	3.27E+00
Q2_Var331	10	0.00E+00	370	7.61E-01
Q2_Var332	20	0.00E+00	370	9.00E-01
Q2_Var333	370	-9.18E+00	370	4.11E+00
Q2_Var334	370	-1.46E+01	370	2.44E+00

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q2_Var335	370	-2.35E+00	370	3.92E+00
Q2_Var336	370	-2.19E+00	370	3.87E+00
Q2_Var337	370	-2.23E+00	370	3.86E+00
Q2_Var338	370	-2.15E+00	370	3.74E+00
Q2_Var339	0	0.00E+00	370	9.89E+02
Q2_Var340	0	0.00E+00	370	1.76E+03
Q2_Var341	0	0.00E+00	370	9.66E+02
Q2_Var342	0	0.00E+00	370	6.69E+01
Q2_Var343	0	0.00E+00	370	6.53E+01
Q2_Var344	0	0.00E+00	370	5.51E+01
Q2_Var345	0	0.00E+00	370	9.44E+02
Q2_Var346	0	0.00E+00	370	2.63E+01
Q2_Var347	0	0.00E+00	370	1.40E+01
Q2_Var348	0	0.00E+00	370	1.01E+03
Q2_Var349	0	0.00E+00	370	1.65E+03
Q2_Var350	0	0.00E+00	370	9.83E+02
Q2_Var351	0	0.00E+00	370	3.13E+01
Q2_Var352	0	0.00E+00	370	1.92E+11
Q2_Var353	0	0.00E+00	370	2.02E+11
Q2_Var354	0	0.00E+00	370	1.01E+01
Q2_Var355	0	0.00E+00	370	1.14E+01
Q2_Var356	0	0.00E+00	370	1.26E+01
Q2_Var357	0	0.00E+00	370	1.43E+01
Q2_Var358	0	0.00E+00	370	1.20E+00
Q2_Var359	0	0.00E+00	370	1.24E+00
Q2_Var360	0	0.00E+00	370	3.49E-02
Q2_Var361	370	-6.84E-01	370	2.23E+00
Q2_Var362	0	0.00E+00	370	1.11E+03
Q2_Var363	0	0.00E+00	370	3.79E+02
Q2_Var364	0	0.00E+00	370	8.65E-01
Q2_Var365	36	0.00E+00	370	8.68E-01
Q2_Var366	289	-2.00E-01	347	2.50E-01
Q2_Var367	0	0.00E+00	370	8.37E-01
Q2_Var368	248	0.00E+00	0	1.00E+00
Q4_Var1	4	0.00E+00	370	1.21E+10
Q4_Var2	1	0.00E+00	370	2.84E+09
Q4_Var3	1	0.00E+00	370	4.77E+09
Q4_Var4	0	0.00E+00	370	1.53E+11
Q4_Var5	0	0.00E+00	370	1.64E+10
Q4_Var6	0	0.00E+00	370	7.79E+09
Q4_Var7	0	0.00E+00	370	1.02E+11
Q4_Var8	0	0.00E+00	370	7.40E+10
Q4_Var9	93	0.00E+00	370	3.49E+09
Q4_Var10	0	0.00E+00	370	6.16E-01

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q4_Var11	0	0.00E+00	370	4.83E-01
Q4_Var12	0	0.00E+00	370	9.21E-01
Q4_Var13	370	-5.52E-01	370	1.04E+00
Q4_Var14	370	-1.97E+09	370	5.60E+09
Q4_Var15	370	-1.81E+09	370	2.21E+09
Q4_Var16	370	-1.18E-02	370	3.76E-01
Q4_Var17	0	0.00E+00	370	2.24E-01
Q4_Var18	370	-3.27E-02	370	3.43E+01
Q4_Var19	370	-1.21E+01	70	0.00E+00
Q4_Var20	0	0.00E+00	370	1.39E+01
Q4_Var21	0	0.00E+00	370	1.27E+01
Q4_Var22	0	0.00E+00	370	1.04E+01
Q4_Var23	50	0.00E+00	370	7.75E+00
Q4_Var24	370	-2.31E+00	370	3.16E+02
Q4_Var25	370	-9.17E+09	370	2.16E+10
Q4_Var26	370	-7.22E-01	370	1.27E+00
Q4_Var27	370	-2.72E+01	370	9.57E+01
Q4_Var28	0	0.00E+00	370	9.21E-01
Q4_Var29	0	0.00E+00	370	3.90E-01
Q4_Var30	0	0.00E+00	370	1.17E+00
Q4_Var31	5	0.00E+00	370	7.92E-01
Q4_Var32	178	0.00E+00	370	8.98E+00
Q4_Var33	0	0.00E+00	370	9.00E+00
Q4_Var34	178	0.00E+00	370	7.98E+00
Q4_Var35	178	0.00E+00	370	1.51E+01
Q4_Var36	147	0.00E+00	370	6.98E-01
Q4_Var37	187	0.00E+00	370	2.21E+00
Q4_Var38	370	-1.45E+01	370	1.32E+01
Q4_Var39	107	0.00E+00	370	1.37E+00
Q4_Var40	370	-5.52E-01	370	1.04E+00
Q4_Var41	370	-1.02E+01	370	3.98E+01
Q4_Var42	0	0.00E+00	370	8.28E-01
Q4_Var43	0	0.00E+00	370	1.06E+01
Q4_Var44	370	-2.77E+07	370	6.12E+08
Q4_Var45	370	-1.41E+08	370	2.74E+09
Q4_Var46	253	0.00E+00	370	3.59E+01
Q4_Var47	370	-2.13E-01	370	1.68E+00
Q4_Var48	195	0.00E+00	367	1.83E+00
Q4_Var49	189	0.00E+00	370	1.82E+00
Q4_Var50	0	0.00E+00	370	9.65E-01
Q4_Var51	50	0.00E+00	370	6.36E-01
Q4_Var52	4	0.00E+00	370	4.96E-01
Q4_Var53	370	-2.00E+00	370	9.28E-01
Q4_Var54	370	-3.93E-01	370	7.93E-01

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q4_Var55	370	-1.81E+00	370	1.85E+00
Q4_Var56	2	0.00E+00	370	8.92E-01
Q4_Var57	0	0.00E+00	370	6.72E-01
Q4_Var58	0	0.00E+00	370	3.47E-01
Q4_Var59	0	0.00E+00	0	1.00E+00
Q4_Var60	178	0.00E+00	370	9.38E-01
Q4_Var61	370	-1.94E+00	370	5.70E-01
Q4_Var62	136	0.00E+00	370	1.10E+02
Q4_Var63	176	0.00E+00	370	3.95E+02
Q4_Var64	0	0.00E+00	33	1.00E+00
Q4_Var65	14	0.00E+00	5	1.00E+00
Q4_Var66	5	0.00E+00	370	9.19E-01
Q4_Var67	32	0.00E+00	370	7.53E-01
Q4_Var68	8	0.00E+00	370	1.02E+00
Q4_Var69	370	-1.70E-02	370	4.91E-01
Q4_Var70	240	0.00E+00	370	3.94E+01
Q4_Var71	370	-4.71E-01	370	4.05E+00
Q4_Var72	370	-2.46E+00	370	1.30E+00
Q4_Var73	370	-3.01E-01	370	3.46E+00
Q4_Var74	60	0.00E+00	370	1.47E-01
Q4_Var75	370	-2.59E-04	370	2.52E-01
Q4_Var76	370	-2.17E-02	370	2.67E+00
Q4_Var77	370	-2.26E-01	370	7.60E-01
Q4_Var78	156	0.00E+00	370	2.11E+00
Q4_Var79	370	-1.11E+00	370	7.83E-01
Q4_Var80	370	-2.73E-01	370	2.02E+00
Q4_Var81	370	-9.95E-01	370	1.27E+00
Q4_Var82	224	0.00E+00	370	1.92E+00
Q4_Var83	370	-9.20E-01	370	8.19E-01
Q4_Var84	116	0.00E+00	137	1.00E+00
Q4_Var85	4	0.00E+00	370	3.58E+00
Q4_Var86	0	0.00E+00	370	5.99E+02
Q4_Var87	0	0.00E+00	370	5.36E+02
Q4_Var88	0	0.00E+00	370	5.21E+02
Q4_Var89	0	0.00E+00	370	8.59E+02
Q4_Var90	15	0.00E+00	370	7.69E+02
Q4_Var91	0	0.00E+00	370	4.46E+02
Q4_Var92	0	0.00E+00	370	4.41E+02
Q4_Var93	0	0.00E+00	370	4.45E+02
Q4_Var94	0	0.00E+00	370	3.73E+02
Q4_Var95	0	0.00E+00	370	1.34E+01
Q4_Var96	1	0.00E+00	370	1.46E+02
Q4_Var97	1	0.00E+00	370	1.50E+02
Q4_Var98	1	0.00E+00	370	1.50E+02

