# Finance Protector - Corporate Bankruptcy Prediction System

**Table of Contents**

**1. Executive Summary** ..... 5
**2. Background** ..... 6
2.1 Introduction to Corporate Bankruptcy ..... 6
2.2 Prevalence of Corporate Bankruptcy ..... 6
2.2.1 Global Trends ..... 6
2.2.2 Local Trends ..... 6
2.2.3 Corporate Bankruptcies Across Industries ..... 6
2.3 Causes of Corporate Bankruptcy ..... 7
2.3.1 Financial Mismanagement ..... 7
2.3.2 Unforeseen Events and Economic Downturns ..... 7
2.4 Significance of Insufficient Information in Stock Markets ..... 7
2.4.1 Investors ..... 7
2.4.2 Financial Institutions and Lenders ..... 7
2.5 Problem Statement, Project Objectives and Targets ..... 8
2.5.1 Problem Statement and Objectives ..... 8
2.5.2 Project Targets ..... 8
**3. Market Analysis and Evaluation** ..... 9
3.1 Evaluation of Current Approaches ..... 9
3.1.1 Lack of Predictive Accuracy ..... 9
3.1.2 Lack of Model Diversity ..... 9
3.2 Business Opportunities ..... 9
3.2.1 Boosting Predictive Power Using a Variety of Models ..... 9
3.2.2 Empowering Decision Making ..... 9
**4. Data Cleaning and Exploratory Data Analysis (EDA)** ..... 10
4.1 Introduction to Dataset ..... 10
4.2 Data Cleaning ..... 10
4.2.1 Missing Values and Duplicates ..... 10
4.2.2 Outlier Handling ..... 10
4.3 Exploratory Data Analysis (EDA) ..... 10
4.4 Data Pre-Processing ..... 12

4.4.1 Oversampling using SMOTE....12
4.4.2 Noise Reduction Using KNN....12
**5. Predictive Modelling**....13
5.1 Variable Selection using Logistic Regression and Correlation....13
5.2 Metrics Used....13
5.2.1 Sensitivity and Precision....13
5.2.2 Balanced Accuracy....13
5.2.3 AUC-ROC Score....13
5.2.4 F1 Score....13
5.3 Model 1: Multivariate Adaptive Regression Splines (MARS)....13
5.3.1 Feature Selection....14
5.3.2 Interaction Analysis using MARS....14
5.3.3 Comparison of MARS Models with Varying Complexity....15
5.3.4 Optimising via Tuning of Parameters....16
5.4 Model 2: Random Forest....17
5.4.1 Model Optimisation....17
5.4.2 ROC Curve Analysis....18
5.4.3 Evaluation of Tuned and Default Models....18
5.5 Model 3: Extreme Gradient Boosting (XGBoost)....19
5.5.1 Model Optimisation....19
5.5.2 Evaluation of Tuned and Default Models....19
5.5.3 Insights from Shapley Additive Explanations (SHAP) Analysis....20
5.6 Final Evaluation of Tuned Models....20
5.6.1 Model Performance....20
5.6.2 Model Explainability....20
5.6.3 Individual Model Usage....20
**6. Solution**....21
6.1 Risk Threshold Model (MARS)....21
6.1.1 Practical Implications....21
6.1.2 Feasibility of Implementation....21
6.2 Bankruptcy Risk Classifier (XGBoost)....21
6.2.1 Practical Implications....21

6.2.2 Feasibility of Implementation..... 22
6.3 Risk Pattern Aggregator (Random Forest).....22
6.3.1 Practical Implications.....22
6.3.2 Feasibility of Implementation..... 22
6.4 Model Integration via Web Application – Finance Protector..... 23
6.4.1 Key Differentiators of Solution.....23
6.5 Benefits, Expected Outcomes, Limitations & Future Development.....23
6.5.1 Benefits..... 23
6.5.2 Expected Outcomes.....24
6.5.3 Limitations..... 24
6.5.4 Future Developments and Expansions..... 24
6.6 Risk Assessments and Mitigation Strategies.....25
6.6.1 Data Handling and Compliance Risks..... 25
6.6.2 Over-reliance on Technological Implementations..... 25
**7. Conclusion**.....25
**8. Appendix**..... 26
**9. References**..... 41

## 1. Executive Summary

Corporate bankruptcy is an increasingly significant problem in the business industry, characterised by the inability of a company to meet its long-term debt obligations. Global bankruptcy prevalence has been rising, with bankruptcy filings increasing by 33.5%. Bankruptcy is most prevalent in firms with a high debt ratio, where the risk of being unable to repay their liabilities increase significantly. In Singapore, companies in compulsory liquidation rose, with the number of applications increasing by 6% from 257 in 2022 to 273 in 2023. Bankruptcy causes massive spillover effects to the industry, affecting stakeholders such as investors, lenders and financial institutions alike. Despite its severity and significance, there is still a lack of anticipatory measures for such stakeholders due to insufficient information given.

The objective of this report is to show how machine learning and our data-driven approach is able to tackle this issue of bankruptcy prediction. We aim to create a predictive model for bankruptcy prediction to aid in the decision-making process of our targeted stakeholders. The focus is to utilise these technologies to increase the accuracy of bankruptcy prediction, along with identifying problematic financial aspects to take note of, thus allowing for early prediction and proactive measures to prevent financial losses.

The dataset we considered contains extensive financial data collected from the Taiwan Economic Journal for the years 1999 to 2009, with 6819 instances and 96 attributes. It includes various financial variables, such as book price, realised sales, and common financial indicators such as liabilities and equity. To predict bankruptcy, we first employed the use of SMOTE to balance our train set for modelling, followed by Logistic Regression and correlation analysis to select the final 21 variables which were used in modelling. We employed the use of Multivariate Adaptive Regression Splines (MARS), Random Forest and Extreme Gradient Boosting (XGBoost) for prediction, and evaluated each model based on their individual strength. Eventually, these 3 models were all used for different aspects of our final solution.

While current measures primarily rely on peer-submitted content and are static approaches to predict bankruptcy, our solution – "Finance Protector" is a one-stop web application which integrates AI and machine learning to offer a more data-driven approach that automates risk detection, minimising reliance on subjective judgement. "Finance Protector" integrates all 3 machine learning models we have discussed for different usages, ensuring that our prediction is well rounded and offers different perspectives for different stakeholders. Stakeholders can gain access to model explainability tools and sector comparisons, all of which are backed by real-time algorithmic insights rather than unstructured peer opinions. Ultimately, this will empower stakeholders to make more informed decisions and avoid financial loss due to asymmetric information.

Our proposed recommendation considers a multifaceted approach in bankruptcy prediction, primarily benefitting financial stakeholders. As our solution serves all 3 stakeholders, we aim to promote data-driven decision making to build a more resilient finance industry. We consider a micro (national) approach initially, with potential future expansion globally to attract more beneficiaries. By leveraging on our solution, we can bring awareness to the significance and severity of corporate bankruptcy, enabling more research and development to go into this area, supporting Singapore's goals of building a resilient economy. Our potential global expansion explores the commercial value of Finance Protector for major overseas financial institutions such as Goldman Sachs, bringing positive impact to the world stage.

## 2. Background

### 2.1 Introduction to Corporate Bankruptcy

Corporate bankruptcy occurs when the market value of a company's asset falls below its debt obligations, limiting its ability in meeting its liabilities while increasing its risk of defaulting (Smart Towkays, 2023). Debt restructuring or asset liquidation may then be involved to help the company continue its operations and repay its creditors respectively (Gibson & Petrick, 2025). In Singapore, corporate bankruptcy is governed by the Insolvency, Restructuring and Dissolution Act 2018 (IRDA), which provides a legal framework for businesses experiencing financial distress (Ministry of Law, 2020).

Although corporate bankruptcy may seem complex, financial indicators like cash flows, interest payments and fixed costs can act as early warnings for impending financial distress, which are further explained in Appendix 1 (ICAEW, n.d.; Patritti, 2024). By monitoring these warning signs, stakeholders can take precautionary measures to mitigate the risks of corporate bankruptcy.

