HEC MONTRÉAL

Understanding Financial Distress :
A Cost-Sensitive, Explainable AI Approach to Bankruptcy Prediction

par

Amal Koodoruth

Dr Denis Larocque
HEC Montréal
Directeur de recherche

Sciences de la gestion
(Spécialisation Data Science and Business Analytics)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences
(M. Sc.)

April 2025

Résumé

Naviguer dans l'incertitude financière d'une entreprise, c'est un peu comme naviguer sur des eaux imprévisibles. Les investisseurs jouent le rôle de navigateurs, les auditeurs celui de vérificateurs, et les régulateurs surveillent la situation depuis la côte. Lorsque des signes de détresse apparaissent, toutes ces parties prenantes ont besoin d'un signal clair, suffisamment tôt pour agir et suffisamment transparent pour inspirer confiance. Cette thèse développe un système d'alerte précoce basé sur l'apprentissage automatique pour détecter les difficultés financières à différents cycles économique, bien avant que la faillite ne survienne. L'étude évalue divers modèles dans des conditions économiques variées, en se concentrant sur des algorithmes de boosting de pointe ainsi que sur des méthodes de référence sans boosting. Le coût d'une mauvaise classification, étant différent selon la partie prenante, les régulateurs craignant les faux négatifs et les prêteurs les faux positifs, les modèles sont entraînés à l'aide de stratégies d'apprentissage sensibles aux coûts qui tiennent compte de ces asymétries. En ajustant la pénalité pour les faux négatifs par rapport aux faux positifs, le système peut refléter les différentes priorités et tolérances au risque des parties prenantes. Les modèles ont conservé d'excellentes performances, même sur des données inédites, de 2020 à 2023, une période marquée par des perturbations économiques, soulignant leur capacité à généraliser au-delà des conditions d'entraînement. Cependant, la prédiction seule ne suffit pas. En eaux incertaines, anticiper l'arrivée d'une tempête n'est qu'une partie du défi. Comprendre sa direction permet une navigation sûre et stratégique. De même, dans la prise de décision financière, les parties prenantes ont besoin de plus qu'un signal binaire; elles doivent com-

prendre le raisonnement derrière. Cette thèse aborde ce problème en mettant l'accent sur l'interprétabilité, en utilisant des techniques telles que les valeurs SHAP et les graphiques ICE pour faire apparaître les caractéristiques financières qui guident chaque prédiction. Ces informations clarifient le raisonnement du modèle, le transformant d'une boîte noire en une aide à la navigation. Ce cadre devient alors plus qu'un outil prédictif: il sert de système d'orientation partagé. Dans un environnement où le coût de l'inaction peut être catastrophique, il offre aux décideurs un signal à la fois opportun et transparent, les aidant non seulement à rester à flot, mais aussi à naviguer de manière décisive dans l'incertitude financière.

Mots-clés

Prévision de faillite, apprentissage automatique en finance, IA explicable, déséquilibre des classes, apprentissage sensible aux coûts, IA centrée sur les parties prenantes, systèmes d'alerte précoce, algorithmes de boosting

Méthodes de recherche

Cette thèse est organisée comme suit : nous commençons par l'introduction, qui décrit la motivation de l'étude, définit la problématique de recherche et présente les objectifs principaux. Le chapitre suivant fournit une revue de la littérature pertinente sur la prédiction des faillites et l'apprentissage automatique explicatif. Le chapitre 1 présente le contexte théorique, détaillant les techniques de prédiction et d'interprétabilité employées dans l'étude. Le chapitre 2 décrit l'ensemble de données, y compris les processus de collecte, de construction et d'étiquetage des données. Le chapitre 3 décrit la méthodologie utilisée pour le développement, l'évaluation et l'analyse d'explicabilité du modèle. Le chapitre 4 présente les résultats sur la performance du modèle dans différentes configurations de prévalence et de coût des faillites. Le chapitre 5 présente les résultats d'explicabilité, en se concentrant sur les modèles les plus performants identifiés par l'évaluation prédictive. Le

chapitre 6 propose une discussion des résultats, les reliant aux objectifs de la recherche. Le chapitre 7 conclut la thèse, souligne les limites et esquisse des pistes de recherche futures.

Abstract

Navigating a company through financial uncertainty is much like steering a ship through unpredictable waters. Investors act as navigators, auditors serve as map-checkers, and regulators monitor from the coast. When signs of distress begin to emerge, all these stakeholders need a clear signal early enough to act, and transparent enough to trust. This thesis develops a machine learning–based early warning system to detect financial distress across different stages of the economic cycle, well before bankruptcy strikes. The study evaluates a range of models under varying economic conditions, focusing on state-of-the-art boosting algorithms as well as baseline non-boosting methods. Since the cost of misclassification carries different weight depending on the stakeholder, with regulators fearing false negatives and lenders worrying about false positives, the models are trained using cost-sensitive learning strategies that account for these asymmetries. By adjusting the penalty for false negatives relative to false positives, the system can reflect diverse stakeholder priorities and risk tolerances. The models maintained strong performance even on unseen data from 2020 to 2023, a period marked by economic disruption, highlighting their ability to generalize beyond training conditions. However, prediction alone is not enough. In uncertain waters, knowing a storm is coming is only part of the challenge. Understanding which direction it's coming from is what allows for safe and strategic navigation. Similarly, in financial decision-making, stakeholders need more than a binary signal; they need to see the reasoning behind it. This thesis addresses this by placing emphasis on interpretability, using techniques such as SHAP values and ICE plots to surface the financial features that guide each prediction. These insights bring

clarity to the model's reasoning, transforming it from a black box into a navigational aid. Taken together, this framework becomes more than a predictive tool: it serves as a shared system of orientation. In an environment where the cost of inaction can be catastrophic, it offers decision-makers a signal that is both timely and transparent, helping them not just stay afloat, but steer decisively through financial uncertainty.

Keywords

Bankruptcy Prediction, Machine Learning in Finance, Explainable AI, Class Imbalance, Cost-sensitive Learning, Stakeholder-centered AI, Early Warning Systems, Boosting Algorithms

Research Methods

This thesis is organized as follows: We start with the introduction, which outlines the motivation for the study, defines the research problem, and presents the core objectives. The next chapter provides a review of the relevant literature on bankruptcy prediction and explainable machine learning. Chapter 1 presents the theoretical background, detailing the predictive and interpretability techniques employed in the study. Chapter 2 describes the dataset, including the processes of data collection, dataset construction, and labeling. Chapter 3 outlines the methodology used for model development, evaluation, and explainability analysis. Chapter 4 reports the results on model performance across different bankruptcy prevalence and cost configurations. Chapter 5 presents the explainability results, focusing on the most performant models as identified through predictive evaluation. Chapter 6 offers a discussion of the findings, linking them back to the research objectives. Chapter 7 concludes the thesis, highlights limitations and outlines avenues for future research.

Contents

List of Tables

List of Figures

List of Acronyms

AdaBoost Adaptative Boosting

CQ Calendar Quarter

DT Decision Tree

EBIT Earnings Before Interest and Tax

FN False Negative

FP False Positive

ICE Individual Conditional Expectation

LightGBM Light Gradient Boosting Machine

LR Logistic Regression

MKTCAP Market Capitalization

ML Machine Learning

MLP Multi Layer Perceptron

PDP Partial Dependence Plot

PFI Permutation-based Feature Importance

RE Retained Earnings

RF Random Forest

SEC Securities and Exchange Commission

TA Total Assets

TL Total Liabilities

TR Total Revenue

WC Working Capital

XAI Explainable AI

XGBoost Extreme Gradient Boosting

Acknowledgements

This thesis would not have been possible without the support and contributions of several individuals, whom I gratefully acknowledge.

First and foremost, my heartfelt thanks are due to my supervisor, Dr Denis Larocque, for his unwavering support throughout the research process. His guidance, constructive feedback, and timely answers to my questions were invaluable in shaping both the direction and quality of this thesis.

I am also deeply grateful to Sebastien Fiaux, CEO and Founder of SFJ Technologies, for his support at multiple stages of this project. He played a key role in helping me navigate the initial literature review and refine the financial theory underpinning this work.

I am truly blessed to have the most supportive parents and sister. No matter how busy they were, they always made time for me, especially when I needed a few encouraging words. Thank you for knowing exactly when to change the subject to football or Mauritian politics, just to help me reset and breathe. You made the most stressful moments more manageable. You are amazing.

A special thank you to my girlfriend for being there for me and bearing with me. From listening patiently to my rambling ideas to keeping me grounded during moments of doubt, your support has meant more than I can express.

Finally, I'd like to thank my roommate, friends, and classmates for being around especially when I needed a break. Whether it was to hang out, share a laugh, or play a much-needed football match, your company made this journey a lot more enjoyable.

Use of Generative AI Tools

During the course of this research, OpenAI's ChatGPT models were used in a limited and supplementary capacity. Preliminary code for data analysis and explainability methods was initially generated using AI and then rigorously tested, validated, and modified by the author to ensure correctness and alignment with the research objectives. Additionally, generative AI was used to assist in rewriting portions of the body text, particularly after the author had completed the initial analysis. In all such cases, the generated text was critically reviewed and edited by the author. All substantive ideas, experimental design decisions, and analytical interpretations remain the intellectual work of the author.

Introduction

Background & Motivation

Corporate bankruptcy is a high-impact event with profound implications for a wide range of stakeholders. For investors, it can result in substantial financial losses and destabilize investment portfolios. Creditors and lenders are exposed to heightened default risk and diminished loan recoverability. Regulatory bodies bear the responsibility of mitigating systemic risks and safeguarding the integrity of financial markets. For corporate management, the timely identification of financial distress can enable proactive measures such as cost restructuring, strategic mergers, or turnaround planning to avoid failure. Given these wide-ranging consequences, the early and accurate detection of financial distress or impending bankruptcy is not only valuable but essential for informed decision-making across the financial ecosystem.

Traditional bankruptcy prediction models, such as Beaver's univariate analysis (Beaver 1966), Altman's Z-score (E. I. Altman 1968), and Ohlson's O-score (Ohlson 1980), have historically relied on a static set of accounting ratios derived from annual financial statements. While these models are interpretable and grounded in economic theory, they suffer from several limitations.

First, they do not use any historical data. These static models use financial information from a single fiscal year or average across multiple years, making them relatively unresponsive to rapid changes in firm health. Moreover, because financial reports are published with a lag, often several months after the reporting period ends, these models

effectively base their predictions on outdated information. In practice, this can result in predictions being delayed, limiting their usefulness as early warning systems for financial distress.

Second, these models often exhibit rigidity due to their reliance on linear relationships between predictors and outcomes. Such assumptions fail to capture the nonlinear, complex interactions among financial indicators that frequently characterize corporate distress in modern economies. As corporate risk dynamics become increasingly complex, linear modeling frameworks risk oversimplifying the relationships underlying bankruptcy events.

Third, generalization remains a significant concern. Many traditional models were developed using datasets from specific industries, regions, or historical periods, limiting their applicability to today's diverse and rapidly evolving market conditions. Changes in industry structures, financial practices, disclosure standards, and macroeconomic environments further erode the predictive reliability of models that are not regularly updated or adapted.

Finally, studies artificially balanced their datasets. While this simplification facilitated statistical modeling, it substantially overestimated predictive performance and failed to replicate the natural class imbalance observed in real-world settings, where bankruptcies are rare events. As a result, such models often perform poorly when deployed in practical environments.

Together, these limitations underscore the need for more dynamic, flexible, and context-aware modeling approaches to bankruptcy prediction—approaches that can account for the timeliness of data, capture nonlinear dependencies, and generalize across different economic contexts. Addressing these gaps forms a central motivation for the modeling strategies adopted in this thesis.

In recent years, machine learning (ML) models have gained prominence in financial risk analytics due to their capacity to learn complex patterns from large datasets. Boosting models such as XGBoost (Chen and Guestrin 2016) and LightGBM (Ke et al. 2017), as well as deep learning approaches, have demonstrated superior performance over tra-

ditional statistical models in various financial applications, including credit scoring and fraud detection. However, these models are frequently criticized for being black boxes (Balcaen and Ooghe 2006), that is, the decision-making processes is not interpretable. This hinders their adoption in high-stakes domains where interpretability is critical for trust, compliance, and accountability.

As such, the intersection of high-performing machine learning techniques and explainable artificial intelligence (XAI) represents a promising yet underexplored avenue for improving bankruptcy prediction. It enables not only accurate forecasts but also actionable insights into the financial and behavioral signals driving firm-level insolvency risk.

Problem Statement

Despite the promising performance of machine learning models in bankruptcy prediction, several key challenges persist. One major concern is class imbalance: in most real-world datasets, bankrupt firms represent only a small minority, often less than 5% of all firms. This imbalance leads to biased models that are prone to misclassifying the rare but critical bankrupt cases as healthy, especially when trained using conventional loss functions.

Another layer of complexity arises from economic cycle fluctuations. During economic expansions, bankruptcy rates tend to decline, while recessions trigger surges in corporate failures. A model trained in one economic regime may not generalize well to another. Hence, there is a need to simulate and evaluate model performance across varying bankruptcy prevalence levels to ensure robustness.

Moreover, many high-performing models operate as black boxes, offering little to no transparency into their internal decision-making processes. However, stakeholders may require clear justifications for why specific firms are flagged as potentially bankrupt. This demand for transparency is especially critical in regulated sectors, where model outputs must often withstand scrutiny in audits, compliance reviews, and regulatory disclosures.

In summary, the challenge lies in building robust, explainable, and cost-sensitive ML

models that can accurately predict bankruptcy under severe class imbalance and evolving economic conditions.

Research Objectives

The primary objective of this thesis is to develop and evaluate a machine learning pipeline for corporate bankruptcy prediction that balances predictive accuracy, interpretability, and practical relevance. The study attempts to address the following goals:

- Model Comparison: Evaluate and compare the performance of boosting models with non-boosting alternatives. The goal is to identify trade-offs between model complexity, interpretability, and performance in the bankruptcy context.

- Economic Cycle Simulation: Simulate different macroeconomic conditions by constructing datasets with varying levels of bankruptcy prevalence: 1% (optimistic scenario), 5% (neutral scenario), and 10% (pessimistic scenario). This allows for robust assessment across different stress conditions.

- Cost-sensitive Learning: Incorporate cost matrices to reflect stakeholder-specific preferences, penalizing false negatives more heavily to account for the asymmetric risks of failing to identify bankrupt firms.

- Explainability and Interpretation: Explain models' predictions using XAI techniques, including:

  - Permutation-based feature importance to quantify global variable influence
  - Individual Conditional Expectation (ICE) plots to visualize feature effects at the instance level,
  - SHAP values for global feature attribution,

Through these objectives, this research aims to contribute a comprehensive framework for explainable, cost-aware bankruptcy prediction using modern machine learning techniques.

Expected Contributions

- Realistic Class Imbalance: Unlike most studies that rely on undersampling, oversampling, or SMOTE to artificially balance datasets, we simulate the naturally occurring class prevalence to reflect optimistic, neutral, and pessimistic economic cycles. This design choice increases the ecological validity of our results and enables a more accurate assessment of model robustness in real-world deployment scenarios, where bankruptcies remain rare events.

- Cost-sensitive Threshold Tuning: While much of the literature emphasizes metrics like AUC or accuracy, these can obscure poor minority-class performance. Our models are calibrated using cost matrices that reflect stakeholder-specific preferences, tuning thresholds based on false negative to false positive cost ratios. This enables the development of decision rules that are economically meaningful and aligned with varying levels of risk aversion, such as that of regulators, lenders, or equity investors.

- Model Comparison and Experimental Rigor: We systematically benchmark boosting and non-boosting models under a unified experimental framework. This includes consistent preprocessing, temporal splitting, hyperparameter tuning, and cost-sensitive evaluation across models.

- Advanced Explainability Pipeline: Our explainability layer combines global and local analysis techniques, including SHAP, permutation-based variable importance plots and ICE plots. We stratify predictions by model confidence and explain the decision pathways for the five most confident bankrupt and healthy predictions per

model. This multi-level interpretability framework provides insights into both average behavior and edge cases.

Literature Review

Distress Prediction: Pioneer Work

The prediction of corporate bankruptcy is a well-established area of research within accounting and econometrics. One of the earliest empirical studies in this field was conducted by Beaver (Beaver 1966), who tested whether financial ratios could predict corporate failure several years in advance. Using a univariate framework, Beaver evaluated the predictive ability of individual financial ratios by comparing the distributions of failed and nonfailed firms over a five-year horizon. His results demonstrated that cash flow to total debt, was a strong indicator of impending distress. In contrast, liquidity-based ratios such as the current ratio and working capital to total assets were among the worst performers, exhibiting much higher misclassification rates. Beaver's study thus highlighted not only the potential of accounting data for distress prediction but also the importance of carefully selecting which financial indicators to rely on. However, his method relied heavily on threshold-based classification and did not integrate multiple variables into a single predictive model.

Building on this foundation, one of the most influential contributions came from E. I. Altman (1968), who introduced the Z-score model using Multiple Discriminant Analysis (MDA). This model combined five financial ratios to construct a linear discriminant function that separated bankrupt from non-bankrupt firms. Despite its empirical success, the model relied on assumptions such as multivariate normality and homoscedasticity across classes, which do not always hold in real financial datasets.

Subsequent refinements were proposed, extending the Z-score framework to more complex firm types, but without addressing its fundamental statistical assumptions (E. I. Altman, Haldeman, and Narayanan 1977). To overcome these limitations, Ohlson introduced the O-score model (Ohlson 1980) based on logistic regression, relaxing the assumption of linear separability and allowing for probabilistic interpretation. However, one practical limitation of these models is that they use an annual snapshot of the data, ignoring historical data and trends (Shumway 2001).

Shumway (2001) advanced the field by developing a hazard model that introduced a dynamic framework to bankruptcy prediction, incorporating both firm-specific financial ratios and market-based variables. This addressed time-variation in default probability and allowed for more timely updates. Campbell, Hilscher, and Szilagyi (2008) extended this approach by adding market indicators to model the default probabilities. However, these models suffer from some limitations that include assumptions of linearity and proportional hazards, dependence on efficient and liquid financial markets, and limited applicability to private firms. Furthermore, their designs are not explicitly cost-sensitive and do not naturally accommodate class imbalance.

A critical methodological contribution was made by Zmijewski (1984), who systematically examined two sources of estimation bias in financial distress prediction: choice-based sample bias and sample selection bias. He demonstrated that many early studies estimated models on samples that disproportionately included distressed firms—far exceeding their actual population rate. This oversampling leads to biased parameter estimates and inflated distress probabilities if not corrected using the appropriate techniques. Additionally, Zmijewski showed that excluding firms with incomplete data introduces sample selection bias.

Balcaen and Ooghe (2006) provide a comprehensive critique of traditional statistical models that includes sampling selectivity, that arises from balancing a naturally unbalanced dataset, failure to account for phase-based deterioration and lack of interpretability.

With evidence from Elkan (2001) that real-world applications often demand unequal treatment of errors, cost-sensitive methods for model selection are increasingly being

used. Yotsawat et al. (2023) demonstrate that embedding misclassification costs directly into the learning process—rather than relying on oversampling—leads to more accurate detection of minority class firms. Similarly, Zou, C. Gao, and H. Gao (2022) show that cost-sensitive boosting frameworks significantly improve model robustness and real-world relevance by minimizing cost-weighted losses and improving interpretability. These findings motivate the use of cost-sensitive objective functions in this study to align model outputs with stakeholder priorities.

Machine Learning Techniques

In recent years, ML methods have increasingly been applied to bankruptcy prediction problems, because of their ability to model nonlinear interactions, manage high-dimensional data, and incorporate flexible regularization schemes. Studies such as Barboza, Kimura, and E. Altman (2017), Yotsawat et al. (2023), Zou, C. Gao, and H. Gao (2022) and Son et al. (2019) have demonstrated the effectiveness of boosting models, random forests, and artificial neural networks.

Zięba, S. K. Tomczak, and J. M. Tomczak (2016) emphasize that ensemble methods, especially boosting, are effective at minimizing both bias and variance and are robust to noisy financial inputs. However, a major limitation in many of these studies is the treatment of class imbalance. Bankruptcy datasets typically feature extreme imbalance , yet many works mitigate this using random undersampling or synthetic oversampling like SMOTE (Chawla et al. 2002). While these techniques simplify training, they distort the natural class prevalence and often yield overly optimistic performance estimates.

Explainable AI (XAI) Techniques in Bankruptcy Prediction

The increasing complexity of ML models has triggered a parallel emphasis on explainability.

As high-performing black-box models such as ensemble methods and deep neural networks become more prevalent in financial applications, concerns around their lack of transparency have grown. Stakeholders in finance and other regulated sectors demand clear, understandable justifications for decisions. Doshi-Velez and Kim (2017) argue that interpretability becomes essential in such high-stakes settings due to inherent incompleteness in the problem formulation—when objectives, constraints, or fairness considerations cannot be fully encoded into the model. This perspective motivates the use of post hoc explainability tools in this study to complement high-performing but black-box models.

Explainable AI (XAI) methods aim to make machine learning model outputs understandable to humans, a requirement that becomes especially critical in high-stakes domains such as finance. Among the tools developed for this purpose, SHapley Additive exPlanations (SHAP), introduced by Lundberg and Lee (2017), has emerged as a leading approach for both global and local feature attribution. Tran et al. (2022) applied SHAP to interpret bankruptcy predictions for listed firms in Vietnam across a range of models, including logistic regression, SVM, decision trees, random forests, MLPs, and XGBoost. They used mean absolute SHAP values to identify consistent drivers of model output and leveraged partial dependence plots (PDPs) to visualize the marginal effects of financial ratios on predicted bankruptcy risk. Similarly, Zhang et al. (2022) proposed an XAI framework that integrates SHAP and PDPs into a whole-process ensemble modeling approach for financial distress prediction. Their study demonstrated that SHAP-based explanations aligned closely with expert knowledge, improving both predictive credibility and stakeholder understanding.

Other interpretability techniques such as PDPs, individual conditional expectation (ICE) plots, permutation importance, and LIME are frequently used to visualize model

behavior (Yeo et al. 2025). However, much of the literature remains focused on global feature importance, overlooking how models behave for specific classes or individual predictions. While LIME provides localized explanations by fitting simple surrogate models, it is often unstable and sensitive to the sampling of the neighborhood, making it less reliable for sensitive applications like financial risk modeling.

A critical limitation in current XAI research is the misalignment between interpretability tools and the practical needs of stakeholders such as investors, regulators, and auditors. As Weber, Carl, and Hinz (2023) argue, many explainability efforts remain developer-centric and fail to deliver actionable insights for non-technical decision-makers. Our study addresses this gap by integrating permutation importance, SHAP value analysis and ICE analysis across models, bankruptcy prevalence levels, and cost configurations. This design enhances interpretability in ways that are both technically robust and meaningful to financial decision-makers, reflecting a growing emphasis on domain-relevant storytelling to support the responsible deployment of ML in finance.

Contribution to Literature

This thesis addresses several critical gaps in the literature on bankruptcy prediction:

- Realistic Class Imbalance: Unlike most studies that rely on undersampling, oversampling, or SMOTE to artificially balance datasets, we retain the naturally occurring class prevalence to simulate optimistic, neutral, and pessimistic economic cycles. This approach enhances the ecological validity of the results and enables a more accurate assessment of model robustness under real-world conditions, where bankruptcies remain rare events.

- Cost-sensitive Threshold Tuning: While much of the literature optimizes models based on metrics such as AUC or accuracy, these measures can obscure poor performance on the minority class. In contrast, this thesis calibrates models using stakeholder-specific cost matrices, adjusting thresholds based on false negative to

false positive cost ratios. This enables the development of decision rules that are economically meaningful and aligned with varying degrees of risk aversion, such as those of regulators, lenders, and equity investors.

- Temporal Granularity and Recency Effects: Feature vectors are constructed from 20 consecutive calendar quarters of financial ratios per firm, allowing a direct examination of the predictive value of recent versus older information. This structure enables testing of the common assumption that newer financial signals carry greater predictive weight—an aspect often overlooked or collapsed in studies that aggregate ratios or rely on single-year snapshots.

- Model Comparison and Experimental Rigor: Boosting and non-boosting models are benchmarked within a unified experimental framework, incorporating consistent preprocessing, temporal data splitting, hyperparameter tuning, and cost-sensitive evaluation. This ensures comparability across model classes and strengthens the robustness of the conclusions drawn.

- Advanced Explainability Pipeline: The thesis integrates global and local interpretability methods, including permutation-based variable importance, mean absolute SHAP values, and ICE plots. Predictions are stratified by model confidence, and decision pathways are analyzed for the five most confident bankrupt and healthy cases per model. This multi-level explainability framework offers insights into both average model behavior and critical edge cases.

Chapter 1

Theoretical Background

In this section, we will provide the conceptual foundations that inform the design and evaluation of the bankruptcy prediction framework. This includes an overview of boosting and non-boosting model families, and a review of explainable AI (XAI) techniques used to interpret model behavior both globally and locally.

1.1 Decision Trees

Decision Trees (DT) are a fundamental type of supervised learning model used for both classification and regression tasks. They are non-parametric, i.e. , they do not assume any fixed functional form or distribution for the data. DT learn a sequence of rules that partition the feature space into smaller and more homogeneous subsets, forming a tree-like structure where each internal node represents a decision rule based on a specific feature, each branch corresponds to an outcome of the rule, and each terminal node (or leaf) represents a predicted class label (for classification) or a numerical value (for regression).

The tree is constructed by recursively splitting the data based on criteria that aim to increase the homogeneity, or "purity," of the resulting subsets. To evaluate the quality of a split, the algorithm relies on impurity measures such as *Gini impurity* and *entropy*. These are defined as follows:

$$\text{Gini} = 1 - \sum_{i=1}^{k} p_i^2$$

$$\text{Entropy} = -\sum_{i=1}^{k} p_i \log_2(p_i)$$

where p_i is the proportion of observations belonging to class i at a particular node. A node is considered "pure" when it contains only samples from a single class, which corresponds to a Gini impurity and entropy of zero. During training, the tree grows by selecting the feature and threshold that yield the greatest reduction in impurity. The process continues until a stopping criterion is reached.

One of the key advantages of decision trees is their interpretability. They provide a clear, rule-based decision path for each prediction, which is particularly valuable in high-stakes domains where model transparency is important. Furthermore, decision trees can naturally capture non-linear relationships between features and do not require feature scaling or transformation. They also handle both numerical and categorical variables well and are robust to outliers.

Despite these strengths, decision trees are prone to overfitting, especially when the tree grows too deep and starts to capture noise in the training data. Additionally, they can be unstable, meaning that small changes in the training data can lead to a completely different tree structure.

To mitigate overfitting and improve generalization, several hyperparameters in the decision tree implementation of the `scikit-learn` library (Pedregosa et al. 2011) must be carefully tuned. The `max_depth` parameter controls the maximum depth of the tree, preventing it from growing excessively complex. The `min_samples_split` parameter sets the minimum number of samples required to split an internal node, while `min_samples_leaf` defines the minimum number of samples that must be present at a leaf node. These hyperparameters act as regularization mechanisms, helping reduce model variance and improve performance on unseen data.

1.2 Random Forest

Random Forest (RF) is a widely used ensemble learning method designed to improve the predictive performance and robustness of individual decision trees. It belongs to the family of bagging algorithms and operates by constructing many decision trees during training, where each tree in the forest is trained on a different bootstrap sample. Additionally, at each node, a random subset of features is considered for splitting, introducing further diversity among trees.

The final prediction for a given input $x^{(i)}$ is obtained by aggregating the predictions of all individual trees. In classification tasks, this is typically done by majority voting:

$$\hat{y}^{(i)} = \text{mode}\{T_1(x^{(i)}), T_2(x^{(i)}), \ldots, T_n(x^{(i)})\}$$

where $T_j(x^{(i)})$ denotes the prediction of the j-th decision tree on sample i.

The intuition behind RF is that while individual decision trees are prone to high variance (overfitting), averaging the predictions of many diverse trees can significantly reduce variance without increasing bias, resulting in a model that is both accurate and more stable.

Random Forests inherit several strengths from decision trees: they can handle high-dimensional data, accommodate both numerical and categorical variables, manage missing values, and capture complex, non-linear relationships. However, because predictions are made by aggregating the outcomes of many trees, the overall model becomes less interpretable. Unlike a single decision tree, which provides a clear sequence of rules leading to each prediction, the forest's collective behavior is more opaque.

To balance model complexity with generalization, several key hyperparameters in `scikit-learn`'s `RandomForestClassifier` are tuned. In addition to decision tree parameters such as `max_depth`, `min_samples_split`, and `min_samples_leaf`, the ensemble introduces `n_estimators`, which defines the number of trees in the forest. Increasing this value typically improves predictive performance but incurs higher computational cost. Another important hyperparameter is `max_features`, which controls the

number of features considered when splitting a node. Lower values promote tree diversity and reduce overfitting, while higher values can reduce bias but increase correlation across trees. Together, these hyperparameters help regulate model variance and bias, ensuring better out-of-sample performance.

1.3 Logistic Regression

Logistic Regression (LR) is one of the most established and widely used statistical models in financial risk prediction, particularly due to its simplicity, interpretability, and solid theoretical foundation. It is well-suited to binary classification tasks, such as predicting whether a firm will go bankrupt.

The basic logistic regression model estimates the probability that an observation belongs to the positive class (e.g., bankruptcy) by applying the sigmoid (logistic) function to a linear combination of input features:

$$P(y = 1 \mid \mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}^T \mathbf{x}}}$$

where $\mathbf{x} \in \mathbb{R}^p$ is the feature vector, $\mathbf{w} \in \mathbb{R}^p$ is the coefficient vector, and $\mathbf{w}^T \mathbf{x}$ is the linear predictor. The parameters $\mathbf{w}$ are estimated by maximizing the log-likelihood function:

$$\mathscr{L}(\mathbf{w}) = \sum_{i=1}^{n} [y_i \log P(y_i = 1 \mid \mathbf{x}_i) + (1 - y_i) \log(1 - P(y_i = 1 \mid \mathbf{x}_i))]$$

While this formulation constitutes the classical logistic regression model, regularized extensions are commonly used to improve generalization, particularly in high-dimensional settings or when multicollinearity is present.

A *regularized logistic regression model* incorporates a penalty term into the objective function to shrink coefficient estimates and reduce overfitting. Two widely used variants are:

- **L1 regularization** (lasso penalty), which penalizes the absolute magnitude of the coefficients and can produce sparse models:

$$\mathscr{L}_{\text{lasso}}(\mathbf{w}) = \mathscr{L}(\mathbf{w}) - \lambda \|\mathbf{w}\|_1$$

  This form of regularization tends to shrink some coefficients to zero, effectively performing feature selection. As a result, L1 regularization can produce sparse models that rely only on a subset of the most relevant features.

- **L2 regularization** (ridge penalty), which penalizes the squared magnitude of the coefficients:

$$\mathscr{L}_{\text{ridge}}(\mathbf{w}) = \mathscr{L}(\mathbf{w}) - \lambda \|\mathbf{w}\|_2^2$$

  Unlike L1, L2 regularization discourages large weights but does not drive them exactly to zero. It distributes the penalty across all coefficients, encouraging small but non-zero values and leading to more stable models when features are correlated.

Here, $\lambda > 0$ controls the strength of the regularization. In practice, λ is treated as a hyperparameter and selected through cross-validation or grid search. Regularization introduces bias into the estimates but often results in improved predictive performance by reducing variance.

Modern software libraries, such as `scikit-learn`, implement these regularized versions of logistic regression by default. For example, in the `sklearn` implementation, the regularization strength is controlled via the parameter `C`, where $\texttt{C} = 1/\lambda$.

In software implementations such as `scikit-learn`, another important hyperparameter is the choice of optimization algorithm, specified via the `solver` argument. Common solvers include `"liblinear"` (a coordinate descent algorithm suitable for small datasets and L1-penalized models), `"lbfgs"` (Limited-memory Broyden–Fletcher–Goldfarb–Shanno), and `"newton-cg"` (Newton-Conjugate Gradient), each offering different trade-offs in terms of speed, memory usage, and support for regularization types.

In summary, regularized logistic regression models combine the interpretability of linear models with improved robustness to overfitting. Their solid statistical grounding

and probabilistic outputs make them valuable tools in financial risk modeling, especially when balancing performance and explainability is essential.