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q4_Var99	1	0.00E+00	370	2.29E+02
Q4_Var100	21	0.00E+00	370	2.33E+02
Q4_Var101	0	0.00E+00	370	4.21E+03
Q4_Var102	0	0.00E+00	370	4.02E+03
Q4_Var103	0	0.00E+00	370	3.94E+03
Q4_Var104	0	0.00E+00	370	2.81E+03
Q4_Var105	0	0.00E+00	370	4.30E+03
Q4_Var106	0	0.00E+00	370	4.16E+03
Q4_Var107	0	0.00E+00	370	4.14E+03
Q4_Var108	0	0.00E+00	370	2.92E+03
Q4_Var109	1	0.00E+00	370	5.46E+01
Q4_Var110	1	0.00E+00	370	5.10E+01
Q4_Var111	1	0.00E+00	370	5.53E+01
Q4_Var112	1	0.00E+00	370	8.31E+01
Q4_Var113	23	0.00E+00	370	8.95E+01
Q4_Var114	0	0.00E+00	370	4.92E+01
Q4_Var115	0	0.00E+00	370	5.36E+01
Q4_Var116	0	0.00E+00	370	5.40E+01
Q4_Var117	0	0.00E+00	370	8.37E+01
Q4_Var118	9	0.00E+00	370	8.30E+01
Q4_Var119	0	0.00E+00	370	4.10E+01
Q4_Var120	0	0.00E+00	370	3.49E+01
Q4_Var121	0	0.00E+00	370	3.45E+01
Q4_Var122	0	0.00E+00	370	5.25E+01
Q4_Var123	16	0.00E+00	370	5.51E+01
Q4_Var124	0	0.00E+00	370	2.97E+01
Q4_Var125	0	0.00E+00	370	3.52E+00
Q4_Var126	0	0.00E+00	370	3.63E+00
Q4_Var127	0	0.00E+00	370	3.69E+00
Q4_Var128	0	0.00E+00	370	4.81E+00
Q4_Var129	16	0.00E+00	370	4.83E+00
Q4_Var130	0	0.00E+00	370	8.50E+00
Q4_Var131	0	0.00E+00	370	1.73E+02
Q4_Var132	0	0.00E+00	370	1.61E+02
Q4_Var133	0	0.00E+00	370	1.63E+02
Q4_Var134	0	0.00E+00	370	2.60E+02
Q4_Var135	15	0.00E+00	370	2.53E+02
Q4_Var136	0	0.00E+00	370	1.37E+01
Q4_Var137	0	0.00E+00	370	1.40E+01
Q4_Var138	0	0.00E+00	370	1.42E+01
Q4_Var139	0	0.00E+00	370	2.11E+01
Q4_Var140	16	0.00E+00	370	2.17E+01
Q4_Var141	0	0.00E+00	370	5.28E+01
Q4_Var142	0	0.00E+00	370	1.86E+00

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q4_Var143	0	0.00E+00	370	1.95E+00
Q4_Var144	0	0.00E+00	370	2.00E+00
Q4_Var145	0	0.00E+00	370	2.63E+00
Q4_Var146	16	0.00E+00	370	2.76E+00
Q4_Var147	0	0.00E+00	370	6.95E+00
Q4_Var148	0	0.00E+00	370	7.38E+00
Q4_Var149	0	0.00E+00	370	7.64E+00
Q4_Var150	0	0.00E+00	370	1.05E+01
Q4_Var151	10	0.00E+00	370	1.08E+01
Q4_Var152	179	0.00E+00	370	2.03E-01
Q4_Var153	179	0.00E+00	370	2.20E-01
Q4_Var154	179	0.00E+00	370	2.23E-01
Q4_Var155	370	-1.12E-01	370	3.20E-01
Q4_Var156	93	0.00E+00	370	1.65E-01
Q4_Var157	93	0.00E+00	370	1.81E-01
Q4_Var158	93	0.00E+00	370	1.84E-01
Q4_Var159	370	-1.27E-01	370	2.64E-01
Q4_Var160	93	0.00E+00	370	4.94E-01
Q4_Var161	93	0.00E+00	370	5.03E-01
Q4_Var162	93	0.00E+00	370	5.07E-01
Q4_Var163	370	-2.75E-01	370	7.45E-01
Q4_Var164	93	0.00E+00	370	2.40E+01
Q4_Var165	93	0.00E+00	370	2.16E+01
Q4_Var166	93	0.00E+00	370	2.13E+01
Q4_Var167	370	-9.70E-01	370	3.63E+01
Q4_Var168	93	0.00E+00	370	3.09E-01
Q4_Var169	93	0.00E+00	370	3.43E-01
Q4_Var170	93	0.00E+00	370	3.52E-01
Q4_Var171	370	-2.73E-01	370	5.20E-01
Q4_Var172	179	0.00E+00	370	5.07E+09
Q4_Var173	370	-3.84E+08	370	7.43E+09
Q4_Var174	239	0.00E+00	370	4.09E+09
Q4_Var175	107	0.00E+00	370	6.32E+09
Q4_Var176	370	-3.26E+08	370	9.07E+09
Q4_Var177	19	0.00E+00	370	1.20E+00
Q4_Var178	121	0.00E+00	370	1.61E+00
Q4_Var179	179	0.00E+00	370	2.02E-01
Q4_Var180	260	0.00E+00	370	2.26E-01
Q4_Var181	177	0.00E+00	370	3.78E-01
Q4_Var182	123	0.00E+00	370	8.47E-01
Q4_Var183	198	0.00E+00	370	8.29E-01
Q4_Var184	1	0.00E+00	370	9.79E-01
Q4_Var185	10	0.00E+00	370	9.83E-01
Q4_Var186	370	-7.07E-02	370	9.38E-01

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q4_Var187	370	-4.76E-01	370	8.55E-01
Q4_Var188	93	0.00E+00	370	1.02E+00
Q4_Var189	370	-3.60E-01	370	9.13E-01
Q4_Var190	20	0.00E+00	370	1.34E+00
Q4_Var191	14	0.00E+00	370	1.61E+00
Q4_Var192	5	0.00E+00	370	4.62E-01
Q4_Var193	173	0.00E+00	370	4.73E-01
Q4_Var194	3	0.00E+00	370	4.76E-01
Q4_Var195	42	0.00E+00	370	4.74E-01
Q4_Var196	370	-5.32E-02	370	2.40E-01
Q4_Var197	370	-4.68E-02	370	2.26E-01
Q4_Var198	75	0.00E+00	370	7.30E-01
Q4_Var199	84	0.00E+00	370	7.41E-01
Q4_Var200	121	0.00E+00	370	1.51E+00
Q4_Var201	370	-3.02E-01	370	1.41E+00
Q4_Var202	370	-4.94E-02	370	6.55E-02
Q4_Var203	370	-5.89E-02	370	1.84E-01
Q4_Var204	107	0.00E+00	370	1.19E+00
Q4_Var205	370	-2.86E-01	370	1.10E+00
Q4_Var206	179	0.00E+00	370	1.31E+00
Q4_Var207	370	-3.12E-01	370	1.16E+00
Q4_Var208	370	-3.00E+01	370	2.46E+01
Q4_Var209	370	-1.06E+01	370	1.73E+01
Q4_Var210	188	0.00E+00	370	3.29E-01
Q4_Var211	184	0.00E+00	370	3.46E-01
Q4_Var212	187	0.00E+00	370	3.75E-01
Q4_Var213	370	-2.74E-01	370	5.40E-01
Q4_Var214	370	-9.98E-03	370	3.03E-01
Q4_Var215	370	-1.01E-02	370	3.39E-01
Q4_Var216	370	-1.02E-02	370	3.46E-01
Q4_Var217	370	-2.48E-01	370	4.89E-01
Q4_Var218	370	-3.61E-01	368	7.80E+00
Q4_Var219	370	-5.02E+01	370	3.81E+01
Q4_Var220	370	-1.40E+01	370	2.40E+01
Q4_Var221	1	0.00E+00	370	2.34E+00
Q4_Var222	15	0.00E+00	370	1.79E+00
Q4_Var223	370	-1.59E+00	370	9.28E-01
Q4_Var224	370	-8.79E-01	370	8.12E-01
Q4_Var225	370	-3.23E+01	370	2.82E+01
Q4_Var226	370	-1.13E+01	370	1.94E+01
Q4_Var227	370	-1.97E+09	370	5.60E+09
Q4_Var228	370	-3.21E+09	370	6.78E+09
Q4_Var229	370	-1.81E+09	370	2.21E+09
Q4_Var230	370	-2.72E+09	370	2.77E+09