### 2.2 Prevalence of Corporate Bankruptcy

#### 2.2.1 Global Trends

Across the globe, there has been a rise in the number of corporate bankruptcies, with filings increasing by a staggering 33.5% (Knauth & Blackwell, 2025). Countries like the United States were also severely impacted, with bankruptcy filings totalling to 694 in 2024, an all-time high since 2010 (Lazzaro, 2025). With greater political instability and economic uncertainties, experts projected a 25% rise in the number of corporate insolvencies, highlighting a cause for concern (Kearney, 2025).

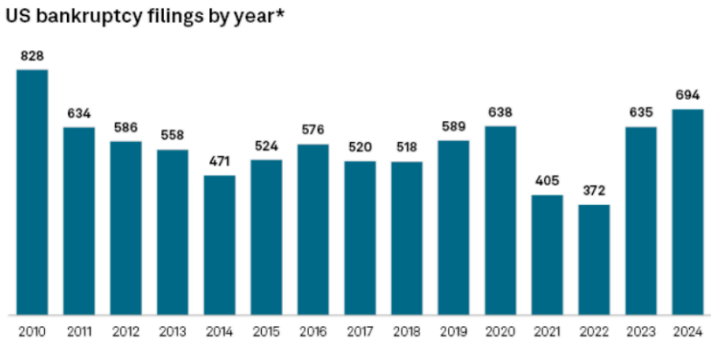


*Figure 1: Total Number of US Bankruptcy Filings by Year (Lazzaro, 2025)*

#### 2.2.2 Local Trends

Closer to home, figures released by the Ministry of Law's Insolvency Office revealed that a total of 307 companies were forced to wind up in 2024, a 52.7% increase from the previous year (Oh, 2025). In addition, companies in compulsory liquidation also rose, with the number of applications increasing by 6% from 257 in 2022 to 273 in 2023 (Peh, 2024). These alarming statistics show that even a renowned financial hub like Singapore is not immune to the threats of corporate bankruptcy, underscoring the urgent need to tackle these issues.

#### 2.2.3 Corporate Bankruptcies Across Industries

Corporate bankruptcy is a multi-faceted issue that spans across industries, ranging from financial institutions to entertainment companies. One of the most notable cases is the collapse of the Lehman

Brothers in 2008, which was attributed to excessive subprime mortgage lending, triggering a severe global recession that brought the economy to a standstill (Hernandez, 2023).

Beyond the financial sector, many other major corporations have also succumbed to bankruptcy. American department store chain J.C. Penney, Chilean airline LATAM Airlines and Chinese electric vehicle start-up WM Motor all filed for bankruptcy due to the economic turmoil caused by the COVID-19 pandemic (Cross, 2025; The Straits Times, 2024; Alfaro, 2020). The entertainment industry was also not spared, with Canadian entertainment giant Cirque de Soleil being forced into restructuring (Valinsky, 2020).

Therefore, these cases are stark reminders of how even the most established of companies can be driven into financial collapse, emphasising on the importance of having robust early warning systems and proactive monitoring frameworks to identify the red flags in an increasingly volatile economic landscape.

### 2.3 Causes of Corporate Bankruptcy

Most cases of corporate bankruptcy are found to stem from one or a combination of factors.

#### 2.3.1 Financial Mismanagement

Corporations often suffer from bankruptcy due to poor financial planning and cost monitoring (Patritti, 2024). Many firms tend to overleverage before realising their incapability of repaying the debt (Stapleton & Kenton). Fraudulent activities like embezzlement also reduces the availability of financial resources, worsening a firm's financial position.

#### 2.3.2 Unforeseen Events and Economic Downturns

Unforeseen events can also have detrimental impacts on businesses, triggering economic downturns and supply shocks that heighten their risk of bankruptcy (Patritti, 2024). These crises often reduce consumer demand and spending, causing firms to experience a sharp decline in their sales and revenues, crippling their ability to cover their financial obligations, and pushing them toward insolvency (Nalley, 2024).

### 2.4 Significance of Insufficient Information in Stock Markets

The devastating impacts of corporate bankruptcy are further magnified when there is a lack of market information as stakeholders can no longer anticipate risks.

#### 2.4.1 Investors

(a) Misjudgement of Stock Value

Insufficient stock market information may compel investors to rely on firms' financial statements which may be skewed, making it difficult for investors to identify firms with underperforming stocks and overestimate their intrinsic value instead (Wang, 2023).

(b) Irrational Herd Behaviours

The lack of limited information channels, coupled with investors' limited investing experience, may cause investors to abandon their own investment strategies and succumb to irrational herd behaviour that exacerbates stock price volatility and market inefficiencies (Ooi, 2020).

#### 2.4.2 Financial Institutions and Lenders

(a) Adverse Selection and Moral Hazard

Companies with higher likelihoods of loan defaults are more likely to source for external financing, with them using their secured capital in ways that misalign with the lenders' interests (World Bank Group, n.d.). Therefore, without sufficient information, financial institutions and lenders may struggle to detect these risks, increasing their susceptibility to defaults and contributing to market instability.

(b) Credit Rationing and Exclusion from Financial Markets

Credit rationing may also occur, whereby financial institutions and lenders are reluctant to extend credit to borrowers as they are overly cautious about the true financial health of companies (Calomiris et al., 2008). Thus, performing firms may be mistakenly excluded from financial assistance, stifling productive investments (Elemes & Filip, 2022). Hence, improving information flow can not only increase transparency but also accessibility to capital.

### 2.5 Problem Statement, Project Objectives and Targets

As discussed in Sections 2.2 and 2.4, with the incidence of corporate bankruptcy rising and its crippling impacts, the need for an accurate and robust prediction of corporate bankruptcy becomes more urgent than ever.

#### 2.5.1 Problem Statement and Objectives

Hence, we have come up with the following problem statement:

***"How can we develop a robust prediction and recommendation system to assist investors, lenders and financial institutions in predicting and navigating corporate bankruptcy?"***

Our project also aims to achieve the following objectives:

**(a) Conduct comparative and competitor analysis** of existing corporate bankruptcy prediction methods and the current implementations of prediction and recommendation systems to identify areas for improvement.

**(b) Develop enhanced predictive models** to help stakeholders predict corporate bankruptcy more accurately using a curated set of indicators.

**(c) Design a functional prototype** tailored for the various stakeholders such as investors, financial institutions and lenders. The prototype will (i) provide them with performance metrics like accuracy, precision and sensitivity and (ii) actionable insights to facilitate informed decision-making.

#### 2.5.2 Project Targets

Building on our problem statement, the following key questions serve to guide the development of a more reliable prediction and recommendation system:

**(a) Key Question 1: What are the significant financial variables that reliably indicates a company's bankruptcy riskiness?**

The selection of variables can be done using the variable importance analysis inherent in the different predictive models such as logistic regression. This process ensures that only relevant variables with high predictive power are included in our analysis, facilitating the implementation of a more accurate and efficient predictive system.

**(b) Key Question 2: What are the optimal thresholds for the aforementioned financial indicators to accurately identify at-risk companies, while minimising instances of false positives and false negatives?**

Models that can analyse non-linear interactions between financial indicators can be implemented, allowing for the identification of thresholds where bankruptcy risks vary. Moreover, hypertuning of parameters cross-validation can be done to minimise false positives and negatives. The models will then be evaluated to ensure that the predictions generated are accurate and reliable for stakeholders.

## 3. Market Analysis and Evaluation

In this section, we will evaluate current bankruptcy prediction models and discuss potential business opportunities.

### 3.1 Evaluation of Current Approaches

#### 3.1.1 Lack of Predictive Accuracy

There are many statistical models developed for corporate bankruptcy prediction, with prominent ones being Altman's Z-Score Model, the Springate Model and the Zmijewski Model (Vukčević et al., 2024). However, these models tend to rely on a limited number of financial ratios (Appendix 2), providing an incomplete view of a company's financial health (Radovanovic & Haas 2023). Additionally, each of the model has its own weaknesses, further undermining its predictive accuracy (Appendix 3). Hence, it is imperative to establish a predictive model that incorporates a broader range of financial metrics for a more nuanced and accurate assessment of a company's financial position.