1.4 Multi-Layer Perceptron

The Multi-Layer Perceptron (MLP) is a class of feedforward artificial neural network that is well-suited to learning complex, non-linear patterns in high-dimensional datasets. Unlike linear models such as LR, MLPs can approximate any continuous function given sufficient depth (number of layers) and width (number of neurons per layer), making them a powerful and flexible choice for classification problems, including bankruptcy prediction.

An MLP consists of an input layer, one or more hidden layers, and an output layer. Each neuron in a given layer computes a weighted sum of its inputs, adds a bias term, and applies a non-linear activation function to introduce non-linearity. Mathematically, the transformation performed by a single neuron can be expressed as:

$$y = f\left(\sum_{i=1}^{d} w_i x_i + b\right)$$

where w_i are the input weights, x_i the input features, b the bias term, and f a non-linear activation function. Common choices for f include the rectified linear unit (ReLU), hyperbolic tangent (tanh), and sigmoid functions. For the binary classification task of bankruptcy prediction, the sigmoid activation is typically applied in the final output layer to map predictions into the interval $[0,1]$, representing the estimated probability of bankruptcy. In our implementation however, we exclude the sigmoid activation function on the final layer of the model. This means that our model outputs logits (log-odds), which is desirable for our loss function, Sigmoid Focal Loss. We apply the sigmoid function only during evaluation to compute the prediction probabilities and associated cost.

The capacity of an MLP is primarily determined by its architecture—that is, the number of hidden layers and the number of neurons in each layer. Deeper networks with more

neurons can model more intricate relationships but are also more susceptible to overfitting, particularly in small or imbalanced datasets.

In this study, architectures consisting of two to three hidden layers were explored. The number of neurons per layer was drawn from uniform distributions over the ranges 64–512 for the layers. The MLP models were trained using the Adam optimizer with early stopping and learning rate decay. Dropout were applied between layers to stabilize training and mitigate overfitting. Hyperparameters such as learning rate, batch size, and dropout probability were tuned using the Optuna framework with minimzing the misclassification cost as objective.

Focal Loss

To address the pronounced class imbalance typical of corporate bankruptcy datasets, where bankrupt firms may constitute only 1% to 10% of the total, our experiments demonstrated that a standard binary cross-entropy loss may not be sufficient. In such cases, the model may become biased toward the majority (non-bankrupt) class. To counter this, we adopted the *Focal Loss*. It is a variant of the binary cross-entropy loss introduced by T.-Y. Lin et al. (2018) which dynamically scales the loss contribution of each sample based on the confidence of its prediction.

The focal loss for a single prediction is defined as:

$$\mathscr{L}_{\text{focal}}(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t)$$

where:

- $y \in \{0, 1\}$ is the true binary label,
- $p = \sigma(z) \in [0, 1]$ is the predicted probability for the positive class, obtained from the sigmoid function applied to the logit z,
- $p_t = \begin{cases} p & \text{if } y = 1 \\ 1 - p & \text{if } y = 0 \end{cases}$

- $\alpha_t = \begin{cases} \alpha & \text{if } y = 1 \\ 1 - \alpha & \text{if } y = 0 \end{cases}$

- $\alpha \in [0, 1]$ is the weighting factor that balances the importance of positive and negative examples,

- $\gamma \geq 0$ is the focusing parameter that reduces the loss contribution from easy examples and extends the range in which an example receives low loss.

When $\gamma = 0$, focal loss reduces to the weighted binary cross-entropy loss. As γ increases, the loss places more focus on hard, misclassified examples. When $\alpha = 0.5$, no class weighting is applied.

In our model, we use the `sigmoid_focal_loss` function from the `torchvision.ops` module in `PyTorch` (Paszke et al. 2019). This implementation expects the model outputs as raw logits rather than probabilities. Therefore, we do not apply a sigmoid activation to the final layer of the neural network during training. This is consistent with the formulation of focal loss in logit space and helps prevent numerical instability when computing the loss.

1.5 AdaBoost (Adaptive Boosting)

AdaBoost, short for Adaptive Boosting, is a boosting learning algorithm first introduced by Freund and Schapire (1997) that constructs a strong classifier by iteratively combining multiple weak learners—typically shallow decision trees. Each weak learner is trained to focus on the errors made by the ensemble up to that point, making the overall model progressively better at handling difficult cases.

AdaBoost assigns weights to training instances and updates them after each round of learning. Initially, all training samples are given equal weight. After each weak learner is trained, the algorithm increases the weights of the misclassified examples and decreases

the weights of the correctly classified ones. As a result, subsequent weak learners focus more on the harder examples that previous learners failed to classify correctly.

The final prediction is obtained through a weighted majority vote (in classification) or weighted sum (in regression) of all the weak learners. For binary classification, the ensemble model can be written as:

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right)$$

where $h_t(x)$ is the prediction of the t-th weak learner, α_t is the weight assigned to that learner based on its performance, and T is the total number of learners. A learner with lower classification error receives a higher weight α_t, contributing more strongly to the final decision.

In our implementation, we use the `AdaBoostClassifier` from the `scikit-learn` library. Three key hyperparameters govern its behavior. The first is *n_estimators*, which determines the number of boosting rounds, or equivalently, the number of weak learners in the ensemble. Increasing this value can enhance predictive performance but also leads to higher computational cost and may increase the risk of overfitting. The second is the weak learner, *estimator*, which is often a decision tree. Since the choice of the base estimator exponentially increases the hyperparameter tuning phase for AdaBoost, we use the default *estimator*, which is a decision tree with *max_depth*$= 1$. The third hyperparameter is the *learning_rate*, a shrinkage parameter that scales the contribution of each weak learner. Lower values of the learning rate make the model more conservative, which can improve generalization but typically requires a larger number of boosting rounds to maintain performance.

1.6 LightGBM

LightGBM (Light Gradient Boosting Machine) is a high-performance gradient boosting framework developed by Ke et al. (2017) to handle large-scale datasets with improved

speed and memory efficiency. It is well-suited to classification and regression tasks involving structured data, and has become a popular choice due to its scalability, accuracy, and flexibility.

LightGBM belongs to the family of gradient boosting algorithms, which build an ensemble of decision trees in a sequential manner. Each new tree is trained to minimize the residual errors made by the ensemble up to that point. The training process involves optimizing a differentiable loss function, typically binary cross-entropy for classification, using gradient descent. Unlike traditional boosting methods, LightGBM introduces several key innovations that accelerate training and reduce memory usage without sacrificing performance.

Unlike conventional gradient boosting frameworks that grow trees by expanding all leaves at the same depth before moving deeper, LightGBM grows trees leaf-wise, choosing the leaf with the highest loss reduction at each step. This often leads to deeper trees and better accuracy, although it increases the risk of overfitting.

The complexity and behavior of LightGBM models are affected by several hyperparameters. The parameter *num_leaves* controls the maximum number of leaves in each tree and directly influences model complexity; a higher value increases flexibility but also the risk of overfitting. The parameter *max_depth* optionally sets a hard limit on tree depth, acting as a regularization mechanism. The *learning_rate* scales the contribution of each new tree and is typically set to a small value to ensure gradual learning. Subsampling techniques are implemented via *subsample*, which determines the fraction of training instances used per tree, and *colsample_bytree*, which specifies the fraction of features considered during split selection.

We use the `LGBMClassifier` from the `lightgbm` Python library for our gradient boosting implementation. The model's complexity and behavior are governed by several key hyperparameters. The `num_leaves` parameter controls the maximum number of leaves per tree and directly affects model capacity; larger values allow more flexible trees but may increase the risk of overfitting. The `max_depth` parameter optionally sets a hard limit on tree depth, serving as a regularization mechanism to prevent exces-

sively deep trees. The `learning_rate` determines the contribution of each new tree to the ensemble; smaller values slow down learning but often yield better generalization. To introduce randomness and improve generalization, LightGBM supports subsampling techniques through `subsample`, which sets the fraction of training data used per tree, and `colsample_bytree`, which controls the fraction of features randomly selected for each tree.

1.7 XGBoost (Extreme Gradient Boosting)

XGBoost, short for Extreme Gradient Boosting, is a highly optimized and scalable implementation of gradient boosting proposed by Chen and Guestrin (2016) that has become a benchmark model in structured data tasks, particularly in finance and tabular machine learning competitions. It builds upon the standard gradient boosting framework by incorporating regularization, native support for sparse input, and efficient system-level enhancements such as parallel computation and cache awareness.

Like other gradient boosting methods, XGBoost builds an ensemble of decision trees in a sequential manner. At each stage, a new tree is trained to correct the residual errors made by the current ensemble. The model optimizes an objective function composed of two components: a training loss that quantifies how well the model fits the data, and a regularization term that discourages overly complex trees. This regularization component helps control model capacity by penalizing tree structures that are too deep or rely on extreme leaf weights. By doing so, XGBoost achieves a balance between predictive performance and model simplicity, reducing the risk of overfitting.

One of XGBoost's distinguishing features is its ability to handle missing values during both training and prediction. Instead of requiring imputation, XGBoost learns the optimal path for missing values during tree construction, treating them as a separate informative branch. The model also supports sparse input natively, making it efficient for datasets with many zero or missing entries. However, since we want to ensure fair comparison across the different selected models, we do not leverage this strength of XGBoost.

We use the `XGBClassifier` from the `xgboost` Python library for our implementation of gradient boosting. XGBoost shares many hyperparameters with other boosting frameworks. The parameter `n_estimators` controls the number of trees in the ensemble, while `learning_rate` scales the contribution of each tree. `max_depth` sets the maximum depth of each tree, affecting model expressiveness and overfitting risk. To introduce randomness, `subsample` specifies the fraction of training data sampled per tree, and `colsample_bytree` controls the fraction of features considered when constructing each tree. The `gamma` parameter imposes a minimum reduction in the loss function required to perform a split, encouraging only the most informative partitions. Finally, `reg_alpha` and `reg_lambda` correspond to ℓ_1 and ℓ_2 regularization penalties on the leaf weights, helping to constrain model complexity and improve generalization.

1.8 Explainable AI

As machine learning models become increasingly complex and opaque, the need for interpretability in high-stakes applications such as bankruptcy prediction becomes crucial. In financial contexts, regulatory bodies, stakeholders, and analysts require transparent models that can provide not only accurate predictions but also insights into the rationale behind each decision. Explainable AI (XAI) encompasses a suite of techniques that help interpret and audit the internal mechanics of predictive models. This section reviews three major approaches used in this thesis: Permutation feature importance (PFI), SHapley Additive exPlanations (SHAP) values, and individual conditional expectation (ICE).

1.8.1 Permutation Feature Importance

Permutation Feature Importance (PFI) is a model-agnostic approach used to assess the contribution of individual input features to a model's predictive performance initially proposed by Breiman (2001) and improved by Fisher, Rudin, and Dominici (2019). The core idea behind PFI is intuitive: if a feature is important, disrupting its association with the

outcome should degrade the model's performance (or increase the loss). Conversely, if permuting a feature has little to no impact on performance, the model is likely not relying on that feature in its decision-making process.

To evaluate the importance of a given feature x_j, we permute the values of that feature across all instances in the dataset, effectively breaking its relationship with the target variable y, while keeping all other features unchanged. The predictions on this modified dataset are then compared to predictions on the original data. The drop in predictive accuracy (or increase in loss) serves as a measure of how important the permuted feature was to the model.

One of the primary advantages of PFI is its broad applicability. It can be used with any machine learning model, including decision trees, gradient boosting machines, neural networks, and even ensemble models. Because it does not rely on model-specific internal structures, it provides a uniform basis for comparing feature importance across different algorithms.

1.8.2 SHAP

SHapley Additive exPlanations (SHAP) is a unified framework for interpreting machine learning model predictions based on concepts from cooperative game theory (Lundberg and Lee 2017). It assigns each input feature a contribution value that reflects its impact on a specific prediction, hence enabling both global and local interpretability. SHAP values are based on the Shapley value framework from cooperative game theory, which provides a principled approach for attributing the overall outcome of a model to individual features. Each feature's contribution is quantified by averaging its marginal impact on the prediction across all possible subsets of features, ensuring a fair and theoretically grounded measure of feature importance.

SHAP satisfies several desirable properties that make it particularly appealing for explanation:

- Local accuracy (additivity): The sum of feature attributions equals the model out-

put.

- Consistency: If a model changes such that the marginal contribution of a feature increases, its SHAP value will not decrease.

- Missingness: Features not used in the model receive a SHAP value of zero.

SHAP values can be visualized in multiple ways to support both instance-level and dataset-level interpretations. In particular, *Mean Absolute SHAP values* measures feature importance in prediction. *Dependence plots* show how a feature's SHAP value varies with its value, optionally highlighting interactions with other features.

1.8.3 ICE Plots

Individual Conditional Expectation (ICE) plots are a model-agnostic visualization technique used to examine how changes in a single feature affect a model's predictions at the instance level. They provide a fine-grained, instance-specific complement to Partial Dependence Plots (PDPs), which show the average effect of a feature across all observations. While PDPs can mask heterogeneity in feature effects, ICE plots reveal this variation explicitly by plotting separate curves for each instance in the dataset.

By generating ICE curves for multiple instances, we obtain a collection of trajectories showing how each prediction responds to hypothetical changes in a specific feature. These curves can be plotted together to visualize both common trends and deviations. If the ICE curves for all instances are roughly parallel, this suggests that the model treats the feature's effect consistently across the dataset, and a global summary like a PDP may suffice. However, if the ICE curves diverge significantly, it indicates feature interactions or non-linear behavior that varies by instance.

A common enhancement is the use of *centered ICE plots*, where each curve is shifted so that all trajectories start at a common baseline. This helps isolate the shape of the response function from differences in model output due to other fixed features.

Chapter 2

Dataset

2.1 Dataset Description

This study is based on a custom dataset comprising publicly listed companies in the United States. The dataset includes five key accounting ratios that serve as predictive features:

- Earnings Before Interest and Taxes to Total Assets (EBIT/TA): This ratio measures the firm's operating profitability relative to its asset base, abstracting from tax and interest effects. It captures the fundamental earning power of a company's assets, with higher values indicating stronger internal cash generation and a lower probability of financial distress.

- Market Capitalization to Total Liabilities (MKTCAP/TL): This solvency measure compares the market value of equity to the book value of total liabilities. It reflects the market's assessment of the firm's ability to cover its debts. Higher values imply a larger equity cushion relative to obligations, reducing the risk of insolvency.

- Retained Earnings to Total Assets (RE/TA): Retained earnings represent cumulative past profits that have been reinvested in the business. This ratio serves as a proxy for the firm's historical profitability and its ability to internally fund opera-

tions. Lower values may indicate a history of operating losses or heavy reliance on external financing, both of which can signal elevated bankruptcy risk.

- Total Revenue to Total Assets (TR/TA): Often interpreted as an asset turnover ratio, this measure captures the firm's ability to generate sales from its asset base. While high turnover generally reflects operational efficiency, extreme values may indicate either insufficient capitalization or aggressive revenue strategies, both of which can pose financial risks if not managed carefully.

- Working Capital to Total Assets (WC/TA): Working capital, defined as current assets minus current liabilities, represents short-term financial health. When scaled by total assets, it provides insight into liquidity relative to the size of the firm. Persistently low or negative working capital ratios may signal potential short-term cash flow difficulties.

These ratios are well-established indicators of financial health and have been widely used in the literature, notably forming the foundation of Altman's seminal work (E. I. Altman 1968) on bankruptcy prediction.

The dataset spans from Calendar Quarter 1 (CQ1) 1998 to CQ4 2023, covering 26 years worth of quarterly data. To collect data for this study, we utilize S&P Capital IQ Pro, a financial intelligence platform developed by S&P Global Market Intelligence. Capital IQ Pro provides detailed financial, accounting, and market data on publicly listed and private companies across global markets. The platform aggregates standardized firm-level data from a variety of regulatory filings, earnings announcements, and audited financial statements. The raw data was retrieved at the firm-quarter level and included all variables necessary to compute the five financial ratios used in this study.

The next sections outline the data construction process, rationale for key design decisions and the structure of the resulting dataset.

2.2 Dataset Construction

2.2.1 Russell 3000 Constituents

For each calendar quarter within the time range, the full list of companies included in the Russell 3000 Index, which represents the 3000 largest US-traded companies by market capitalization, was retrieved. We then constructed a master set of firms by taking the union of all constituent lists across quarters from CQ1 1998 to CQ4 2023.

This approach captures firms that entered or exited the index at different times, ensuring that their historical data is included even if their index membership was not continuous. In contrast, selecting companies based only on quarterly index membership, that is, including companies only when they form part of the index, would lead to systematic missingness in the financial records of firms that temporarily exited the index, causing their removal during the data cleaning phase. Taking the union allows us to maximize the retention of firms with complete temporal coverage.

2.2.2 Corporate Hierarchies

For each company in our universe, we retrieved its associated *parent* and *ultimate parent* firms using definitions provided by S&P Capital IQ (SPCIQ):

- The parent refers to the immediate controlling entity—it is usually the company that directly holds a majority ownership stake.

- The ultimate parent is the top-level legal entity in the ownership hierarchy that consolidates financial control across subsidiaries.

Incorporating these entities allows us to better capture bankrupt firms as in some cases, it is the parent company or the ultimate parent company that gets labelled as "bankrupt" as per our definition (see Section 2.3).

2.2.3 Bankrupt Companies

To enhance the representation of distressed firms, we supplemented the dataset with all US public companies that met our definition of bankruptcy and had available financial data in S&P Capital IQ Pro (SPCIQ). These companies were included regardless of their Russell 3000 membership status.

2.2.4 "Originally Reported" Financial Data.

All financial statement values used in this study are *originally reported* values, that is, they reflect the original values filed by companies in each quarter, without incorporating subsequent restatements.

Relying on *"originally reported"* financial data strengthens the external validity of this study by aligning the modeling process with the real-world conditions under which financial decisions are made. Stakeholders such as analysts, lenders, and investors typically base their assessments on the financial statements available at the time of filing, not on figures revised months or years later.

In contrast, incorporating restated financial statements can introduce structural bias into the learning process. Firms that revise their reports more frequently—due to internal audits, regulatory enforcement, or delayed recognition of accounting issues—may appear more financially stable in hindsight. However, these corrections reflect information that was unavailable at the time of decision-making. Consequently, models trained on restated data risk learning patterns informed by future knowledge, rather than by the signals of financial health available at the time. This creates a biased learning environment in which firms with post-filing corrections may be systematically favored, as their historical data has been retroactively improved—making them appear less prone to bankruptcy than they were in real-time.

2.3 Bankruptcy Definition

The target variable in this study is a binary indicator: *bankrupt* versus *healthy*. A company is labeled as bankrupt if, according to SPCIQ, it satisfies at least one of the following conditions:

1. Filed for bankruptcy, under Chapter 7 or Chapter 11 of the US Bankruptcy Code, as disclosed in SEC filings, court documents, or press releases.

2. Key Development: Potential red flag – debt outstanding, indicating that the company was flagged by SPCIQ analysts as having unresolved debt or liquidity concerns.

This definition captures both formal insolvency and serious signs of financial distress, enabling the model to detect a broader spectrum of risky cases.

2.4 Forecast Horizon and Labeling Strategy

The objective of this study is to predict bankruptcy in advance. As outlined in the Literature Review section, several studies use financial data for a year to predict bankruptcy in the next. This, however, would not be practical since there would be other indicators (for example market indicators), hinting about bankruptcy. Moreover, as outlined in Ohlson (1980), there is often a lead between a bankruptcy event and the release of the relevant financial reports. Our goal therefore is to find models that identify patterns in the data available before a company goes bankrupt as per our definition. Accordingly, a company is labeled as bankrupt for 8 calendar quarters (CQs) before the actual bankruptcy or red-flag event.

2.5 Temporal Structure

Each company-quarter instance in the dataset is represented by a unique *(SPCIQID$_i$, date)* tuple, where *date* denotes the calendar quarter at which we want to make the prediction (e.g., *CQ3_2021* is the third calendar quarter of the year 2021). In particular, the feature set consists of ratios reported in a fixed-length window of 20 CQs preceeding the *date*. A column such as *tr_ta_cq_1* refers to the *TR/TA* ratio calculated from figures disclosed one calendar quarters before the target date. Such a column would be denoted by *TR/TA CQ_1* in this thesis.

This study focuses on U.S. public companies, which are required by the Securities Exchange Act of 1934 to file quarterly financial reports (Form 10-Q) and an annual report (Form 10-K) with the Securities and Exchange Commission (SEC). As such, quarterly data should be consistently available for compliant firms. Following standard practice in the literature (Ohlson 1980, Shumway 2001), we exclude firms with incomplete reporting windows from the sample to maintain temporal consistency and model comparability. While this filtering ensures the integrity of input sequences, it may introduce sample selection bias—particularly if firms with missing data are systematically more likely to be in distress. Zmijewski (1984) cautions that excluding such firms can bias parameter estimates and understate bankruptcy probabilities. However, given that financial disclosures are a regulatory requirement for listed firms, persistent missing data is rare. We adopt this exclusion criterion as a pragmatic trade-off, acknowledging the potential bias while preserving model validity across a uniform 20-quarter input horizon.

2.6 Temporal Splitting and Undersampling Strategy

A structured data sampling strategy was implemented to ensure both the validity and robustness of the predictive models. First, the dataset was temporally split to prevent data leakage and to more accurately reflect real-world forecasting scenarios. The training set included only observations from CQ1 1998 through CQ4 2019, while the remaining data,

spanning CQ1 2020 to CQ4 2023, was evenly divided into validation and test sets. This chronological separation preserves the temporal integrity of model evaluation and ensures that future data is not inadvertently used to inform past predictions.

To address the severe class imbalance inherent in bankruptcy prediction, we applied a targeted undersampling strategy to the majority class (non-bankrupt firms). Specifically, for a given bankruptcy prevalence level (e.g. 1%), all bankrupt observations were retained, and a random subset of healthy ones was sampled such that the final proportion of bankrupt to healthy firms matched the specified ratio. This procedure was applied independently to the training, validation, and test sets, allowing us to simulate different economic conditions while preserving the minority class in full. This design enhances both generalizability and comparability across prevalence scenarios.

2.7 Dataset Exploratory Data Analysis

This section presents a descriptive analysis of the financial ratios used in the study, based on firm-quarter level data that has been winsorized to reduce the influence of extreme values. Winsorization was applied separately within each calendar quarter to preserve temporal comparability.

For each financial ratio, we report summary statistics and visualize the distribution through box plots for five representative calendar quarters: CQ4 2004, CQ4 2008, CQ4 2012, CQ4 2016, and CQ4 2020. The summary statistics for all the calendar quarters can be found in Appendix A (6.7). In addition, Figure 2.1 displays histograms of the full distribution of each ratio across all quarters. To assess the asymmetry of the distributions over time, Figure 2.2 shows the average skewness of each ratio, computed across all calendar quarters.

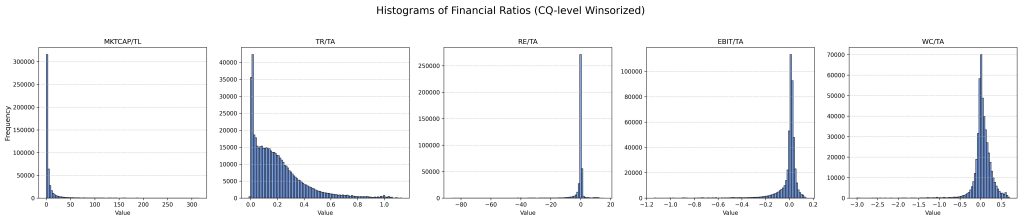


Figure 2.1: Histogram of Ratios Winsorized at CQ-level

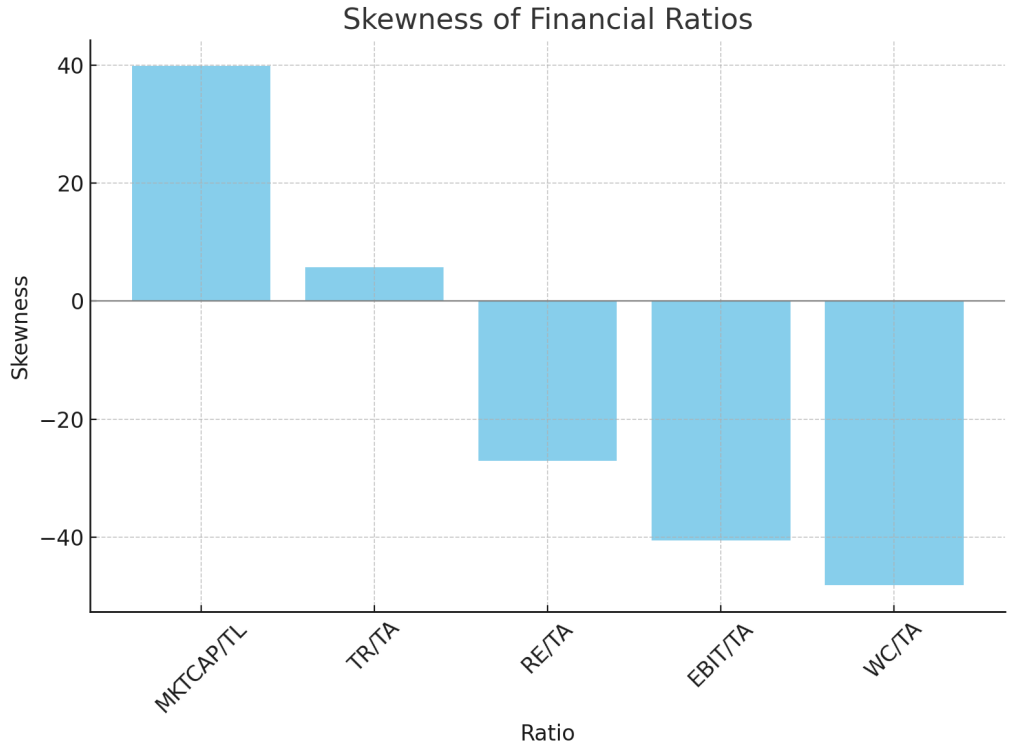


Figure 2.2: Mean Skewness per Accounting Ratio over all CQs

Table 2.1: Summary Statistics of EBIT/TA for Selected Calendar Quarters

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ4 2020	-0.0944	-0.0139	0.0097	0.0234	-164.0000	0.4741	-51.02
CQ4 2016	-0.0857	-0.0051	0.0126	0.0253	-99.5554	0.8870	-40.81
CQ4 2012	-0.1477	0.0008	0.0150	0.0287	-285.9835	1.3875	-51.92
CQ4 2008	-0.1768	-0.0103	0.0123	0.0278	-631.5087	0.9664	-67.14
CQ4 2004	-0.0123	0.0007	0.0167	0.0324	-54.2202	203.9198	52.49

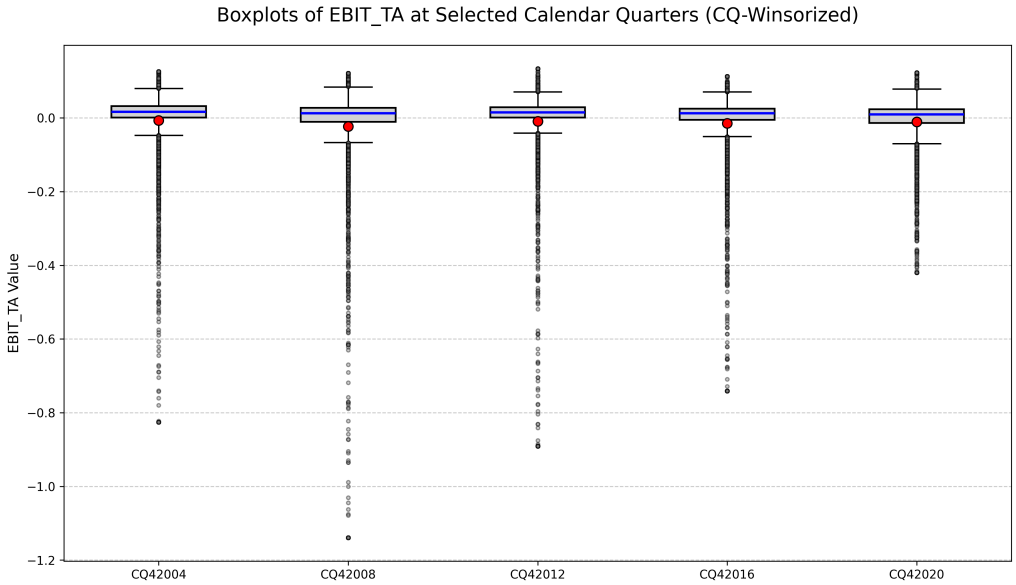


Figure 2.3: EBIT/TA Box Plots for Selected CQs

2.7.1 EBIT/TA

Boxplots for selected calendar quarters in Figure 2.3 reveal that the distribution of EBIT/TA remains consistently centered slightly above zero. However, the presence of a large number of negative outliers pulls the mean downwards. The median EBIT/TA stays positive across periods, suggesting that the majority of firms maintain modest profitability relative to their assets, while a minority experiences severe operational losses.

The histogram confirms a strong left-skewed distribution, with a peak close to zero and a long negative tail extending to approximately -1.2. The mean skewness for EBIT/TA is around -40, reflecting the asymmetry caused by distressed firms with large negative EBIT relative to their asset base.

2.7.2 WC/TA

The WC/TA ratio exhibits a similar pattern to EBIT/TA. Boxplots indicate that although the median WC/TA is slightly positive, the mean is consistently lower due to numerous extreme negative outliers. This behavior suggests that while most firms maintain positive or neutral working capital positions, a significant minority operates with substantial

Table 2.2: Summary Statistics of WC/TA for Selected Calendar Quarters

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ4 2020	-3.5247	-0.0385	0.0115	0.0891	-8456.0000	0.9197	-50.01
CQ4 2016	-0.7106	-0.0380	0.0194	0.1079	-595.4490	0.9454	-28.93
CQ4 2012	-0.3998	-0.0295	0.0349	0.1439	-577.9490	0.9570	-39.68
CQ4 2008	-0.2460	-0.0282	0.0440	0.1524	-774.8564	0.9602	-58.05
CQ4 2004	-1.5851	-0.0297	0.0402	0.1588	-3870.9580	0.9997	-49.83

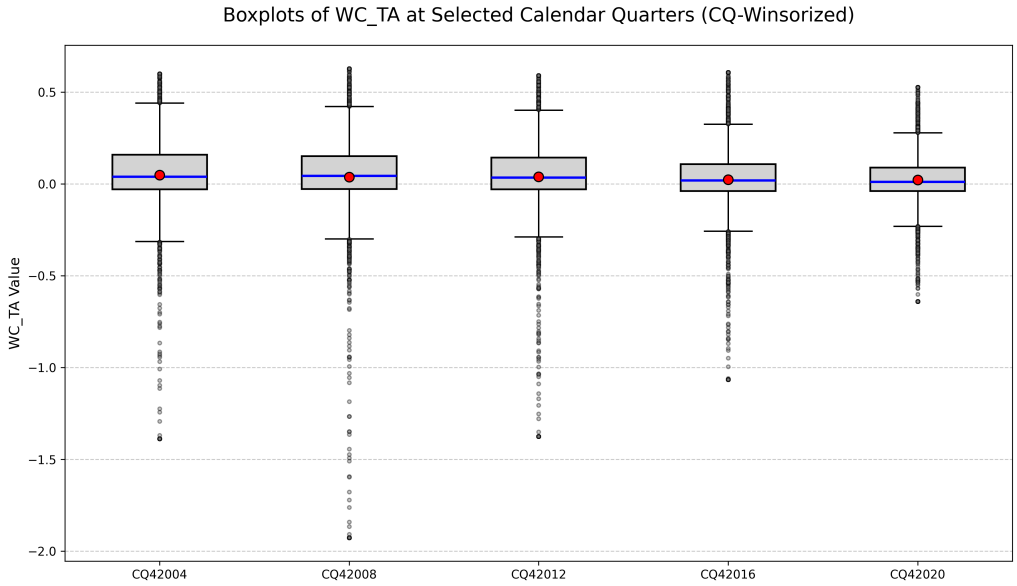


Figure 2.4: WC/TA Box Plots for Selected CQs

deficits.

The histogram displays a clear concentration around zero with a pronounced left tail, stretching as far as -3.0. The mean skewness for WC/TA is the most negative among the five ratios, at approximately -50, highlighting the extreme working capital deficiencies.