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q4_Var231	370	-2.56E+10	370	6.17E+09
Q4_Var232	370	-5.39E+09	370	3.15E+09
Q4_Var233	370	-2.29E+11	171	0.00E+00
Q4_Var234	370	-6.41E+09	370	4.60E+09
Q4_Var235	370	-4.66E+09	370	4.67E+09
Q4_Var236	370	-2.46E+09	370	3.78E+09
Q4_Var237	370	-3.07E+10	370	2.73E+09
Q4_Var238	370	-2.30E+09	370	5.28E+09
Q4_Var239	370	-7.26E+09	370	6.42E+09
Q4_Var240	370	-2.60E+10	370	2.18E+09
Q4_Var241	370	-1.60E-01	370	2.15E-01
Q4_Var242	370	-5.84E+01	370	4.63E+01
Q4_Var243	370	-2.46E+01	370	2.28E+01
Q4_Var244	370	-7.78E-01	370	9.04E-01
Q4_Var245	370	-5.38E+00	370	6.02E+00
Q4_Var246	370	-2.65E+10	370	1.50E+09
Q4_Var247	370	-6.53E+09	370	4.47E+09
Q4_Var248	179	0.00E+00	370	9.78E+00
Q4_Var249	1	0.00E+00	370	4.03E+00
Q4_Var250	107	0.00E+00	370	1.66E+01
Q4_Var251	3	0.00E+00	370	3.54E+00
Q4_Var252	22	0.00E+00	370	6.09E+00
Q4_Var253	25	0.00E+00	370	6.29E+00
Q4_Var254	370	-1.27E-01	370	2.54E+00
Q4_Var255	370	-2.90E-01	370	5.09E+00
Q4_Var256	370	-2.84E-01	370	5.29E+00
Q4_Var257	370	-4.15E-01	370	2.28E+00
Q4_Var258	370	-7.14E-01	370	6.73E+00
Q4_Var259	370	-2.23E-01	370	1.46E+00
Q4_Var260	370	-3.39E-01	370	4.43E+00
Q4_Var261	370	-3.08E+00	370	1.86E+01
Q4_Var262	370	-9.79E-01	370	1.18E+01
Q4_Var263	369	-2.60E+00	370	1.54E+01
Q4_Var264	361	-9.50E-01	370	1.28E+01
Q4_Var265	370	-2.74E+00	369	1.50E+01
Q4_Var266	370	-9.62E-01	370	1.25E+01
Q4_Var267	370	-2.62E+00	370	2.15E+01
Q4_Var268	370	-9.45E-01	370	1.70E+01
Q4_Var269	370	-2.41E+00	370	2.16E+01
Q4_Var270	370	-9.28E-01	370	1.68E+01
Q4_Var271	370	-3.51E+00	370	1.98E+01
Q4_Var272	370	-1.29E+00	370	1.74E+01
Q4_Var273	370	-9.44E-01	370	1.77E+01
Q4_Var274	370	-7.70E-01	370	5.32E+00

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q4_Var275	370	-6.29E-01	370	7.98E+00
Q4_Var276	370	-6.27E-01	370	8.04E+00
Q4_Var277	370	-6.39E-01	370	5.69E+00
Q4_Var278	370	-8.70E-01	370	5.95E+00
Q4_Var279	370	-5.98E-01	370	2.79E+00
Q4_Var280	370	-5.81E+09	370	5.07E+09
Q4_Var281	370	-1.18E-02	370	3.76E-01
Q4_Var282	370	-1.27E-01	370	2.54E+00
Q4_Var283	370	-2.90E-01	370	5.09E+00
Q4_Var284	370	-6.85E-01	370	2.40E+00
Q4_Var285	370	-5.64E-01	370	1.53E+00
Q4_Var286	216	0.00E+00	370	3.19E-01
Q4_Var287	93	0.00E+00	370	2.13E+00
Q4_Var288	370	-6.77E-01	370	2.93E+00
Q4_Var289	93	0.00E+00	370	1.96E+00
Q4_Var290	370	-6.65E-01	370	2.58E+00
Q4_Var291	370	-1.12E-01	370	2.08E+00
Q4_Var292	370	-7.13E-01	370	2.85E+00
Q4_Var293	370	-1.12E-01	370	1.91E+00
Q4_Var294	370	-7.06E-01	370	2.51E+00
Q4_Var295	370	-2.32E-02	370	2.17E+00
Q4_Var296	370	-6.41E-01	370	3.00E+00
Q4_Var297	370	-2.32E-02	370	1.99E+00
Q4_Var298	370	-6.36E-01	370	2.66E+00
Q4_Var299	193	0.00E+00	370	1.96E+00
Q4_Var300	370	-6.49E-01	370	2.69E+00
Q4_Var301	370	-2.65E-02	370	1.99E+00
Q4_Var302	370	-6.21E-01	370	2.75E+00
Q4_Var303	0	0.00E+00	370	3.55E+01
Q4_Var304	16	0.00E+00	370	5.37E+01
Q4_Var305	0	0.00E+00	370	3.55E+01
Q4_Var306	16	0.00E+00	370	5.37E+01
Q4_Var307	179	0.00E+00	370	2.70E+00
Q4_Var308	370	-5.25E-01	370	3.62E+00
Q4_Var309	107	0.00E+00	370	3.04E+00
Q4_Var310	370	-4.74E-01	370	4.00E+00
Q4_Var311	370	-8.85E-02	370	2.48E+00
Q4_Var312	370	-7.22E-01	370	3.47E+00
Q4_Var313	178	0.00E+00	370	1.85E+01
Q4_Var314	0	0.00E+00	370	4.42E+01
Q4_Var315	0	0.00E+00	370	3.41E+01
Q4_Var316	198	0.00E+00	369	9.41E+00
Q4_Var317	0	0.00E+00	369	1.21E+00
Q4_Var318	370	-2.90E+00	370	9.38E+00

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q4_Var319	370	-2.73E+00	370	1.00E+01
Q4_Var320	184	0.00E+00	370	1.71E+01
Q4_Var321	370	-2.16E+00	370	2.87E+00
Q4_Var322	370	-2.15E+00	370	4.01E+00
Q4_Var323	370	-3.61E+00	370	9.38E-01
Q4_Var324	370	-4.75E+00	370	1.12E+00
Q4_Var325	370	-1.89E+00	370	5.48E+00
Q4_Var326	370	-2.59E+00	370	6.46E+00
Q4_Var327	370	-3.87E+00	370	2.31E+00
Q4_Var328	370	-4.30E+00	370	3.19E+00
Q4_Var329	370	-2.84E+00	370	2.47E+00
Q4_Var330	370	-2.70E+00	370	3.16E+00
Q4_Var331	10	0.00E+00	370	7.37E-01
Q4_Var332	20	0.00E+00	370	8.80E-01
Q4_Var333	370	-9.06E+00	370	3.95E+00
Q4_Var334	370	-1.43E+01	370	2.31E+00
Q4_Var335	370	-2.30E+00	370	3.87E+00
Q4_Var336	370	-2.15E+00	370	3.81E+00
Q4_Var337	370	-2.20E+00	370	3.82E+00
Q4_Var338	370	-2.07E+00	370	3.70E+00
Q4_Var339	0	0.00E+00	370	9.60E+02
Q4_Var340	0	0.00E+00	370	1.67E+03
Q4_Var341	0	0.00E+00	370	9.21E+02
Q4_Var342	0	0.00E+00	370	6.54E+01
Q4_Var343	0	0.00E+00	370	6.19E+01
Q4_Var344	0	0.00E+00	370	5.30E+01
Q4_Var345	0	0.00E+00	370	8.99E+02
Q4_Var346	0	0.00E+00	370	2.50E+01
Q4_Var347	0	0.00E+00	370	1.38E+01
Q4_Var348	0	0.00E+00	370	9.92E+02
Q4_Var349	0	0.00E+00	370	1.54E+03
Q4_Var350	0	0.00E+00	370	9.43E+02
Q4_Var351	0	0.00E+00	370	2.95E+01
Q4_Var352	0	0.00E+00	370	1.76E+11
Q4_Var353	0	0.00E+00	370	1.86E+11
Q4_Var354	0	0.00E+00	370	9.89E+00
Q4_Var355	0	0.00E+00	370	1.12E+01
Q4_Var356	0	0.00E+00	370	1.25E+01
Q4_Var357	0	0.00E+00	370	1.41E+01
Q4_Var358	0	0.00E+00	370	1.19E+00
Q4_Var359	0	0.00E+00	370	1.23E+00
Q4_Var360	0	0.00E+00	370	3.14E-02
Q4_Var361	370	-6.83E-01	370	2.22E+00
Q4_Var362	0	0.00E+00	370	1.06E+03

Variables	Num capped low	Low threshold	Num capped high	High threshold
Q4_Var363	0	0.00E+00	370	3.63E+02
Q4_Var364	0	0.00E+00	370	8.37E-01
Q4_Var365	33	0.00E+00	370	8.37E-01
Q4_Var366	370	-1.85E-01	370	2.28E-01
Q4_Var367	0	0.00E+00	370	8.10E-01
Q4_Var368	230	0.00E+00	0	1.00E+00

Table B.2 (a) Feature Importance Ranking (Rank 1–126)