#### 3.1.2 Lack of Model Diversity

Although many models were developed for bankruptcy prediction over the years, many of them are adaptations of the foundational Altman Z-Score model, limiting the dimensionality of insights that can be drawn (Vukčević et al., 2024). Furthermore, while machine learning techniques like neural networks display higher prediction accuracies, they remain underutilised, with traditional statistical methods dominating the field for bankruptcy prediction (Máté et al., 2023; Barboza et al., 2017). This highlights significant missed opportunities to harness the full predictive power of machine learning, which could enhance forecasting capabilities for stakeholders.

### 3.2 Business Opportunities

#### 3.2.1 Boosting Predictive Power Using a Variety of Models

To strengthen the predictive accuracy and reliability of corporate bankruptcy forecasts, we will machine learning and predictive models like MARS, Random Forest and Extreme Gradient Boosting (XGBoost). These models are selected for their ability to examine non-linear relationships, increasing the robustness of financial risk assessment while addressing the problem of their underutilisation.

#### 3.2.2 Empowering Decision Making

Building on our predictive framework, we can offer our solution of bankruptcy prediction as a Software-as-a-Service (SaaS) platform. Such an implementation can increase the accessibility of advanced prediction models to stakeholders by lowering implementation costs, a key deterrent to adopting these models (Kenge, 2020).

In addition, the platform can be equipped with AI-driven early warning mechanisms to alert stakeholders of imminent financial threats. Personalised recommendation engines can also be provided to help stakeholders formulate tailored strategies that align with their specific risk profiles and objectives, whether they are investors, financial institutions or lenders (Basu, 2021).

These features serve as extensions to current measures, which often lack timely alerts and actionable insights, leaving stakeholders to manually monitor markets in real-time and piece together fragmented sources of information when encountering undesirable market fluctuations, which is especially time-consuming (Buchholz, 2022). Therefore, by incorporating these features into our solution, we can empower stakeholders to make more informed decisions and better prepare them for financial risks.

## 4. Data Cleaning and Exploratory Data Analysis (EDA)

### 4.1 Introduction to Dataset

Our Kaggle dataset contains extensive financial data from the Taiwan Economic Journal for the years 1999 to 2009, with 6819 instances and 96 attributes. It includes various financial variables commonly found in financial statements, such as book price, liabilities and equity. Although this dataset is specific to the Taiwanese economy, these metrics are widely used across economies to assess corporate health and is thus, suitable for our use in developing an advanced predictive model.

### 4.2 Data Cleaning

#### 4.2.1 Missing Values and Duplicates

There are no missing nor duplicate values in this dataset.

#### 4.2.2 Outlier Handling

In our case, outliers could represent edge cases of bankruptcy. Thus, they are not removed as it may result in the loss of information crucial for the development for a model that is resilient against significant market fluctuations.

### 4.3 Exploratory Data Analysis (EDA)

We conducted exploratory data analysis to draw preliminary observations and highlight certain key findings from the dataset.

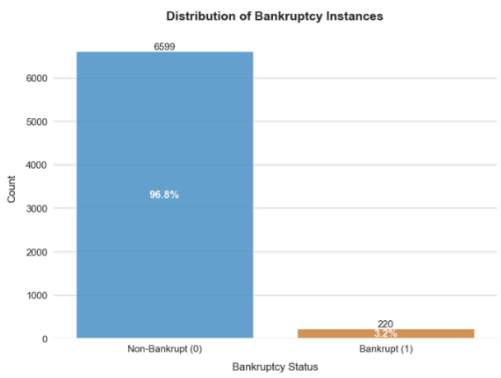


*Figure 2: Bar Chart Showing the Distribution of Bankruptcy Cases*

There are significantly more cases of non-bankrupt companies than bankrupt companies, creating a "class imbalance". Although this reflects the real-world scenario of bankruptcy being a rare event, this distribution could lead to the development of biased and unreliable models. To resolve this, data preprocessing using Synthetic Minority Over-Sampling Technique (SMOTE) and K-Nearest Neighbours (KNN) are needed.

Furthermore, we also investigated key differentiators between bankrupt and non-bankrupt firms using Standard Mean Difference (SMD) for a fairer comparison between the variables. The approach highlighted the importance of net worth and liabilities in predicting bankruptcy, with Net worth assets, Debt ratio and Net Value Per Share emerging as top variables (Figure 3).

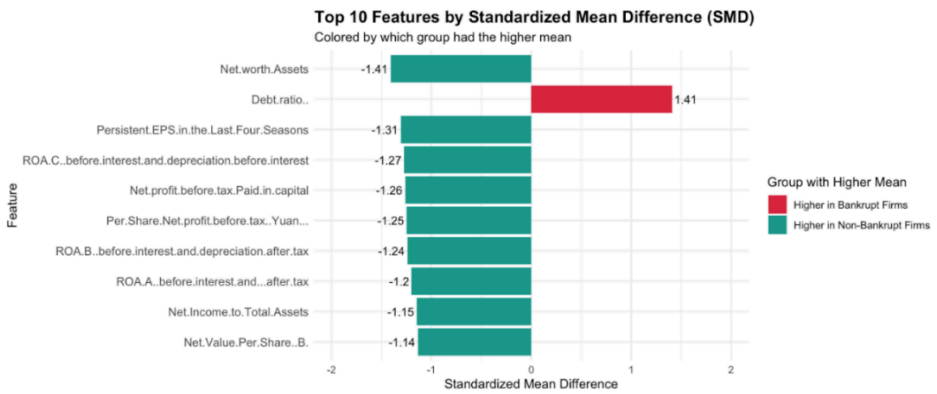


*Figure 3: Chart of top 10 features by SMD*

The chart displays the top 10 features ranked by Standardized Mean Difference (SMD), highlighting their ability to differentiate between bankrupt and non-bankrupt firms. **Debt Ratio** stands out as the only feature with a significantly higher mean in bankrupt firms, reinforcing its role as a key risk indicator. All other features—such as **Net Worth of Assets**, **Persistent EPS in the Last Four Seasons**, and various **profitability and ROA metrics**—have higher means in non-bankrupt firms, suggesting that stronger financial health and profitability are closely linked to solvency. The high negative SMD values for these variables indicate they are strongly associated with financial stability. This insight supports the conclusion that **low debt burden and solid asset-based performance** are reliable indicators of reduced bankruptcy risk.

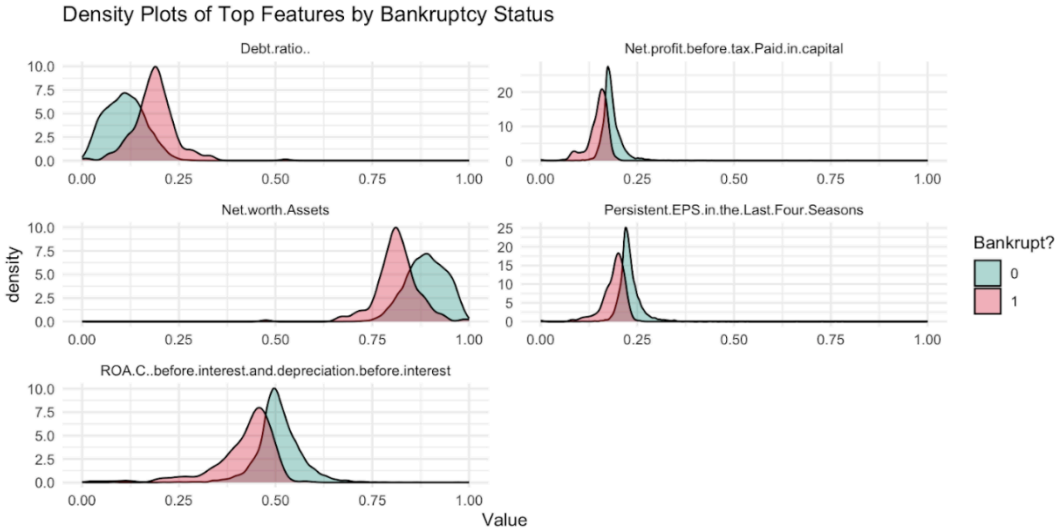


*Figure 4: Density Plots of top 5 Variables by SMD*

Density distributions of the top five variables were then plotted to assess their discriminatory power through analysing their overall spread, skewness, and shape of distributions. Zooming in on ROA vs Bankruptcy, non-bankrupt firms (in red) generally have higher ROA, with their distribution concentrated to the right of the bankrupt firms' curve (Figure 4 & Appendix 4). However, the overlap between the two curves is significant, and some bankrupt firms appear to have positive or even strong ROA values. While the graph shows a general trend of higher ROA correlating with survival, it also suggests that not all bankrupt firms are unprofitable on paper. Some may still appear healthy in terms of asset returns, indicating that profitability alone does not guarantee financial survival. This reinforces the idea that bankruptcy is often the result of a complex mix of factors.