2.7.3 RE/TA

The RE/TA ratio consistently has negative mean across all periods. The boxplots in Figure 2.5 demonstrate that while the median remains close to zero, the distributions are dominated by negative retained earnings.

The histogram reveals a sharp peak near zero but an extensive negative tail reaching

Table 2.3: Summary Statistics of RE/TA for Selected Calendar Quarters

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ4 2020	-37.9017	-0.2913	0.0266	0.2458	-44612.0000	7612.7562	-29.48
CQ4 2016	-34.8125	-0.2100	0.0417	0.2801	-80324.7359	5080.6578	-53.06
CQ4 2012	-5.8633	-0.1622	0.0552	0.2976	-9149.0340	3357.6159	-30.09
CQ4 2008	-6.6734	-0.2256	0.0467	0.2849	-11344.9335	4844.7424	-29.69
CQ4 2004	-16.5056	-0.1952	0.0513	0.2574	-22602.9760	1885.9358	-36.66

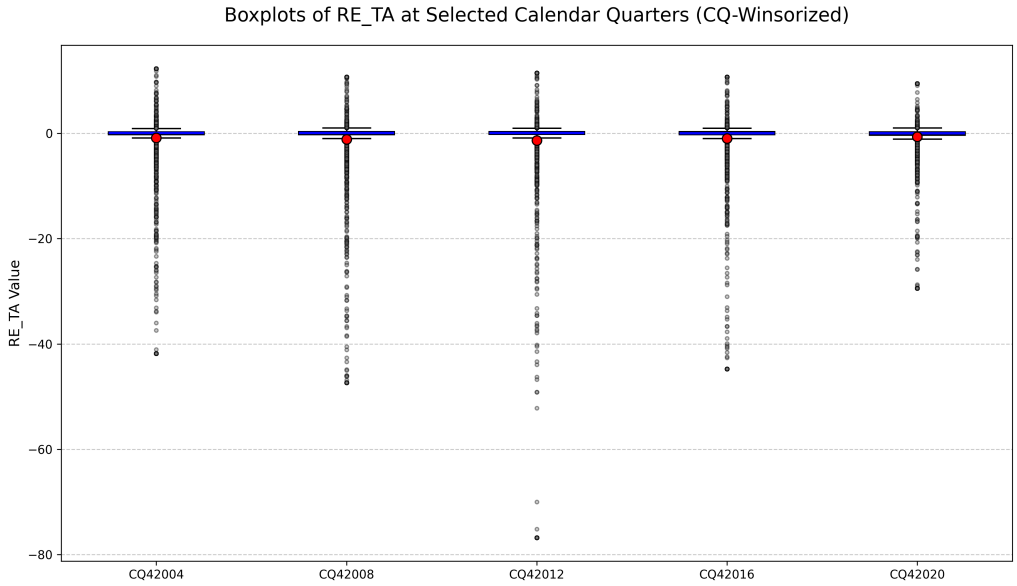


Figure 2.5: RE/TA Box Plots for Selected CQs

beyond -80. The mean skewness across all periods for RE/TA is approximately -30, showing that the distribution is very negatively skewed.

2.7.4 TR/TA

TR/TA displays similar patterns, but flipped. Boxplots in Figure 2.6 show that while both the mean and median TR/TA remain positive across all periods, there are significant positive outliers. There is a notable decrease in the mean of TR/TA values post-2020, coinciding with the COVID-19 pandemic.

The histogram is moderately right-skewed, with most observations concentrated between 0.1 and 0.3, and a few firms achieving exceptionally high revenue-to-asset ratios.

Table 2.4: Summary Statistics of TR/TA for Selected Calendar Quarters

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ4 2020	0.1547	0.0375	0.1158	0.2146	-0.3970	2.5159	2.85
CQ4 2016	0.1871	0.0449	0.1409	0.2558	-0.3178	3.9598	3.79
CQ4 2012	0.2123	0.0574	0.1605	0.2855	-1.3254	5.0997	4.84
CQ4 2008	0.2256	0.0604	0.1696	0.2920	-1.1523	28.2789	39.33
CQ4 2004	0.2132	0.0655	0.1783	0.3106	-90.1257	14.0823	-63.66

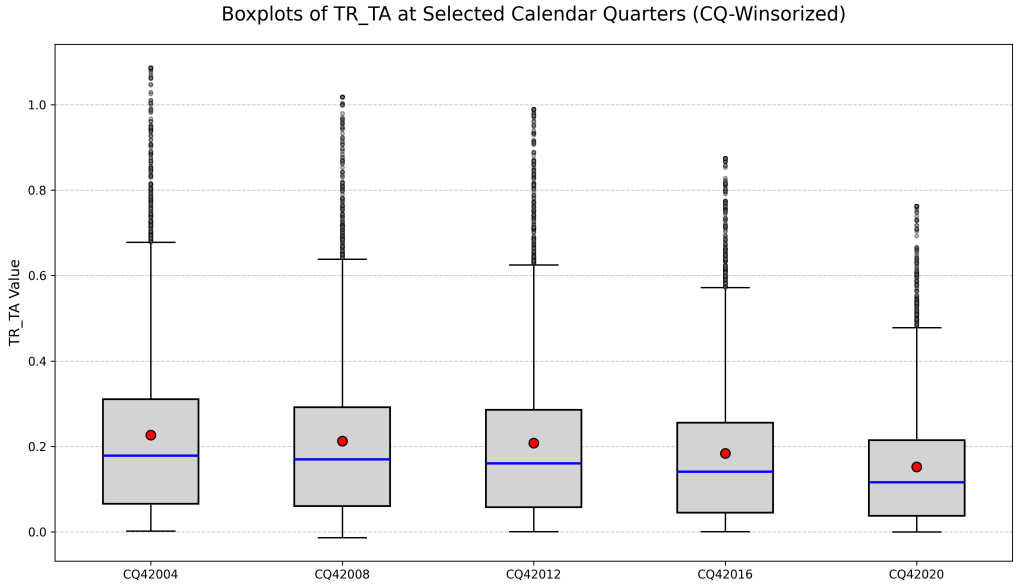


Figure 2.6: TR/TA Box Plots for Selected CQs

The mean skewness is positive, around +5, suggesting that while the majority of firms operate within a typical range, a small number exhibit exceptionally high turnover efficiency.

2.7.5 MKTCAP/TL

MKTCAP/TL stands out with very different distributional characteristics compared to the other ratios. Boxplots illustrate extremely high positive skewness, with the majority of firms clustered near low MKTCAP/TL values and a small fraction reaching extraordinarily high levels.

The histogram shows an extremely right-skewed distribution, extending beyond 100

Table 2.5: Summary Statistics of MKTCAP/TL for Selected Calendar Quarters

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ4 2020	19.6408	0.5603	1.7960	6.3120	-0.6598	5789.0248	24.39
CQ4 2016	16.8373	0.6659	1.6889	4.3161	0.0027	11996.4814	41.72
CQ4 2012	13.6387	0.5480	1.5547	4.1278	0.0001	10374.2394	42.29
CQ4 2008	9.9257	0.2729	0.9612	2.8788	0.0000	15665.4350	63.99
CQ4 2004	18.6686	0.7570	2.1755	6.5569	0.0000	14567.7854	41.47

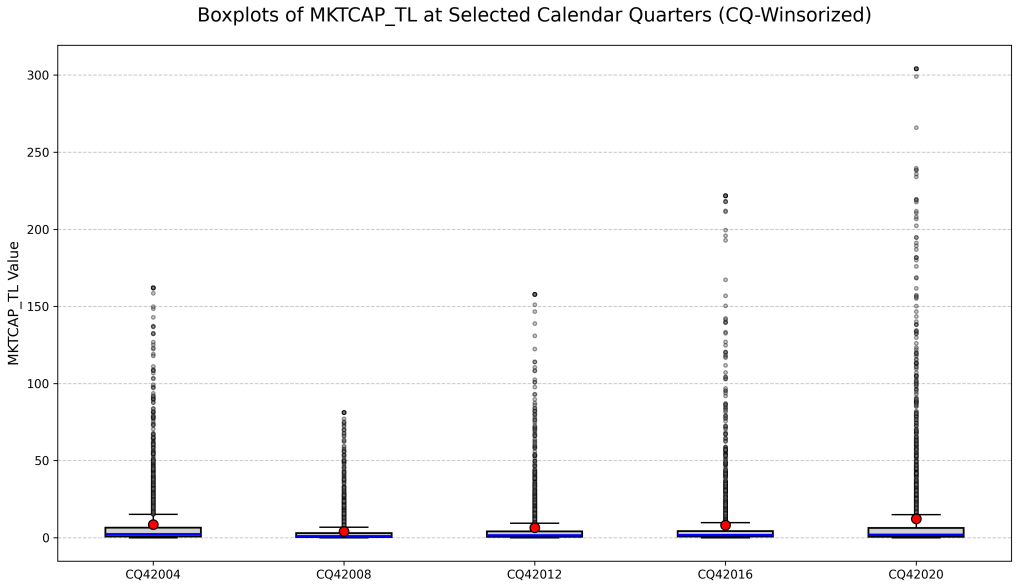


Figure 2.7: MKTCAP/TL Box Plots for Selected CQs

in some periods. The mean skewness for MKTCAP/TL is approximately $+40$, the highest among all ratios.

2.7.6 Ratios Summary

Overall, the financial ratios analyzed exhibit skewed distributions, highlighting important patterns in firms' financial health across economic cycles. EBIT/TA, WC/TA, and RE/TA display strong left-skewness, indicating that although the central mass of firms achieves near-neutral or slightly positive financial outcomes, a substantial fraction consistently records significant negative values.

In contrast, TR/TA displays a relatively balanced distribution, with a slight right skew.

The majority of firms cluster around typical TR/TA ratios, but a small number achieve exceptionally high turnover efficiencies. MKTCAP/TL stands out from the other ratios by exhibiting extreme positive skewness, as reflected in the box plots and histograms. Most firms maintain relatively low market capitalization relative to liabilities, while a minority of firms have extraordinarily high market valuations, leading to a heavy right-tailed distribution.

Together, the results highlight that firm-level financial health is highly uneven. While most firms maintain typical financial ratios, a non-negligible portion consistently falls into the extreme tails of the distributions. These patterns reflect the inherent asymmetries in firms' financial characteristics over time.

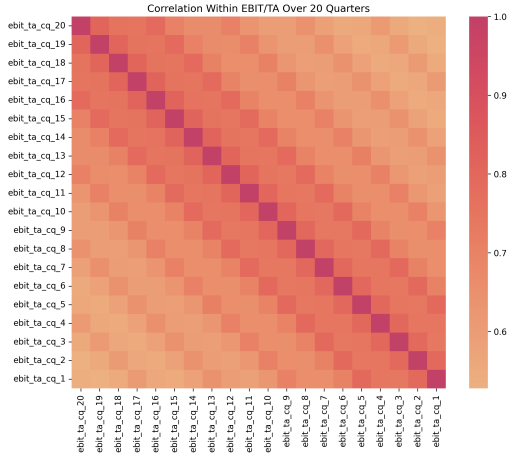
2.8 Correlation Analysis

To examine redundancy across the temporal features used for bankruptcy prediction, we performed a two-part correlation analysis: first within each accounting ratio across time (intra-ratio), and then across ratios at a fixed point in time (inter-ratio). These analyses provide insight into multicollinearity, which has direct implications for both model training and interpretability.

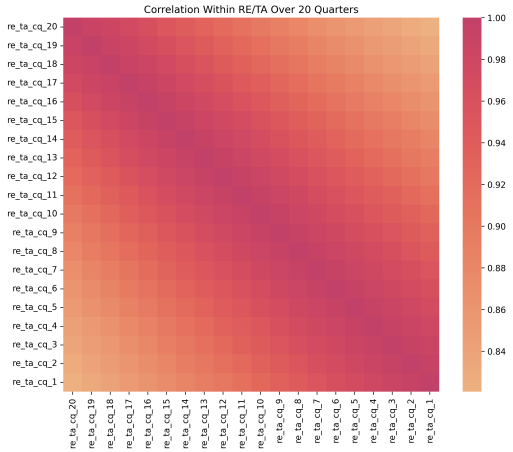
2.8.1 Intra-Ratio Correlation Over 20 Quarters

Figures 2.8a through 2.8d present heatmaps of Pearson correlations between 20 consecutive CQs for each financial ratio. The patterns reveal high autocorrelation within each ratio, particularly for *RE/TA* and *WC/TA*, where values remain above 0.90 across wide time lags. Even more volatile ratios such as *EBIT/TA* and *MKTCAP/TL* show substantial persistence, with most correlations above 0.70.

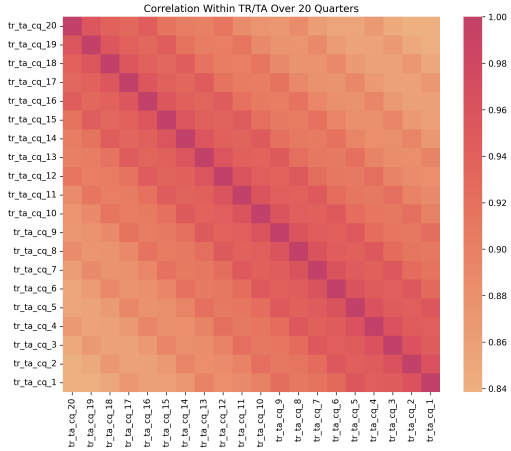
Although the high correlation observed within each financial ratio across calendar quarters is technically a form of temporal autocorrelation, our models treat all input features as flat, unordered variables without temporal structure. Consequently, this intra-ratio



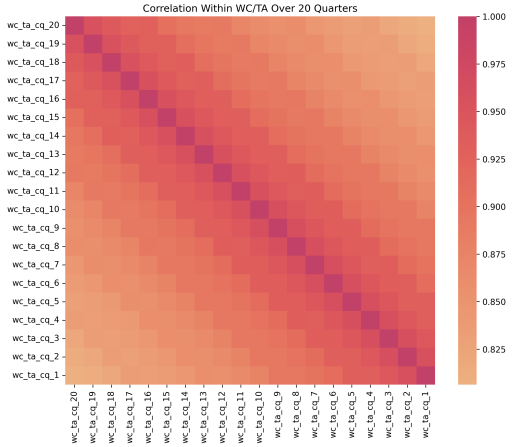
(a) EBIT/TA over 20 quarters



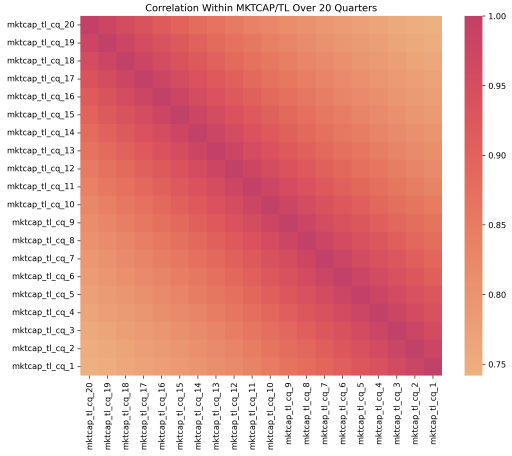
(b) RE/TA over 20 quarters



(c) TR/TA over 20 quarters



(d) WC/TA over 20 quarters



(e) MKTCAP/TL over 20 quarters

Figure 2.8: Intra-ratio correlation heatmaps over 20 calendar quarters. Each plot shows the degree of autocorrelation within a financial ratio.

correlation introduces feature redundancy that behaves analogously to multicollinearity, particularly in its impact on feature attribution in post-hoc explainability methods.

This strong temporal correlation implies considerable redundancy in the 20-CQ input representation. Such multicollinearity can obscure feature attribution in post-hoc explainability techniques like PFI and SHAP. When multiple features convey overlapping information, these methods often distribute importance arbitrarily across them, making it difficult to isolate the true temporal origin of the predictive signal. Consequently, interpretability degrades even when prediction performance remains high.

2.8.2 Inter-Ratio Correlation at CQ_1

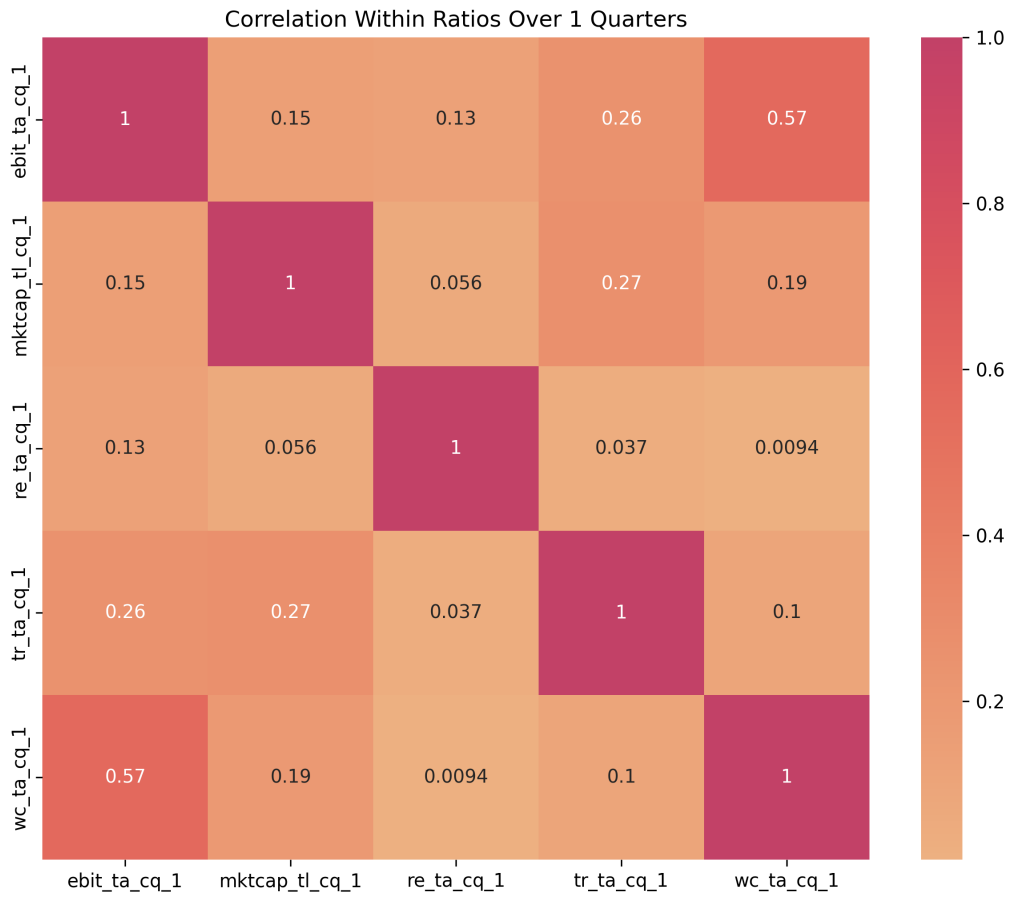


Figure 2.9: Cross Ratio Correlation at *CQ_1*

To address this, we explore a simplified input structure that retains only the most

recent value for each ratio, i.e., calendar quarter 1 (CQ_1). Figure 2.9 shows the correlation heatmap among the five financial ratios at CQ_1. The inter-ratio correlations are notably lower, with most values below 0.30. The highest observed correlation is between *EBIT/TA* and *WC/TA* (0.57). Other pairs exhibit weak associations, suggesting minimal redundancy. This reduction in multicollinearity offers a more interpretable feature set.

2.9 Data Transformations

To enhance the stability and predictive power of the models, a series of data transformations were applied uniformly across all features in the dataset. These preprocessing steps were consistently performed for every model and across all experimental settings. Importantly, transformations were applied at the CQ level to maintain the temporal structure and ensure no information from future quarters leaked into earlier ones. This consistent application ensured fair and comparable input representations across the full modeling pipeline.

2.9.1 Winsorization of Outliers

Outliers in financial ratios can severely distort model training, particularly in datasets with significant inter-firm heterogeneity or during particular economic periods. To reduce the influence of extreme values while preserving the overall structure of the data, we applied *per-quarter Winsorization*. Specifically, for each feature within a given calendar quarter, values were clipped at the 1st and 99th percentiles. Let $x_i^{(q)}$ denote the value of a given feature for instance i in quarter q. Define:

$$x_{i,\text{wins}}^{(q)} = \begin{cases} P_1^{(q)} & \text{if } x_i^{(q)} < P_1^{(q)} \\ x_i^{(q)} & \text{if } P_1^{(q)} \leq x_i^{(q)} \leq P_{99}^{(q)} \\ P_{99}^{(q)} & \text{if } x_i^{(q)} > P_{99}^{(q)} \end{cases}$$

where $P_1^{(q)}$ and $P_{99}^{(q)}$ represent the 1st and 99th percentiles of the feature values within quarter q. This transformation preserves the temporal context of the data while dampening extreme variations.

2.9.2 Logarithmic Transformation for Skewness

Many financial ratios exhibit skewed distributions due to extreme values. To reduce skewness and stabilize variance, a *signed modulus-log transformation* was applied:

$$x_{i,\log} = \text{sign}(x_i) \cdot \log(1 + |x_i|)$$

This transformation is monotonic and preserves the sign of the original data while compressing the range of extreme values. It is particularly useful for dealing with heavy-tailed distributions and reducing the impact of large-magnitude observations.

2.9.3 Robust Scaling for Temporal Normalization

To ensure that the magnitude of features is comparable across different quarters, we performed *robust normalization per quarter*. Unlike z-score normalization, which uses the mean and standard deviation, robust scaling is less sensitive to outliers as it uses the median and interquartile range (IQR). For each feature $x_i^{(q)}$ in quarter q, the normalized value is computed as:

$$\tilde{x}_i^{(q)} = \frac{x_i^{(q)} - \text{median}^{(q)}}{\text{IQR}^{(q)}}$$

Here, $\text{IQR}^{(q)} = Q_{75}^{(q)} - Q_{25}^{(q)}$, where $Q_{75}^{(q)}$ and $Q_{25}^{(q)}$ denote the 75th and 25th percentiles of the feature distribution within the quarter. This transformation centers each quarterly group around zero and scales it based on its spread, allowing the model to learn from relative rather than absolute values across time.

2.10 Bankruptcies Distribution

This section provides an overview of the dataset used to develop and evaluate the bankruptcy prediction models. The dataset is structured at the firm-quarter level, with each instance representing the financial profile of a company over a sequence of 20 consecutive calendar quarters.

Table 2.6 presents the number of unique companies, bankrupt companies, and the total number of healthy and bankrupt observations in the training, validation, and test sets for each of the three bankruptcy prevalence levels (1%, 5%, and 10%). The number of bankrupt observations was held constant across splits to ensure a controlled experimental setup, facilitating a fair comparison of model behavior under varying class imbalance and false negative cost configurations.

Table 2.6: Dataset Composition by Bankruptcy Prevalence Level

Bankruptcy Rate	Split	Companies	Bankrupt Companies	Healthy Observations	Bankrupt Observations
1%	Training	3,757	325	503,118	5,082
1%	Validation	1,825	42	29,799	301
1%	Test	1,852	39	31,185	315
5%	Training	3,350	325	96,558	5,082
5%	Validation	952	42	5,719	301
5%	Test	988	39	5,985	315
10%	Training	2,936	325	45,738	5,082
10%	Validation	716	42	2,709	301
10%	Test	733	39	2,835	315

It is important to note that since firms with incomplete financial histories were excluded, resulting in the omission of bankruptcy events occurring before the first quarter of 2004 (*CQ1_2004*). Table 2.7 details the number of bankruptcies observed per calendar quarter between 2004 and 2023. This distribution highlights temporal fluctuations in bankruptcy incidence, reflecting both macroeconomic cycles and firm-specific vulnerabilities.

Figure 2.10 illustrates the annual distribution of bankruptcies between 2004 and 2023. The trend reveals two prominent peaks: one in 2009, following the global financial crisis,

Table 2.7: Number of Bankruptcies per Quarter (2004–2023)

Year	CQ1	CQ2	CQ3	CQ4	Year	CQ1	CQ2	CQ3	CQ4
2004	26	11	15	12	2014	14	11	15	13
2005	15	13	26	10	2015	18	16	25	20
2006	21	10	21	11	2016	19	11	40	27
2007	20	18	33	20	2017	17	14	27	17
2008	33	27	45	36	2018	11	10	16	11
2009	17	11	43	15	2019	15	19	36	17
2010	13	12	70	6	2020	9	7	106	20
2011	7	10	40	12	2021	4	5	5	4
2012	12	10	52	7	2022	7	3	20	10
2013	14	15	18	12	2023	3	5	81	26

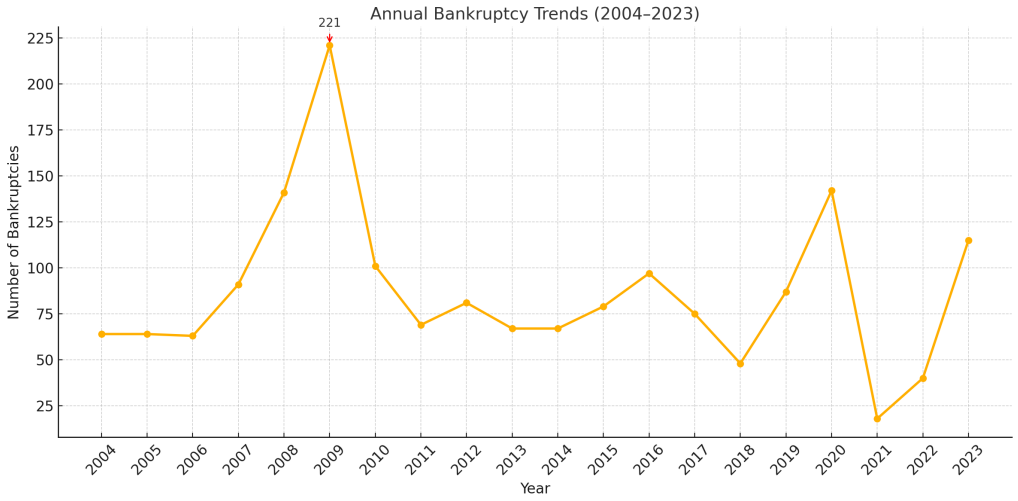


Figure 2.10: Annual number of bankruptcies from 2004 to 2023

and another in 2020, coinciding with the COVID-19 pandemic. These peaks align with periods of significant macroeconomic instability. Interestingly, we observe a decline in bankruptcies in 2021 and 2022. This may be partly attributed to temporary relief measures deployed at the time, raising concerns about a potential backlog of distressed firms.

Chapter 3

Methodology

This chapter presents the methodology adopted to develop, evaluate, and interpret machine learning models for corporate bankruptcy prediction. The process consists of multiple stages, outlined as follows:

1. Constructing datasets with varying bankruptcy prevalence levels (1%, 5%, and 10%) to simulate different economic environments,

2. Training a range of machine learning models, including both non-boosting and boosting approaches,

3. Optimizing model hyperparameters using the Optuna framework, with a cost-sensitive objective that reflects stakeholder-specific asymmetries between false negatives and false positives,

4. Evaluating model performance primarily based on total misclassification cost, with secondary consideration given to conventional metrics such as precision, recall, F1-score, and AUC,

5. Comparing model performance under two temporal input configurations: full historical data (20 consecutive CQs) versus the most recent quarter (CQ_1),

6. Applying Explainable AI (XAI) techniques to models trained on CQ_1 inputs in order to enhance interpretability. Global interpretability is assessed using Permutation Feature Importance (PFI) and Mean Absolute SHAP values , while local explanations are provided using Individual Conditional Expectation (ICE) plots and SHAP waterfall plots.

3.1 Model Classes and Setup

The machine learning models used in this study are categorized into two broad families based on their learning mechanisms:

- **Boosting Models**: These ensemble learners iteratively improve predictions by focusing on previously misclassified observations, making them particularly effective for structured and imbalanced data. The models in this group include AdaBoost, XGBoost, and LightGBM.

- **Non-Boosting Models**: To establish meaningful baselines and evaluate performance trade-offs, we include Logistic Regression, Decision Tree, Random Forest, and a Multilayer Perceptron (MLP).

All models were implemented in Python using the `scikit-learn`, `xgboost`, `lightgbm`, and `pytorch` libraries. The training and evaluation pipelines were designed to ensure reproducibility, consistent preprocessing, and fair comparison across model classes.

3.2 Input Configurations and Temporal Window Comparison

To evaluate the influence of historical information on predictive performance, two distinct input configurations were explored:

- **20-Quarter Input Window**: The primary setup uses lagged financial ratios spanning the 20 quarters preceding each firm's bankruptcy evaluation point.
- **Most Recent Quarter (CQ_1)**: A secondary configuration simplifies the input by retaining only the most recent value of each financial ratio, aiming to improve explainability and reduce redundancy.

To maintain a fair comparison, both configurations were applied to the same dataset sample, without regenerating the dataset for shorter history windows. That is, firms included in the 20-quarter configuration were retained in the CQ_1 setting, excluding those with no historical data. This consistency in the training dataset ensures that any observed differences in model performance stem from the temporal input structure rather than sample variation.

3.3 Bankruptcy Level Thresholds

Bankruptcy events are rare, particularly in developed economies. However, their prevalence can vary significantly under different macroeconomic conditions. To evaluate model robustness under varying class imbalance levels, three datasets were created with different bankruptcy prevalence levels:

A 1% bankruptcy rate reflects an optimistic, low-risk environment, consistent with periods of economic expansion or stable credit markets. A 5% bankruptcy rate represents a neutral or average scenario. Finally, a 10% bankruptcy rate simulates a stressed or pessimistic market condition, capturing the dynamics of financial downturns or recessionary periods with elevated firm default rates.

3.4 Cost Matrices and Stakeholder Perspectives

Binary classification typically assumes uniform misclassification costs. However, in the context of bankruptcy prediction, FN cases where a bankrupt firm is misclassified as

healthy, carry significantly higher consequences than FP.

To reflect stakeholder-specific risk tolerances, we defined three asymmetric cost matrices, parameterized by the FN:FP ratio:

Table 3.1: False Negative and False Positive Cost Settings by Scenario

Scenario	FN Cost	FP Cost
Moderate	5	1
Balanced	10	1
Conservative	20	1

Different stakeholders are likely to favor different cost configurations based on their exposure to risk and tolerance for false alarms. Investors, for instance, may opt for a moderate cost ratio of 5:1 to avoid unnecessary alerts that could disrupt portfolio strategy. Regulators and lenders, on the other hand, are more likely to adopt a conservative stance, assigning a much higher cost to false negatives—up to 20 times that of a false positive—to ensure that potential bankruptcies are not overlooked. Meanwhile, auditors and rating agencies might favor a more balanced configuration (10:1), aiming to strike a compromise between the cost of misreporting and the operational burden of issuing false warnings.

To reflect these varying risk preferences, stakeholder-informed cost matrices were incorporated throughout the modeling pipeline: during hyperparameter tuning, threshold selection, and model performance evaluation under each scenario. In this thesis, we will refer to these cost configurations using shorthand such as *FN cost = 5x* or *FN penalty = 5* to indicate the relative weighting applied to false negatives.

3.5 Hyperparameter Optimization

To ensure fair and optimal model comparisons, hyperparameter tuning was performed using Optuna, which is a state-of-the-art Bayesian optimization framework. This allowed efficient exploration of high-dimensional, non-convex search spaces and reduced the number of trials needed to converge on high-performing configurations.

3.5.1 Optimization Objective

Unlike conventional hyperparameter tuning processes that aim to maximize generic metrics such as AUC or accuracy, this study adopts a cost-sensitive approach grounded in the relative penalties of misclassification. Specifically, hyperparameters are selected by minimizing the following misclassification cost function on the validation set:

$$\text{Total Cost} = (\text{FP} \times \text{CostFP}) + (\text{FN} \times \text{CostFN}) \tag{3.1}$$

The models are trained with gradients computed and backpropagated in the usual manner. However, during validation, model predictions are evaluated in terms of the total cost defined above. Hyperparameter configurations that yield the lowest total misclassification cost are retained. This approach ensures that model selection aligns with the asymmetric cost structure associated with false positives and false negatives, which is central to the application context.

3.5.2 Search Spaces

The search spaces presented in table 3.2 were used for hyperparameter tuning.