Rank	Feature	Importance	Rank	Feature	Importance	Rank	Feature	Importance
1	Var318	0.01952	43	Var352	0.002669	85	Var84	0.001654
2	Var319	0.019387	44	Var351	0.002642	86	Var181	0.001649
3	Var313	0.017673	45	Q2_Var288	0.002623	87	Var63	0.001631
4	Var61	0.016414	46	Q2_Var189	0.002581	88	Var316	0.001627
5	Var320	0.012514	47	Var25	0.002545	89	Q2_Var159	0.001617
6	Q2_Var319	0.009856	48	Var168	0.002469	90	Var278	0.001609
7	Var17	0.007841	49	Var54	0.002463	91	Var255	0.001601
8	Q2_Var313	0.00739	50	Var33	0.002441	92	Var286	0.001584
9	Q2_Var61	0.007104	51	Q4_Var191	0.002331	93	Q4_Var62	0.001578
10	Var314	0.007074	52	Q2_Var171	0.002275	94	Var213	0.00157
11	Var253	0.007031	53	Var323	0.002228	95	Var118	0.00156
12	Var191	0.006685	54	Var353	0.002186	96	Q2_Var207	0.001558
13	Var251	0.006545	55	Q2_Var346	0.002174	97	Var151	0.001557
14	Q2_Var318	0.006458	56	Q2_Var213	0.002172	98	Q4_Var212	0.001556
15	Q2_Var312	0.006122	57	Var190	0.002161	99	Var53	0.001555
16	Var260	0.005741	58	Var47	0.002155	100	Q4_Var63	0.001554
17	Var187	0.005636	59	Var257	0.002108	101	Var57	0.001548
18	Q2_Var320	0.005459	60	Var324	0.002106	102	Q2_Var155	0.001532
19	Var252	0.004827	61	Var24	0.00208	103	Var361	0.001528
20	Var170	0.004566	62	Q2_Var47	0.001996	104	Q2_Var22	0.001516
21	Var150	0.004537	63	Q4_Var320	0.001946	105	Var218	0.001507
22	Var62	0.004391	64	Q2_Var252	0.00193	106	Q4_Var313	0.001501
23	Var148	0.004205	65	Var18	0.001918	107	Var1	0.001499
24	Q2_Var292	0.004128	66	Var312	0.001916	108	Var300	0.001495
25	Q2_Var348	0.004043	67	Q4_Var58	0.001882	109	Q2_Var308	0.001494
26	Var12	0.003706	68	Var212	0.001871	110	Q2_Var300	0.001491
27	Var149	0.003623	69	Var171	0.00187	111	Var279	0.001486
28	Var284	0.00341	70	Q2_Var163	0.001864	112	Var189	0.001485
29	Q2_Var191	0.003312	71	Q2_Var35	0.001829	113	Q4_Var251	0.001482
30	Var348	0.003277	72	Var259	0.001828	114	Var5	0.001468
31	Q2_Var187	0.003205	73	Q2_Var352	0.001768	115	Var147	0.001457
32	Q4_Var319	0.003053	74	Var195	0.001768	116	Q4_Var182	0.001446
33	Q4_Var318	0.002996	75	Q2_Var290	0.001754	117	Var163	0.001444
34	Q4_Var61	0.002996	76	Var70	0.001746	118	Var55	0.001443
35	Q2_Var45	0.002904	77	Var346	0.001744	119	Q4_Var5	0.001443
36	Q2_Var12	0.002879	78	Q4_Var316	0.001743	120	Var159	0.001429
37	Var4	0.002833	79	Var258	0.001727	121	Var276	0.001425
38	Q2_Var339	0.00281	80	Var339	0.001722	122	Var237	0.001418
39	Var45	0.002809	81	Q2_Var351	0.001708	123	Q2_Var294	0.001414
40	Var30	0.002808	82	Q2_Var32	0.001702	124	Var233	0.001412
41	Var35	0.002759	83	Var180	0.00169	125	Var21	0.001404
42	Q2_Var253	0.002696	84	Var58	0.001655	126	Var277	0.0014

Table B.3 (b) Feature Importance Ranking (Rank 127–252)

Rank	Feature	Importance	Rank	Feature	Importance	Rank	Feature	Importance
127	Q2_Var201	0.001392	169	Q2_Var217	0.001186	211	Q4_Var99	0.001021
128	Var155	0.001385	170	Var363	0.001185	212	Var217	0.001021
129	Var203	0.001374	171	Var201	0.001172	213	Q2_Var168	0.00102
130	Var197	0.001367	172	Q2_Var167	0.001162	214	Q2_Var195	0.001017
131	Q4_Var23	0.001365	173	Var169	0.001154	215	Var357	0.001015
132	Var22	0.001364	174	Q4_Var59	0.001148	216	Var343	0.001015
133	Var290	0.00135	175	Var362	0.001143	217	Q2_Var296	0.001012
134	Q2_Var353	0.001344	176	Var356	0.001141	218	Q4_Var47	0.001011
135	Q2_Var310	0.001339	177	Var56	0.001133	219	Q4_Var211	0.001011
136	Var73	0.001337	178	Var304	0.001133	220	Var185	0.001009
137	Q2_Var284	0.001336	179	Q2_Var251	0.001129	221	Q2_Var30	0.001008
138	Q2_Var316	0.001334	180	Var8	0.001124	222	Var52	0.001002
139	Var207	0.001333	181	Q4_Var57	0.001121	223	Q2_Var31	0.001002
140	Q2_Var63	0.00133	182	Var179	0.001121	224	Var162	0.001001
141	Var294	0.001328	183	Var256	0.001121	225	Q2_Var69	0.000998
142	Q2_Var58	0.001327	184	Var293	0.001116	226	Var359	0.000996
143	Var44	0.001301	185	Var285	0.0011	227	Var164	0.000988
144	Q2_Var20	0.001299	186	Q2_Var84	0.0011	228	Q2_Var176	0.000987
145	Q2_Var256	0.001284	187	Var205	0.0011	229	Q2_Var23	0.000987
146	Var59	0.001277	188	Var134	0.001094	230	Var184	0.000987
147	Var69	0.001271	189	Q2_Var25	0.001093	231	Q2_Var1	0.000981
148	Var254	0.001265	190	Q2_Var57	0.001093	232	Q2_Var298	0.000976
149	Q2_Var53	0.001258	191	Var231	0.001089	233	Q4_Var21	0.000975
150	Q4_Var150	0.001254	192	Q4_Var22	0.001088	234	Q2_Var257	0.000974
151	Var354	0.001252	193	Var157	0.001081	235	Var122	0.000969
152	Var135	0.001251	194	Var342	0.001079	236	Var211	0.000966
153	Var20	0.001251	195	Var121	0.001078	237	Var337	0.000963
154	Q2_Var59	0.001245	196	Var153	0.001075	238	Q2_Var56	0.000961
155	Q4_Var60	0.001241	197	Var317	0.001075	239	Var292	0.00096
156	Q2_Var51	0.001233	198	Q2_Var314	0.001068	240	Q2_Var55	0.000959
157	Q2_Var62	0.00123	199	Var167	0.001067	241	Var43	0.000951
158	Var355	0.001223	200	Var349	0.001065	242	Q2_Var205	0.000951
159	Q4_Var351	0.001218	201	Q2_Var211	0.001064	243	Var311	0.00095
160	Var46	0.001218	202	Var78	0.00106	244	Var193	0.000946
161	Q4_Var187	0.001216	203	Var7	0.001052	245	Q4_Var51	0.000944
162	Var275	0.001212	204	Var315	0.001051	246	Var199	0.000944
163	Var3	0.001211	205	Var245	0.001044	247	Var77	0.000943
164	Q2_Var24	0.00121	206	Var19	0.001039	248	Var172	0.000943
165	Q2_Var173	0.001205	207	Q2_Var169	0.001038	249	Var196	0.000942
166	Var221	0.001203	208	Var358	0.00103	250	Var326	0.000941
167	Var16	0.001201	209	Var160	0.001029	251	Q2_Var259	0.000939
168	Var117	0.001186	210	Var32	0.001029	252	Var66	0.000938

Table B.4 (c) Feature Importance Ranking (Rank 253–378)