Moreover, bankrupt companies (in blue) tend to have much higher debt-to-equity and borrowing dependency ratios than the median of non-bankrupt companies (in red), with some overleveraging till their interest obligations eat up all operating income (Figure 4 & Appendix 4). The distribution of the former is shifted to the right, peaking later and spanning a broader range, while the latter show a tighter cluster of lower debt ratios. This indicates that higher reliance on debt is more common among high-risk firms, reinforcing the idea that over-leveraging increases financial vulnerability. We can also see how visually separated the two distributions are, making debt ratio one of the clearest standalone indicators of bankruptcy risk.

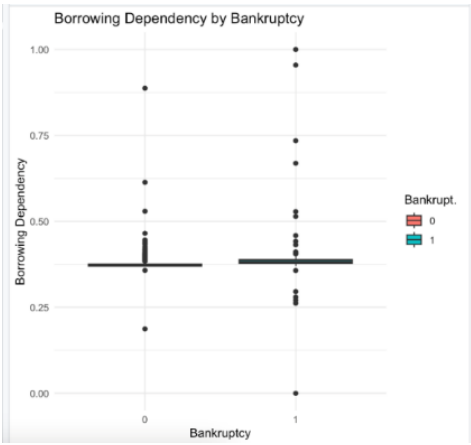

*Figure 5: Boxplot of Borrowing Dependency*

Bankrupt firms also tend to have a wider and higher spread of borrowing dependency compared to non-bankrupt firms (Figure 5). While the medians may be relatively close, the bankrupt group has a visibly broader interquartile range and more extreme upper values. This suggests that bankrupt firms are more reliant on external borrowing, and that their borrowing patterns are less consistent, with some firms being exceptionally over-leveraged. Non-bankrupt firms, in contrast, appear to maintain more moderate and stable borrowing behaviour. Interestingly, it highlights that borrowing dependency not just as a risk factor, but also as a source of instability. The inconsistency itself may reflect financial strain, risky decision-making or last resort attempts to stay afloat, all of which may precede bankruptcy.

### 4.4 Data Pre-Processing

#### 4.4.1 Oversampling using SMOTE

The dataset was stratified by bankruptcy outcome and split into training and test sets using a 70-30 ratio. To address the significant class imbalance in the dataset, we used SMOTE to generate synthetic samples, thereby balancing the class distribution and improving model performance. It was noted that certain data points may be generated in regions with sparse minority class representation, leading to potential noise generated.

#### 4.4.2 Noise Reduction Using KNN

To ensure the validity of all data points, we used KNN with k=5, meaning the 5 nearest neighbours, to compare the distances between each generated synthetic sample and the real minority class samples. Thereafter, we set a noise threshold of 95%, whereby synthetic samples whose distance to their 5th nearest neighbour exceeded this threshold were considered as “noisy” and thus removed from the dataset. This resulted in over 230 rows being removed and only high-quality synthetic data points being retained, minimising noise while ensuring that the dataset is more representative of the true distribution of the minority class.

## 5. Predictive Modelling

MARS, XGBoost and Random Forest were then used to build our bankruptcy prediction models and are thereafter evaluated.

### 5.1 Variable Selection using Logistic Regression and Correlation

Since 96 variables displayed high collinearity, logistic regression was used to shortlist more important variables for modelling. Using the inbuilt variable importance function, we declared an importance threshold of 0.4, which was around the median of the importances. Despite this, certain variables were still highly collinear, particularly for seemingly related variables. Hence, variable pairs in the correlation matrix with correlations above 0.8 were examined, with only those showing higher importance being retained. This ensures that each variable contributes distinctly to bankruptcy prediction while reducing overfitting, allowing the model to be generalisable to new unseen data. With this, the final dimension of our trainset is 9016 x 21 (Appendix 5).

### 5.2 Metrics Used

The metrics below are used in the evaluation of our models for bankruptcy prediction. Details of their calculations and their significance in the context of bankruptcy prediction are outlined in Appendix 6.

### 5.2.1 Sensitivity and Precision

In bankruptcy prediction, the costs of incorrect predictions are unequal and hence, certain evaluation metrics may be prioritised before others. For instance, the goal is to not only maximise accuracy but also ensure the correct identification of as many actual bankruptcies are possible. This makes sensitivity a critical performance metric – missing a bankrupt company (a false negative) could result in severe financial consequences for an organisation relying on such predictions for credit risk or investment decisions. High precision is also required, as it measures the accuracy of the model's prediction.

### 5.2.2 Balanced Accuracy

Balanced accuracy is used instead of the overall classification accuracy so that equal weights are given to both classes, which is useful for imbalanced datasets like our test set.

### 5.2.3 AUC-ROC Score

The Area Under the ROC curve (AUC) evaluates how well the model separates bankrupt from non-bankrupt firms across different classification thresholds. This is particularly important in bankruptcy prediction, where the decision threshold can be adjusted to balance false positives and negatives according to business requirements.

### 5.2.4 F1 Score

The F1 Score combines precision and sensitivity using their harmonic mean, providing a balance between the two. Since bankruptcy prediction is an imbalance classification problem, the F1 Score ensures that the model does not overly favour the majority class, with strong performance in recognising positive cases while minimising false positives and negatives, thus making it a suitable metric for use (Encord, 2023).

## 5.3 Model 1: Multivariate Adaptive Regression Splines (MARS)

MARS creates a piecewise linear model that captures non-linear interactions using multiple linear regression and can hence achieve good performance on challenging regression problems (Brownlee, 2021). While MARS is fundamentally a regression technique, it can be adapted to classification tasks by using the glm fit for categorical y.

For MARS, its performance changes as more terms are added. The vertical line of 18 indicates the optimal number terms retained based on 6 predictors. Beyond this point, any additional terms retained results in less than 0.001 improvement in GCV R2, suggesting that the inclusion of more terms does not enhance the model's performance and may instead cause overfitting (Appendix 7).

### 5.3.1 Feature Selection

MARS Model includes a backwards elimination feature selection routine that looks at reductions in the GCV estimate of error as each predictor is added to the model. We can also measure the change in residual sums of squares (RSS) as terms are added. However, there is little difference between the 2 methods used (Appendix 8).

Hence, a Partial Dependence Plot (PDP) is generated to visualise the relationship between the variables and Bankruptcy. Two top-ranked features, **ROA.A..before.interest.and...after.tax** and **Net.worth.Assets,** exhibit distinct, non-linear relationships with the following responses (Figure 6a):

- **Return on Assets (ROA):** Initially, changes in ROA have little impact, but beyond a value of approximately 0.5, there is a sharp decrease. Past 0.7, it plateaus and remains steady
- **Net worth Assets:** The predicted risk remains roughly the same until the feature value reaches about **0.75**, before decreasing sharply reflecting non-linear risk dynamics at extreme values.

An interaction plot between these two features further highlights their combined effect (Figure 6b). The interaction intensifies the model's response, indicating that the joint influence of ROA and Net Worth on bankruptcy is greater than their individual effects alone — a behaviour the MARS model is well-suited to capture due to its ability in handling of interactions and non-linearity.