3.6 Model Evaluation and Interpretability

This section describes the evaluation procedures used to assess model performance and interpretability. Given the high cost of misclassification in bankruptcy prediction, our evaluation framework prioritizes cost-sensitive analysis while also reporting standard classification metrics for completeness and comparability with existing literature. Moreover, we applied both global and local explainability techniques to better understand model behavior and feature importance.

Table 3.2: Hyperparameter Search Space for Each Model

Note. "Range" denotes the set of values searched during hyperparameter tuning. Intervals such as (a, b) indicate continuous ranges, while "step" refers to uniform sampling within a discrete grid. Parameters with "log scale" were searched over a logarithmic scale.

Model	Hyperparameters (Range)
XGBoost	*n_estimators*: (100, 1000, step 100) *max_depth*: (3, 15, step 1) *learning_rate*: (0.01, 0.3, log scale) *subsample*: (0.6, 1.0) *colsample_bytree*: (0.6, 1.0) *reg_alpha*: (0, 10) *reg_lambda*: (0, 10)
LightGBM	*n_estimators*: (100, 1000, step 100) *max_depth*: (3, 15) *learning_rate*: (0.01, 0.3, log scale) *num_leaves*: (20, 150, step 10) *subsample*: (0.6, 1.0) *colsample_bytree*: (0.6, 1.0) *reg_alpha*: (0, 10) *reg_lambda*: (0, 10)
Random Forest	*n_estimators*: (100, 600, step 100) *max_features*: (1, *num_features*, step 1) *max_depth*: (5, 20, step 1) *min_samples_split*: (2, 10, step 1) *min_samples_leaf*: (1, 10, step 1)
Decision Tree	*max_depth*: (3, 20) *min_samples_split*: (2, 10, step 1) *min_samples_leaf*: (1, 10, step 1)
Logistic Regression	*C*: (0.001, 10, log scale) *solver*: {lbfgs, newton-cg, liblinear}
AdaBoost	*n_estimators*: (50, 500, step 50) *learning_rate*: (0.001, 1.0, log scale)
MLP	*learning_rate_cls*: (1e-8, 1e-4, log scale) *batch_size*: {512, 1024} *num_cls_hidden_layers*: (1, 4, step 1) *cls_hidden_size_layer_i*: (64, 512, step 64) *alpha*: (0.5, 0.99) *gamma*: {2, 3}

3.6.1 Evaluation Metrics

Primary Metric: Total Misclassification Cost

The central evaluation metric in this study is the total cost associated with misclassifications, calculated using the asymmetric cost matrices. This cost reflects stakeholder-specific preferences and penalizes FN more heavily than FP. It is given by equation 3.1

This formulation ensures that model selection and threshold tuning are aligned with the economic and strategic priorities of different stakeholder groups. For each cost scenario, the optimal classification threshold was identified by minimizing the cost on the validation set.

Secondary Metrics: Traditional Classification Metrics

While total cost served as the primary decision criterion for model selection and evaluation, several traditional classification metrics were also reported to facilitate comparison with prior studies and to provide a more comprehensive assessment of model behavior. We calculated Recall (Sensitivity), which measures the proportion of correctly identified bankrupt firms. Precision was also evaluated, quantifying the proportion of predicted bankruptcies that were correct and therefore offering insight into the model's false alarm rate. To balance the trade-off between recall and precision, we reported the F1-Score, the harmonic mean of the two, which is particularly informative in imbalanced settings. Finally, we included the Area Under the Receiver Operating Characteristic Curve (AUC), a threshold-independent metric that captures the model's overall ability to discriminate between bankrupt and healthy firms. All secondary metrics were computed on the test set predictions, using the optimal threshold identified for each model and cost configuration. The formula for each metric is given below:

- **Recall (Sensitivity)**:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

- **Precision**:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **F1-Score**:

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

- **Area Under the ROC Curve (AUC)**:

  AUC summarizes the Receiver Operating Characteristic curve, which plots the True Positive Rate (Recall) against the False Positive Rate across different threshold settings. Formally, it can be expressed as:

$$\text{AUC} = \int_0^1 \text{TPR}(\text{FPR})\, d(\text{FPR})$$

  where TPR is the True Positive Rate and FPR is the False Positive Rate.

3.6.2 Explainable AI Techniques

To enhance transparency and increase trust in model predictions, we applied Explainable AI (XAI) techniques to the best-performing model from each model family, across all bankruptcy prevalence levels and cost scenarios. These interpretability methods were applied specifically to models trained using the CQ_1 input configuration, a design choice that minimizes multicollinearity and improves the reliability of feature attribution. The selected XAI tools provide both global and local insights into feature influence and model decision-making.

Global Interpretability

Global interpretability techniques help to identify the overall drivers of model predictions across the dataset, providing stakeholders with an understanding of which features matter most and how they interact.

- Permutation Feature Importance: For each feature, its values are randomly permuted to break the association with the target. The resulting drop in performance,

or increase in the loss function, indicates the importance of that feature. This is a model-agnostic that was applied on all classifiers. In this study, we will use the false negative penalty as loss function to the PFI calculation.

- Mean Absolute SHAP Values: SHAP values decompose each prediction into additive contributions from each feature. Aggregating the absolute values across all instances provides a global importance ranking that is consistent and theoretically grounded in cooperative game theory.

These tools were used to assess whether boosting models and non-boosting models relied on similar features, and whether their decision logic was robust under different prevalence and cost conditions.

Local Interpretability

Local interpretability focuses on explaining individual predictions, which is critical in high-stakes applications where users need to understand why a particular firm was classified as high-risk or low-risk.

ICE plots visualize the marginal effect of a feature on the model's predicted probability for individual instances. We favour ICE plots over partial dependence plots (PDP) since ICE captures heterogeneity by plotting one curve per instance, thus enabling the detection of non-monotonic effects and interactions.

Chapter 4

Results – Model Comparison

This chapter presents the empirical results of the bankruptcy prediction models developed in this study. The primary goal is to compare the performance of boosting and non-boosting model classes across different experimental conditions that simulate varying economic environments and stakeholder cost preferences.

Model evaluation is centered on minimizing total misclassification cost, which captures the asymmetric penalties associated with false negatives and false positives. This cost-sensitive metric reflects the priorities of key stakeholders, such as investors, auditors, and regulators. To complement this primary criterion, traditional classification metrics including precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC), are reported as secondary indicators of model performance.

Results are organized by bankruptcy prevalence level (1%, 5%, and 10%). Within each scenario, models are evaluated under three cost configurations that assign increasing weight to false negatives (5x, 10x, and 20x the cost of a false positive). For each combination of model, prevalence, and cost setting, we report results under two input configurations: a full 20-quarter historical input window and a reduced 1-quarter setup using only the most recent financial data. To facilitate direct comparison, results for the 1 CQ configuration are shown in parentheses directly below the corresponding 20 CQ row for each model.

4.1 Evaluation at 1% Bankruptcy Level

The results are summarized in tables 4.1, 4.2 and 4.3. Figure 4.1 and Figure 4.2 illustrate the precision-recall trade-offs and threshold-dependent cost behavior of all models trained with 20 CQ data.

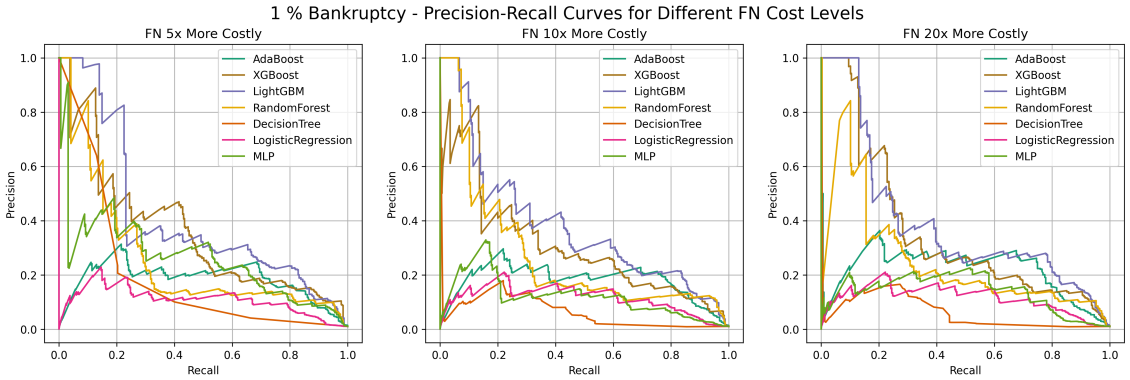


Figure 4.1: Precision-Recall Curves at 1% Bankruptcy Rate under FN cost ratios of 5x, 10x, and 20x. (20 CQ)

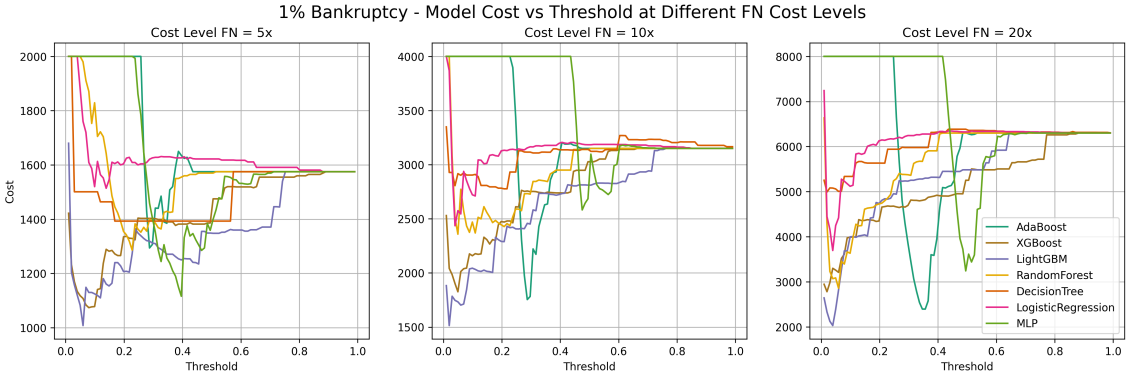


Figure 4.2: Threshold-cost tradeoffs at 1% Bankruptcy Rate for all models across FN cost ratios. (20 CQ)

4.1.1 FN Cost = 5x

Under the FN cost setting of 5x at 1% bankruptcy prevalence, XGBoost emerges as the best-performing boosting model, achieving the lowest total cost (1145) and the highest F1-score (0.400) among boosting models. The inclusion of 20 calendar quarters (CQ)

Table 4.1: Performance at 1% Bankruptcy Rate with FN Cost = 5x. The first row for each model shows results for 20 CQ input, and the second row (in parentheses) shows results for 1 CQ input.

Model	Cost	AUC	Precision	Recall	F1
XGBoost	**1145**	0.965	0.442	0.366	0.400
	(1202)	(0.954)	(0.375)	(0.356)	(0.365)
LightGBM	1146	0.968	0.291	0.531	0.376
	(1266)	(0.939)	(0.292)	(0.381)	(0.331)
AdaBoost	1176	0.955	0.240	0.689	0.356
	(**1159**)	(0.942)	(0.240)	(0.721)	(0.360)
Random Forest	1371	0.951	0.356	0.204	0.259
	(**1355**)	(0.928)	(0.248)	(0.356)	(0.292)
Decision Tree	1393	0.776	0.641	0.130	0.216
	(1371)	(0.797)	(0.229)	(0.397)	(0.290)
Logistic Regression	1545	0.897	0.193	0.117	0.146
	(1581)	(0.924)	(0)	(0)	(0)
MLP	**1264**	0.919	0.497	0.248	0.331
	(1389)	(0.949)	(0.425)	(0.162)	(0.234)

of input history significantly enhances its performance compared to the 1-CQ variant, reducing cost and improving all key metrics, underscoring the value of temporal context in detecting rare bankruptcy signals.

Among non-boosting models, the MLP delivers the best results, with a cost of 1264, outperforming all other traditional models. We also note that the Logistic Regression model was overfit in this configuration with our methodology. The optimal cost was attained when the model converged towards the trivial solution, marking all observations as "*healthy*". The best non-boosting model trained with 1-CQ data was Random Forest with a cost of 1355.

4.1.2 FN Cost = 10x

Under the 10x FN cost scenario, LightGBM is the best-performing boosting model, achieving the lowest total cost (1581) and the highest recall (0.839) among all models. The substantial cost reduction compared to its 1-CQ counterpart (1910) suggests the contribution

Table 4.2: Performance at 1% Bankruptcy Rate with FN Cost = 10x. The first row for each model shows results for 20 CQ input, and the second row (in parentheses) shows results for 1 CQ input.

Model	Cost	AUC	Precision	Recall	F1
XGBoost	1878	0.965	0.304	0.558	0.392
	(2204)	(0.953)	(0.173)	(0.575)	(0.266)
LightGBM	**1581**	0.968	0.259	0.839	0.395
	(1910)	(0.952)	(0.276)	(0.533)	(0.364)
AdaBoost	1709	0.955	0.226	0.695	0.341
	(**1852**)	(0.951)	(0.188)	(0.724)	(0.299)
Random Forest	2513	0.951	0.273	0.664	0.387
	(**2262**)	(0.929)	(0.250)	(0.403)	(0.308)
Decision Tree	2777	0.776	0.433	0.207	0.278
	(2438)	(0.803)	(0.155)	(0.498)	(0.236)
Logistic Regression	2476	0.897	0.264	0.646	0.375
	(2474)	(0.924)	(0.131)	(0.635)	(0.218)
MLP	**2419**	0.919	0.281	0.666	0.395
	(2432)	(0.916)	(0.156)	(0.495)	(0.238)

of long-term financial history in enhancing its predictive performance.

Among non-boosting models, the MLP delivers the best results, with a cost of 2419 and a matched F1-score of 0.395. The best-performing model trained with 1-CQ data was again Random Forest with a cost of 2262.

4.1.3 FN Cost = 20x

Under the most conservative cost configuration (FN cost = 20x), LightGBM is the top-performing boosting model with the lowest total cost (2126). The large cost gap between the 20-CQ and 1-CQ inputs (2126 vs. 3643) again highlights the value of long-term financial histories in improving predictive accuracy under extreme cost asymmetry and data imbalance.

For non-boosting models, the MLP continues to be the strongest performer, achieving a cost of 2925. While not as cost-efficient as LightGBM, the MLP presents a compelling alternative among traditional and deep learning baselines. In this setting as well, Random

Table 4.3: Performance at 1% Bankruptcy Rate with FN Cost = 20x. The first row for each model shows results for 20 CQ input, and the second row (in parentheses) shows results for 1 CQ input.

Model	Cost	AUC	Precision	Recall	F1
XGBoost	2992	0.965	0.268	0.619	0.377
	(3300)	(0.954)	(0.152)	(0.660)	(0.247)
LightGBM	**2126**	0.968	0.223	0.796	0.351
	(3643)	(0.951)	(0.140)	(0.610)	(0.227)
AdaBoost	2273	0.955	0.259	0.746	0.384
	(**2674**)	(0.946)	(0.211)	(0.708)	(0.325)
Random Forest	3091	0.951	0.273	0.660	0.384
	(**3605**)	(0.925)	(0.133)	(0.635)	(0.220)
Decision Tree	5014	0.776	0.241	0.263	0.252
	(3948)	(0.811)	(0.166)	(0.498)	(0.249)
Logistic Regression	4173	0.897	0.253	0.630	0.361
	(3624)	(0.924)	(0.131)	(0.635)	(0.218)
MLP	**2925**	0.919	0.256	0.667	0.372
	(4426)	(0.861)	(0.104)	(0.524)	(0.173)

Forest was the best performing model when trained with 1-CQ features, with a cost of 3605.

4.2 Evaluation at 5% Bankruptcy Level

This section evaluates model performance at a moderate bankruptcy prevalence of 5%, where the data distribution reflects more frequent corporate distress events compared to the 1% setting. Performance is analyzed under the same cost levels and input data size as in the previous section.

4.2.1 FN Cost = 5x

At the 5% bankruptcy level with FN cost = 5x, LightGBM is the best-performing boosting model, achieving the lowest total cost (589) and the highest F1-score (0.616) among all models. Notably, the cost increases substantially when restricted to a 1-CQ input (907),

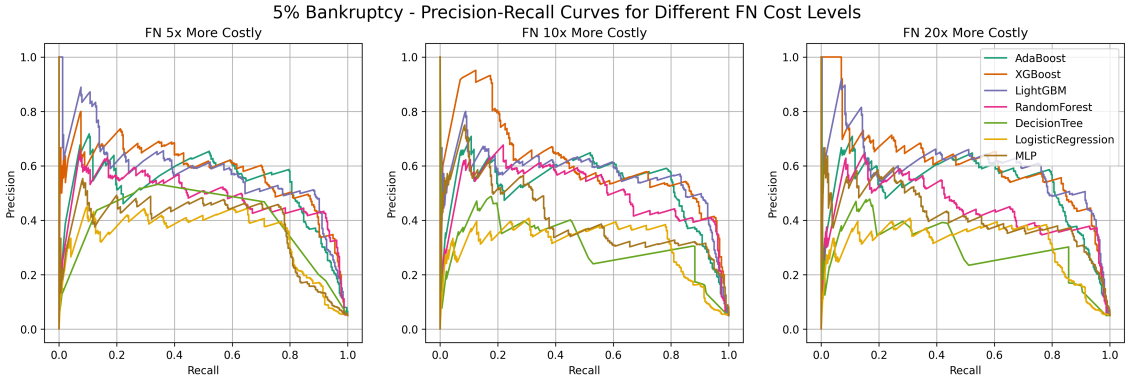


Figure 4.3: Precision-Recall Curves at 5% Bankruptcy Rate under FN cost ratios of 5x, 10x, and 20x. (20 CQ)

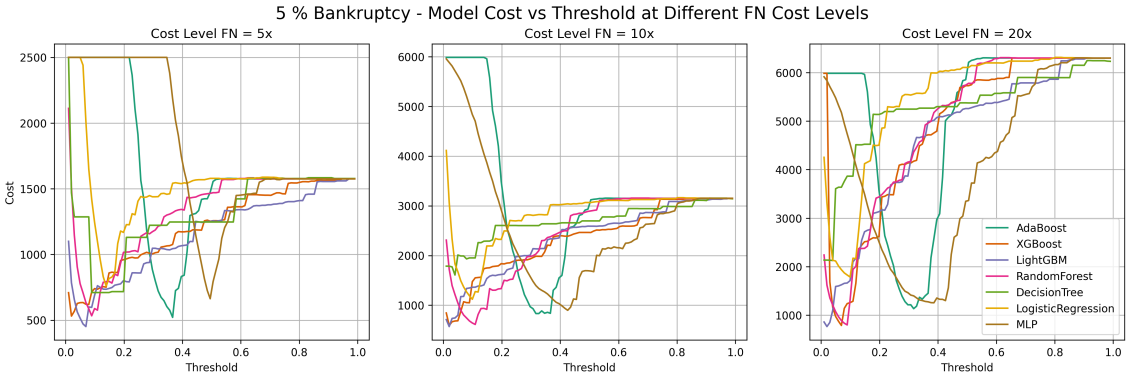


Figure 4.4: Threshold-cost tradeoffs at 5% Bankruptcy Rate for all models across FN cost ratios. (20 CQ)

confirming the importance of leveraging long-term financial trajectories.

Among non-boosting models, Random Forest achieves the lowest cost (591) and delivers a competitive F1-score (0.568), performing nearly on par with LightGBM and AdaBoost. It also benefits clearly from the 20-CQ input window, with both recall and cost improving significantly compared to the 1-CQ setup.

4.2.2 FN Cost = 10x

Under the 10x false negative penalty at 5% bankruptcy prevalence, XGBoost achieves the lowest cost (690), outperforming all models in terms of cost-efficiency while maintaining a high recall (0.916). AdaBoost is the best performing boosting model when trained with

Table 4.4: Performance at 5% Bankruptcy Rate with FN Cost = 5x. The first row for each model shows results for 20 CQ input, and the second row (in parentheses) shows results for 1 CQ input.

Model	Cost	AUC	Precision	Recall	F1
XGBoost	604	0.959	0.389	0.899	0.543
	(726)	(0.947)	(0.331)	(0.905)	(0.485)
LightGBM	**589**	0.958	0.511	0.775	0.616
	(907)	(0.935)	(0.422)	(0.584)	(0.490)
AdaBoost	599	0.956	0.429	0.844	0.569
	(**643**)	(0.952)	(0.401)	(0.844)	(0.543)
Random Forest	**591**	0.953	0.425	0.857	0.568
	(848)	(0.919)	(0.420)	(0.638)	(0.506)
Decision Tree	710	0.913	0.468	0.711	0.564
	(**789**)	(0.912)	(0.531)	(0.606)	(0.566)
Logistic Regression	859	0.893	0.391	0.660	0.491
	(818)	(0.916)	(0.324)	(0.825)	(0.456)
MLP	663	0.903	0.445	0.769	0.567
	(827)	(0.911)	(0.462)	(0.619)	(0.529)

Table 4.5: Performance at 5% Bankruptcy Rate with FN Cost = 10x. The first row for each model shows results for 20 CQ input, and the second row (in parentheses) shows results for 1 CQ input.

Model	Cost	AUC	Precision	Recall	F1
XGBoost	**690**	0.959	0.350	0.916	0.509
	(1051)	(0.939)	(0.261)	(0.930)	(0.407)
LightGBM	711	0.958	0.467	0.838	0.602
	(922)	(0.950)	(0.309)	(0.911)	(0.461)
AdaBoost	840	0.956	0.432	0.844	0.571
	(**920**)	(0.952)	(0.392)	(0.838)	(0.534)
Random Forest	**691**	0.953	0.365	0.907	0.520
	(1444)	(0.936)	(0.334)	(0.676)	(0.447)
Decision Tree	1019	0.913	0.359	0.779	0.478
	(1371)	(0.912)	(0.508)	(0.625)	(0.560)
Logistic Regression	1169	0.893	0.310	0.765	0.440
	(**1098**)	(0.916)	(0.313)	(0.835)	(0.455)
MLP	896	0.903	0.364	0.817	0.502
	(1144)	(0.918)	(0.268)	(0.876)	(0.410)

1-CQ data with a cost of 920.

Among non-boosting models, Random Forest matches XGBoost in cost (691) and delivers a nearly identical F1-score (0.520), making it the best-performing non-boosting model in this setting. Logistic Regression is the best model of this category when trained with 1-CQ data, achieving a cost of 1098.

4.2.3 FN Cost = 20x

Table 4.6: Performance at 5% Bankruptcy Rate with FN Cost = 20x. The first row for each model shows results for 20 CQ input, and the second row (in parentheses) shows results for 1 CQ input.

Model	Cost	AUC	Precision	Recall	F1
XGBoost	790	0.959	0.312	0.920	0.468
	(1258)	(0.946)	(0.286)	(0.914)	(0.436)
LightGBM	**766**	0.958	0.414	0.866	0.558
	(**1158**)	(0.952)	(0.300)	(0.924)	(0.453)
AdaBoost	1139	0.956	0.338	0.908	0.493
	(1378)	(0.948)	(0.337)	(0.867)	(0.485)
Random Forest	**968**	0.953	0.338	0.901	0.493
	(**1199**)	(0.943)	(0.265)	(0.940)	(0.414)
Decision Tree	1525	0.913	0.330	0.774	0.461
	(1539)	(0.905)	(0.345)	(0.835)	(0.488)
Logistic Regression	1850	0.893	0.287	0.768	0.423
	(1697)	(0.916)	(0.266)	(0.848)	(0.405)
MLP	1252	0.903	0.329	0.825	0.468
	(1809)	(0.920)	(0.226)	(0.860)	(0.358)

In the most conservative cost scenario (FN cost = 20x) at 5% bankruptcy prevalence, LightGBM achieves the lowest cost (766) among all models, making it the best-performing boosting model. It delivers a strong F1-score (0.558) and combines high recall (0.866) with respectable precision (0.414), favoring broad bankruptcy detection without excessive false alarms.

Among non-boosting models, Random Forest performs best with a cost of 968 and a high recall of 0.901.

LightGBM and Random Forest are also the best performing models when trained with 1-CQ data, with a cost of 1158 and 1199 respectively.

4.3 Evaluation at 10% Bankruptcy Level

This section examines model performance under a higher bankruptcy prevalence of 10%, where financially distressed firms are more frequent.

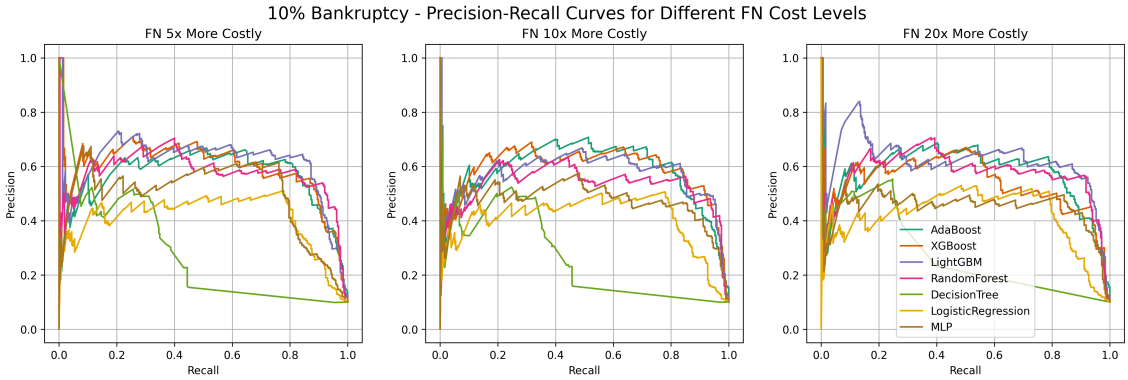


Figure 4.5: Precision-Recall Curves at 10% Bankruptcy Rate under FN cost ratios of 5x, 10x, and 20x. (20 CQ)

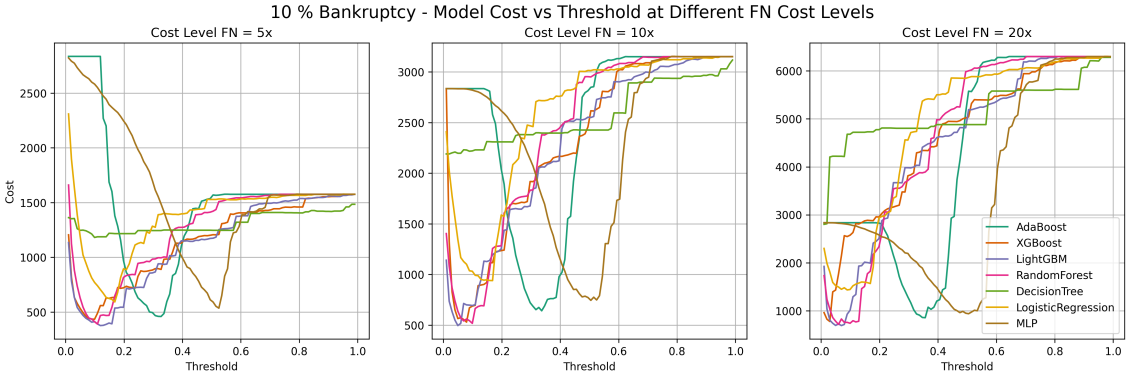


Figure 4.6: Threshold-cost tradeoffs at 10% Bankruptcy Rate for all models across FN cost ratios. (20 CQ)

Table 4.7: Performance at 10% Bankruptcy Rate with FN Cost = 5x. The first row for each model shows results for 20 CQ input, and the second row (in parentheses) shows results for 1 CQ input.

Model	Cost	AUC	Precision	Recall	F1
XGBoost	437	0.941	0.529	0.879	0.660
	(527)	(0.927)	(0.493)	(0.838)	(0.620)
LightGBM	**418**	0.946	0.554	0.876	0.678
	(**468**)	(0.934)	(0.470)	(0.908)	(0.619)
AdaBoost	453	0.943	0.529	0.867	0.657
	(530)	(0.932)	(0.433)	(0.898)	(0.585)
Random Forest	**471**	0.942	0.524	0.857	0.650
	(786)	(0.898)	(0.339)	(0.822)	(0.480)
Decision Tree	1248	0.619	0.301	0.387	0.339
	(1033)	(0.838)	(0.415)	(0.479)	(0.445)
Logistic Regression	601	0.879	0.486	0.784	0.600
	(**618**)	(0.902)	(0.431)	(0.825)	(0.566)
MLP	566	0.901	0.537	0.774	0.634
	(716)	(0.887)	(0.358)	(0.851)	(0.504)

4.3.1 FN Cost = 5x

At the 10% bankruptcy level with FN cost = 5x, LightGBM stands out as the top-performing boosting model, achieving the lowest total cost (418) and the highest F1-score (0.678). It offers a strong balance between precision (0.554) and recall (0.876), making it highly effective in moderately imbalanced settings. It is also the best model when 1-CQ data are fed as inputs.

Among non-boosting models, Random Forest is the best performer with a cost of 471 and an F1-score of 0.650. Logistic Regression emerges as the best performing model of this category when trained with 1-CQ data.

4.3.2 FN Cost = 10x

At a 10% bankruptcy prevalence and a 10x false negative cost, LightGBM achieves the lowest total cost (491) and the highest F1-score (0.645) among all models.

Among non-boosting models, Random Forest performs best with a cost of 686 and an

Table 4.8: Performance at 10% Bankruptcy Rate with FN Cost = 10x. The first row for each model shows results for 20 CQ input, and the second row (in parentheses) shows results for 1 CQ input.

Model	Cost	AUC	Precision	Recall	F1
XGBoost	574	0.941	0.470	0.901	0.614
	(**646**)	(0.933)	(0.403)	(0.933)	(0.563)
LightGBM	**491**	0.946	0.504	0.904	0.645
	(691)	(0.933)	(0.433)	(0.898)	(0.584)
AdaBoost	646	0.945	0.445	0.908	0.598
	(720)	(0.934)	(0.409)	(0.902)	(0.563)
Random Forest	**686**	0.942	0.452	0.887	0.598
	(**804**)	(0.902)	(0.317)	(0.949)	(0.475)
Decision Tree	2190	0.619	0.261	0.443	0.327
	(1050)	(0.838)	(0.266)	(0.921)	(0.413)
Logistic Regression	935	0.879	0.406	0.835	0.546
	(876)	(0.903)	(0.328)	(0.908)	(0.482)
MLP	715	0.901	0.456	0.821	0.584
	(816)	(0.897)	(0.356)	(0.905)	(0.511)

F1-score of 0.598 when trained with 20-CQ data, and with a cost of 804 and F1-score of 0.475 when trained with 1-CQ data.

4.3.3 FN Cost = 20x

Under the most conservative cost setting (FN cost = 20x) at 10% bankruptcy prevalence, LightGBM achieves the lowest total cost (742) among boosting models and delivers a strong F1-score (0.550) with very high recall (0.959). While its precision (0.385) is moderate, its cost advantage and sensitivity to bankruptcies make it the most effective model in this high-stakes scenario. Performance declines with 1-CQ input (cost rises to 903), reaffirming that access to long-term financial history is essential to optimizing detection under heavy false negative penalties. Yet, it remains the best performing boosting-model when trained with 1-CQ data.

Among non-boosting models, Random Forest delivers the lowest cost (708) and the highest F1-score (0.570), outperforming even the best boosting model in this configura-

Table 4.9: Performance at 10% Bankruptcy Rate with FN Cost = 20x. The first row for each model shows results for 20 CQ input, and the second row (in parentheses) shows results for 1 CQ input.

Model	**Cost**	**AUC**	**Precision**	**Recall**	**F1**
XGBoost	850	0.931	0.388	0.914	0.545
	(943)	(0.923)	(0.360)	(0.933)	(0.519)
LightGBM	**742**	0.946	0.385	0.959	0.550
	(**903**)	(0.924)	(0.339)	(0.949)	(0.500)
AdaBoost	893	0.944	0.358	0.943	0.519
	(908)	(0.933)	(0.405)	(0.924)	(0.563)
Random Forest	**708**	0.942	0.420	0.892	0.570
	(**879**)	(0.910)	(0.348)	(0.949)	(0.509)
Decision Tree	2814	0.619	0.260	0.460	0.328
	(1233)	(0.837)	(0.266)	(0.933)	(0.414)
Logistic Regression	1438	0.879	0.366	0.836	0.499
	(1149)	(0.903)	(0.321)	(0.914)	(0.475)
MLP	981	0.901	0.418	0.824	0.559
	(1007)	(0.894)	(0.372)	(0.917)	(0.530)

tion. It maintains a strong balance of precision (0.420) and recall (0.892), and, similar to LightGBM, suffers a performance drop in the 1-CQ case (cost rises to 879, F1 drops to 0.509). Interestingly, Random Forest beats the boosting models when trained with 1-CQ data.