Rank	Feature	Importance	Rank	Feature	Importance	Rank	Feature	Importance
253	Q4_Var100	0.000935	295	Q2_Var117	0.000851	337	Var120	0.000807
254	Var192	0.000931	296	Var10	0.000851	338	Var222	0.000806
255	Q2_Var199	0.00092	297	Q4_Var195	0.000851	339	Var131	0.000806
256	Q4_Var288	0.000913	298	Q2_Var324	0.00085	340	Q2_Var75	0.000805
257	Q4_Var140	0.000913	299	Var288	0.00085	341	Q4_Var221	0.000804
258	Q2_Var21	0.000913	300	Var129	0.000849	342	Var42	0.000804
259	Var36	0.000906	301	Q2_Var354	0.000847	343	Q2_Var64	0.000802
260	Q4_Var24	0.000906	302	Q2_Var258	0.000847	344	Q4_Var18	0.000798
261	Var132	0.000901	303	Var158	0.000847	345	Q2_Var5	0.000798
262	Q2_Var224	0.000899	304	Q4_Var352	0.000844	346	Q4_Var354	0.000797
263	Var216	0.000897	305	Var248	0.000844	347	Q2_Var245	0.000796
264	Var23	0.000894	306	Q4_Var190	0.000843	348	Q2_Var68	0.000796
265	Var280	0.000893	307	Var48	0.000842	349	Var230	0.000795
266	Q2_Var182	0.000891	308	Q2_Var140	0.000838	350	Var49	0.000793
267	Q4_Var317	0.000891	309	Q2_Var317	0.000834	351	Var92	0.00079
268	Var130	0.00089	310	Var27	0.000833	352	Q2_Var361	0.00079
269	Var15	0.000889	311	Q2_Var323	0.000832	353	Var74	0.000789
270	Q2_Var33	0.000889	312	Var289	0.000831	354	Q4_Var52	0.000789
271	Q2_Var54	0.000887	313	Q4_Var184	0.00083	355	Var338	0.000788
272	Var139	0.000886	314	Var75	0.00083	356	Var154	0.000788
273	Q2_Var221	0.000885	315	Q2_Var315	0.000829	357	Q4_Var20	0.000786
274	Var51	0.000884	316	Q2_Var194	0.000828	358	Var183	0.000785
275	Var204	0.000882	317	Q2_Var185	0.000828	359	Q4_Var279	0.000784
276	Q2_Var209	0.000877	318	Q4_Var56	0.000827	360	Var327	0.000784
277	Q4_Var312	0.000877	319	Var186	0.000826	361	Var206	0.000783
278	Q4_Var213	0.000876	320	Var308	0.000826	362	Var344	0.000782
279	Var116	0.000875	321	Q2_Var19	0.000825	363	Var144	0.000782
280	Q2_Var277	0.000874	322	Q4_Var294	0.000824	364	Q2_Var349	0.000781
281	Q4_Var252	0.000872	323	Q2_Var122	0.000823	365	Q2_Var128	0.000775
282	Q2_Var129	0.000871	324	Var113	0.000823	366	Q4_Var222	0.000774
283	Q2_Var218	0.000871	325	Var146	0.000822	367	Var6	0.000772
284	Var210	0.000868	326	Var95	0.000821	368	Q2_Var222	0.000772
285	Var123	0.000867	327	Q2_Var278	0.000817	369	Q4_Var125	0.00077
286	Q4_Var1	0.000866	328	Var291	0.000817	370	Q4_Var255	0.000769
287	Var89	0.000864	329	Q4_Var4	0.000817	371	Var145	0.000768
288	Q2_Var43	0.00086	330	Q2_Var203	0.000816	372	Q4_Var122	0.000765
289	Q2_Var72	0.000859	331	Q4_Var314	0.000816	373	Var208	0.000765
290	Q2_Var260	0.000858	332	Var140	0.000816	374	Var38	0.000763
291	Var336	0.000858	333	Var11	0.000816	375	Var156	0.000762
292	Q2_Var52	0.000856	334	Var105	0.000811	376	Q2_Var220	0.000762
293	Q2_Var60	0.000855	335	Q4_Var54	0.00081	377	Q4_Var159	0.000759
294	Q2_Var74	0.000854	336	Q2_Var123	0.000809	378	Var13	0.000759

Table B.5 (d) Feature Importance Ranking (Rank 379–504)

Rank	Feature	Importance	Rank	Feature	Importance	Rank	Feature	Importance
379	Var228	0.000758	421	Q4_Var38	0.000725	463	Var241	0.00069
380	Var202	0.000758	422	Q2_Var112	0.000724	464	Var99	0.000689
381	Q2_Var279	0.000758	423	Q4_Var310	0.000723	465	Q4_Var256	0.000688
382	Var329	0.000755	424	Q4_Var166	0.000722	466	Q4_Var37	0.000687
383	Q2_Var190	0.000754	425	Q4_Var130	0.000721	467	Var37	0.000686
384	Q4_Var347	0.000754	426	Var41	0.00072	468	Var67	0.000685
385	Var176	0.000752	427	Q2_Var145	0.000717	469	Q2_Var120	0.000684
386	Var101	0.000748	428	Q4_Var19	0.000716	470	Q4_Var112	0.000684
387	Q4_Var135	0.000747	429	Var31	0.000715	471	Q2_Var322	0.000683
388	Q4_Var149	0.000745	430	Var223	0.000714	472	Q2_Var341	0.000682
389	Q4_Var173	0.000745	431	Var106	0.000713	473	Q2_Var11	0.000682
390	Q4_Var336	0.000745	432	Q4_Var104	0.000713	474	Q4_Var123	0.000682
391	Var9	0.000744	433	Var14	0.000713	475	Q2_Var86	0.000681
392	Q2_Var150	0.000744	434	Var93	0.000712	476	Q2_Var358	0.000681
393	Q2_Var50	0.000743	435	Var234	0.000712	477	Q2_Var197	0.000679
394	Q4_Var30	0.000743	436	Q2_Var344	0.000712	478	Q2_Var184	0.000678
395	Var39	0.000741	437	Q4_Var308	0.000711	479	Var341	0.000678
396	Q2_Var350	0.00074	438	Q2_Var198	0.000711	480	Q2_Var67	0.000677
397	Q4_Var171	0.000739	439	Q4_Var278	0.00071	481	Q4_Var342	0.000676
398	Var269	0.000738	440	Q2_Var342	0.00071	482	Var119	0.000676
399	Q2_Var178	0.000737	441	Var90	0.000706	483	Var325	0.000675
400	Var198	0.000737	442	Q4_Var226	0.000706	484	Var104	0.000674
401	Var115	0.000736	443	Q2_Var73	0.000706	485	Q2_Var166	0.000674
402	Q2_Var285	0.000735	444	Q2_Var121	0.000704	486	Var127	0.000674
403	Q4_Var119	0.000735	445	Q2_Var99	0.000703	487	Q4_Var55	0.000673
404	Q2_Var183	0.000735	446	Q4_Var34	0.000703	488	Q4_Var346	0.000672
405	Var126	0.000733	447	Q4_Var92	0.000702	489	Var236	0.00067
406	Q2_Var286	0.000732	448	Q2_Var70	0.000701	490	Q2_Var302	0.000669
407	Q2_Var135	0.000731	449	Q2_Var356	0.0007	491	Var250	0.000669
408	Q2_Var46	0.000731	450	Q4_Var361	0.0007	492	Var166	0.000667
409	Q2_Var192	0.000731	451	Var328	0.000698	493	Q4_Var134	0.000665
410	Var303	0.000731	452	Q2_Var3	0.000698	494	Q2_Var177	0.000665
411	Q4_Var284	0.00073	453	Var182	0.000698	495	Q2_Var8	0.000664
412	Var224	0.000729	454	Q2_Var233	0.000696	496	Var347	0.000664
413	Q4_Var17	0.000727	455	Q2_Var4	0.000696	497	Var128	0.000662
414	Var244	0.000727	456	Q2_Var16	0.000696	498	Q2_Var37	0.000662
415	Q2_Var255	0.000727	457	Q2_Var134	0.000695	499	Q2_Var108	0.000661
416	Var88	0.000726	458	Q2_Var212	0.000694	500	Q4_Var43	0.00066
417	Q2_Var26	0.000726	459	Q2_Var139	0.000692	501	Var272	0.00066
418	Q4_Var53	0.000725	460	Var309	0.000691	502	Q2_Var41	0.000656
419	Q4_Var292	0.000725	461	Q4_Var167	0.000691	503	Var242	0.000656
420	Var340	0.000725	462	Var102	0.000691	504	Q4_Var359	0.000656

Table B.6 (e) Feature Importance Ranking (Rank 505–630)