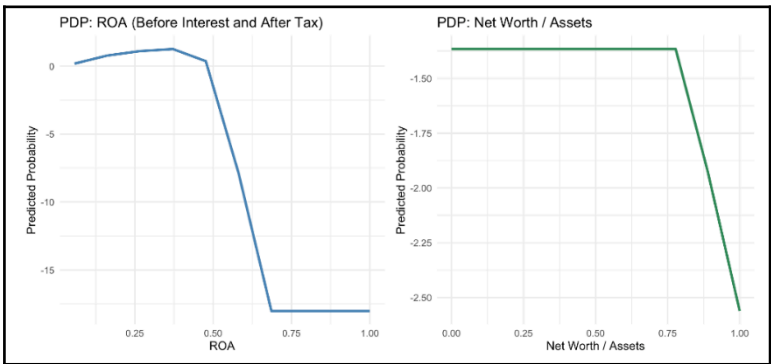


*Figure 6a: PDP for Examining Relationships Between Variables*

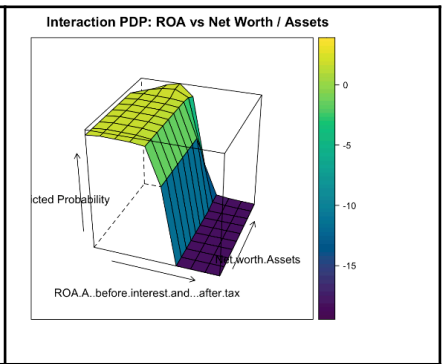


*Figure 6b: PDP for interaction between ROA and Net Worth*

### 5.3.2 Interaction Analysis using MARS

Variance importance only measures the impact of prediction error from the inclusion of features and neglects the impact of specific hinge functions for a particular feature. Hence, it does not provide us with insights on the possibly important non-linear relationships between variables. To resolve this, MARS was used to detect such complex and meaningful interactions between the different financial indicators and bankruptcy risk.

From the examination of hinge terms, there are several significant findings from the hinge interactions, that companies can use to determine their bankruptcy risk and identify the root cause, and they are as follows (Appendix 9):

**(a) Borrowing.dependency * Net.worth.Assets:** Firms with high borrowing dependency (> 37% - 39%) and low net worth face significant bankruptcy risk. The negative coefficient for one of the terms could indicate a subgroup where high borrowing with low net worth is less disastrous than expected, but such a combination is still risky, as the negative interaction may be counterbalancing the large positive effect where Borrowing.dependency > 0.37. This implies that lenders and regulators should scrutinise firms in this zone for higher risk of bankruptcy, flagging them for debt restructuring.

**(b) ROA * Net.Worth.Assets:** When both ROA and net worth relative to assets reaches critical thresholds (ROA: between 53% - 57%, Net Worth to Assets: 77%), the risk of bankruptcy increases, reflecting how weak profitability in companies erodes equity reserves and thus accelerates financial distress. Hence, investors should monitor these trends for early warning signs.

**(c) Net.worth.assets * Equity.to.liability:** High net worth and low equity to liability ratio is shown to decrease bankruptcy risks, as the valuable assets provide protection for the debt heavy firm. However, it may not be sustainable if firms are using asset sales to service their debt temporarily, highlighting the need for regulators to be mindful of debt heavy firms and ensure their solvency when needed.

**(d) ROA * Equity.to.Liability:** When ROA reaches a certain threshold (>0.43) while equity to liability ratio is low, the bankruptcy risk will still decrease as the high ROA could be sufficient to counter impact the risk introduced by low equity to liability ratio. However, creditors should still demand higher covenants for firms near these equity to liability thresholds, pushing for such firms to engage in turnaround strategies which address both profitability and debt reduction.

These findings capture the combined effects of financial health indicators, particularly highlighting the significance of ROA and net worth in determining bankruptcy risk, and how businesses with weak profitability and high liabilities may be more likely to face bankruptcy. Hence, the early identification of these tell-tale signs can better position stakeholders to take proactive financial interventions.

### 5.3.3 Comparison of MARS Models with Varying Complexity

We evaluated MARS classification models with different complexities ranging from Degrees 1 to 3 (Degree 1: No Interaction, Degree 2: Pairwise Interactions, Degree 3: 3-Way Interactions) by testing them on a common test dataset and analysing the metrics mentioned in 5.2.

| Model | Balanced Accuracy | Precision | Sensitivity | F1 | AUC-ROC |
|---|---|---|---|---|---|
| **Degree1** | 0.8633 | 0.1738 | 0.8636 | 0.2893 | 0.9221 |
| **Degree2** | 0.8416 | 0.2008 | 0.7879 | 0.3200 | 0.9052 |
| **Degree3** | 0.8043 | 0.1865 | 0.7121 | 0.2956 | 0.9049 |

*Table 1: Evaluation of MARS Model*

All three models exhibit high overall accuracy and strong AUC-ROC values, indicating good discriminatory ability. However, **precision** is consistently low across the models due to the difficulty in correctly identifying the minority class, which in turn pulls down the **F1 score**.

The **Degree 2** model strikes the best balance, achieving the highest **Precision**, and **F1 Score**, making it the most well-rounded in terms of general performance. On the other hand, **Degree 1** leads in **Balanced Accuracy**, **Recall** and **AUC-ROC**, making it better at detecting actual bankruptcies but at the cost of significantly more false positives.

Since the task is to **predict bankruptcy**, and false positives and false negatives both carry significant implications, the **F1 score**—which balances precision and sensitivity—is particularly important. Therefore, the **Degree 2** model is considered the most suitable among the three as the best compromise between catching bankruptcies and minimizing false alarms

### 5.3.4 Optimising via Tuning of Parameters

There are 2 important tuning parameters associated with the MARS model: The **maximum degree of interactions** and the **number of terms retained** in the final model. We can perform a grid search to identify the optimal combination of these hyperparameters that minimize prediction error.

To further optimise our model, we used Caret to perform a Grid Search using 10-fold CV. The model that provides the optimal combination includes second degree interactions effects and retains 23 terms. The Cross-Validated F1 score for these models are displayed below. The Optimal model's cross validated F1 score was 0.928495, as shown in Figure 7.

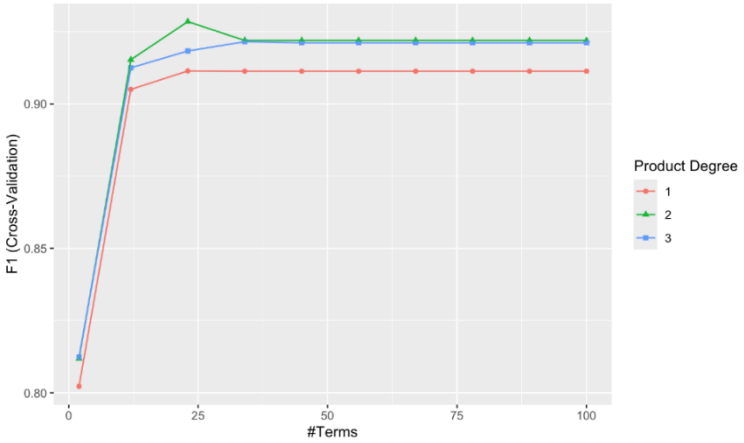


*Figure 7: Optimisation of MARS model*

The optimal MARS model of second-degree interactions (Degree 2) effect and retains 23 terms (nprune = 23). When evaluating on the test set, the model yields the following performance (Appendix 9).

| Metrics | True Positives (TP) | False Positives (FP) | True Negatives (TN) | False Negatives (FN) |
|---|---|---|---|---|
| **Values** | 48 | 182 | 1797 | 18 |

*Table 2: Summary of Confusion Matrix for MARS (Appendix 10)*

| | Balanced Accuracy | Precision | Sensitivity | F1 | AUC-ROC |
|---|---|---|---|---|---|
| **Tuned** | 0.8177 | 0.2087 | 0.7273 | 0.3243 | 0.9066 |

*Table 3: Evaluation of Optimum MARS Model*

When comparing the default MARS model (Degree 1, 2 and 3), our Optimal model shows improvements in Balanced Accuracy, Precision and F1 scores. However, it underperforms in terms of Recall and AUC-ROC. This trade-off is expected, as the tuning process prioritises F1 score. In contexts like bankruptcy prediction, where both types of errors carry consequences, this balanced approach aligns well with real-world objectives. Thus, the tuned model demonstrates superior practical utility and statistical performance over its default counterparts.