Chapter 5

Results – Explainability Analysis

This chapter investigates the factors driving model predictions using XAI techniques. While predictive performance indicates how accurately models classify firms, explainability is essential to understanding the rationale behind these decisions, particularly in high-stakes applications such as financial risk assessment.

We focus on the best-performing model type across bankruptcy prevalence and cost settings to analyze feature importance and feature effects using complementary interpretability methods. These include permutation feature importance, mean absolute SHAP values and ICE plots, offering both global and local perspectives on model behavior. However, because of the high level of multicollinearity in our 20 CQ dataset, we perform the XAI analyses on the best models that were trained only with 1 CQ data, which has considerably less correlation.

The analysis is organized by bankruptcy prevalence level and examines both consistent predictors of financial distress and variations in feature relevance and attribution patterns across models and false negative cost configurations.

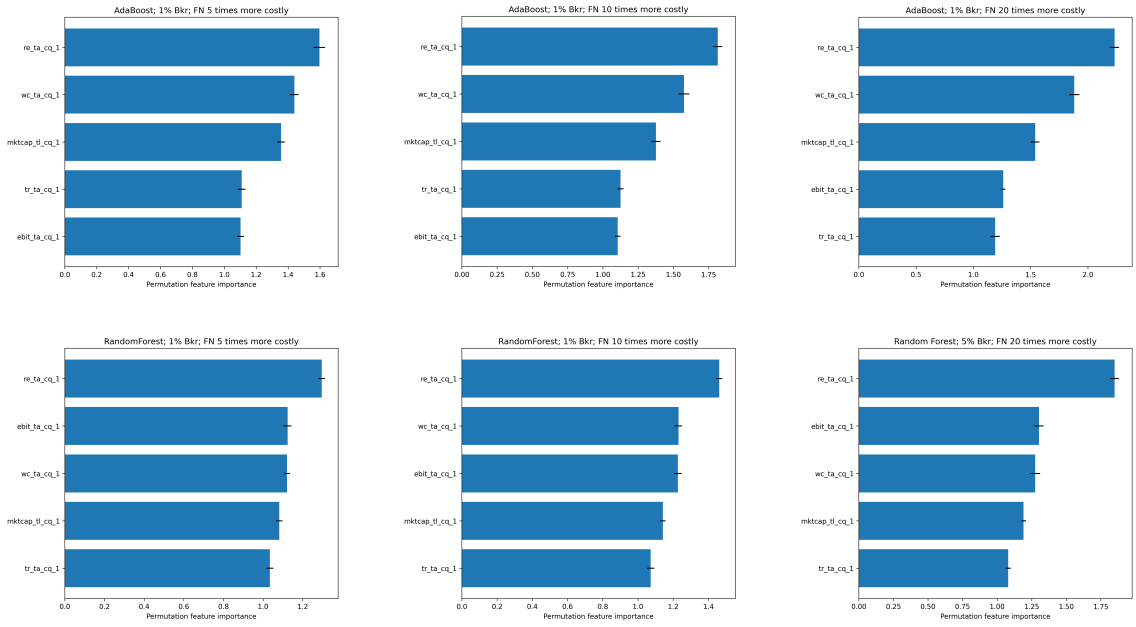


Figure 5.1: 1% Bankruptcy; PFI

5.1 PFI Results

5.1.1 1% Bankruptcy Level

Figure 5.1 (top) displays the permutation feature importance rankings for AdaBoost at a 1% bankruptcy prevalence level. Across all three cost configurations, the model consistently relies on *RE/TA*, *WC/TA*, and *MKTCAP/TL* as the primary drivers of bankruptcy prediction.

As the penalty for false negatives increases, AdaBoost becomes more selective, concentrating its reliance on a narrower subset of features. The importance scores for *RE/TA* and *WC/TA* increase from 5x FN cost to 20x FN cost. In contrast, the relative importance of *TR/TA* and *EBIT/TA* remains more stable or declines slightly relative to the other ratios.

Figure 5.1 (bottom) presents the permutation feature importance rankings for Random Forest. In all configurations, the model consistently ranks *RE/TA*, *EBIT/TA*, and *WC/TA* as the most influential predictors.

As the cost of false negatives increases, similar to AdaBoost, the importance scores for *RE/TA* increases, indicating the model's growing emphasis on retained earnings as a

robust signal of financial resilience. The relative importance of *EBIT/TA* also increases slightly, while *WC/TA* maintains a stable contribution across cost settings. In contrast, *MKTCAP/TL* and *TR/TA* consistently rank lower in terms of importance.

5.1.2 5% Bankruptcy Level

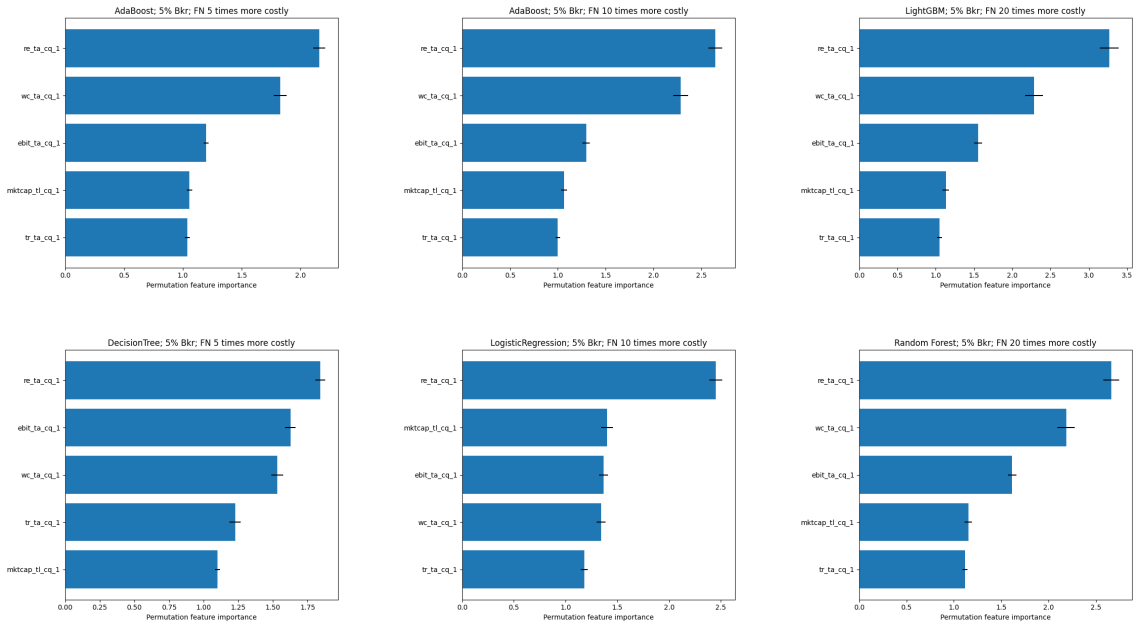


Figure 5.2: 5% Bankruptcy; PFI

Figure 5.2 presents permutation feature importance rankings for selected models under a 5% bankruptcy prevalence level. Across both boosting and non-boosting models, there is strong convergence in feature prioritization: *RE/TA* and *WC/TA* consistently dominate the attribution profiles. In contrast, *MKTCAP/TL* and *TR/TA* generally rank lower.

Among boosting models, AdaBoost and LightGBM focus on *RE/TA* and *WC/TA* as FN cost increases from 5x to 20x. The LightGBM at FN penalty$= 20\times$, in particular, shows the highest magnitude of feature importance values, suggesting stronger reliance on these key ratios to minimize high-stakes misclassification.

Random Forest displays a similar trend but with a more balanced feature distribution. The model still attributes meaningful weight to *EBIT/TA* and *WC/TA*, indicating a more diffuse reliance pattern.

Decision Tree and Logistic Regression follow comparable patterns, with *RE/TA* remaining the dominant signal. We also note that the PFI of the Logistic Regression model is diffused among the secondary ratios.

5.1.3 10% Bankruptcy Level

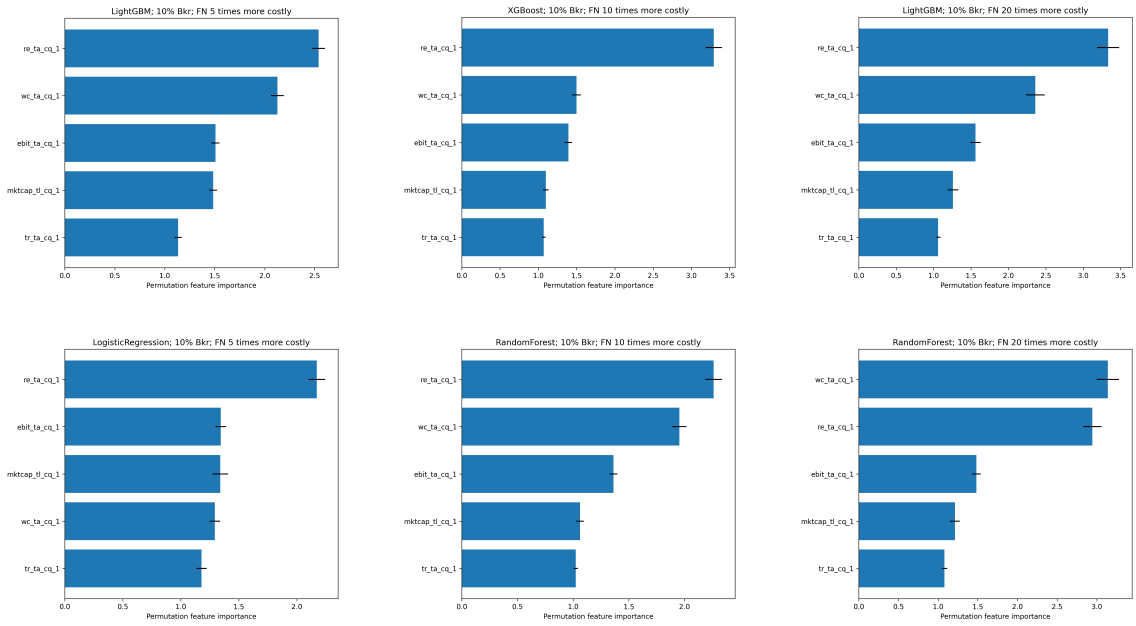


Figure 5.3: 10% Bankruptcy; PFI

Figure 5.3 summarizes the permutation feature importance rankings across selected models at a 10% bankruptcy prevalence level. As in previous settings, *RE/TA* and *WC/TA* consistently emerge as the most influential features across all models and cost scenarios.

For boosting models, *RE/TA* dominates in all configurations, with importance scores peaking at FN cost = 20x. The role of *WC/TA* also strengthens as FN penalties rise, while the contribution of *MKTCAP/TL* and *TR/TA* diminishes.

Random Forest also prioritizes *RE/TA* and *WC/TA*, but exhibits more balanced reliance across features than boosting models. Interestingly, at FN cost = 20x, *WC/TA* surpasses *RE/TA* in importance.

Logistic Regression demonstrates a similar ranking, with *RE/TA* as the top predictor. However, the separation between features is less pronounced, and the relative importance

of *WC/TA*, *EBIT/TA*, and *MKTCAP/TL* remains closer in magnitude, similar to its behaviour at 5% bankruptcy prevalence.

Overall, these results suggest the dominant role of *RE/TA* and *WC/TA* as early indicators of financial distress and show that boosting models intensify their focus on these ratios under conservative cost assumptions. Random Forest offers a more balanced feature importance, and logistic regression tends to have more diffused feature importance among secondary indicators.

5.2 SHAP-Based Analysis

To analyze feature importance at a specific point in time, we computed the mean absolute SHAP values for each of the five core financial ratios at the most recent calendar quarter. SHAP values were calculated on a log scale to account for variation in feature magnitudes. The analysis was performed across model classes (boosting and non-boosting) and false negative cost penalties under varying bankruptcy prevalence levels.

5.2.1 1% Bankruptcy Level

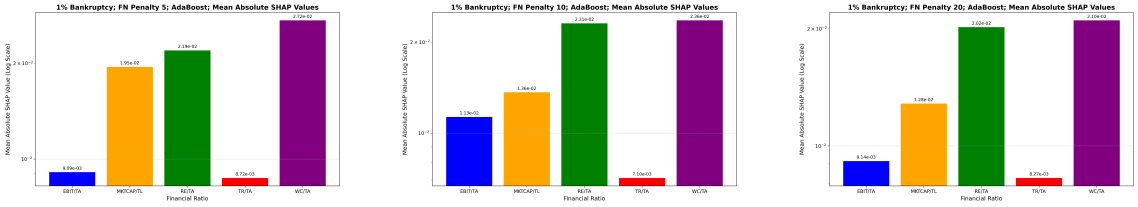


Figure 5.4: 1% Bankruptcy; Boosting; Mean Absolute SHAP Values

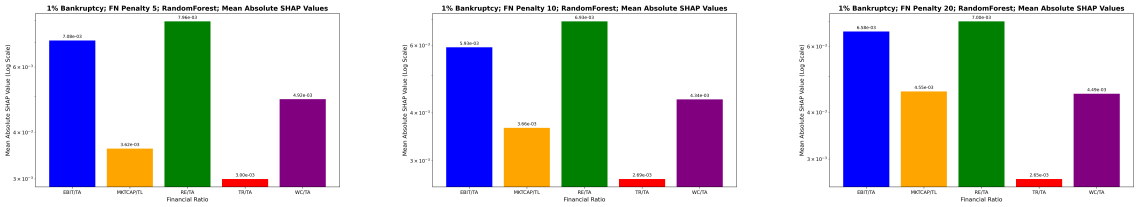


Figure 5.5: 1% Bankruptcy; Non-Boosting; Mean Absolute SHAP Values

Figures 5.4 and 5.5 present the mean absolute SHAP values for each financial ratio across different false negative (FN) penalty levels. These values reflect the global importance of each feature, computed as the average magnitude of its SHAP contribution across all test samples.

In the boosting models (Figure 5.4), *RE/TA* and *WC/TA* consistently emerge as the most influential predictors. At FN penalty 5, *WC/TA* dominates with a SHAP value of 2.72×10^{-2}, followed by *RE/TA* at 2.19×10^{-2}. As the FN penalty increases, both features maintain their prominence. This indicates that these two ratios are robust indicators of distress when minimizing costly misclassifications is prioritized. Meanwhile, *EBIT/TA* and *TR/TA* exhibit substantially lower SHAP values, suggesting reduced model reliance on profitability and revenue metrics under this setting.

In contrast, the Random Forest models (Figure 5.5) yield a more distributed attribution across features. While *RE/TA* still ranks highest at each FN penalty level, the gap between the top and bottom features is narrower than in boosting. *EBIT/TA* receives more consistent weight in Random Forest models, suggesting that this ratio plays a more stable role in this architecture. Similar to what we saw in the boosting models, *TR/TA* is the least influential feature.

5.2.2 5% Bankruptcy Level

Figures 5.6 and 5.7 display mean absolute SHAP values for boosting and non-boosting models under the 5% bankruptcy setting. In all boosting models, *RE/TA* and *WC/TA* consistently dominate, with sharply increasing SHAP values as the FN penalty rises. For instance, under LightGBM with FN cost 20, their values exceed 0.45, while *TR/TA* remains minimally influential throughout.

Non-boosting models show more varied attribution. Logistic regression heavily weights *RE/TA* (up to 1.32 at FN 10), while decision trees and random forests offer more balanced distributions. *MKTCAP/TL* has smaller influence in these models, especially at FN penalties 5 and 20.

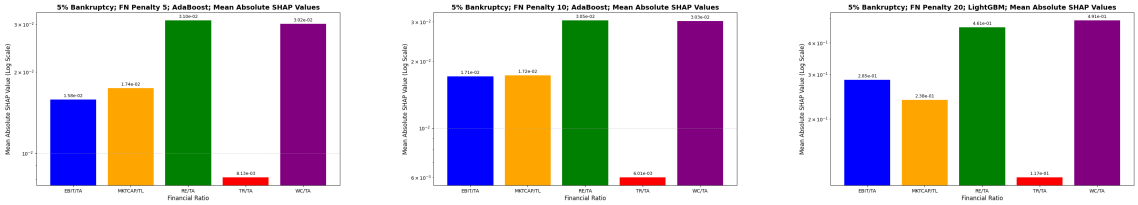


Figure 5.6: 5% Bankruptcy; Boosting; Mean Absolute SHAP Values

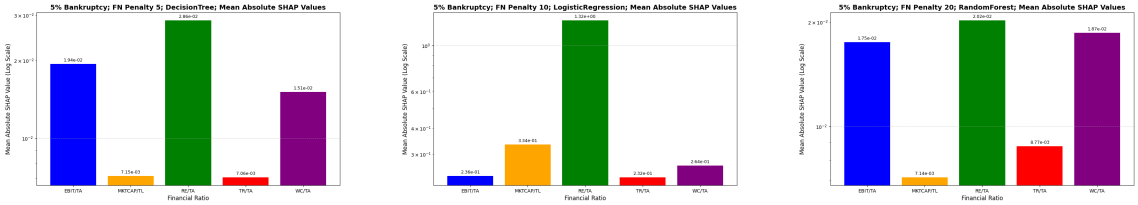


Figure 5.7: 5% Bankruptcy; Non-Boosting; Mean Absolute SHAP Values

5.2.3 10% Bankruptcy Level

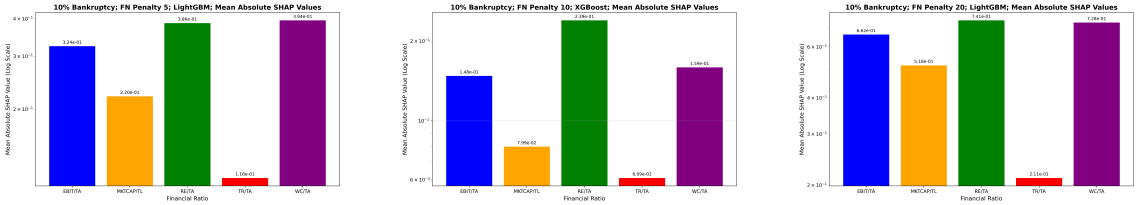


Figure 5.8: 10% Bankruptcy; Boosting; Mean Absolute SHAP Values

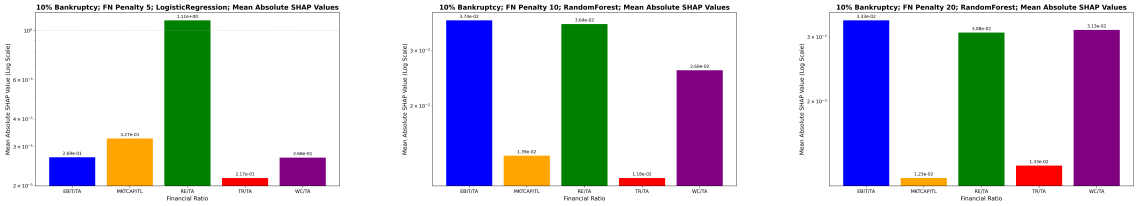


Figure 5.9: 10% Bankruptcy; Non-Boosting; Mean Absolute SHAP Values

Figures 5.8 and 5.9 display the mean absolute SHAP values for boosting and non-boosting models under the 10% bankruptcy setting across varying FN cost penalties.

In the boosting models, *RE/TA* and *WC/TA* consistently emerge as the most influential features. At FN penalty 5, both ratios register SHAP values close to 0.39, indicating strong and balanced influence. As the FN penalty increases, their importance intensifies further and reaching above 0.7 under LightGBM at FN 20. *EBIT/TA* also becomes more

prominent at higher penalties. In contrast, *TR/TA* remains the weakest contributor across all boosting configurations.

Logistic regression heavily weights *RE/TA* (SHAP = 1.11 at FN 5), while decision trees and random forests assign more balanced importance to *RE/TA*, *EBIT/TA*, and *WC/TA*. Across all non-boosting models, *TR/TA* and *MKTCAP/TL* show consistently lower influence.

These results confirm the robustness of *RE/TA* and *WC/TA* in higher-prevalence distress settings, while highlighting the sharper cost sensitivity in boosting architectures compared to the more stable attribution patterns of non-boosting models.

5.3 ICE Plots

This section uses Individual Conditional Expectation (ICE) plots to analyze how the *CQ_1* ratios influence model predictions at the instance level. ICE plots provide a fine-grained view of model behavior by illustrating how changes in a given feature affect the predicted probability of bankruptcy for individual samples, while holding other features fixed. We focus on the best-performing model trained with *CQ_1* accounting ratios within each family across all bankruptcy prevalence levels and cost ratios.

5.3.1 1% Bankruptcy Level

Boosting

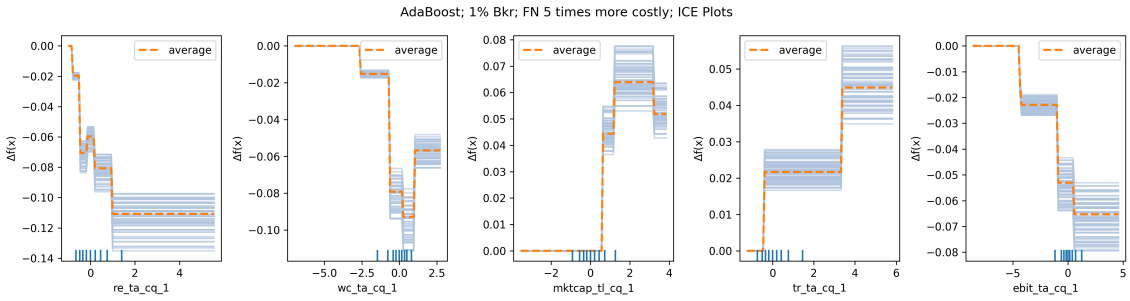


Figure 5.10: ICE Plots; 1% Bankruptcy Level; Boosting; FN Penalty 5

FN Cost 5 Figure 5.10 presents ICE plots for AdaBoost under the 1% bankruptcy scenario with a false negative cost five times higher than a false positive. Each subplot illustrates the marginal effect of a specific financial ratio on the model's predicted bankruptcy risk, with 100 blue lines representing 100 randomly selected individual instances and the orange dashed line showing the partial dependence (average effect).

The y-axis shows $\Delta f(x)$, which represents the change in the model's predicted output when the value of a single feature is varied, holding all other features fixed. A negative $\Delta f(x)$ indicates that increasing the feature reduces the model's predicted probability of bankruptcy, while a positive value reflects a rising predicted risk. These effects are computed for each observation, producing a family of curves that reveal both global trends and individual-level variation.

The features *RE/TA*, *WC/TA*, and *EBIT/TA* all exhibit strong negative marginal effects: as values increase, the model's predicted risk of bankruptcy decreases. For *RE/TA* and *EBIT/TA*, this effect is monotonic, with the average prediction sharply dropping near zero and plateauing at higher values. This suggests that firms with stronger retained earnings or earnings before interest and taxes are perceived by the model as significantly less likely to default.

WC/TA shows a steep drop in prediction risk around zero, indicating a threshold-like behavior where firms with positive working capital are viewed as substantially safer than those with negative liquidity positions. This aligns with theoretical expectations about solvency and short-term operational risk.

In contrast, *MKTCAP/TL* and *TR/TA* exhibit positive marginal effects, with higher values increasing the model's predicted risk. For *MKTCAP/TL*, the upward trend may reflect the model's learned association between high equity volatility or inflated market valuations and potential distress. The positive relationship in *TR/TA* is weaker but still notable, suggesting that revenue alone may not mitigate bankruptcy risk.

FN Cost 10 Figure 5.11 displays ICE plots for AdaBoost under the 1% bankruptcy scenario with a false negative cost ten times greater than a false positive.

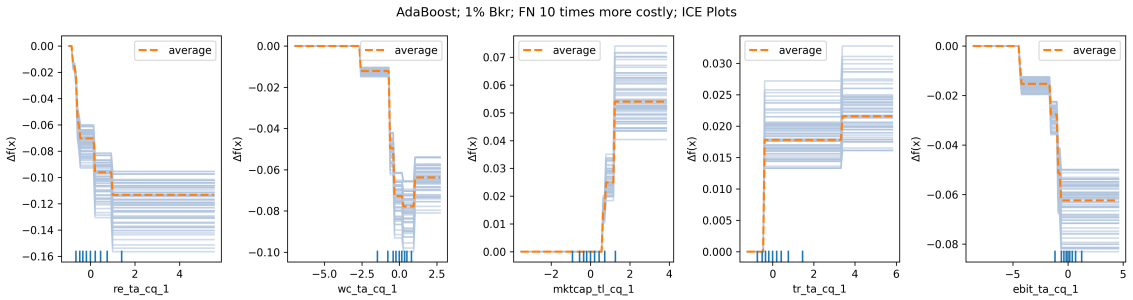


Figure 5.11: ICE Plots; 1% Bankruptcy Level; Boosting; FN Penalty 10

The plots show a sharpening of feature effects compared to the 5x cost setting. *RE/TA* exhibits a steep, monotonic decline in prediction risk as values increase from negative to positive territory, with a saturation point beyond which additional retained earnings no longer reduce risk. *WC/TA* follows a similar threshold pattern near zero.

The marginal effect of *MKTCAP/TL* remains positive, with a sharp upward shift at low values.

TR/TA and *EBIT/TA* show consistent effects with the prior configuration. *TR/TA* only slightly increases predicted risk, and *EBIT/TA* shows a clear downward trend.

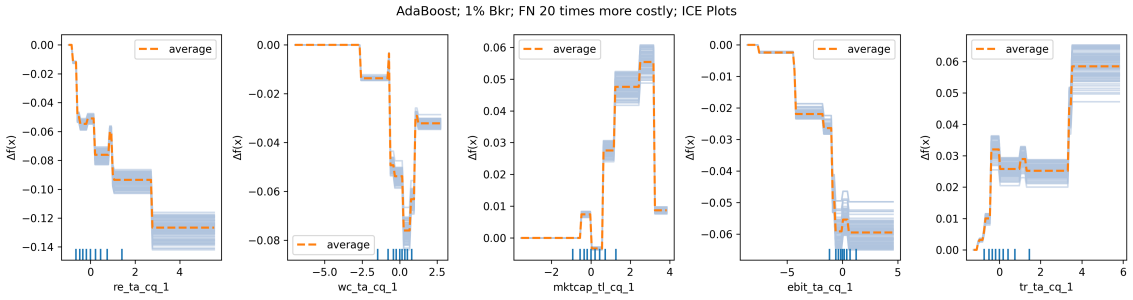


Figure 5.12: ICE Plots; 1% Bankruptcy Level; Boosting; FN Penalty 20

FN Cost 20 Figure 5.12 displays the individual conditional expectation (ICE) plots for AdaBoost under the 1% bankruptcy setting with a false negative cost twenty times greater than a false positive.

RE/TA produces a sharply decreasing marginal effect on bankruptcy risk, with a pronounced decline for values below zero and saturation at around 2. This indicates that even

moderate levels of retained earnings provide strong protective signals under conservative risk preferences.

WC/TA shows a steep drop in predicted risk around the zero threshold. The individual curves are tightly clustered, indicating consistent behavior across firms.

MKTCAP/TL and *TR/TA* continue to show positive effects, suggesting that higher values in these ratios correspond to higher predicted bankruptcy probabilities. Notably, the effect of *TR/TA* intensifies compared to earlier configurations, with a clear nonlinear jump around the value of 3.

EBIT/TA retains a negative marginal effect, but the slope is less steep than *RE/TA*, and individual effects remain more dispersed.

Non-Boosting

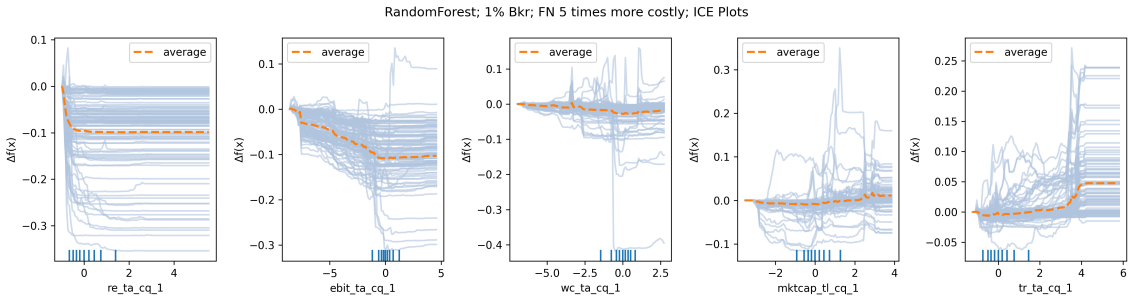


Figure 5.13: ICE Plots; 1% Bankruptcy Level; Non-Boosting; FN Penalty 5

FN Cost 5 Figure 5.13 presents ICE plots for the Random Forest model under the 1% bankruptcy scenario with a false negative cost five times higher than a false positive.

The features *RE/TA* and *EBIT/TA* exhibit strong and consistent negative marginal effects: as retained earnings and profitability increase, predicted bankruptcy risk declines. In particular, *RE/TA* shows a steep decline followed by a plateau. *EBIT/TA* demonstrates a similarly negative slope with more heterogeneity among individual effects.

WC/TA displays a modest negative effect centered around zero. While the average trend is slightly downward-sloping, the individual ICE curves suggest greater dispersion,

reflecting that working capital may interact with other features in Random Forest's ensemble structure.

In contrast, *MKTCAP/TL* and *TR/TA* exhibit positive marginal effects. Higher values of *TR/TA*, in particular, are associated with increased predicted risk, especially beyond a threshold of 3.

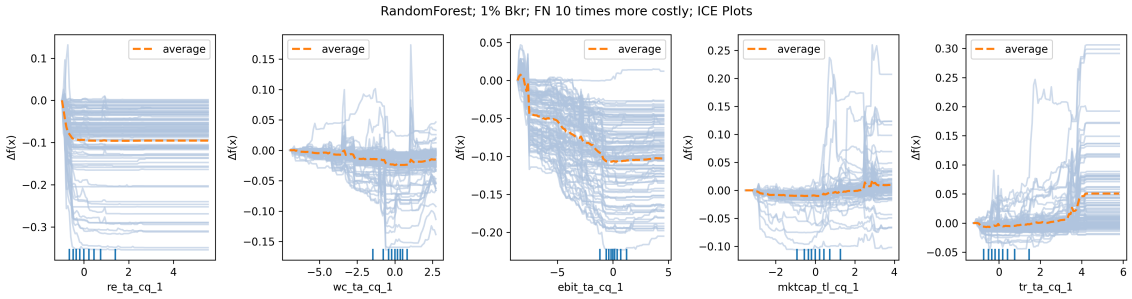


Figure 5.14: ICE Plots; 1% Bankruptcy Level; Non-Boosting; FN Penalty 10

FN Cost 10 Figure 5.14 illustrates the effect of increasing the false negative penalty on Random Forest's feature behavior. The response to *RE/TA* becomes even more decisive, with a sharp and monotonic decrease in predicted bankruptcy risk as retained earnings increase. The average trend flattens beyond zero, indicating that even low but positive earnings are protective against bankruptcy under high-cost settings.

The marginal effects of *EBIT/TA* remain negative but show increased variability across observations. While the average curve continues to decline, some individual instances exhibit nonlinear or even flat effects.

For *MKTCAP/TL*, the model response remains mild and noisy. The average curve increases only slightly at higher values, with substantial dispersion across individual ICE lines, suggesting this feature is still viewed as less predictive in this setting.

TR/TA shows a clearer pattern of increasing predicted risk at higher values. The effect becomes more convex above a ratio of 3, indicating that very high turnover relative to assets is flagged more aggressively as a risk indicator.