Rank	Feature	Importance	Rank	Feature	Importance	Rank	Feature	Importance
505	Q4_Var233	0.000656	547	Q4_Var25	0.00063	589	Q4_Var209	0.000591
506	Q2_Var244	0.000655	548	Q2_Var355	0.000629	590	Var298	0.000591
507	Q4_Var108	0.000652	549	Q2_Var275	0.000627	591	Q4_Var9	0.000591
508	Var247	0.000652	550	Q4_Var192	0.000627	592	Var301	0.000589
509	Q4_Var66	0.000651	551	Q4_Var290	0.000624	593	Q4_Var185	0.000588
510	Q2_Var17	0.000651	552	Q2_Var147	0.000623	594	Var345	0.000588
511	Q2_Var136	0.00065	553	Q2_Var119	0.00062	595	Q4_Var75	0.000587
512	Q4_Var98	0.00065	554	Q2_Var196	0.00062	596	Var83	0.000586
513	Q4_Var315	0.000649	555	Q4_Var176	0.00062	597	Var133	0.000586
514	Q2_Var132	0.000648	556	Q4_Var10	0.000619	598	Var322	0.000586
515	Q2_Var343	0.000648	557	Var238	0.000618	599	Var100	0.000585
516	Var200	0.000647	558	Var107	0.000617	600	Var333	0.000584
517	Q4_Var96	0.000646	559	Q4_Var363	0.000616	601	Q4_Var32	0.000583
518	Var240	0.000646	560	Q4_Var299	0.000615	602	Q2_Var93	0.000583
519	Q2_Var234	0.000645	561	Q2_Var363	0.000615	603	Q2_Var13	0.000582
520	Q4_Var3	0.000644	562	Var178	0.000613	604	Var2	0.00058
521	Var26	0.000644	563	Q4_Var224	0.000612	605	Q2_Var246	0.00058
522	Q4_Var84	0.000644	564	Q4_Var338	0.000612	606	Q2_Var76	0.00058
523	Q2_Var34	0.000644	565	Q2_Var71	0.000611	607	Q2_Var357	0.00058
524	Var330	0.000642	566	Var112	0.000611	608	Var220	0.000578
525	Q2_Var276	0.000642	567	Var194	0.00061	609	Q4_Var276	0.000578
526	Q2_Var237	0.000641	568	Q2_Var149	0.000608	610	Q4_Var7	0.000578
527	Var243	0.00064	569	Q2_Var181	0.000608	611	Var239	0.000576
528	Var299	0.00064	570	Q2_Var337	0.000607	612	Q4_Var87	0.000576
529	Q2_Var102	0.000639	571	Q2_Var304	0.000607	613	Q4_Var88	0.000575
530	Var274	0.000638	572	Q4_Var77	0.000607	614	Q2_Var92	0.000575
531	Q2_Var113	0.000638	573	Q4_Var353	0.000606	615	Var98	0.000575
532	Q2_Var214	0.000636	574	Q2_Var170	0.000605	616	Q2_Var248	0.000574
533	Q4_Var245	0.000636	575	Q4_Var74	0.000604	617	Q2_Var38	0.000574
534	Q4_Var183	0.000635	576	Q4_Var78	0.000604	618	Var246	0.000573
535	Q4_Var142	0.000635	577	Q2_Var133	0.000603	619	Q4_Var145	0.000572
536	Q2_Var18	0.000633	578	Var173	0.000602	620	Q2_Var336	0.000572
537	Q2_Var104	0.000633	579	Var335	0.000601	621	Q2_Var154	0.000572
538	Var142	0.000633	580	Var91	0.0006	622	Var262	0.000572
539	Q4_Var189	0.000633	581	Var85	0.000597	623	Var235	0.000571
540	Q2_Var359	0.000631	582	Var350	0.000596	624	Q4_Var64	0.000571
541	Q2_Var243	0.000631	583	Q4_Var324	0.000595	625	Var334	0.000569
542	Q2_Var130	0.000631	584	Q2_Var116	0.000595	626	Q4_Var89	0.000568
543	Q2_Var151	0.000631	585	Var94	0.000593	627	Var82	0.000567
544	Q4_Var69	0.000631	586	Var307	0.000592	628	Q4_Var79	0.000567
545	Q4_Var155	0.00063	587	Q2_Var88	0.000592	629	Q4_Var68	0.000567
546	Q2_Var44	0.00063	588	Var103	0.000591	630	Q4_Var137	0.000566

Table B.7 (f) Feature Importance Ranking (Rank 631–756)

Rank	Feature	Importance	Rank	Feature	Importance	Rank	Feature	Importance
631	Var136	0.000566	673	Var109	0.000534	715	Q2_Var82	0.000506
632	Q4_Var196	0.000564	674	Q4_Var71	0.000532	716	Q2_Var162	0.000506
633	Q2_Var193	0.000564	675	Var165	0.000532	717	Q2_Var172	0.000506
634	Var263	0.000564	676	Q4_Var197	0.000532	718	Q4_Var86	0.000505
635	Q2_Var65	0.000564	677	Var111	0.000532	719	Q2_Var326	0.000505
636	Q2_Var242	0.000563	678	Q4_Var323	0.000531	720	Q4_Var208	0.000504
637	Var161	0.000562	679	Q2_Var210	0.000531	721	Q4_Var172	0.000503
638	Q2_Var77	0.000561	680	Q4_Var300	0.000529	722	Q2_Var174	0.000502
639	Q2_Var250	0.000561	681	Var321	0.000529	723	Var177	0.000502
640	Q4_Var120	0.000559	682	Var138	0.000529	724	Q4_Var238	0.000502
641	Q4_Var177	0.000559	683	Q4_Var198	0.000527	725	Q4_Var107	0.000502
642	Var273	0.000558	684	Q2_Var66	0.000527	726	Q2_Var303	0.0005
643	Q2_Var287	0.000558	685	Q4_Var103	0.000527	727	Q2_Var94	0.0005
644	Q4_Var97	0.000557	686	Q4_Var277	0.000526	728	Q4_Var258	0.0005
645	Q4_Var201	0.000557	687	Var137	0.000525	729	Q4_Var234	0.000499
646	Q2_Var106	0.000556	688	Q2_Var164	0.000524	730	Q4_Var36	0.000499
647	Q4_Var193	0.000556	689	Q2_Var146	0.000524	731	Q2_Var143	0.000498
648	Q4_Var73	0.000555	690	Q4_Var26	0.000524	732	Q4_Var296	0.000498
649	Q2_Var103	0.000554	691	Var219	0.000523	733	Q4_Var94	0.000498
650	Q4_Var203	0.000554	692	Q2_Var165	0.000523	734	Q4_Var206	0.000497
651	Q2_Var36	0.000551	693	Q2_Var27	0.000523	735	Q4_Var325	0.000497
652	Q2_Var241	0.00055	694	Q4_Var241	0.000522	736	Q4_Var83	0.000497
653	Var108	0.00055	695	Q2_Var110	0.000522	737	Q2_Var91	0.000497
654	Q4_Var50	0.000548	696	Var143	0.000522	738	Q4_Var244	0.000497
655	Q2_Var329	0.000548	697	Q2_Var206	0.00052	739	Q4_Var31	0.000496
656	Q2_Var247	0.000548	698	Q4_Var253	0.00052	740	Q4_Var260	0.000496
657	Q4_Var146	0.000547	699	Q4_Var45	0.000519	741	Q2_Var347	0.000496
658	Q2_Var107	0.000547	700	Q4_Var341	0.000519	742	Var232	0.000496
659	Q4_Var357	0.000546	701	Q4_Var91	0.000518	743	Q4_Var117	0.000496
660	Q4_Var199	0.000543	702	Var188	0.000517	744	Q2_Var202	0.000495
661	Q4_Var214	0.000542	703	Q4_Var343	0.000517	745	Var97	0.000495
662	Q2_Var225	0.000542	704	Q4_Var128	0.000516	746	Q2_Var87	0.000494
663	Q4_Var218	0.00054	705	Q2_Var270	0.000516	747	Var225	0.000494
664	Q4_Var275	0.000539	706	Q2_Var228	0.000515	748	Q4_Var67	0.000494
665	Var214	0.000539	707	Q4_Var65	0.000513	749	Q2_Var254	0.000493
666	Q4_Var33	0.000539	708	Q2_Var83	0.000513	750	Q2_Var335	0.000492
667	Q4_Var13	0.000539	709	Var71	0.000511	751	Var87	0.000492
668	Q2_Var78	0.000538	710	Q4_Var131	0.000508	752	Q2_Var280	0.000492
669	Q2_Var289	0.000538	711	Q2_Var90	0.000508	753	Var268	0.000492
670	Q4_Var27	0.000536	712	Q2_Var14	0.000507	754	Q4_Var247	0.000491
671	Q4_Var76	0.000535	713	Var310	0.000507	755	Q4_Var105	0.000491
672	Q4_Var358	0.000534	714	Q4_Var280	0.000506	756	Q4_Var102	0.00049

Table B.8 (g) Feature Importance Ranking (Rank 757–882)