### 5.4 Model 2: Random Forest

Random Forest is an ensemble machine learning method that constructs multiple decision trees during training, merges their predictions through majority voting and reduces variance using bootstrap aggregation. Given these capabilities, it makes the model ideal for problems like bankruptcy prediction, where data is high-dimensional and often imbalanced.

### 5.4.1 Model Optimisation

For the Random Forest model, we employed a two-pronged optimisation approach: hyperparameter tuning and threshold adjustment.

Firstly, we optimised the model's hyperparameters using grid search to tune the mtry hyperparameter from 2 to 20 at each decision tree split, as its default value may be undesirable for our dataset. We then trained the model on the imbalanced cleaned dataset, before assessing model performance at each level using 5-fold CV, repeated thrice for reliability, before evaluating it using AUC-ROC. The configuration with the mtry value that yielded the highest average AUC score was then used to retrain the final model with 500 decision trees (ntree = 500), ensuring a robust ensemble.

However, the Random Forest only recognises a positive instance when its predicted probability exceeds 0.5. Hence tuning hyperparameters alone does not fully address the limitations imposed by class imbalance, as it tends to favour the majority non-bankrupt class in imbalanced datasets, leading to poor sensitivity. Thus, threshold optimisation was conducted by analysing the tuned model's Receiver Operating Characteristic (ROC) curve to find the ideal trade-off between true positives and false positives, identifying 0.249 as the optimal probability threshold. This significantly increased the model's ability to identify bankrupt companies from 51.52% to 86.36%, which is crucial in financial applications where the cost of a missed bankruptcy is far greater than the cost of a false alert.

Therefore, by combining hyperparameter tuning with custom threshold calibration, we were able to significantly enhance the Random Forest model's practical effectiveness for bankruptcy prediction.

| **Metrics** | **True Positives (TP)** | **False Positives (FP)** | **True Negatives (TN)** | **False Negatives (FN)** |
|---|---|---|---|---|
| **Values** | 57 | 220 | 1759 | 9 |

*Table 4: Summary of Confusion Matrix from Random Forest (Appendix 10)*

The confusion matrix confirms that the tuned Random Forest model is highly effective at detecting bankruptcies with minimal disruption to healthy companies. With **57 correct bankruptcy predictions**, the model captures **23 more bankruptcies** than the default model (Appendix 9). While it produces more false positives (i.e., healthy companies being flagged), the cost of these is far outweighed by the value of identifying truly distressed firms early, making the tuned model particularly well-suited for investors, financial institutions and lenders.

### 5.4.2 ROC Curve Analysis

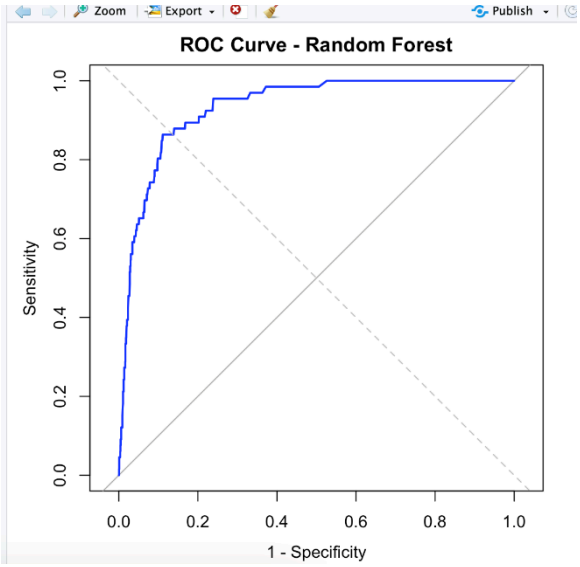


*Figure 8: ROC Curve for Random Forest*

Figure 8 illustrates the trade-off between sensitivity and specificity at various thresholds. The model demonstrates strong classification performance, with the curve bending sharply towards the top-left corner - an indicator of high true positive rates with low false positive rates. The AUC-ROC score of **0.9324** confirms the model's ability to distinguish between bankrupt and non-bankrupt companies across all thresholds, justifying our use of a custom threshold for optimal sensitivity in bankruptcy prediction.

### 5.4.3 Evaluation of Tuned and Default Models

| Model | Balanced Accuracy | Sensitivity | Precision | AUC-ROC | F1-Score |
|---|---|---|---|---|---|
| **Tuned** | 0.8762 | 0.8636 | 0.2058 | 0.9324 | 0.3324 |
| **Default** | 0.7427 | 0.5152 | 0.3656 | 0.9278 | 0.4277 |

*Table 5: Evaluation of Tuned and Default Random Forest Models*

While the default model shows better precision (36.56%) compared to the tuned model (20.58%), this only means that it is more often accurate when predicting bankruptcy. However, this comes at a significant cost - its sensitivity is much lower at 51.52%, meaning it misses the detection of nearly half of all actual bankruptcies. In high-stakes financial contexts, such missed cases can result in severe consequences, making this trade-off unacceptable.

The tuned Random Forest model, on the other hand, shows better overall performance: it achieves a higher sensitivity of 86.36%, identifying a much higher proportion of at-risk companies and outperforms the default in terms of balanced accuracy (87.62% vs. 74.27%) and AUC-ROC (93.24% vs. 92.78%). Even though its precision is lower (20.588% vs 36.56%), meaning it raises more false alarms, this is often a tolerable trade-off in financial contexts - it is better to manually review a few healthy companies flagged by mistake than to miss a truly failing one.

Hence, the tuned Random Forest Model becomes significantly more aggressive in flagging potential bankruptcies, with a lower F1 Score (33.24% vs 42.77%), to represent the trade-off made to optimise for higher sensitivity. This allows the model to better align with risk management and financial forecasting, whose goal is to prioritise minimising false negatives by focusing on identifying as many bankruptcies as possible, even if that results in a few more false alarms. Therefore, the tuned Random Forest model is a more suitable model to be employed in bankruptcy prediction.

### 5.5 Model 3: Extreme Gradient Boosting (XGBoost)

XGBoost is a supervised leaning boosting algorithm that makes use of gradient boosting for its sequential tree construction (Kavlakoglu & Russi, 2024). This iterative process makes XGBoost more precise in capturing complex patterns in this dataset. The model was also trained using the default hyperparameters.

#### 5.5.1 Model Optimisation

Similar to optimisation procedures of Random Forest mentioned in Section 5.4.1, hyperparameters were tuned using GridSearchCV, before extracting their optimal combinations (Figure 9).

```
Best Parameters: {'colsample_bytree': 1, 'learning_rate': 0.2, 'max_depth': 7, 'n_estimators': 300, 'subsample': 0.8}
```

*Figure 9: Optimal combinations of hyperparameters used for XGBoost*

The optimal threshold value of 0.0004 was also obtained from the ROC curve, with the threshold performing better than the default model in predicting bankrupt and non-bankrupt cases and boasting an impressive increase in sensitivity.

#### 5.5.2 Evaluation of Tuned and Default Models

| Metrics | True Positives (TP) | False Positives (FP) | True Negatives (TN) | False Negatives (FN) |
|---|---|---|---|---|
| **Values** | 53 | 247 | 1732 | 13 |

*Table 6: Summary of Confusion Matrix from XGBoost (Appendix 10)*

| Model | Balanced Accuracy | Sensitivity | Precision | AUC-ROC | F1-Score |
|---|---|---|---|---|---|
| **Default** | 0.7351 | 0.5000 | 0.3587 | 0.9350 | 0.4177 |
| **Tuned** | 0.8391 | 0.8030 | 0.1767 | 0.9282 | 0.2896 |

*Table 7: Evaluation of tuned and default models of XGBoost*

Similarly, the results of the 2 models show the trade-off between precision and sensitivity in bankruptcy prediction. While the default model achieves a substantially higher precision (35.87% vs 17.67%), AUC-ROC (93.50% vs 92.82%) and F1 score (41.77% vs 28.96%), the tuned model demonstrates substantially improved performance in detecting bankruptcy cases, as shown by its significantly higher sensitivity (80.30% vs 50.00%) and hence its lower false negative rate of 20%. In addition, it also had a higher balanced accuracy (83.91% vs 73.51%). This means that the tuned model correctly identifies 30.30% more actual bankruptcy cases while reducing missed bankruptcy cases by the same extent, despite classifying more solvent companies as risky. Although the F1 Score of the

tuned model is significantly lower (28.96% vs 41.77%), this trade off was made due to the higher sensitivity and balanced accuracy, which are the prioritised metrics in this case.