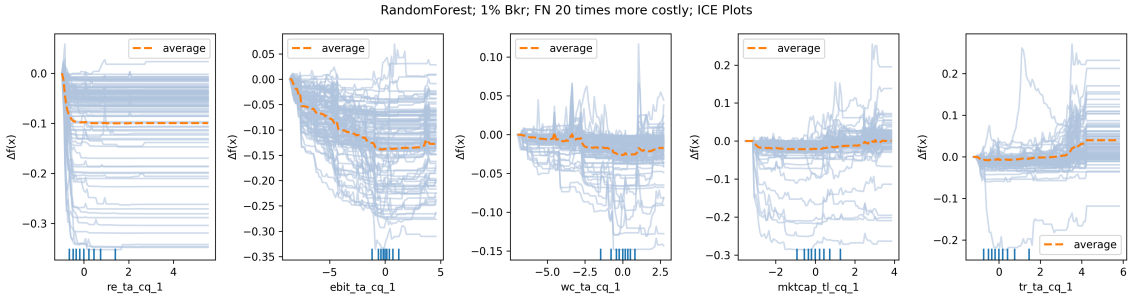


Figure 5.15: ICE Plots; 1% Bankruptcy Level; Non-Boosting; FN Penalty 20

FN Cost 20 The marginal effect of *RE/TA* remains strongly negative and stabilizes around -0.1 for most of the observed range. The steep decline in predictions for negative values is consistent with the model's increased aversion to missed bankruptcies.

The effect of *WC/TA* remains negative on average, but its influence appears to plateau. While individual ICE curves remain volatile, the average marginal effect flattens compared to prior settings, suggesting that working capital is treated as a secondary discriminator.

Both *MKTCAP/TL* and *TR/TA* show minimal average impact. While some outlier trajectories exist, the bulk of individual predictions are centered around zero.

5.3.2 5% Bankruptcy Level

Boosting

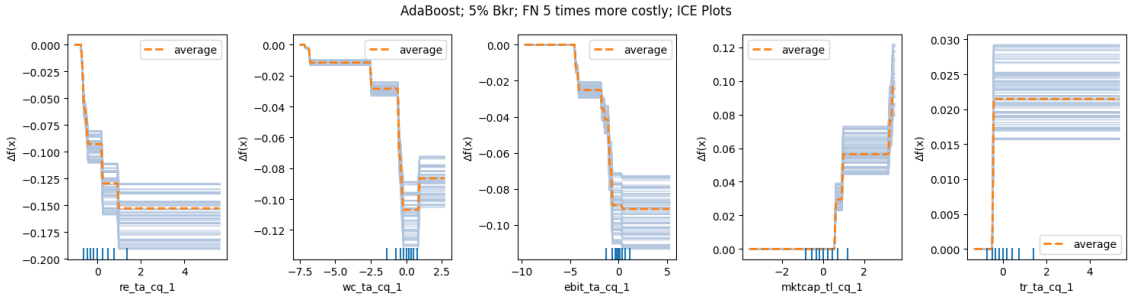


Figure 5.16: ICE Plots; 5% Bankruptcy Level; Boosting; FN Penalty 5

FN Cost 5 The marginal effect of *RE/TA* is steep and consistently negative, with a drop near zero, similar to the 1% bankruptcy level.

WC/TA exhibits a similar step-like pattern. The average prediction drops steeply as working capital becomes negative, then plateaus.

EBIT/TA also shows a strongly negative relationship. The average effect steadily declines across the observed range.

In contrast, *MKTCAP/TL* presents a sharp increase in predictions beyond the zero mark. This inversion is counterintuitive and suggests interaction effects or outlier sensitivity in how AdaBoost processes market-based features.

TR/TA remains mostly flat, with a positive shift. The average response increases at higher values, but the effect is small, implying this feature contributes marginally under this setting.

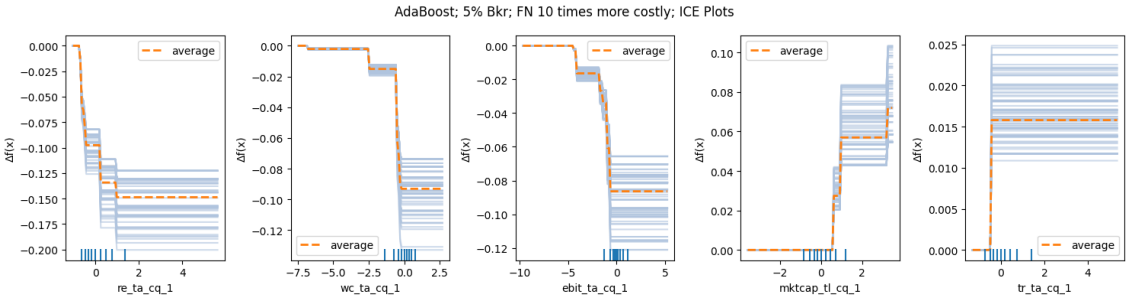


Figure 5.17: ICE Plots; 5% Bankruptcy Level; Boosting; FN Penalty 10

FN Cost 10 *RE/TA* exhibits a pronounced drop in prediction values as the ratio approaches zero. Notably, the ICE curves flatten quickly beyond this cutoff, indicating diminishing marginal effects once the retained earnings reach positive territory.

WC/TA follows a similar binary-like behavior. Once working capital surpasses zero, its contribution to the model prediction remains mostly stable. Below this threshold, however, the average bankruptcy prediction drops steeply.

The treatment of *MKTCAP/TL* remains counterintuitive. The model still increases predictions for higher values of this market-based ratio.

TR/TA continues to exert only a mild positive influence, with low variability across individuals and a nearly flat average response. Its marginal role is consistent with its lower importance ranking in permutation-based analyses.

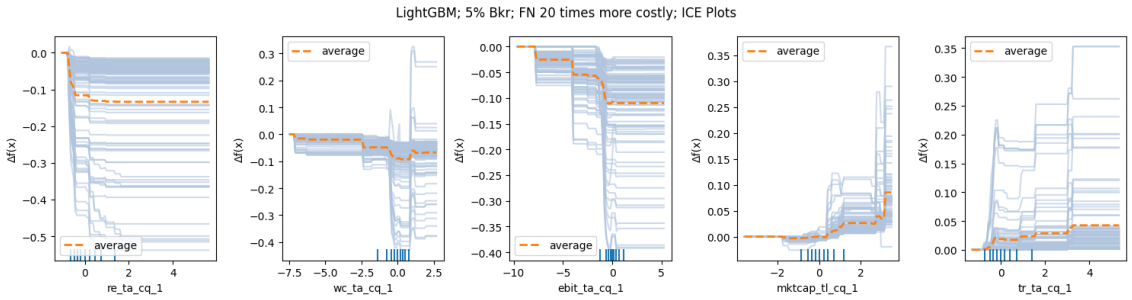


Figure 5.18: ICE Plots; 5% Bankruptcy Level; Boosting; FN Penalty 20

FN Cost 20 Figure 5.18 presents the ICE plots for LightGBM under the most conservative cost configuration at the 5% bankruptcy level.

The effect of *RE/TA* is particularly strong and stable: across the majority of observations, the model assigns significantly lower risk scores as retained earnings increase. The ICE curves are tightly aligned below the zero threshold, highlighting LightGBM's discrimination between negative and positive profitability history.

WC/TA displays similarly concentrated effects. The ICE curves show a steep decline near zero, after which the effect plateaus, indicating that the model strongly penalizes negative working capital levels.

For *EBIT/TA*, the curves reflect a sharp nonlinear drop in predicted risk for values increasing beyond zero.

The model's response to *MKTCAP/TL* and *TR/TA* remains positive but more subdued. Both features exhibit greater variability and weaker average slopes, suggesting that while they contribute to the prediction, their influence is secondary.

Non-Boosting

FN Cost 5 The most prominent marginal effects appear for *RE/TA* and *EBIT/TA*. Both features display strong negative influence on the predicted bankruptcy risk, with the av-

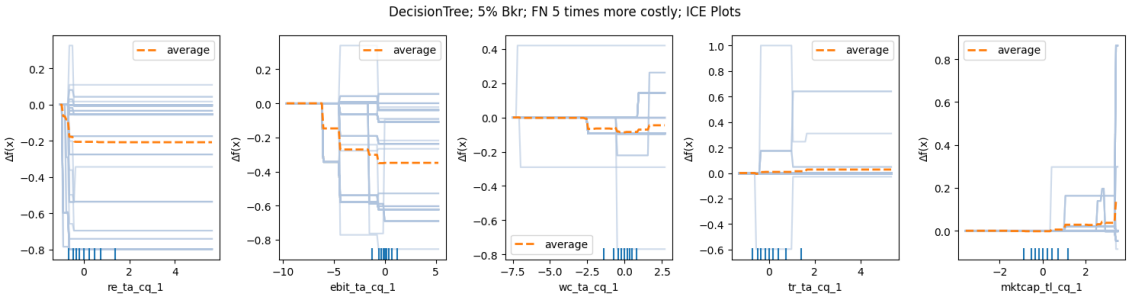


Figure 5.19: ICE Plots; 5% Bankruptcy Level; Non-Boosting; FN Penalty 5

erage ICE curve dropping sharply as values increase above zero. However, the individual curves vary considerably in magnitude, with several samples showing abrupt shifts. Such discontinuities are typical in tree-based models, where predictions change abruptly at decision boundaries rather than evolving gradually.

WC/TA shows a relatively flat response for most of the feature range, with slight drops in predicted risk near the zero threshold. Nonetheless, the uneven trajectory of the ICE curves and the absence of a consistent trend across observations reduce interpretability.

TR/TA and *MKTCAP/TL* demonstrate the weakest and most erratic marginal effects. Many ICE lines remain constant or fluctuate without clear patterns, indicating that these features are either underutilized in the tree structure or only activated in a narrow subset of the decision paths.

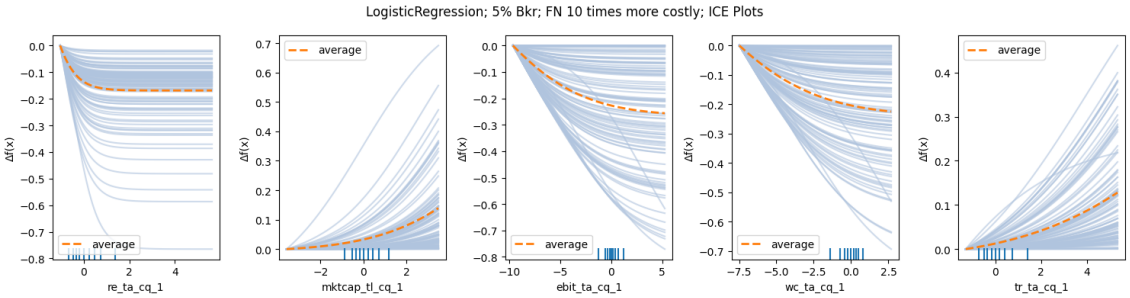


Figure 5.20: ICE Plots; 5% Bankruptcy Level; Non-Boosting; FN Penalty 10

FN Cost 10 Figure 5.20 displays the ICE plots for Logistic Regression under a conservative cost setting. As expected from a linear model, the marginal effects are smooth and monotonic across all features, with consistent directionality across individual samples.

RE/TA, *EBIT/TA*, and *WC/TA* exhibit strong, negative slopes, indicating that higher values of these profitability and liquidity ratios are associated with a pronounced decrease in predicted bankruptcy risk. The parallel structure of the ICE curves across these features reinforces the model's global linear behavior.

MKTCAP/TL and *TR/TA* show positive marginal effects in this setting as well. Increases in these ratios correspond to higher predicted financial health.

The ICE curves reveal minimal individual variation, and the average response closely tracks the individual trends. This uniformity enhances transparency but may also limit the model's ability to capture nuanced interactions or nonlinear thresholds observed in more complex classifiers.

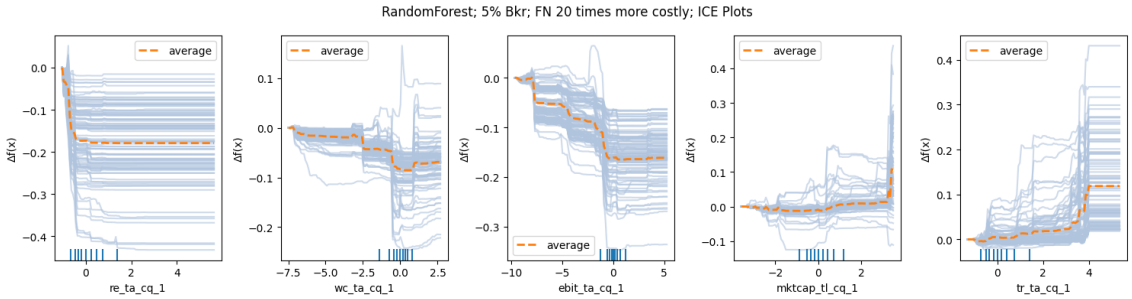


Figure 5.21: ICE Plots; 5% Bankruptcy Level; Non-Boosting; FN Penalty 20

FN Cost 20 *RE/TA*, *EBIT/TA*, and *WC/TA* exhibit a consistent downward trend, with increasing values associated with a lower predicted bankruptcy probability.

In contrast, *MKTCAP/TL* and *TR/TA* show positive associations with financial health, particularly in the upper range of their observed values. The patterns are similar to Random Forest's behaviour at 1% bankruptcy level.

5.3.3 10% Bankruptcy Level

Boosting

FN Cost 5 Figure 5.22 displays the ICE plots for LightGBM under a 5x FN penalty.

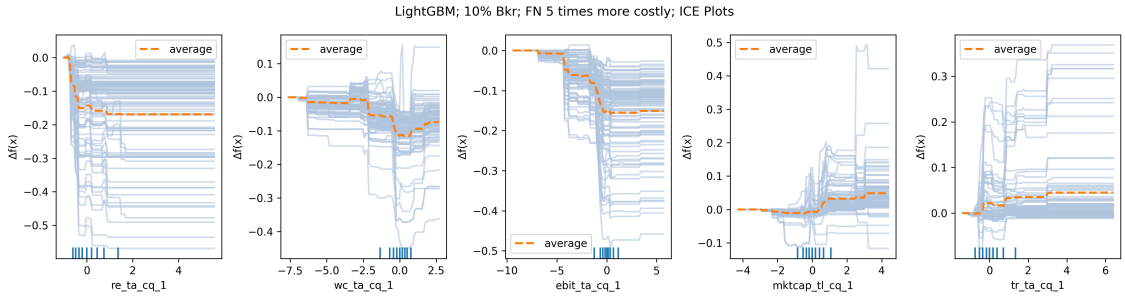


Figure 5.22: ICE Plots; 10% Bankruptcy Level; Boosting; FN Penalty 5

The model captures a consistent negative relationship between profitability and bankruptcy risk. Both *RE/TA* and *EBIT/TA* exhibit strong monotonic effects, with risk declining sharply as retained earnings and operating profits increase—stabilizing near zero, consistent with their roles as financial buffers. *WC/TA* follows a similar but weaker trend, with greater variability across firms suggesting conditional effects based on broader financial context.

In contrast, the leverage-related ratios, *MKTCAP/TL* and *TR/TA*, show nonlinear and positive marginal effects. Their influence becomes pronounced only after surpassing key thresholds (approximately 1).

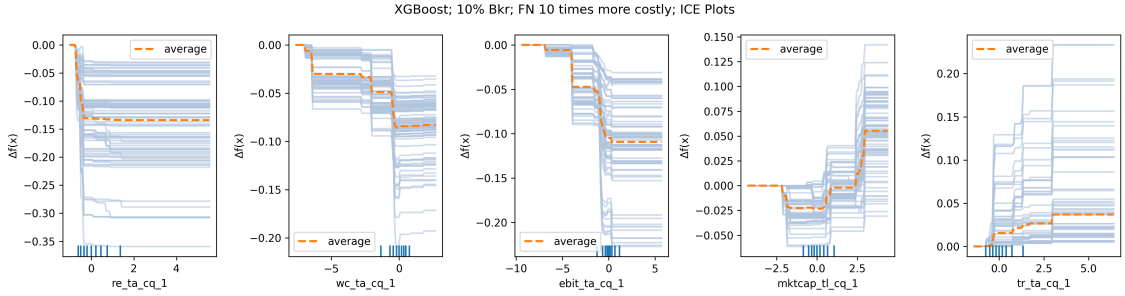


Figure 5.23: ICE Plots; 10% Bankruptcy Level; Boosting; FN Penalty 10

FN Cost 10 The strongest marginal effect is observed for *RE/TA*, where even modest improvements from negative to slightly positive values sharply reduce predicted bankruptcy risk, with diminishing gains thereafter. Similar threshold behavior appears for *WC/TA* and *EBIT/TA*, where risk drops markedly around zero, indicating the model treats a return to positive liquidity or earnings as a key turning point.

By contrast, *MKTCAP/TL* and *TR/TA* exhibit more gradual nonlinear increases in the prediction. The greater variation in *TR/TA* effects across observations suggests interaction effects that the model captures through tree-based splits.

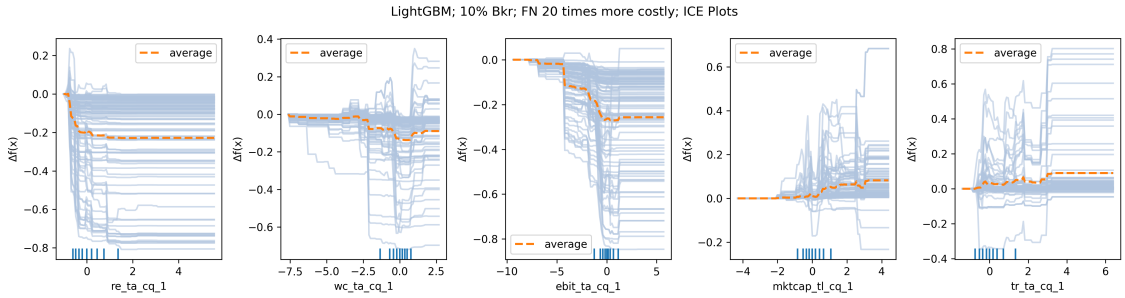


Figure 5.24: ICE Plots; 10% Bankruptcy Level; Boosting; FN Penalty 20

FN Cost 20 *RE/TA* shows the most pronounced effect, with both individual and average curves indicating a sharp increase in predicted risk as the ratio fall below zero. Similar threshold patterns are seen for *WC/TA* and *EBIT/TA*, particularly around zero, where the model sharply penalizes negative values.

In contrast, *MKTCAP/TL* and *TR/TA* display more moderate, upward-sloping effects with greater heterogeneity across observations. Their contribution to reducing bankruptcy risk appears less decisive and more conditional.

Non-Boosting

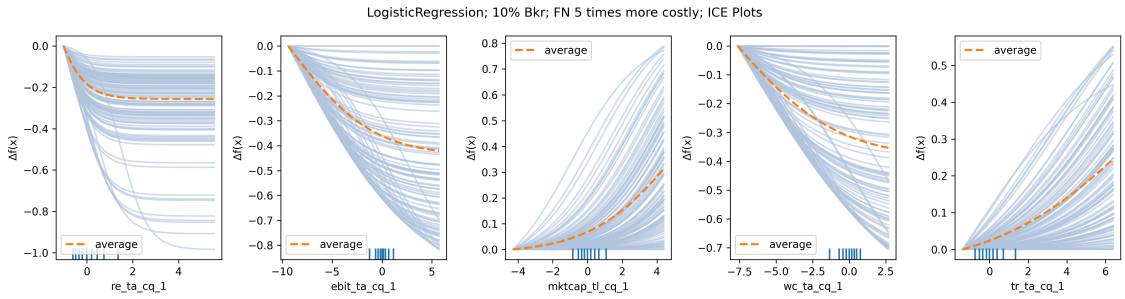


Figure 5.25: ICE Plots; 10% Bankruptcy Level; Non-Boosting; FN Penalty 5

FN Cost 5 *RE/TA* and *EBIT/TA* both show strong negative slopes, indicating that higher values in these ratios reduce the predicted bankruptcy risk. These effects are steep for values near zero.

MKTCAP/TL and *TR/TA* both demonstrate positive marginal effects, with nearly linear increases in model output as the values rise.

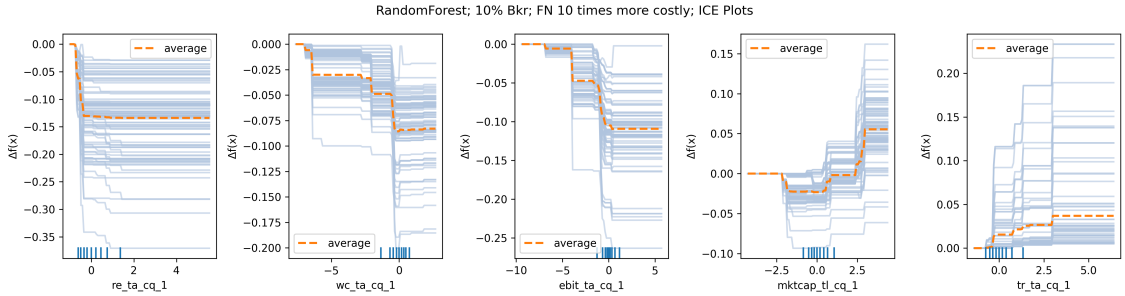


Figure 5.26: ICE Plots; 10% Bankruptcy Level; Non-Boosting; FN Penalty 10

FN Cost 10 For *RE/TA*, the ICE curves exhibit a sharply negative slope for low values, with saturation occurring near zero. A similar pattern is observed for *EBIT/TA*, where the marginal effect flattens past moderate profitability levels, indicating diminishing returns on further gains.

MKTCAP/TL and *TR/TA* both show upward trends, but with scattered and abrupt changes.

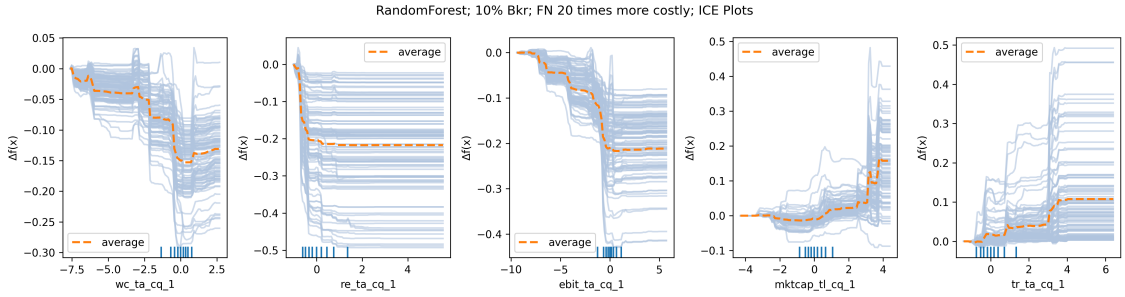


Figure 5.27: ICE Plots; 10% Bankruptcy Level; Non-Boosting; FN Penalty 20

FN Cost 20 *RE/TA* and *EBIT/TA* show sharp risk reductions as values cross above zero, with effects tapering off at higher levels. *WC/TA* follows a similar threshold pattern but with more variability, indicating firm-specific responses to liquidity shortfalls.

MKTCAP/TL and *TR/TA* contribute positively in a stepwise manner, though with greater dispersion.

Chapter 6

Discussion

6.1 Model Performance Across Cost Settings and Prevalence Levels

The experimental results reveal several consistent patterns across varying bankruptcy prevalence levels and cost configurations. Notably, boosting-based ensemble methods, particularly LightGBM and AdaBoost, demonstrate superior adaptability under different stakeholder-driven cost asymmetries.

Among non-boosting models, Random Forest emerges as the most competitive baseline, occasionally surpassing boosting models at higher prevalence levels and extreme FN cost settings. Its relatively strong recall, combined with moderate precision, makes it a viable interpretable alternative in operational contexts demanding transparency. Logistic Regression and Decision Tree models, by contrast, demonstrate inconsistent performance and a limited ability to adapt to increasing cost asymmetry, often sacrificing either recall or precision without achieving corresponding cost benefits and often overfitting.

A noteworthy outcome is the models' ability to generalize effectively to the test set spanning CQ1 2020 to CQ4 2023. This period was marked by heightened uncertainty following the COVID-19 pandemic. Despite being trained exclusively on data up to 2019, the models performed surprisingly well. This might be because of our labeling strategy

that forced the model to learn early warning patterns better.

6.2 Temporal Input Importance

Across all models, the integration of a 20-quarter input window yields substantial performance gains over a single-quarter snapshot. The cost reduction observed between 20-CQ and 1-CQ inputs is particularly pronounced under 10x and 20x FN cost scenarios, highlighting the critical importance of temporal dynamics in financial distress prediction. Models utilizing longer historical sequences not only improve recall but also stabilize precision. This finding validates the hypothesis that bankruptcy is not a point-in-time event but the culmination of long-term deterioration, best captured through extended financial histories.

6.3 Feature Importance and Interpretability Insights

The XAI techniques used in this study provide converging evidence on the relative influence of financial ratios and the model-specific nature of their effects.

Across both PFI and SHAP analyses, *RE/TA* consistently emerges as the most influential predictor. It ranks highest in nearly all configurations, reaffirming the centrality of retained earnings in capturing accumulated financial stress. *WC/TA* also plays a key role, often ranking second. In contrast, *TR/TA* is systematically the least important ratio across all models, prevalence levels, and FN cost scenarios. *MKTCAP/TL* displays modest and inconsistent importance, indicating limited predictive value in comparison to accounting-based features.

A clear pattern emerges with increasing FN cost: the absolute magnitude of both PFI and mean SHAP values rises. This trend reflects the model's growing reliance on high-signal features such as *RE/TA* and *WC/TA* under conservative risk preferences. As the cost of false negatives increases, models increasingly prioritize these core ratios, reducing reliance on less discriminative features like *TR/TA* and *MKTCAP/TL*.

ICE plots offer further insight into how models respond to marginal changes in input features. Across all settings and model types, *RE/TA* exhibits a steep decline in predicted bankruptcy probability as values cross from negative to positive. *EBIT/TA* and *WC/TA* show similar behavior, with threshold-like drops and eventual flattening, indicating diminishing marginal gains.

In contrast, *MKTCAP/TL* and *TR/TA* display positive marginal effects, where higher values are sometimes associated with increased predicted bankruptcy risk. This counter-intuitive pattern, especially for *MKTCAP/TL*, may reflect outlier sensitivity or interaction effects not captured in global rankings.

The structure of ICE plots also reveals model-specific interpretability dynamics. In decision trees, the ICE lines are highly separated, consistent with abrupt decision boundaries. LightGBM and Random Forest models exhibit more clustered ICE curves, indicating smoother and more stable feature effects across instances. Logistic Regression produce monotonic and near-parallel ICE lines, reflecting continuous and linear treatment of features.

Notably, these interpretability patterns remain stable across bankruptcy prevalence levels and FN cost configurations. This consistency reinforces the robustness of profitability and liquidity ratios as drivers of predictive confidence and strengthens the case for their inclusion in early-warning frameworks.

6.4 Link to Research Objectives

This study set out to build a machine learning framework for bankruptcy prediction that balances predictive accuracy, stakeholder-relevant cost sensitivity, and interpretability. The key research objectives outlined at the start of this thesis were fully addressed through the empirical design and analysis:

- Model Comparison: The thesis compared boosting models with non-boosting alternatives across a unified pipeline. The results confirmed that boosting models, par-

ticularly LightGBM, offer superior performance, while Random Forest emerges as a competitive interpretable baseline. Logistic Regression and Decision Tree models were shown to be less performant under extreme class imbalance and cost asymmetries.

- Economic Cycle Simulation: Bankruptcy prevalence levels were varied across 1%, 5%, and 10% to simulate optimistic, neutral, and pessimistic macroeconomic scenarios. Model behavior was evaluated under each setting, revealing stable feature prioritization (especially for *RE/TA* and *WC/TA*) and consistent performance trends, thus confirming the framework's robustness across economic conditions.

- Cost-sensitive Learning: Stakeholder-specific cost matrices were incorporated with FN:FP ratios of 5:1, 10:1, and 20:1. The experiments demonstrated that increasing FN cost sharpened model focus on the most predictive features and shifted decision thresholds to minimize high-cost errors. The framework successfully produced decision rules that are economically meaningful for regulators, lenders, and investors.

- Explainability and Interpretation: A multi-level interpretability layer was implemented using PFI, SHAP values, and ICE plots. *RE/TA* consistently emerged as the most influential feature across all models and evaluation methods. ICE plots revealed model-specific behavioral patterns, including smooth monotonic responses in linear models, clustered responses in ensemble methods, and sharp discontinuities in decision trees. These results enhance model transparency and provide actionable insights into the financial drivers of distress.

Together, these outcomes fulfill the thesis objectives and establish a replicable, interpretable, and cost-aware methodology for corporate bankruptcy prediction. The pipeline not only supports rigorous model benchmarking but also aligns predictive outputs with real-world decision-making needs.

6.5 Stakeholder-Centered Interpretation

A central objective of this thesis was to align model outputs with the specific needs and risk tolerances of key stakeholders involved in bankruptcy prediction. In practice, different actors such as regulators, lenders, auditors, and investors face varying costs for false negatives and false positives, and their interpretability requirements differ accordingly.

The integration of cost-sensitive learning addressed these differences by allowing the models to prioritize recall or precision depending on stakeholder preferences. For instance, under conservative cost settings that penalize false negatives more heavily, the models emphasized recall, reducing the likelihood of missing high-risk firms. This is particularly relevant for regulators and creditors, who bear substantial losses when bankruptcies go undetected. In contrast, under moderate cost configurations, where the cost of false positives is also a concern, models' predictions showed a more balanced trade-off—suitable for investors seeking to avoid overreacting to noise while still flagging potential risks.

Interpretability tools such as SHAP, permutation importance, and ICE plots enhanced the practical usability of the models in stakeholder contexts. These techniques provided not only rankings of influential features but also case-specific insights into prediction drivers. For example, a regulator could use the model to identify firms flagged as high-risk due to negative retained earnings and liquidity shortfalls, while a lender could inspect the ICE trajectory of a specific firm's working capital to understand whether its improvement trajectory is sufficient to downgrade risk. The clustered and monotonic patterns observed in ensemble and linear models, respectively, support decision-making processes where consistency and transparency are paramount.

By making the model's logic more transparent and its risk signals more aligned with stakeholder priorities, this work supports the deployment of machine learning systems in real-world financial settings. Ultimately, stakeholder-centered interpretation bridges the gap between predictive analytics and responsible action, ensuring that AI-driven decisions are not only accurate, but also intelligible, actionable, and justifiable.

6.6 Limitations

First, the study is restricted to publicly listed U.S. firms. As a result, the models may not generalize to private companies or firms in other jurisdictions, where financial reporting standards, data availability, and economic conditions differ significantly.

Second, the feature space is intentionally constrained to five core accounting ratios to ensure interpretability and alignment with prior literature. While this design improves transparency, it may exclude complementary signals such as qualitative disclosures, macroeconomic indicators, or market sentiment that could enhance predictive power in more complex settings.

Third, while the highest-performing models were trained on 20-quarter sequences, the interpretability analysis was conducted on models using only the most recent quarter as input. This decision was made to avoid the effects of multicollinearity, as accounting ratios across consecutive quarters were found to be highly correlated. While this simplification enabled clearer attribution and reduced interpretability noise, it introduces a disconnect: the explanations reflect a simplified version of the input structure used in the best-performing models. Future research could explore time-aware interpretability techniques capable of handling multivariate temporal inputs directly.

Fourth, the use of fixed bankruptcy prevalence levels (1%, 5%, and 10%) to simulate different economic regimes constitutes an experimental approximation rather than a reflection of observed empirical distributions. These prevalence levels were selected to approximate plausible conditions under expansionary, neutral, and recessionary periods, respectively. While this design enables cost-sensitive evaluation under controlled class imbalance, it does not capture the endogenous dynamics of economic cycles. As such, the findings should be interpreted within the context of these constructed scenarios rather than as representations of firms operating under actual economic conditions corresponding to the assumed bankruptcy prevalence.

Finally, while the interpretability methods employed—SHAP, permutation importance, and ICE plots—provide valuable insights, they are primarily post hoc and do not offer

formal guarantees about causality or fairness. Additional tools such as counterfactual explanations or fairness audits would strengthen the interpretability and ethical reliability of the framework.

Addressing these limitations presents opportunities for future research to extend the model's generalizability, temporal sophistication, and feature richness, while maintaining the core commitment to stakeholder relevance and transparency.