Rank	Feature	Importance	Rank	Feature	Importance	Rank	Feature	Importance
757	Q2_Var208	0.000489	799	Q4_Var14	0.000469	841	Q4_Var39	0.00044
758	Q2_Var89	0.000488	800	Q2_Var231	0.000469	842	Q4_Var240	0.00044
759	Q4_Var223	0.000488	801	Var215	0.000468	843	Q2_Var274	0.00044
760	Q2_Var79	0.000485	802	Q2_Var204	0.000468	844	Var110	0.00044
761	Q4_Var42	0.000485	803	Q4_Var204	0.000468	845	Q4_Var165	0.000436
762	Q2_Var7	0.000484	804	Q4_Var175	0.000466	846	Q4_Var205	0.000435
763	Q4_Var243	0.000484	805	Var125	0.000466	847	Q2_Var291	0.000434
764	Var141	0.000484	806	Q4_Var46	0.000466	848	Q4_Var339	0.000434
765	Var124	0.000483	807	Q4_Var11	0.000465	849	Q2_Var240	0.000433
766	Q4_Var82	0.000483	808	Q2_Var131	0.000465	850	Q2_Var100	0.000432
767	Q4_Var85	0.000483	809	Q2_Var328	0.000464	851	Q4_Var15	0.000431
768	Var29	0.000483	810	Q2_Var295	0.000463	852	Q2_Var48	0.000431
769	Q2_Var188	0.000482	811	Var76	0.000463	853	Q4_Var327	0.000431
770	Q4_Var356	0.00048	812	Q4_Var194	0.000463	854	Q2_Var156	0.000431
771	Q4_Var126	0.00048	813	Var271	0.000463	855	Q2_Var330	0.000431
772	Var267	0.00048	814	Q4_Var335	0.000462	856	Q4_Var216	0.00043
773	Var175	0.00048	815	Q2_Var160	0.000462	857	Q2_Var180	0.000429
774	Q2_Var200	0.00048	816	Q4_Var127	0.000459	858	Q4_Var228	0.000427
775	Q4_Var114	0.000478	817	Var86	0.000459	859	Var261	0.000426
776	Q4_Var106	0.000478	818	Q4_Var207	0.000458	860	Q2_Var142	0.000426
777	Q2_Var301	0.000478	819	Q2_Var97	0.000456	861	Q4_Var115	0.000425
778	Q4_Var90	0.000478	820	Q4_Var285	0.000455	862	Q4_Var232	0.000425
779	Q2_Var179	0.000477	821	Q2_Var269	0.000455	863	Q2_Var235	0.000424
780	Q4_Var8	0.000477	822	Q2_Var307	0.000455	864	Var265	0.000424
781	Q2_Var219	0.000477	823	Q2_Var144	0.000454	865	Q4_Var70	0.000423
782	Q4_Var344	0.000476	824	Q4_Var250	0.000454	866	Q4_Var41	0.000423
783	Q4_Var48	0.000476	825	Q4_Var254	0.000452	867	Q4_Var231	0.000423
784	Q4_Var349	0.000476	826	Q2_Var2	0.000451	868	Q4_Var163	0.000422
785	Var114	0.000476	827	Q4_Var168	0.00045	869	Q4_Var49	0.000422
786	Q4_Var289	0.000474	828	Q4_Var186	0.000449	870	Q2_Var325	0.000421
787	Q4_Var328	0.000474	829	Q4_Var121	0.000448	871	Q4_Var29	0.000421
788	Q4_Var93	0.000474	830	Q4_Var72	0.000448	872	Var296	0.000419
789	Q4_Var355	0.000474	831	Q2_Var309	0.000448	873	Q4_Var246	0.000418
790	Q4_Var217	0.000473	832	Q4_Var124	0.000447	874	Q4_Var202	0.000418
791	Q2_Var297	0.000473	833	Q4_Var348	0.000446	875	Q4_Var298	0.000417
792	Q4_Var129	0.000473	834	Q2_Var215	0.000445	876	Q2_Var216	0.000417
793	Q4_Var109	0.000473	835	Q2_Var42	0.000445	877	Var295	0.000417
794	Q4_Var180	0.000472	836	Q4_Var237	0.000445	878	Q4_Var113	0.000416
795	Q4_Var345	0.000471	837	Var226	0.000443	879	Q4_Var295	0.000416
796	Var96	0.000471	838	Q4_Var141	0.000443	880	Q4_Var220	0.000415
797	Q4_Var322	0.000469	839	Var287	0.000443	881	Q2_Var148	0.000414
798	Var209	0.000469	840	Q4_Var286	0.000442	882	Q4_Var147	0.000413

Table B.9 (h) Feature Importance Ranking (Rank 883–1008)

Rank	Feature	Importance	Rank	Feature	Importance	Rank	Feature	Importance
883	Q2_Var232	0.000412	925	Q4_Var309	0.000385	967	Q4_Var301	0.000357
884	Q4_Var257	0.000411	926	Q4_Var236	0.000385	968	Q2_Var238	0.000356
885	Var302	0.000411	927	Var249	0.000384	969	Q4_Var293	0.000355
886	Q2_Var299	0.00041	928	Q4_Var174	0.000382	970	Var80	0.000355
887	Q4_Var329	0.00041	929	Q4_Var200	0.000382	971	Var331	0.000355
888	Q2_Var157	0.000409	930	Q4_Var340	0.000381	972	Q2_Var261	0.000355
889	Q4_Var111	0.000409	931	Q4_Var291	0.000381	973	Q2_Var327	0.000353
890	Q2_Var273	0.000409	932	Q2_Var223	0.00038	974	Var264	0.000352
891	Q4_Var230	0.000408	933	Q2_Var10	0.000379	975	Q2_Var101	0.000351
892	Q2_Var125	0.000407	934	Q2_Var362	0.000378	976	Var332	0.00035
893	Q2_Var111	0.000407	935	Q4_Var210	0.000377	977	Q4_Var144	0.00035
894	Q2_Var226	0.000407	936	Q4_Var362	0.000377	978	Q4_Var188	0.00035
895	Q2_Var98	0.000407	937	Q2_Var109	0.000377	979	Q4_Var6	0.000349
896	Q4_Var132	0.000407	938	Q2_Var239	0.000377	980	Q2_Var262	0.000346
897	Q4_Var248	0.000407	939	Q4_Var16	0.000376	981	Q4_Var311	0.000345
898	Q2_Var338	0.000406	940	Q2_Var230	0.000375	982	Q2_Var236	0.000345
899	Var174	0.000406	941	Q4_Var304	0.000374	983	Q4_Var215	0.000344
900	Q4_Var169	0.000406	942	Q2_Var161	0.000374	984	Q4_Var337	0.000339
901	Var270	0.000405	943	Q2_Var265	0.000373	985	Q4_Var235	0.000336
902	Q4_Var136	0.000404	944	Q2_Var272	0.000373	986	Q4_Var350	0.000334
903	Q4_Var225	0.000404	945	Q4_Var219	0.000373	987	Q4_Var157	0.000333
904	Q4_Var154	0.000403	946	Q4_Var161	0.000372	988	Q2_Var267	0.000331
905	Q2_Var153	0.000402	947	Q4_Var2	0.000372	989	Q4_Var151	0.000329
906	Q4_Var95	0.000401	948	Q4_Var333	0.000372	990	Q4_Var297	0.000327
907	Q2_Var15	0.0004	949	Var297	0.000372	991	Q4_Var303	0.000327
908	Q2_Var311	0.0004	950	Q2_Var158	0.000372	992	Q2_Var105	0.000326
909	Q2_Var126	0.0004	951	Q2_Var141	0.000371	993	Var81	0.000326
910	Q4_Var270	0.000399	952	Q4_Var330	0.000371	994	Var266	0.000326
911	Q4_Var181	0.000399	953	Q4_Var321	0.000366	995	Q4_Var162	0.000325
912	Q4_Var35	0.000398	954	Q4_Var271	0.000365	996	Q4_Var44	0.000323
913	Q4_Var148	0.000398	955	Q4_Var272	0.000364	997	Q2_Var95	0.000323
914	Q4_Var274	0.000397	956	Q4_Var302	0.000364	998	Q4_Var81	0.000322
915	Q2_Var293	0.000397	957	Q2_Var9	0.000364	999	Q4_Var133	0.000321
916	Q4_Var334	0.000396	958	Q4_Var239	0.000364	1000	Q4_Var326	0.00032
917	Q4_Var110	0.000394	959	Q2_Var345	0.000362	1001	Q2_Var118	0.000318
918	Q2_Var321	0.000392	960	Q2_Var271	0.000361	1002	Q4_Var261	0.000318
919	Q4_Var160	0.000392	961	Q4_Var101	0.00036	1003	Q4_Var139	0.000318
920	Q2_Var340	0.000391	962	Q4_Var287	0.00036	1004	Q4_Var259	0.000316
921	Q2_Var334	0.000389	963	Q2_Var263	0.00036	1005	Q4_Var262	0.000315
922	Q4_Var118	0.000386	964	Q2_Var114	0.00036	1006	Q4_Var158	0.000313
923	Q4_Var242	0.000386	965	Q2_Var124	0.000359	1007	Q4_Var269	0.000311
924	Q2_Var186	0.000386	966	Q4_Var178	0.000358	1008	Q4_Var80	0.000309

Table B.10 (i) Feature Importance Ranking (Rank 1009–1066)

Rank	Feature	Importance	Rank	Feature	Importance	Rank	Feature	Importance
1009	Q4_Var249	0.000306	1051	Q2_Var366	0.000108			
1010	Q2_Var49	0.000305	1052	Q2_Var368	0.000107			
1011	Q4_Var12	0.000305	1053	Q4_Var365	0.000106			
1012	Q2_Var175	0.000303	1054	Q4_Var360	0.000084			
1013	Q2_Var96	0.000302	1055	Q2_Var365	0.000083			
1014	Q2_Var264	0.000297	1056	Var367	0.000074			
1015	Q4_Var153	0.000297	1057	Var366	0.000072			
1016	Q2_Var6	0.000295	1058	Q2_Var367	0.000062			
1017	Q4_Var143	0.000292	1059	Q2_Var364	0.000062			
1018	Q2_Var249	0.000292	1060	Var368	0.000061			
1019	Q4_Var138	0.000291	1061	Q4_Var367	0.000058			
1020	Q2_Var127	0.000287	1062	Q4_Var364	0.000053			
1021	Q4_Var156	0.000286	1063	Var364	0.000033			
1022	Q4_Var116	0.000286	1064	Var365	0.000032			
1023	Q4_Var179	0.000285	1065	Var360	0.000021			
1024	Q2_Var137	0.000284	1066	Q2_Var360	0.000016			
1025	Q2_Var138	0.000272						
1026	Q2_Var39	0.000271						
1027	Q4_Var273	0.000269						
1028	Q4_Var265	0.000268						
1029	Q2_Var85	0.000267						
1030	Q2_Var268	0.000267						
1031	Q2_Var333	0.000264						
1032	Q4_Var266	0.000261						
1033	Q2_Var81	0.000261						
1034	Q2_Var266	0.00026						
1035	Q4_Var307	0.000259						
1036	Q4_Var263	0.000252						
1037	Q4_Var267	0.000242						
1038	Q2_Var29	0.000231						
1039	Q4_Var164	0.00023						
1040	Q2_Var80	0.000226						
1041	Q4_Var264	0.00022						
1042	Q4_Var170	0.000217						
1043	Q2_Var115	0.000213						
1044	Q2_Var332	0.000207						
1045	Q4_Var331	0.0002						
1046	Q4_Var268	0.000186						
1047	Q4_Var332	0.000183						
1048	Q2_Var331	0.000175						
1049	Q4_Var368	0.000163						
1050	Q4_Var366	0.000125						