As missing true positives is costlier than false alarms in the business context, the tuned model's superior sensitivity outweighs its minor drawbacks in precision. The 80.30% detection rate represents a significant 30.30% improvement over the default model which could potentially prevent financial losses for its ability to minimise false negatives when predicting bankruptcy.

### 5.5.3 Insights from Shapley Additive Explanations (SHAP) Analysis

The SHAP analysis was also used to quantify the contributions of variables in XGBoost (Appendix 11), with firms displaying a high reliance on borrowing found to experience greater bankruptcy risk, highlighting the need for stakeholders to closely monitor their loans to reduce solvency risks. Moreover, variables like ROA, Net Worth and Total Asset Turnover served as asset-related protective variables, underscoring the importance of high profitability and strong equity buffers in reducing bankruptcy risk. SHAP also includes interaction insights, with high borrowing dependency found to negatively affect low ROA firms most, highlighting the influence of liabilities. Overall, SHAP values can offer our stakeholders a deeper understanding of how specific variables influence bankruptcy risk, allowing them to adopt more targeted solutions.

## 5.6 Final Evaluation of Tuned Models

| Model | Balanced Accuracy | Sensitivity | Precision | AUC-ROC | F1-Score |
|---|---|---|---|---|---|
| **MARS** | 0.8177 | 0.7273 | 0.2087 | 0.9066 | 0.3243 |
| **Random Forest** | 0.8762 | 0.8636 | 0.2058 | 0.9324 | 0.3324 |
| **XGBoost** | 0.8391 | 0.8030 | 0.1767 | 0.9282 | 0.2896 |

*Table 8: Evaluation of all tuned models*

### 5.6.1 Model Performance

Random Forest is the best performing model, outperforming all other models in all metrics except for precision, which it lost narrowly to MARS by 0.29% (Table 8). Its higher sensitivity indicates that it correctly detects 86.36% of bankruptcies, and higher balanced accuracy of 87.26% implies that it has the most optimal predictions. Moreover, the higher AUC-ROC and F1 Scores show its ability to rank positive instances higher than negative ones, making it desirable for bankruptcy prediction. Although the precision scores for all models are unfavourable, a trade-off between sensitivity and precision is preferred to maximise the number of accurate bankruptcy diagnosis.

### 5.6.2 Model Explainability

Model explainability is imperative to bankruptcy predictions as it affects the ability of stakeholders to make informed decisions based on the data gathered. Among the three models, MARS, equipped with interaction analysis, stands out as the most explainable. Its transparent thresholds and interaction rules simplify complex interactions, offering stakeholders clear insights of how different factors influence bankruptcy risk. Furthermore, its ability to conduct hypothetical analysis whereby users can manipulate data inputs and project financial outcomes reduces business uncertainty, making it an effective tool for decision-making. XGBoost, despite being less intuitive than MARS, can also be interpreted by stakeholders using SHAP analysis, allowing them to understand the relative importance

of factors. Random Forest emerges as the least user-friendly due to its nature of being a "black-box model".

#### 5.6.3 Individual Model Usage

All 3 models can complement one another by serving different purposes in our solutions. MAR's high explainability and transparency makes it ideal to report threshold values and generate actionable insights, while XGBoost's fast prediction time can be harnessed for quick, rough predictions on live data. Lastly, Random Forest's high sensitivity can be used to increase accuracy in risk analytics and bankruptcy prediction. These ideas will be further elaborated on in Section 6.

## 6. Solution

### 6.1 Risk Threshold Model (MARS)

This model aims to help stakeholders understand when and why a firm becomes financially distressed by assessing bankruptcy variable thresholds.

#### 6.1.1 Practical Implications

This model helps investors pinpoint when exactly a firm becomes financially risky, allowing them to set threshold filters to avoid investing in firms with high default risk, which is especially beneficial for investors' decision making in volatile markets. This, coupled with the model's customisable rules, can serve as an edge during ESG investing, whereby investors often avoiding complex models that lack clarity (Jones et al., 2012). This also means that the model allows for "what-if" analysis, enabling investors to test different scenarios and see how bankruptcy risk will change. Lastly, with its interaction analysis, investors can be aware of how different combination of financial indicators like profitability-solvency and liquidity-efficiency interact with bankruptcy risk.

For lenders, especially those in Small and Medium Enterprises (SME) or retail banking, the model can be used to screen borrowers based on financial thresholds linked to historical defaults. This allows loan applications to be triaged more efficiently, speeding up the approval process and reducing human bias. Besides, since the model produces rule-based outputs, credit analysts can clearly justify their lending decisions to clients. This would greatly benefit SMEs with inconsistent financial records, as the model handles missing or irregular data relatively well.

Meanwhile, financial institutions like regulators or banks can use the model to define risk thresholds across entire sectors by identifying which financial indicators highlight potential distress, such as declining cash flow or rising short-term debt. These would become valuable insights during times of economic uncertainty, enabling proactive policy-making decisions like tightening credit standards or increasing capital buffers before risks escalate. Using the available interaction data, financial institutions can take note on interactions which will contribute to bankruptcy prediction and ensure they do not reach the variable threshold, on variables mainly associated with debt and liability.

#### 6.1.2 Feasibility of Implementation

MARS is highly feasible for stakeholders that require quick and actionable insights, such as SME lenders or investors who often need to assess bankruptcy risk with limited resources. Since the model operates on rule-based thresholds, it is easy to integrate it into existing risk management systems without requiring significant technical resources. Besides, its reliance on straightforward decision

rules based on financial thresholds means it can be deployed relatively fast with minimal training data, making it a cost-effective solution for stakeholders to streamline risk assessments.

### 6.2 Bankruptcy Risk Classifier (XGBoost)

This model provides accurate, real-time bankruptcy prediction from multiple variables, especially for stakeholders who require precise forecasting in high-risk environments.

#### 6.2.1 Practical Implications

For investors managing large portfolios, this model can flag firms with an increasing risk of bankruptcy, even when their profitability still looks strong. It can detect early warning signs, such as unusual patterns in cash flow volatility that precede market awareness. With these signals from the onset, investors can act before the market reacts and reduce potential losses. Investors can also use SHAP values to rank portfolio companies by risk, providing them with a quantifiable feature in diversifying their portfolio.

Lenders can integrate this model into their real-time credit scoring systems to improve credit assessments at scale. By identifying variable interactions, the model supports more accurate decisions. Lenders can instantly adjust loan terms and interest rates based on each borrower's predicted risk level. This improves both the speed and quality of credit approvals, especially in high-volume lending environments. Lenders can also leverage on SHAP values to understand which specific variables they have to take note of before lending money to a firm.

For financial institutions, the model serves as a scalable monitoring tool by scanning thousands of firms for signs of financial weakness. It allows for early interventions by helping to identify systemic risk areas across industries. As the model can be retrained quickly in response to macroeconomic changes, risk assessments are kept current and relevant. While the model uses advanced algorithms, its results can be visualised, which makes it easier to share insights with board members or regulatory partners. This helps to align risk governance with real-time market trends.

#### 6.2.2 Feasibility of Implementation

XGBoost is ideal for large financial institutions such as hedge funds and investment banks that have access to high-performance computing resources. These institutions benefit from the model's ability to analyse complex financial data in real-time and provide precise bankruptcy predictions. The model can be integrated such that stakeholders receive up-to-date predictions, and they can adjust their strategies accordingly. This means that even in volatile market conditions, stakeholders can exercise long-term strategic decision-making.