6.7 Future Work

Several avenues remain open for extending the contributions of this thesis. First, future work should explore time-aware interpretability methods that can handle multivariate sequential inputs directly to better align interpretability with the high-performing 20-quarter models. Dimensionality reduction techniques could also be explored to decorrelate highly collinear historical features prior to explainability analysis. Expanding the feature space beyond core accounting ratios to include macroeconomic indicators, ESG metrics, qualitative disclosures, or alternative data sources like sentiment and analyst forecasts could also enhance model performance and early-warning capabilities. Applying the framework to private firms, emerging markets, or different regulatory environments would help assess its generalizability across contexts with varying data transparency. Future research may also incorporate causal and counterfactual explanation techniques to enhance the actionability and fairness of model outputs.

Conclusion

This thesis addressed the challenge of designing machine learning models for corporate bankruptcy prediction that are not only accurate, but also interpretable and aligned with the cost sensitivities and transparency needs of real-world stakeholders. The study tackled three key constraints in this domain: severe class imbalance, macroeconomic variability, and the practical need for model interpretability in high-stakes financial decision-making.

In response, a structured, cost-sensitive modeling pipeline was developed and evaluated across a range of bankruptcy prevalence levels (1%, 5%, and 10%) and false negative cost ratios (5x, 10x, and 20x). Both boosting and non-boosting model families were systematically compared under consistent preprocessing, hyperparameter tuning, and evaluation protocols.

The empirical results demonstrate that boosting algorithms consistently achieve the lowest misclassification costs and highest recall rates, especially under conservative cost scenarios where false negatives carry severe consequences. Random Forest, among non-boosting models, emerged as a competitive alternative, offering a balance between performance and interpretability. Notably, all models generalized well to the out-of-sample test period spanning 2020 to 2023, despite being trained only on pre-2020 data. This robustness suggests that the models successfully internalized early warning patterns that remained valid amid economic disruptions.

Beyond predictive performance, this research contributes a multi-level explainability framework that combines permutation-based feature importance, SHAP values, and individual conditional expectation (ICE) plots. These techniques revealed that *RE/TA* and

WC/TA are the most influential financial ratios across models, configurations, and cost settings. *EBIT/TA* also played a secondary but consistent role. In contrast, *TR/TA* and *MKTCAP/TL* were repeatedly found to have limited predictive value. ICE plots showed steep risk reductions near zero for *RE/TA*, *EBIT/TA* and *WC/TA*, with model-specific patterns in how individual cases were treated, ranging from sharp splits in decision trees to smooth, monotonic behavior in logistic regression.

A key contribution of this work lies in aligning technical modeling choices with the asymmetric risk tolerances of real-world stakeholders. By tuning thresholds based on stakeholder-specific cost matrices and incorporating interpretability at both global and local levels, the framework demonstrates that high-performing machine learning models can be both practical and accountable.

Ultimately, this thesis offers a rigorous and extensible methodology for bankruptcy prediction under uncertainty. It provides theoretical clarity on feature influence, empirical insight into model behavior under varying conditions, and practical tools for explainable, cost-aware deployment. By bridging predictive performance with interpretability and stakeholder alignment, this work contributes to the development of trustworthy AI in high-stakes financial contexts.

Bibliography

Altman, Edward I. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy". In: *The Journal of Finance* 23.4, pp. 589–609. DOI: 10.1111/j.1540-6261.1968.tb00843.x.

Altman, Edward I., Robert G. Haldeman, and P. Narayanan (1977). "ZETATM analysis A new model to identify bankruptcy risk of corporations". In: *Journal of Banking & Finance* 1.1, pp. 29–54. ISSN: 0378-4266. DOI: https://doi.org/10.1016/0378-4266(77)90017-6. URL: https://www.sciencedirect.com/science/article/pii/0378426677900176.

Balcaen, Sofie and Hubert Ooghe (2006). "35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems". In: *The British Accounting Review* 38.1, pp. 63–93. DOI: 10.1016/j.bar.2005.09.001.

Barboza, Flavio, Herbert Kimura, and Edward Altman (Oct. 2017). "Machine learning models and bankruptcy prediction". In: *Expert Systems with Applications* 83, pp. 405–417. DOI: 10.1016/j.eswa.2017.04.006.

Beaver, William H. (1966). "Financial Ratios As Predictors of Failure". In: *Journal of Accounting Research* 4, pp. 71–111. ISSN: 00218456, 1475679X. URL: http://www.jstor.org/stable/2490171 (visited on 04/25/2025).

Breiman, Leo (2001). "Random forests". In: *Machine learning* 45, pp. 5–32.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi (2008). "In Search of Distress Risk". In: *The Journal of Finance* 63.6, pp. 2899–2939. DOI: 10.1111/j.1540-6261.2008.01416.x.

Chawla, Nitesh V et al. (2002). "SMOTE: synthetic minority over-sampling technique". In: *Journal of artificial intelligence research* 16, pp. 321–357.

Chen, Tianqi and Carlos Guestrin (Aug. 2016). "XGBoost: A Scalable Tree Boosting System". In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. ACM, pp. 785–794. DOI: 10.1145/2939672.2939785. URL: http://dx.doi.org/10.1145/2939672.2939785.

Doshi-Velez, Finale and Been Kim (2017). "Towards a rigorous science of interpretable machine learning". In: *arXiv preprint arXiv:1702.08608*.

Elkan, Charles (2001). "The Foundations of Cost-Sensitive Learning". In: *Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 973–978. URL: https://dl.acm.org/doi/10.5555/1642194.1642224.

Fisher, Aaron, Cynthia Rudin, and Francesca Dominici (2019). *All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously*. arXiv: 1801.01489 [stat.ME]. URL: https://arxiv.org/abs/1801.01489.

Freund, Yoav and Robert E Schapire (1997). "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1, pp. 119–139. ISSN: 0022-0000. DOI: https://doi.org/10.1006/jcss.1997.1504. URL: https://www.sciencedirect.com/science/article/pii/S002200009791504X.

Ke, Guolin et al. (2017). "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". In: *Advances in Neural Information Processing Systems*. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc. URL: https://proceedings.neurips.cc/paper%5C_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf.

Lin, Tsung-Yi et al. (2018). *Focal Loss for Dense Object Detection*. arXiv: 1708.02002 [cs.CV]. URL: https://arxiv.org/abs/1708.02002.

Lundberg, Scott and Su-In Lee (2017). *A Unified Approach to Interpreting Model Predictions*. arXiv: 1705.07874 [cs.AI]. URL: https://arxiv.org/abs/1705.07874.

Ohlson, James A. (1980). "Financial Ratios and the Probabilistic Prediction of Bankruptcy". In: *Journal of Accounting Research* 18.1, pp. 109–131. DOI: 10.2307/2490395.

Paszke, Adam et al. (2019). "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc.

Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Shumway, Tyler (2001). "Forecasting Bankruptcy More Accurately: A Simple Hazard Model". In: *The Journal of Business* 74.1, pp. 101–124. DOI: 10.1086/209665.

Son, H. et al. (Dec. 2019). "Data Analytic Approach for bankruptcy prediction". In: *Expert Systems with Applications* 138, p. 112816. DOI: 10.1016/j.eswa.2019.07.033.

Tran, Kim Long et al. (2022). "Explainable Machine Learning for Financial Distress Prediction: Evidence from Vietnam". In: *Data* 7.11. ISSN: 2306-5729. DOI: 10.3390/data7110160. URL: https://www.mdpi.com/2306-5729/7/11/160.

Weber, Patrick, K. Valerie Carl, and Oliver Hinz (Feb. 2023). "Applications of explainable artificial intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science Literature". In: *Management Review Quarterly* 74.2, pp. 867–907. DOI: 10.1007/s11301-023-00320-0.

Yeo, Wei Jie et al. (Mar. 2025). "A comprehensive review on financial explainable ai". In: *Artificial Intelligence Review* 58.6. DOI: 10.1007/s10462-024-11077-7.

Yotsawat, Wirot et al. (Aug. 2023). "Bankruptcy prediction model using cost-sensitive extreme gradient boosting in the context of imbalanced datasets". In: *International Journal of Electrical and Computer Engineering (IJECE)* 13.4, p. 4683. DOI: 10.11591/ijece.v13i4.pp4683-4691.

Zhang, Zijiao et al. (2022). "An explainable artificial intelligence approach for financial distress prediction". In: *Information Processing & Management* 59.4, p. 102988. ISSN: 0306-4573. DOI: https://doi.org/10.1016/j.ipm.2022.102988. URL: https://www.sciencedirect.com/science/article/pii/S0306457322001030.

Zięba, Maciej, Sebastian K. Tomczak, and Jakub M. Tomczak (Oct. 2016). "Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction". In: *Expert Systems with Applications* 58, pp. 93–101. DOI: 10.1016/j.eswa.2016.04.001.

Zmijewski, Mark E. (1984). "Methodological Issues Related to the Estimation of Financial Distress Prediction Models". In: *Journal of Accounting Research* 22, pp. 59–82. ISSN: 00218456, 1475679X. URL: http://www.jstor.org/stable/2490859 (visited on 04/24/2025).

Zou, Yao, Changchun Gao, and Han Gao (2022). "Business Failure Prediction Based on a Cost-Sensitive Extreme Gradient Boosting Machine". In: *IEEE Access* 10, pp. 42623–42639. DOI: 10.1109/ACCESS.2022.3168857.

Appendix A – Features Summary Statistics

Table 1: Summary Statistics of EBIT/TA Across All Calendar Quarters

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ42023	-0.0297	-0.0121	0.0105	0.0234	-23.4599	2.5876	-36.89
CQ32023	-0.0358	-0.0165	0.0102	0.0238	-38.4190	0.5463	-52.88
CQ22023	-0.0457	-0.0192	0.0103	0.0240	-77.6337	0.2781	-61.10
CQ12023	-0.0913	-0.0235	0.0080	0.0218	-275.6200	0.8268	-65.22
CQ42022	-0.0346	-0.0225	0.0095	0.0239	-34.2940	1.0375	-47.96
CQ32022	-0.0307	-0.0258	0.0099	0.0246	-18.9820	1.0250	-33.63
CQ22022	-0.0266	-0.0259	0.0094	0.0248	-20.8920	0.4567	-45.27
CQ12022	-0.0535	-0.0258	0.0080	0.0227	-58.2300	0.8800	-40.68
CQ42021	-0.0629	-0.0223	0.0098	0.0244	-103.5190	0.7113	-52.38
CQ32021	-0.0312	-0.0218	0.0095	0.0240	-34.5287	0.4292	-43.34
CQ22021	-0.0325	-0.0180	0.0099	0.0249	-21.7644	0.4626	-30.66
CQ12021	-0.0661	-0.0195	0.0089	0.0224	-110.0000	0.7781	-52.78
CQ42020	-0.0944	-0.0139	0.0097	0.0234	-164.0000	0.4741	-51.02
CQ32020	-0.0645	-0.0135	0.0093	0.0229	-144.0980	0.4065	-61.17
CQ22020	-0.0313	-0.0240	0.0049	0.0174	-21.0080	0.4460	-38.48
CQ12020	-0.0680	-0.0197	0.0063	0.0176	-57.4071	0.6052	-39.12
CQ42019	-0.5638	-0.0075	0.0104	0.0218	-1612.1509	0.2826	-60.76

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ32019	-0.0638	-0.0052	0.0112	0.0226	-78.1860	0.5698	-39.86
CQ22019	-0.0376	-0.0057	0.0110	0.0227	-21.0079	0.4983	-29.63
CQ12019	-0.0469	-0.0096	0.0094	0.0203	-47.4246	1.5850	-42.99
CQ42018	-0.0792	-0.0041	0.0125	0.0258	-88.7970	0.6607	-37.94
CQ32018	-0.0854	-0.0034	0.0129	0.0257	-85.4054	2.5666	-33.72
CQ22018	-0.2117	-0.0033	0.0127	0.0261	-597.4470	0.6064	-62.13
CQ12018	-0.0741	-0.0055	0.0109	0.0233	-90.0206	0.4513	-41.15
CQ42017	-0.1092	-0.0032	0.0130	0.0253	-174.4682	27.3303	-41.23
CQ32017	-0.0376	-0.0047	0.0119	0.0247	-24.9020	1.0399	-28.31
CQ22017	-0.1291	-0.0030	0.0122	0.0249	-333.2910	0.5493	-62.53
CQ12017	-0.0580	-0.0061	0.0106	0.0225	-33.8933	0.4579	-26.36
CQ42016	-0.0857	-0.0051	0.0126	0.0253	-99.5554	0.8870	-40.81
CQ32016	-0.1174	-0.0027	0.0132	0.0259	-110.0354	1.2455	-35.95
CQ22016	-0.1575	-0.0047	0.0125	0.0259	-343.9520	0.5556	-57.77
CQ12016	-0.1282	-0.0089	0.0104	0.0227	-245.4810	0.5314	-52.32
CQ42015	-0.1063	-0.0044	0.0127	0.0264	-104.5909	2.1918	-32.84
CQ32015	-0.0574	-0.0035	0.0134	0.0265	-58.5030	0.9003	-38.25
CQ22015	-0.0344	-0.0045	0.0129	0.0259	-18.0448	0.3786	-25.19
CQ12015	-0.4045	-0.0050	0.0113	0.0237	-1685.9923	1.9484	-66.65
CQ42014	-1.4090	-0.0007	0.0150	0.0288	-5394.7088	0.6260	-65.56
CQ32014	-0.0583	0.0001	0.0157	0.0285	-95.2640	0.4753	-49.70
CQ22014	-0.0705	-0.0004	0.0145	0.0267	-156.4790	1.5992	-62.47
CQ12014	-0.0729	-0.0027	0.0122	0.0242	-138.8200	0.5564	-59.93
CQ42013	-0.0864	-0.0001	0.0148	0.0278	-162.3430	4.2462	-58.11
CQ32013	-0.0663	0.0008	0.0153	0.0277	-62.4605	0.9562	-33.68
CQ22013	-0.1206	0.0012	0.0152	0.0282	-286.2560	1.0273	-58.78
CQ12013	-0.1233	-0.0016	0.0126	0.0252	-216.9055	0.7153	-52.42

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ42012	-0.1477	0.0008	0.0150	0.0287	-285.9835	1.3875	-51.92
CQ32012	-0.1332	0.0017	0.0156	0.0285	-123.1170	0.5000	-32.10
CQ22012	-0.7988	0.0025	0.0159	0.0300	-3166.6312	0.4845	-65.80
CQ12012	-0.4513	0.0007	0.0142	0.0278	-1356.8995	0.7127	-59.44
CQ42011	-1.5160	0.0018	0.0158	0.0299	-4847.2593	9.8857	-61.94
CQ32011	-1.4175	0.0027	0.0172	0.0316	-5344.3400	0.8406	-65.02
CQ22011	-0.5084	0.0030	0.0166	0.0316	-2132.8699	0.6089	-66.33
CQ12011	-0.0634	-0.0010	0.0143	0.0288	-58.5948	0.5428	-31.78
CQ42010	-0.0487	0.0022	0.0163	0.0312	-53.1148	1.0000	-35.58
CQ32010	-0.0438	0.0025	0.0169	0.0304	-46.0250	1.0188	-36.85
CQ22010	-0.0320	0.0026	0.0164	0.0302	-19.1183	1.5983	-22.13
CQ12010	-0.0419	-0.0007	0.0138	0.0273	-36.6020	0.4561	-34.45
CQ42009	-0.0343	-0.0002	0.0150	0.0298	-33.8058	1.3788	-34.61
CQ32009	-0.0318	-0.0022	0.0142	0.0272	-20.3856	0.6883	-27.78
CQ22009	-0.0519	-0.0056	0.0121	0.0251	-58.4570	0.7506	-43.32
CQ12009	-0.0431	-0.0125	0.0093	0.0232	-14.2706	0.8224	-18.20
CQ42008	-0.1768	-0.0103	0.0123	0.0278	-631.5087	0.9664	-67.14
CQ32008	-0.0868	-0.0035	0.0155	0.0313	-310.7394	6.6019	-67.05
CQ22008	-0.0777	-0.0016	0.0157	0.0307	-258.8889	0.5149	-67.56
CQ12008	-0.0724	-0.0051	0.0135	0.0282	-219.2135	0.3485	-67.57
CQ42007	-0.0208	-0.0020	0.0161	0.0313	-20.1038	1.4260	-35.52
CQ32007	-9.6514	-0.0010	0.0166	0.0310	-45595.3200	0.7137	-68.87
CQ22007	-0.4721	-0.0015	0.0163	0.0313	-2122.3220	2.1437	-69.17
CQ12007	-0.0303	-0.0028	0.0148	0.0289	-26.5735	0.4796	-34.12
CQ42006	-0.0406	0.0006	0.0172	0.0322	-40.7690	0.7490	-36.92
CQ32006	-0.0334	0.0008	0.0171	0.0322	-55.5909	0.3000	-52.48
CQ22006	-0.0287	0.0019	0.0171	0.0324	-27.8010	0.4214	-35.24

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ12006	-0.0474	-0.0003	0.0159	0.0308	-108.0385	4.9741	-58.86
CQ42005	-0.1955	0.0010	0.0181	0.0340	-654.4810	1.8343	-67.08
CQ32005	-0.0373	-0.0004	0.0168	0.0320	-24.0598	10.0535	-23.47
CQ22005	-0.0393	0.0015	0.0170	0.0328	-66.7193	6.0000	-52.20
CQ12005	-0.0384	-0.0007	0.0153	0.0301	-48.5581	0.3633	-42.53
CQ42004	-0.0123	0.0007	0.0167	0.0324	-54.2202	203.9198	52.49
CQ32004	-0.0493	0.0004	0.0169	0.0321	-40.5568	0.5666	-29.89
CQ22004	-0.1833	0.0026	0.0171	0.0320	-581.1429	2.7928	-65.52
CQ12004	-0.2737	0.0001	0.0152	0.0298	-806.5590	0.3422	-64.03
CQ42003	-0.0863	-0.0003	0.0158	0.0305	-121.5319	0.9713	-43.40
CQ32003	-0.8135	-0.0015	0.0150	0.0297	-3435.5714	0.5999	-68.26
CQ22003	-0.3924	-0.0036	0.0141	0.0274	-1584.2500	1.5130	-68.37
CQ12003	-0.1011	-0.0078	0.0115	0.0260	-190.8056	0.5352	-47.71
CQ42002	-0.0325	-0.0086	0.0133	0.0282	-27.3738	3.2713	-32.44
CQ32002	0.0087	-0.0080	0.0129	0.0276	-10.2274	120.2710	66.45
CQ22002	-0.0759	-0.0101	0.0133	0.0271	-204.9850	0.8642	-62.59
CQ12002	-0.2297	-0.0114	0.0110	0.0249	-1032.6990	1.8188	-69.50
CQ42001	0.0492	-0.0198	0.0101	0.0252	-14.3400	333.6103	68.23
CQ32001	-0.0154	-0.0208	0.0107	0.0263	-1.8623	2.6954	-3.86
CQ22001	-0.0148	-0.0187	0.0124	0.0274	-4.7264	22.3234	43.92
CQ12001	-0.0178	-0.0176	0.0121	0.0271	-30.6679	24.0047	-16.13
CQ42000	2.5582	-0.0201	0.0145	0.0311	-11.0122	12519.6370	69.61
CQ32000	-0.0195	-0.0122	0.0161	0.0324	-17.1574	0.4972	-34.05
CQ22000	-0.0110	-0.0090	0.0168	0.0326	-8.8825	12.2351	6.34
CQ12000	-0.0294	-0.0094	0.0154	0.0307	-66.8918	0.4551	-61.38
CQ41999	-0.0185	-0.0102	0.0169	0.0336	-13.0166	2.3967	-27.52
CQ31999	-0.0122	-0.0039	0.0176	0.0336	-8.1043	0.3643	-23.85

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ21999	-0.0117	-0.0016	0.0178	0.0330	-7.8750	0.9885	-20.02
CQ11999	-0.0097	-0.0006	0.0166	0.0307	-16.3601	0.4779	-49.77
CQ41998	-0.0170	-0.0031	0.0180	0.0347	-13.9412	0.7930	-32.60
CQ31998	-0.0044	0.0017	0.0191	0.0350	-15.8242	0.5402	-48.78
CQ21998	-0.0022	0.0040	0.0200	0.0358	-11.7270	0.3255	-36.41
CQ11998	-0.0008	0.0043	0.0194	0.0341	-9.7010	0.2163	-34.00

Table 2: Summary Statistics of TR/TA Across All Calendar Quarters

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ42023	0.1602	0.0481	0.1249	0.2213	-0.4136	2.5490	2.89
CQ32023	0.1612	0.0481	0.1267	0.2229	-0.1261	2.8721	3.27
CQ22023	0.1610	0.0470	0.1242	0.2245	-0.1137	2.5538	3.27
CQ12023	0.1585	0.0461	0.1216	0.2192	-0.0891	2.0480	3.03
CQ42022	0.1613	0.0455	0.1240	0.2227	-1.5590	2.4455	2.40
CQ32022	0.1634	0.0475	0.1272	0.2282	-0.1736	2.8103	3.24
CQ22022	0.1628	0.0444	0.1245	0.2263	-0.1335	2.7886	3.16
CQ12022	0.1562	0.0428	0.1192	0.2182	-0.0789	2.0128	2.54
CQ42021	0.1591	0.0416	0.1199	0.2214	-0.6873	2.6025	2.73
CQ32021	0.1559	0.0397	0.1184	0.2170	-0.1856	2.9098	3.23
CQ22021	0.1534	0.0396	0.1176	0.2185	-24.2316	2.5305	-50.49
CQ12021	0.1483	0.0390	0.1144	0.2127	-24.2316	1.9618	-51.83
CQ42020	0.1547	0.0375	0.1158	0.2146	-0.3970	2.5159	2.85
CQ32020	0.1528	0.0393	0.1129	0.2085	-0.0141	2.4290	3.26
CQ22020	0.1403	0.0341	0.1016	0.1866	-0.2238	8.0252	16.83
CQ12020	0.1508	0.0415	0.1159	0.2071	-0.2811	7.3926	14.13
CQ42019	0.1690	0.0462	0.1268	0.2307	-0.7328	12.6926	27.23

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ32019	0.1692	0.0445	0.1277	0.2332	-0.0525	10.2932	20.45
CQ22019	0.1705	0.0459	0.1256	0.2334	-0.0553	8.0044	13.40
CQ12019	0.1649	0.0468	0.1249	0.2306	-0.0258	2.2241	2.80
CQ42018	0.1839	0.0464	0.1371	0.2515	-0.8248	2.6932	2.44
CQ32018	0.1828	0.0471	0.1360	0.2516	-0.0611	3.2776	3.63
CQ22018	0.1875	0.0481	0.1385	0.2564	-0.0153	2.9606	2.81
CQ12018	0.1821	0.0479	0.1363	0.2489	-0.0153	2.8458	3.03
CQ42017	0.1891	0.0472	0.1388	0.2513	-0.0918	6.8012	8.03
CQ32017	0.1836	0.0470	0.1353	0.2501	-0.0028	4.1890	4.58
CQ22017	0.1832	0.0460	0.1356	0.2513	-0.0081	3.1294	3.12
CQ12017	0.1800	0.0455	0.1352	0.2484	-0.0030	3.3094	3.50
CQ42016	0.1871	0.0449	0.1409	0.2558	-0.3178	3.9598	3.79
CQ32016	0.1853	0.0457	0.1386	0.2540	-0.0309	4.4162	4.85
CQ22016	0.1878	0.0462	0.1384	0.2548	-0.0442	9.2145	13.93
CQ12016	0.1797	0.0442	0.1342	0.2449	-1.0200	8.1900	12.33
CQ42015	0.1893	0.0453	0.1445	0.2565	-2.1868	5.9781	7.24
CQ32015	0.1865	0.0447	0.1407	0.2565	-0.5412	5.1566	7.18
CQ22015	0.1882	0.0455	0.1422	0.2576	-0.0138	5.1133	6.56
CQ12015	0.1869	0.0447	0.1429	0.2524	-0.0073	4.8625	6.51
CQ42014	0.2016	0.0513	0.1521	0.2692	-0.9440	5.2215	6.74
CQ32014	0.2016	0.0545	0.1517	0.2699	-0.0190	5.4824	7.40
CQ22014	0.2030	0.0555	0.1484	0.2676	-0.2229	5.5181	6.64
CQ12014	0.1988	0.0538	0.1444	0.2634	-0.1008	5.2723	7.01
CQ42013	0.2082	0.0558	0.1559	0.2785	-14.6898	7.5364	-11.03
CQ32013	0.2079	0.0551	0.1536	0.2774	-0.0233	6.8061	8.52
CQ22013	0.2089	0.0562	0.1581	0.2798	-0.0837	6.8377	8.36
CQ12013	0.2038	0.0554	0.1528	0.2739	-1.3877	6.3384	7.03

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ42012	0.2123	0.0574	0.1605	0.2855	-1.3254	5.0997	4.84
CQ32012	0.2072	0.0582	0.1598	0.2820	-0.0821	5.9701	6.57
CQ22012	0.2135	0.0579	0.1639	0.2862	-0.7465	7.8929	10.05
CQ12012	0.2106	0.0571	0.1596	0.2871	-0.4690	6.6949	7.48
CQ42011	0.2186	0.0594	0.1656	0.2936	-4.1676	8.2664	9.07
CQ32011	0.2169	0.0591	0.1675	0.2974	-0.4723	9.1316	10.74
CQ22011	0.2189	0.0605	0.1690	0.2965	-0.0972	7.2672	8.81
CQ12011	0.2153	0.0598	0.1627	0.2890	-0.0442	6.4174	6.85
CQ42010	0.2208	0.0582	0.1707	0.2954	-2.4481	8.0003	8.03
CQ32010	0.2155	0.0599	0.1686	0.2920	-0.0475	4.9895	5.40
CQ22010	0.2182	0.0618	0.1696	0.2963	-0.0893	9.3439	13.36
CQ12010	0.2121	0.0608	0.1622	0.2839	-0.5689	16.3180	25.46
CQ42009	0.2190	0.0622	0.1658	0.2924	-1.1717	16.1340	23.57
CQ32009	0.2100	0.0624	0.1602	0.2829	-0.1940	6.4322	8.21
CQ22009	0.2065	0.0603	0.1547	0.2773	-0.1502	21.2138	39.66
CQ12009	0.2072	0.0594	0.1546	0.2729	-0.6851	42.2145	59.80
CQ42008	0.2256	0.0604	0.1696	0.2920	-1.1523	28.2789	39.33
CQ32008	0.2223	0.0658	0.1750	0.3026	-0.4068	10.5146	13.60
CQ22008	0.2200	0.0652	0.1717	0.3013	-1.4522	11.2719	14.96
CQ12008	0.2113	0.0616	0.1644	0.2894	-0.7254	8.7398	10.41
CQ42007	0.2214	0.0616	0.1707	0.3020	-0.8298	9.1816	10.75
CQ32007	0.2142	0.0634	0.1707	0.2960	-10.6019	4.9276	-7.96
CQ22007	0.2188	0.0629	0.1700	0.3048	-0.1014	5.6184	5.40
CQ12007	0.2165	0.0627	0.1720	0.2990	-0.9833	4.7051	4.23
CQ42006	0.2265	0.0653	0.1782	0.3088	-0.4283	4.2945	3.45
CQ32006	0.2230	0.0642	0.1752	0.3098	-0.3940	4.8242	4.06
CQ22006	0.2235	0.0658	0.1740	0.3116	-4.2642	6.7361	4.27

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ12006	0.2237	0.0657	0.1778	0.3070	-0.5124	6.2649	6.99
CQ42005	0.2337	0.0675	0.1811	0.3166	-0.7504	4.6704	4.12
CQ32005	0.2296	0.0660	0.1792	0.3100	-0.0391	12.0225	14.96
CQ22005	0.2265	0.0671	0.1805	0.3122	-1.2025	4.8583	3.73
CQ12005	0.2193	0.0644	0.1747	0.3039	-0.5432	4.2482	3.53
CQ42004	0.2132	0.0655	0.1783	0.3106	-90.1257	14.0823	-63.66
CQ32004	0.2259	0.0634	0.1774	0.3096	-0.1342	4.3140	3.74
CQ22004	0.2288	0.0637	0.1764	0.3096	-0.0399	17.5420	27.41
CQ12004	0.2260	0.0624	0.1765	0.3059	-0.0366	17.1592	27.51
CQ42003	0.2309	0.0631	0.1765	0.3115	-1.8371	5.4654	4.48
CQ32003	0.2236	0.0630	0.1735	0.3095	-0.2667	3.3676	3.07
CQ22003	0.2249	0.0650	0.1752	0.3050	-0.0103	5.8640	5.30
CQ12003	0.2207	0.0641	0.1717	0.2992	-0.0486	3.8329	4.09
CQ42002	0.2276	0.0641	0.1728	0.3108	-7.9407	5.2927	0.01
CQ32002	0.2198	0.0612	0.1676	0.3016	-0.1293	4.5402	4.71
CQ22002	0.2202	0.0606	0.1671	0.2997	-0.0751	8.0190	8.97
CQ12002	0.2148	0.0569	0.1601	0.2917	-0.0973	9.2080	10.98
CQ42001	0.2143	0.0562	0.1611	0.2940	-22.8680	4.5669	-34.68
CQ32001	0.2167	0.0564	0.1622	0.2961	-0.3946	10.4743	14.73
CQ22001	0.2246	0.0595	0.1671	0.3039	-1.4479	23.3343	35.08
CQ12001	0.2305	0.0569	0.1738	0.3032	-0.2499	32.1030	40.45
CQ42000	-10.2073	0.0563	0.1748	0.3130	-56816.0190	7.5529	-73.76
CQ32000	0.2188	0.0545	0.1723	0.3084	-0.8421	5.1885	5.04
CQ22000	0.2217	0.0546	0.1743	0.3114	-3.9189	5.1998	4.11
CQ12000	0.2192	0.0523	0.1740	0.3107	-0.4325	5.6850	5.59
CQ41999	0.2292	0.0586	0.1876	0.3244	-17.7723	5.5757	-23.14
CQ31999	0.2317	0.0651	0.1906	0.3210	-0.3389	5.9957	6.21

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ21999	0.2326	0.0651	0.1923	0.3255	-0.2196	4.6193	3.69
CQ11999	0.2298	0.0632	0.1897	0.3204	-0.2659	7.2411	7.19
CQ41998	0.2447	0.0673	0.1937	0.3346	-0.5474	7.7895	7.26
CQ31998	0.2350	0.0647	0.1943	0.3274	-1.4440	4.6849	4.05
CQ21998	0.2392	0.0676	0.2019	0.3342	-0.0344	5.2351	5.36
CQ11998	0.2352	0.0668	0.2018	0.3324	-0.1784	4.3476	3.39

Table 3: Summary Statistics of RE/TA Across All Calendar Quarters

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ42023	-2.3370	-0.2870	0.0445	0.2840	-14789.8070	9335.9692	-25.33
CQ32023	-2.5775	-0.3485	0.0405	0.2808	-29555.5100	14891.8541	-29.31
CQ22023	-8.3918	-0.3593	0.0382	0.2708	-29517.0910	8831.3542	-48.19
CQ12023	-8.0828	-0.3629	0.0355	0.2684	-29505.1460	3764.2463	-57.23
CQ42022	-8.6324	-0.3693	0.0371	0.2593	-29229.5260	10775.8086	-44.29
CQ32022	-7.0804	-0.3659	0.0313	0.2530	-29195.2320	15285.0054	-35.32
CQ22022	-9.8591	-0.3348	0.0295	0.2493	-29176.2500	1726.6411	-59.36
CQ12022	-10.6042	-0.3266	0.0282	0.2484	-29101.8270	14615.1279	-32.04
CQ42021	-6.8090	-0.3140	0.0262	0.2370	-29002.4300	10102.2906	-49.23
CQ32021	5.0098	-0.3143	0.0242	0.2346	-14447.8430	40324.1512	49.00
CQ22021	-20.1907	-0.3105	0.0211	0.2326	-28796.3780	1693.1415	-36.14
CQ12021	-39.1321	-0.3121	0.0218	0.2389	-45922.0000	3456.4823	-31.81
CQ42020	-37.9017	-0.2913	0.0266	0.2458	-44612.0000	7612.7562	-29.48
CQ32020	-15.8580	-0.3104	0.0274	0.2448	-24168.6255	6607.0946	-32.94
CQ22020	-10.5055	-0.3117	0.0304	0.2443	-28389.2110	5678.0971	-41.42
CQ12020	-23.6576	-0.2931	0.0357	0.2486	-55968.6236	59441.2107	-8.83
CQ42019	-196.3059	-0.2558	0.0410	0.2614	-564590.4839	6045.8662	-58.36