Table B.11 Base Random Forest Parameter Sensitivity Analysis

Method	N_ESTIMATORS	MAX_DEPTH	MIN_SAMPLES_SPLIT	MIN_SAMPLES_LEAF	CV_F1_Mean	CV_F1_Std	CV_Precision_Mean	CV_Precision_Std	CV_Recall_Mean	CV_Recall_Std	CV_MCC_Mean	CV_MCC_Std	CV_Time_Mean
RF1	200	20	15	15	0.61	0.0135	0.48	0.0114	0.84	0.0220	0.61	0.0153	511.56
Base_RF_2	200	5	10	5	0.56	0.0124	0.43	0.0106	0.82	0.0244	0.56	0.0144	292.69
Base_RF_3	200	10	5	10	0.62	0.0142	0.49	0.0118	0.85	0.0231	0.62	0.0162	459.52
Base_RF_4	100	5	2	10	0.56	0.0117	0.43	0.0108	0.82	0.0204	0.56	0.0129	147.54
Base_RF_5	100	10	10	15	0.61	0.0165	0.47	0.0137	0.84	0.0243	0.60	0.0185	222.45
Base_RF_6	200	10	15	5	0.63	0.0108	0.50	0.0080	0.85	0.0239	0.63	0.0131	482.81
Base_RF_7	100	20	2	10	0.62	0.0124	0.49	0.0088	0.85	0.0227	0.62	0.0147	282.60
Base_RF_8	100	10	15	1	0.64	0.0128	0.52	0.0114	0.85	0.0211	0.64	0.0144	259.13
Base_RF_9	150	20	10	15	0.61	0.0155	0.48	0.0132	0.84	0.0206	0.61	0.0172	394.89
Base_RF_10	150	10	2	15	0.61	0.0178	0.47	0.0180	0.84	0.0225	0.60	0.0187	342.15
Base_RF_11	150	10	10	5	0.64	0.0165	0.51	0.0146	0.85	0.0232	0.63	0.0183	377.79
Base_RF_12	150	10	10	5	0.64	0.0165	0.51	0.0174	0.86	0.0211	0.64	0.0173	378.96
Base_RF_13	200	15	10	15	0.61	0.0146	0.48	0.0131	0.84	0.0233	0.61	0.0164	518.77
Base_RF_14	200	20	5	5	0.64	0.0108	0.51	0.0104	0.87	0.0239	0.64	0.0127	634.47
Base_RF_15	150	15	5	10	0.63	0.0122	0.49	0.0100	0.85	0.0202	0.62	0.0140	414.89

Base_RF_16	100	15	15	10	0.63	0.0178	0.50	0.0171	0.86	0.0223	0.63	0.0190	280.49
Base_RF_17	200	10	10	1	0.64	0.0169	0.51	0.0142	0.85	0.0264	0.64	0.0190	522.65
Base_RF_18	100	15	2	10	0.62	0.0123	0.49	0.0095	0.85	0.0216	0.62	0.0143	279.99
Base_RF_19	100	15	10	15	0.61	0.0154	0.48	0.0153	0.84	0.0201	0.61	0.0163	262.13
Base_RF_20	100	15	15	1	0.64	0.0124	0.51	0.0129	0.87	0.0213	0.64	0.0135	323.01

Table B.12 Overall Performance Across Parameter Combinations

NESTIMATORS	MAX_DEPTH	MIN_SAMPLES_SPLIT	MIN_SAMPLES_LEAF	CV_F1_Mean	CV_Precision_Mean	CV_Recall_Mean	CV_MCC_Mean
50	5	2	1	0.62	0.53	0.75	0.61
50	5	2	5	0.62	0.53	0.76	0.61
50	5	2	10	0.62	0.53	0.75	0.61
50	5	5	1	0.62	0.53	0.76	0.61
50	5	5	5	0.62	0.53	0.76	0.61
50	5	5	10	0.62	0.53	0.76	0.61
50	5	10	1	0.63	0.54	0.76	0.61
50	5	10	5	0.63	0.54	0.75	0.61
50	5	10	10	0.63	0.53	0.76	0.61
50	10	2	1	0.72	0.65	0.80	0.71
50	10	2	5	0.71	0.64	0.79	0.69
50	10	2	10	0.70	0.63	0.79	0.69
50	10	5	1	0.72	0.66	0.80	0.71
50	10	5	5	0.71	0.64	0.80	0.70
50	10	5	10	0.70	0.62	0.80	0.68
50	10	10	1	0.72	0.65	0.80	0.70
50	10	10	5	0.71	0.63	0.80	0.69
50	10	10	10	0.69	0.62	0.79	0.68
50	15	2	1	0.73	0.67	0.82	0.72
50	15	2	5	0.72	0.64	0.81	0.70
50	15	2	10	0.70	0.62	0.80	0.69
50	15	5	1	0.73	0.66	0.82	0.72
50	15	5	5	0.72	0.64	0.82	0.70
50	15	5	10	0.70	0.62	0.81	0.69
50	15	10	1	0.73	0.65	0.82	0.71
50	15	10	5	0.72	0.64	0.81	0.70
50	15	10	10	0.70	0.63	0.80	0.69
100	5	2	1	0.63	0.54	0.76	0.62
100	5	2	5	0.62	0.53	0.75	0.60
100	5	2	10	0.63	0.54	0.75	0.61
100	5	5	1	0.63	0.54	0.75	0.61
100	5	5	5	0.63	0.54	0.76	0.61
100	5	5	10	0.62	0.53	0.76	0.61
100	5	10	1	0.63	0.54	0.75	0.61
100	5	10	5	0.63	0.54	0.76	0.61
100	5	10	10	0.62	0.53	0.75	0.61
100	10	2	1	0.72	0.65	0.80	0.70
100	10	2	5	0.72	0.65	0.80	0.70
100	10	2	10	0.70	0.63	0.79	0.69
100	10	5	1	0.72	0.65	0.81	0.71
100	10	5	5	0.72	0.65	0.80	0.70
100	10	5	10	0.70	0.62	0.79	0.68
100	10	10	1	0.71	0.65	0.80	0.70
100	10	10	5	0.71	0.64	0.80	0.70

100	10	10	10	0.70	0.63	0.79	0.69
100	15	2	1	0.74	0.68	0.82	0.73
100	15	2	5	0.72	0.66	0.81	0.71
100	15	2	10	0.71	0.63	0.80	0.69
100	15	5	1	0.73	0.66	0.82	0.72
100	15	5	5	0.72	0.65	0.81	0.71
100	15	5	10	0.71	0.64	0.80	0.69
100	15	10	1	0.73	0.66	0.82	0.72
100	15	10	5	0.72	0.65	0.81	0.71
100	15	10	10	0.71	0.63	0.80	0.69
150	5	2	1	0.63	0.55	0.76	0.62
150	5	2	5	0.63	0.54	0.75	0.61
150	5	2	10	0.63	0.54	0.75	0.61
150	5	5	1	0.64	0.55	0.75	0.62
150	5	5	5	0.63	0.54	0.75	0.61
150	5	5	10	0.63	0.54	0.76	0.61
150	5	10	1	0.63	0.54	0.75	0.61
150	5	10	5	0.63	0.54	0.75	0.61
150	5	10	10	0.63	0.55	0.75	0.62
150	10	2	1	0.72	0.65	0.80	0.70
150	10	2	5	0.71	0.64	0.80	0.70
150	10	2	10	0.70	0.63	0.80	0.69
150	10	5	1	0.72	0.66	0.80	0.71
150	10	5	5	0.71	0.63	0.81	0.70
150	10	5	10	0.69	0.61	0.80	0.68
150	10	10	1	0.72	0.65	0.80	0.70
150	10	10	5	0.71	0.64	0.80	0.70
150	10	10	10	0.70	0.62	0.80	0.68
150	15	2	1	0.74	0.67	0.83	0.73
150	15	2	5	0.72	0.65	0.81	0.71
150	15	2	10	0.71	0.63	0.80	0.69
150	15	5	1	0.74	0.67	0.83	0.73
150	15	5	5	0.73	0.65	0.82	0.71
150	15	5	10	0.70	0.63	0.81	0.69
150	15	10	1	0.73	0.65	0.83	0.72
150	15	10	5	0.72	0.65	0.81	0.71
150	15	10	10	0.71	0.63	0.81	0.70