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ32019	-38.2472	-0.2379	0.0430	0.2625	-45535.9619	5246.8223	-34.07
CQ22019	-18.3081	-0.2301	0.0443	0.2636	-27997.4750	5770.2088	-36.14
CQ12019	-13.8280	-0.2162	0.0447	0.2645	-27897.2840	2155.2180	-50.80
CQ42018	-67.1967	-0.2105	0.0461	0.2800	-211932.5860	7614.9279	-59.04
CQ32018	-7.9637	-0.2219	0.0468	0.2827	-27616.4940	46426.2821	21.24
CQ22018	-65.1528	-0.2263	0.0469	0.2812	-211678.6920	24505.8617	-57.55
CQ12018	-12.4785	-0.2190	0.0453	0.2786	-27354.9870	17924.1623	-22.33
CQ42017	-2.7190	-0.2214	0.0468	0.2800	-35151.7915	91372.7273	35.30
CQ32017	-11.6027	-0.2224	0.0406	0.2710	-35175.9127	6256.2450	-52.03
CQ22017	-17.1969	-0.2189	0.0418	0.2732	-21650.4280	5754.7319	-32.83
CQ12017	22.2015	-0.2296	0.0387	0.2760	-26524.5709	122729.9600	55.24
CQ42016	-34.8125	-0.2100	0.0417	0.2801	-80324.7359	5080.6578	-53.06
CQ32016	-17.9984	-0.2314	0.0387	0.2753	-15920.5587	5268.0578	-25.65
CQ22016	-18.5921	-0.2132	0.0407	0.2781	-20770.3735	4072.4726	-32.14
CQ12016	-10.9330	-0.2172	0.0393	0.2808	-14812.9850	4282.5677	-33.57
CQ42015	-8.1757	-0.2095	0.0413	0.2829	-8486.0089	4420.3774	-25.94
CQ32015	-4.6070	-0.2161	0.0397	0.2794	-6464.1285	2587.0516	-30.58
CQ22015	-6.0588	-0.2200	0.0441	0.2835	-7029.1395	1894.6872	-31.53
CQ12015	-1.9536	-0.2051	0.0467	0.2862	-3162.2962	3130.2110	-2.56
CQ42014	-53.1871	-0.1830	0.0477	0.2854	-179564.7439	5347.9208	-61.17
CQ32014	-6.8751	-0.2078	0.0447	0.2843	-15572.0780	5336.9794	-38.82
CQ22014	-7.4429	-0.2109	0.0457	0.2875	-15441.1820	5173.5962	-36.66
CQ12014	-7.3288	-0.2221	0.0446	0.2865	-15232.8550	5836.3499	-34.13
CQ42013	-10.3501	-0.1819	0.0505	0.2923	-15032.6600	6271.4364	-30.15
CQ32013	-11.1072	-0.1995	0.0487	0.2965	-20008.5610	7011.0619	-36.67
CQ22013	-16.1692	-0.1901	0.0527	0.3011	-26740.0470	3461.7248	-42.44
CQ12013	-31.5284	-0.1824	0.0531	0.3004	-78161.2793	3231.4241	-56.70

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ42012	-5.8633	-0.1622	0.0552	0.2976	-9149.0340	3357.6159	-30.09
CQ32012	-23.5761	-0.1705	0.0547	0.3013	-33301.5150	3332.5523	-38.32
CQ22012	-4.3323	-0.1647	0.0559	0.3034	-6650.0160	3710.3852	-14.40
CQ12012	-1.6960	-0.1731	0.0531	0.2996	-3793.0425	5686.3726	8.19
CQ42011	-46.2646	-0.1654	0.0524	0.3004	-170896.2812	7929.2358	-61.20
CQ32011	-1.8140	-0.1801	0.0538	0.2999	-9806.9650	16552.0504	21.70
CQ22011	2.4456	-0.1707	0.0543	0.2990	-7366.2727	16500.2096	37.03
CQ12011	0.9605	-0.1844	0.0500	0.2971	-4950.0613	9437.3111	28.80
CQ42010	-3.4155	-0.1772	0.0524	0.3015	-7776.8030	3555.1134	-26.17
CQ32010	-236.7361	-0.2048	0.0485	0.2976	-910752.2124	13160.9785	-61.81
CQ22010	-34.1695	-0.2038	0.0483	0.3044	-129035.4090	6724.5915	-61.39
CQ12010	-5.8890	-0.2244	0.0471	0.2993	-21959.0490	12056.8297	-30.64
CQ42009	-16.4944	-0.2067	0.0491	0.2958	-49973.4000	12883.0098	-46.74
CQ32009	-7.4920	-0.2235	0.0482	0.2885	-18560.1481	8010.2300	-33.34
CQ22009	-5.4797	-0.2278	0.0486	0.2852	-11866.1488	9658.3131	-10.74
CQ12009	-4.2329	-0.2431	0.0452	0.2842	-12818.8737	7504.0246	-23.56
CQ42008	-6.6734	-0.2256	0.0467	0.2849	-11344.9335	4844.7424	-29.69
CQ32008	-1.6456	-0.1535	0.0537	0.2828	-4949.9505	3022.3869	-19.83
CQ22008	-223.2066	-0.1564	0.0555	0.2792	-898301.7778	2277.7092	-63.61
CQ12008	-21.2566	-0.1603	0.0518	0.2759	-81479.3371	4093.5742	-63.09
CQ42007	-3.8015	-0.1512	0.0541	0.2710	-8776.1320	2361.0753	-42.30
CQ32007	-34.0580	-0.1614	0.0521	0.2724	-68461.8000	1677.9476	-44.82
CQ22007	-11.8937	-0.1671	0.0536	0.2683	-17115.5480	1878.9664	-38.15
CQ12007	-3.5217	-0.1654	0.0555	0.2761	-8434.9205	1498.4011	-39.83
CQ42006	-2.1962	-0.1509	0.0571	0.2744	-6236.4105	13799.2216	26.66
CQ32006	-7.5214	-0.1753	0.0544	0.2681	-19804.1903	6563.6534	-44.55
CQ22006	-6.1019	-0.1773	0.0517	0.2634	-11158.1172	6539.2577	-25.52

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ12006	-5.9083	-0.1762	0.0500	0.2647	-12841.0870	5624.6101	-35.92
CQ42005	-5.2625	-0.1748	0.0523	0.2687	-11511.3920	2833.1012	-47.55
CQ32005	-4.2898	-0.1929	0.0490	0.2620	-5281.8077	3492.4019	-19.38
CQ22005	-5.6926	-0.1986	0.0487	0.2574	-9063.3548	3730.5455	-32.69
CQ12005	-7.6371	-0.2033	0.0485	0.2571	-21825.2308	2962.2627	-57.23
CQ42004	-16.5056	-0.1952	0.0513	0.2574	-22602.9760	1885.9358	-36.66
CQ32004	-15.1924	-0.1907	0.0505	0.2510	-19329.2578	355.6737	-36.87
CQ22004	-11.4582	-0.1930	0.0517	0.2525	-19297.5530	2346.5714	-45.30
CQ12004	-84.3382	-0.1982	0.0489	0.2426	-294593.1130	1320.1584	-63.56
CQ42003	-274.8128	-0.1934	0.0507	0.2516	-1069363.8936	482.8803	-64.50
CQ32003	-43.1239	-0.2148	0.0483	0.2427	-88472.4773	1287.4177	-51.00
CQ22003	-2959.2069	-0.2204	0.0476	0.2432	-11936588.5000	14615.8445	-63.59
CQ12003	-34.8058	-0.2299	0.0489	0.2442	-93220.5276	11945.0176	-55.82
CQ42002	-5.0901	-0.2412	0.0468	0.2362	-8115.8926	7016.8590	-13.95
CQ32002	-2.2175	-0.2397	0.0466	0.2323	-8129.6822	5796.8836	-19.79
CQ22002	-3.4374	-0.2427	0.0434	0.2313	-11782.0210	5760.4527	-37.64
CQ12002	-6.0424	-0.2245	0.0441	0.2389	-11560.8250	3682.4052	-33.48
CQ42001	-1.4743	-0.1857	0.0463	0.2404	-7249.9930	2342.2216	-50.68
CQ32001	0.5719	-0.1902	0.0445	0.2343	-539.5934	2192.3811	38.36
CQ22001	-8.9841	-0.1790	0.0475	0.2391	-42932.6327	2448.0013	-64.12
CQ12001	-9.5963	-0.1564	0.0502	0.2446	-44906.8480	2560.7440	-64.38
CQ42000	-16.8844	-0.1444	0.0534	0.2486	-39507.8900	1395.3059	-46.43
CQ32000	-3.5026	-0.1409	0.0497	0.2352	-16869.3677	1586.8202	-61.60
CQ22000	-1.0538	-0.1401	0.0484	0.2383	-7797.3111	914.6033	-61.18
CQ12000	-20.0356	-0.1339	0.0476	0.2393	-84134.2313	959.8762	-65.29
CQ41999	0.9074	-0.1090	0.0528	0.2466	-1247.9974	2467.5992	30.44
CQ31999	0.8797	-0.0913	0.0543	0.2482	-895.6390	2309.7951	37.53

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ21999	3.3522	-0.0799	0.0561	0.2524	-569.3544	10779.5923	59.58
CQ11999	0.9246	-0.0621	0.0578	0.2552	-311.5980	2352.3286	53.18
CQ41998	0.8466	-0.0481	0.0584	0.2621	-1133.0915	2678.3275	40.46
CQ31998	4.1103	-0.0370	0.0662	0.2647	-1178.5808	14757.3859	61.12
CQ21998	1.3643	-0.0344	0.0662	0.2698	-1883.8528	6003.9889	49.46
CQ11998	1.0223	-0.0288	0.0685	0.2733	-1782.7441	4510.0159	42.69

Table 4: Summary Statistics of WC/TA Across All Calendar Quarters

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ42023	-0.0303	-0.0393	0.0183	0.1065	-78.4185	0.9405	-37.40
CQ32023	-0.1718	-0.0373	0.0205	0.1130	-352.1058	0.9431	-45.45
CQ22023	-0.1695	-0.0373	0.0213	0.1150	-284.6129	0.9464	-37.76
CQ12023	-0.0975	-0.0369	0.0214	0.1161	-180.5118	0.9512	-37.26
CQ42022	-0.1758	-0.0391	0.0183	0.1087	-391.2046	0.9221	-47.95
CQ32022	-0.0879	-0.0394	0.0191	0.1144	-173.1612	0.9073	-40.41
CQ22022	-0.3050	-0.0352	0.0212	0.1131	-1108.0740	0.9170	-64.07
CQ12022	-1.7270	-0.0369	0.0195	0.1063	-5273.9880	0.9133	-58.06
CQ42021	-1.3807	-0.0403	0.0142	0.0923	-5174.5910	0.9063	-66.17
CQ32021	-0.6475	-0.0388	0.0134	0.0942	-2533.9235	0.9043	-66.39
CQ22021	-1.4522	-0.0380	0.0133	0.0928	-4968.5390	0.9161	-64.75
CQ12021	-3.4724	-0.0351	0.0134	0.0923	-9113.0000	0.9198	-52.20
CQ42020	-3.5247	-0.0385	0.0115	0.0891	-8456.0000	0.9197	-50.01
CQ32020	-0.8026	-0.0368	0.0142	0.0968	-1555.3637	0.9291	-45.27
CQ22020	-1.2201	-0.0358	0.0163	0.1000	-4561.3720	0.8792	-64.76
CQ12020	-1.7349	-0.0320	0.0212	0.1103	-4460.3420	0.8869	-53.67
CQ42019	-4.6037	-0.0360	0.0180	0.1030	-10838.1613	0.9553	-52.34

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ32019	-1.6813	-0.0338	0.0218	0.1122	-4257.4900	0.9199	-56.10
CQ22019	-1.2329	-0.0310	0.0232	0.1174	-4169.6360	0.9206	-63.74
CQ12019	-1.2936	-0.0327	0.0236	0.1191	-4069.4450	0.9201	-61.50
CQ42018	-1.5653	-0.0386	0.0174	0.1121	-3958.5740	0.9439	-56.03
CQ32018	-2.1195	-0.0387	0.0226	0.1162	-3788.6550	0.9433	-44.36
CQ22018	-2.0678	-0.0353	0.0249	0.1185	-3658.6400	0.9668	-41.47
CQ12018	-1.3339	-0.0330	0.0258	0.1182	-3527.1480	0.9338	-54.99
CQ42017	-1.6027	-0.0402	0.0189	0.1094	-3398.6090	0.9429	-50.35
CQ32017	-0.3258	-0.0380	0.0215	0.1133	-360.3683	0.9487	-32.30
CQ22017	-0.5176	-0.0358	0.0224	0.1153	-1169.3030	0.9357	-56.20
CQ12017	-0.4517	-0.0375	0.0223	0.1147	-745.4812	0.9379	-44.72
CQ42016	-0.7106	-0.0380	0.0194	0.1079	-595.4490	0.9454	-28.93
CQ32016	-0.9760	-0.0337	0.0238	0.1201	-2524.5870	0.9453	-59.16
CQ22016	-1.4056	-0.0317	0.0261	0.1228	-2661.6180	0.9446	-45.80
CQ12016	-1.2467	-0.0305	0.0278	0.1269	-2164.2490	0.9955	-44.02
CQ42015	-0.6357	-0.0337	0.0240	0.1196	-965.5370	0.9407	-43.19
CQ32015	-0.3517	-0.0291	0.0281	0.1310	-625.9725	0.9374	-44.19
CQ22015	-0.3099	-0.0274	0.0324	0.1326	-704.9042	0.9370	-51.30
CQ12015	-0.1123	-0.0269	0.0330	0.1343	-301.3770	0.9415	-53.39
CQ42014	-0.5352	-0.0320	0.0295	0.1235	-1483.6540	0.9483	-59.32
CQ32014	-0.5757	-0.0344	0.0342	0.1326	-1629.7780	0.9412	-57.05
CQ22014	-0.2931	-0.0313	0.0340	0.1379	-570.4940	0.9496	-47.79
CQ12014	-0.3056	-0.0318	0.0340	0.1408	-571.3640	0.9498	-47.33
CQ42013	-0.4269	-0.0335	0.0312	0.1370	-595.8090	0.9565	-37.38
CQ32013	-0.7030	-0.0304	0.0366	0.1456	-1244.0360	0.9398	-44.90
CQ22013	-0.4816	-0.0262	0.0410	0.1461	-803.8532	0.9403	-41.89
CQ12013	-1.7128	-0.0249	0.0417	0.1516	-5490.7248	0.9547	-63.85

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ42012	-0.3998	-0.0295	0.0349	0.1439	-577.9490	0.9570	-39.68
CQ32012	-1.3609	-0.0255	0.0404	0.1492	-3275.5480	0.9402	-58.73
CQ22012	-1.4400	-0.0244	0.0425	0.1497	-4388.0567	0.9363	-62.66
CQ12012	-1.1663	-0.0257	0.0428	0.1470	-3485.7143	0.9201	-61.74
CQ42011	-3.4879	-0.0289	0.0368	0.1429	-9579.4630	0.9536	-57.29
CQ32011	-2.3259	-0.0259	0.0411	0.1469	-7435.5000	0.9418	-61.18
CQ22011	-0.4724	-0.0255	0.0451	0.1485	-1571.0616	0.9399	-62.14
CQ12011	-0.3156	-0.0269	0.0426	0.1469	-463.9183	0.9424	-39.17
CQ42010	-0.2601	-0.0306	0.0377	0.1387	-319.5094	1.0000	-36.68
CQ32010	-0.1302	-0.0269	0.0423	0.1419	-299.1000	0.9511	-52.83
CQ22010	-0.0983	-0.0264	0.0413	0.1394	-221.0137	0.9325	-51.48
CQ12010	-0.1759	-0.0269	0.0408	0.1394	-342.7813	0.9578	-49.81
CQ42009	-0.2084	-0.0284	0.0384	0.1383	-314.6833	0.9300	-41.44
CQ32009	-0.2081	-0.0255	0.0432	0.1444	-565.7316	0.9743	-62.41
CQ22009	-0.5263	-0.0244	0.0452	0.1481	-1640.1107	0.9636	-61.71
CQ12009	-0.3612	-0.0219	0.0476	0.1586	-1024.3874	0.9630	-56.57
CQ42008	-0.2460	-0.0282	0.0440	0.1524	-774.8564	0.9602	-58.05
CQ32008	-0.0033	-0.0268	0.0495	0.1605	-103.8079	0.9676	-51.72
CQ22008	-0.3399	-0.0269	0.0483	0.1589	-1580.8889	0.9733	-67.32
CQ12008	-0.0804	-0.0298	0.0461	0.1588	-241.6983	0.9721	-43.55
CQ42007	-0.0386	-0.0320	0.0436	0.1504	-233.3443	0.9727	-56.99
CQ32007	-5.5698	-0.0267	0.0490	0.1625	-18501.1600	0.9912	-57.59
CQ22007	-1.1309	-0.0255	0.0465	0.1612	-2585.6180	0.9932	-45.63
CQ12007	-0.2812	-0.0251	0.0465	0.1616	-1256.4545	0.9931	-66.98
CQ42006	-0.5824	-0.0295	0.0407	0.1515	-2321.8315	0.9954	-66.15
CQ32006	-0.3661	-0.0264	0.0474	0.1604	-909.9233	0.9926	-46.68
CQ22006	-0.3785	-0.0255	0.0479	0.1592	-1074.7287	0.9918	-51.38

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ12006	-0.5008	-0.0269	0.0459	0.1598	-1242.6907	0.9897	-48.37
CQ42005	-0.5564	-0.0298	0.0418	0.1515	-1474.5170	0.9894	-51.07
CQ32005	-0.3003	-0.0273	0.0454	0.1601	-673.9222	0.9905	-47.79
CQ22005	-0.2236	-0.0261	0.0431	0.1605	-663.8778	0.9751	-53.02
CQ12005	-0.0883	-0.0268	0.0429	0.1590	-237.6320	0.9649	-42.62
CQ42004	-1.5851	-0.0297	0.0402	0.1588	-3870.9580	0.9997	-49.83
CQ32004	-0.6791	-0.0284	0.0445	0.1678	-2869.9220	0.9221	-68.12
CQ22004	-1.0137	-0.0256	0.0452	0.1636	-2852.4286	0.9771	-55.49
CQ12004	-0.5563	-0.0287	0.0423	0.1643	-772.6057	0.9639	-36.26
CQ42003	-0.3752	-0.0294	0.0411	0.1586	-799.6170	0.9330	-46.68
CQ32003	-1.9742	-0.0276	0.0451	0.1662	-4417.5470	0.9975	-47.86
CQ22003	-1.4080	-0.0283	0.0435	0.1739	-6266.0000	0.9975	-68.32
CQ12003	-0.7436	-0.0272	0.0463	0.1713	-3473.4960	0.9976	-68.16
CQ42002	0.0124	-0.0304	0.0420	0.1631	-43.3501	0.9975	-29.46
CQ32002	0.0153	-0.0273	0.0454	0.1734	-97.0974	0.9296	-52.48
CQ22002	-0.5258	-0.0241	0.0489	0.1774	-2434.0410	0.9392	-68.06
CQ12002	-0.4164	-0.0214	0.0500	0.1802	-2231.5450	0.9060	-69.26
CQ42001	0.0499	-0.0253	0.0460	0.1753	-15.6432	0.8948	-21.38
CQ32001	0.0658	-0.0226	0.0553	0.1947	-23.1931	0.9219	-31.48
CQ22001	0.0687	-0.0187	0.0568	0.2003	-18.8516	0.9278	-23.35
CQ12001	0.0598	-0.0165	0.0610	0.2037	-41.5542	0.9020	-37.57
CQ42000	0.0528	-0.0183	0.0567	0.1992	-61.6675	0.9026	-49.70
CQ32000	0.0849	-0.0138	0.0621	0.2008	-11.9341	0.9183	-17.73
CQ22000	0.0852	-0.0127	0.0636	0.2036	-9.4600	0.9267	-16.44
CQ12000	0.0850	-0.0128	0.0641	0.2032	-22.1458	0.9490	-33.75
CQ41999	0.0778	-0.0157	0.0646	0.1969	-30.0775	0.9579	-41.19
CQ31999	0.0948	-0.0118	0.0689	0.2105	-10.7629	0.9518	-17.21

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ21999	0.0918	-0.0118	0.0746	0.2125	-22.0075	0.9658	-32.23
CQ11999	0.0957	-0.0093	0.0787	0.2199	-10.9351	0.9789	-14.62
CQ41998	0.0875	-0.0102	0.0757	0.2160	-16.6912	0.9801	-21.43
CQ31998	0.1104	-0.0042	0.0880	0.2270	-6.1079	0.9114	-8.87
CQ21998	0.1091	-0.0027	0.0918	0.2324	-24.0000	0.9027	-38.22
CQ11998	0.1149	-0.0022	0.0924	0.2333	-12.9542	0.9712	-20.93

Table 5: Summary Statistics of MKTCAP/TL Across All Calendar Quarters

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ42023	14.8944	0.5348	1.5763	4.1796	-706.6714	7309.8908	29.97
CQ32023	14.6659	0.4954	1.4556	3.7965	0.0001	6735.4787	28.20
CQ22023	15.1589	0.5325	1.5456	4.2471	0.0004	7101.0689	30.54
CQ12023	14.7485	0.5193	1.5464	4.0848	0.0004	5896.5097	26.51
CQ42022	17.4976	0.5056	1.4732	3.8386	0.0004	12270.8742	38.68
CQ32022	15.9975	0.4812	1.4211	3.9176	0.0005	7473.9366	29.80
CQ22022	16.0005	0.5556	1.5703	4.0732	0.0000	6813.9056	27.89
CQ12022	18.0535	0.7049	1.9515	5.5426	0.0008	6159.6124	25.47
CQ42021	21.5177	0.7175	2.1322	6.2730	0.0000	8523.0592	27.47
CQ32021	23.7628	0.7363	2.1856	6.9870	0.0000	8951.2879	28.24
CQ22021	21.3086	0.7642	2.2331	7.6420	0.0000	8303.1758	30.10
CQ12021	22.1091	0.6927	2.1061	7.2486	0.0024	7837.9103	29.19
CQ42020	19.6408	0.5603	1.7960	6.3120	-0.6598	5789.0248	24.39
CQ32020	16.8559	0.4068	1.4770	4.9631	0.0007	6240.7555	26.39
CQ22020	15.9833	0.3786	1.3583	4.7303	0.0004	6930.0903	30.29
CQ12020	12.4325	0.2850	1.0322	3.6470	0.0002	4745.3773	28.82
CQ42019	16.6980	0.5154	1.6034	4.5730	0.0002	5403.5313	23.12

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ32019	14.2482	0.4954	1.5015	4.2267	0.0006	5098.1352	25.33
CQ22019	14.4906	0.5221	1.5617	4.4975	0.0000	4238.5802	22.32
CQ12019	15.1244	0.5332	1.5615	4.6000	0.0011	4252.5367	21.32
CQ42018	15.1025	0.5121	1.4625	4.2097	0.0000	5049.2933	22.03
CQ32018	18.8505	0.6504	1.8463	5.3323	0.0007	5754.1336	21.38
CQ22018	18.6605	0.6704	1.8691	5.1841	0.0007	8592.9956	32.02
CQ12018	18.1473	0.6375	1.7989	4.9583	0.0022	8101.2155	30.01
CQ42017	17.7746	0.6608	1.8469	4.8831	0.0007	9864.6160	35.72
CQ32017	17.9157	0.6321	1.7682	4.8113	0.0017	9905.8913	35.76
CQ22017	19.1996	0.6384	1.7314	4.5480	0.0027	12262.2520	36.78
CQ12017	19.9719	0.6829	1.7365	4.5713	0.0039	12681.4513	34.81
CQ42016	16.8373	0.6659	1.6889	4.3161	0.0027	11996.4814	41.72
CQ32016	18.1670	0.6075	1.6141	4.3186	0.0004	16038.8591	49.70
CQ22016	17.0323	0.5633	1.5800	4.0684	0.0004	10422.3822	34.81
CQ12016	20.1368	0.5557	1.5842	4.0996	0.0009	9881.3474	26.53
CQ42015	20.8677	0.5719	1.5885	4.3192	0.0000	10763.7663	28.97
CQ32015	19.4825	0.5652	1.5707	4.2858	0.0000	9506.6575	28.23
CQ22015	24.0427	0.6902	1.8229	4.9021	0.0000	12123.0031	28.66
CQ12015	21.1133	0.7140	1.8943	5.0918	0.0000	13304.7044	38.24
CQ42014	23.6368	0.6912	1.8467	4.8694	0.0001	14622.2222	35.20
CQ32014	22.0216	0.6943	1.8586	4.9458	0.0002	17260.4879	39.33
CQ22014	22.2425	0.7737	1.9978	5.4776	0.0000	18646.9882	43.85
CQ12014	20.2046	0.7610	1.9913	5.7476	0.0001	19156.6172	50.24
CQ42013	21.2375	0.7336	1.9936	5.3827	0.0048	18536.5779	44.64
CQ32013	18.9979	0.6841	1.8487	5.0785	0.0003	16863.1963	46.60
CQ22013	14.4131	0.6410	1.7534	4.7623	0.0000	4006.1339	22.33
CQ12013	17.6782	0.6232	1.7497	4.6204	0.0001	12057.0661	35.54

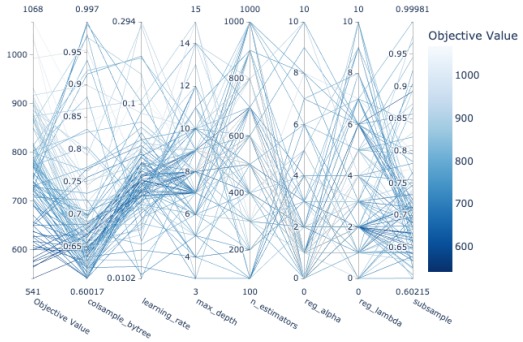
CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ42012	13.6387	0.5480	1.5547	4.1278	0.0001	10374.2394	42.29
CQ32012	14.8909	0.5418	1.5726	4.3294	0.0000	14962.9400	50.70
CQ22012	13.7021	0.5401	1.5715	4.3443	0.0000	13157.4061	50.34
CQ12012	15.9272	0.5696	1.6816	4.6937	0.0000	16262.5227	52.44
CQ42011	13.5407	0.5133	1.4667	4.0013	0.0000	13785.5202	50.59
CQ32011	14.6164	0.4578	1.3598	3.8375	0.0000	16556.9500	53.93
CQ22011	17.5762	0.6020	1.7844	4.9414	0.0000	12390.8622	36.72
CQ12011	18.2521	0.6369	1.8503	5.2579	0.0000	13107.0209	38.02
CQ42010	17.8765	0.5875	1.7296	4.8609	0.0000	19015.8634	48.34
CQ32010	14.3399	0.5301	1.5860	4.4046	0.0000	7844.1432	33.18
CQ22010	14.6647	0.4965	1.4674	4.3511	0.0000	6606.3995	27.40
CQ12010	14.4699	0.5402	1.6431	4.8479	0.0000	5237.6952	25.87
CQ42009	14.1361	0.4764	1.4842	4.5088	0.0000	19828.5273	60.32
CQ32009	154.7687	0.4549	1.4552	4.2434	0.0000	669300.0000	67.52
CQ22009	7.6420	0.3470	1.1788	3.6170	0.0000	1443.6667	18.32
CQ12009	6.7484	0.2438	0.8884	2.8703	0.0000	2651.0970	29.74
CQ42008	9.9257	0.2729	0.9612	2.8788	0.0000	15665.4350	63.99
CQ32008	16.6563	0.4464	1.3328	3.9442	0.0000	35483.6364	65.92
CQ22008	12.4387	0.4985	1.5437	4.6112	0.0000	13673.9220	61.48
CQ12008	29.2999	0.5350	1.6166	4.7054	0.0000	78167.6917	67.61
CQ42007	35.3846	0.6373	1.8473	5.4096	0.0000	89127.6825	62.32
CQ32007	28.6461	0.7366	2.1012	5.9914	0.0000	74608.9277	68.74
CQ22007	27.5503	0.8202	2.2526	6.2315	0.0000	77750.1905	68.91
CQ12007	14.2568	0.8015	2.2523	6.1919	0.0000	3399.0402	19.68
CQ42006	17.5553	0.7710	2.1943	5.8890	0.0000	24540.6403	64.30
CQ32006	14.6780	0.7191	2.0817	5.7335	0.0000	6465.8103	31.46
CQ22006	13.6313	0.7272	2.1350	6.0138	0.0000	5859.2691	33.22

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ12006	25.0680	0.7555	2.3731	6.7204	0.0000	30848.6542	49.63
CQ42005	18.8839	0.6987	2.1152	6.0215	0.0000	25324.4474	61.80
CQ32005	15.6966	0.7033	2.1385	6.2036	0.0000	14144.5672	52.45
CQ22005	10.7779	0.7098	2.0794	5.9564	0.0000	1549.6242	16.27
CQ12005	14.3540	0.7167	2.0756	6.0788	0.0000	14731.2844	60.49
CQ42004	18.6686	0.7570	2.1755	6.5569	0.0000	14567.7854	41.47
CQ32004	12.9521	0.6685	1.9255	5.9025	0.0000	13132.8495	60.43
CQ22004	15.1790	0.6798	2.1000	6.5457	0.0000	18645.4171	64.00
CQ12004	13.4741	0.6512	2.1293	6.7619	0.0000	11397.9551	60.00
CQ42003	19.5238	0.6053	1.9896	6.1758	0.0000	32867.9365	63.62
CQ32003	32.7920	0.5042	1.6963	5.4942	0.0000	107264.9587	68.68
CQ22003	10.3259	0.4508	1.5005	4.7659	0.0000	5747.2846	36.13
CQ12003	8.3042	0.3512	1.2225	3.7745	0.0001	4562.4293	37.06
CQ42002	13.5203	0.3936	1.3230	4.0470	0.0001	19896.9434	61.57
CQ32002	7.4226	0.3624	1.2185	3.6989	0.0000	4857.4749	44.15
CQ22002	10.7445	0.4822	1.5949	4.8308	0.0000	7917.9902	41.19
CQ12002	14.0717	0.5152	1.8109	5.9561	0.0005	6718.3777	31.83
CQ42001	12.7500	0.4835	1.6903	5.8516	0.0003	4012.1197	25.14
CQ32001	8.3906	0.3950	1.3361	4.3487	0.0002	3337.4952	33.35
CQ22001	11.9559	0.5051	1.6216	6.0912	0.0000	4339.7531	32.49
CQ12001	10.6248	0.4423	1.4837	5.2542	0.0000	2030.3785	18.85
CQ42000	13.1458	0.4192	1.4673	5.8841	0.0011	4050.7438	29.04
CQ32000	20.6926	0.4615	1.6613	7.9023	0.0003	15132.7219	55.17
CQ22000	20.8561	0.4644	1.6570	8.5284	0.0003	6912.0078	32.17
CQ12000	23.3311	0.5013	1.7947	9.8515	0.0029	6749.7604	31.54
CQ41999	23.9436	0.5188	1.6766	8.3182	0.0000	14618.7956	51.31
CQ31999	13.5440	0.5469	1.5697	6.0509	0.0004	7603.0158	51.93

CQ	Mean	Q1	Median	Q3	Min	Max	Skewness
CQ21999	15.2936	0.6181	1.7456	6.1702	0.0033	12097.7181	58.18
CQ11999	12.4542	0.5372	1.5120	5.3237	0.0002	7892.1503	54.17
CQ41998	9.4530	0.5943	1.6534	5.4068	0.0000	1454.3942	18.36
CQ31998	17.6195	0.5811	1.5438	4.7043	0.0000	22373.4599	45.30
CQ21998	14.3257	0.7948	2.1239	6.4984	0.0000	10598.5965	45.26
CQ11998	11.9842	0.8734	2.4211	7.1156	0.0000	2690.4507	25.84

Appendix B – Hyperparameter Tuning

5 % Bankruptcy; 10 FN Penalty; XGBoost; Parallel Coordinate Plot



5 % Bankruptcy; 5 FN Penalty; LogisticRegression; Parallel Coordinate Plot