### 6.3 Risk Pattern Aggregator (Random Forest)

This model helps stakeholders identify hidden bankruptcy risk patterns based on historical data.

#### 6.3.1 Practical Implications

For investors, this model groups firms based on similar characteristics to those that previously defaulted. Even if a specific firm has not shown clear warning signs yet, its similarity to high-risk clusters enables investors to proactively investigate and adjust their positions before problems surface.

Lenders can leverage Random Forest to automate the pre-screening of borrowers, even when financial data is missing or incomplete. The model is reliable and can segment applicants into risk tiers and tailor loan products accordingly. For instance, flexible repayment plans can be given to borderline-risk

borrowers, while stricter terms can be applied to high-risk ones. This makes it easier for firms to expand credit access without sacrificing risk standards.

Financial institutions can use Random Forest as part of their risk segmentation and monitoring processes. It provides feature importance scores, which can be used to detect which variables are becoming more important in predicting bankruptcy. These insights can guide policy updates or changes in lending incentives. Overall, Random Forest's ease of use and flexibility make it a good fit for risk flagging in day-to-day operations.

### 6.3.2 Feasibility of Implementation

Random Forest is highly feasible for stakeholders seeking a flexible and reliable risk analysis tool across diverse business contexts. This model can be easily adopted by both small and large-scale financial institutions, as it does not require a high degree of technical expertise to implement effectively. Moreover, its low implementation cost and scalability make it a suitable choice for institutions looking to enhance their credit risk assessments without investing heavily in complicated systems. Random Forest also boasts the highest sensitivity among the three models, enabling it to be the most accurate in catching bankruptcies.

## 6.4 Model Integration via Web Application – Finance Protector

*Figure 10: Prototype of the "Finance Protector" Interface for Investors*

We have come up with a unified web-based recommendation system called "Finance Protector", serving as a one-stop platform for investors, lenders and financial institutions (Figure 10). This comprehensive system aids them in navigating corporate bankruptcy risks by combining risk thresholds (MARS), real-time predictive insights (XGBoost) and pattern-based firm segmentation (Random Forest). With the addition of AI-driven early warning mechanisms, the platform alerts stakeholders to imminent financial threats and provides additional indicators to watch out for. We aim to implement this solution locally, with future expansion opportunities targeting global expansion.

### 6.4.1 Key Differentiators of Solution

While platforms like SumZero and Value Investors Club foster valuable idea-sharing within a professional network, they primarily rely on peer-submitted content and lack model-driven analytics (Goldstein et al., 2024). Our solution builds on this by offering a more data-driven approach that automates risk detection, minimising reliance on subjective judgement. Stakeholders can gain access to model explainability tools and sector comparisons, all of which are backed by real-time algorithmic insights rather than unstructured peer opinions. Ultimately, this will empower stakeholders to make more informed decisions.

With the precision and scalability of advanced analytics, our web application offers a reliable and efficient outlet for stakeholders to assess their financial health and manage bankruptcy risks effectively.

## 6.5 Benefits, Expected Outcomes, Limitations & Future Development

### 6.5.1 Benefits

Unlike generic solutions, our models provide insights tailored according to the risk profiles of our stakeholders. This allows investors, lenders and financial institutions to devise targeted preventive strategies that are responsive to prevailing market conditions. By identifying signals that matter most to each stakeholder, our models empower users to make decision that align with their objectives and constraints.

### 6.5.2 Expected Outcomes

With an accurate prediction model, lenders will be able to identify risky companies early and avoid doing business with them, decreasing loan defaults by a good amount. Furthermore, this also solves the problem of exclusion due to insufficient data mentioned in Section 2.4.2. For financial institutions, they will be able avoid high-risk firms and hence improve their financial resilience as a result. Investors will also have the required information to divest from firms showing high bankruptcy risk early and engage in diversification measures to protect their portfolios. Overall, this model provides sufficient information to the various stakeholders, which allows them to make informed financial decisions.

### 6.5.3 Limitations

However, each model also has its limitations. MARS relies on predefined, static thresholds that need manual updates, which limits its adaptability to sudden economic changes. XGBoost requires significant computational resources and can be challenging for non-technical stakeholders to interpret. Random Forest may be prone to overfitting, and its decision trees can be difficult to interpret for those seeking deeper insights into the model's behaviour.

### 6.5.4 Future Developments and Expansions

To improve these models' performance, future developments will focus on the following:

(i) Integrating real-time updates and economic indicators to improve accuracy in volatile markets

(ii) Incorporating more dynamic and individualised features to increase the practicality and adaptability of the models for different stakeholders and market scenarios

(iii) Improving model interpretability by integrating advanced visualisation tools, enabling stakeholders to better understand and act on predictions

(iv) Refining the models to reduce false positives, giving more reliable risk assessments and minimising unnecessary interventions

(v) Strengthen the scalability of the web application so that it can quickly respond to market changes and business-specific needs

Major financial institutions like HSBC, Goldman Sachs and JPMorgan Chase & Co., can greatly benefit from these improvements by refining their risk management frameworks and credit scoring systems, leading to optimal decision-making in the ever-changing market. For SME lenders like Kabbage, the models act as a valuable tool to streamline loan approval processes, thereby reducing the risk of defaults. Finally, hedge funds like Blackrock can leverage these models to enhance their investment strategies, enabling them to adapt swiftly to market fluctuations and minimise their exposure to potential bankruptcy risks. By embedding these models into their operations, these stakeholders will be better positioned to navigate bankruptcy risks effectively while solidifying their competitive edge within the financial landscape. This also highlights a potential macro global expansion opportunity, allowing major financial institutions all over the world to be beneficiaries of Finance Protector.

In addition, we can also include important insights for companies themselves. Using the three models, our solution offers actionable insights for individual companies to monitor their real-time bankruptcy risk and take proactive and preventative measures to mitigate this risk. This allows companies to avoid collapse, optimise capital and improve financial planning for years to come.

A summary of the benefits, limitations and future improvements of each model can be found in Appendix 12.

### 6.6 Risk Assessments and Mitigation Strategies

This section aims to discuss about the different risks and mitigations for our web-based application.

#### 6.6.1 Data Handling and Compliance Risks

Since sensitive financial data is made available on the web application, this may raise concerns related to data privacy and regulatory compliance. Mishandling such data could lead to legal repercussions, damaging user confidence. To resolve this, data governance protocols like encryption and anonymisation will be established, with stakeholders being required to operate in a sandbox environment for data testing.

#### 6.6.2 Over-reliance on Technological Implementations

Stakeholders may be excessively dependent on automated tools, like our AI warning system, which may result in them neglecting contextual nuances not presented in financials, such as macroeconomic shifts. Hence, we will explicitly position the models as decision-support tools and incorporate human-in-the-loop features like news articles to ensure stakeholder decisions remain well-rounded.

By addressing these risks through continuous improvements, investors, lenders and financial institutions can remain resilient and better manage financial instability during periods of uncertainty.

## 7. Conclusion

Corporate Bankruptcy remains a significant financial issue for companies and stakeholders alike. Its detrimental and serious spillover effects on industries highlight the utmost importance of early detection. By identifying at-risk companies in time, we can ensure that they receive the necessary

guidance and regulatory support, while stakeholders can avoid investing or incurring losses on such companies.

Our solution provides a robust, data-driven approach to assess the risk of corporate bankruptcy for various companies in a manner that is quick, assessable and accurate. We leverage the 3 models – MARS, XGBoost and Random Forest – each serving a distinct purpose, with tailored interfaces designed for all 3 target stakeholders. This approach ensures that each stakeholder receives the most relevant insights based on their financial motivations, and the combination of the models allows us to address different aspects of corporate bankruptcy risks more comprehensively.

In essence, our solution integrates a proactive approach to detecting bankruptcy risks. Moving forward, we believe our solution can serve as a stepping stone to building a more financially resilient industry, one where risks are better anticipated, companies can be guided towards recovery and sustainability, and stakeholders are equipped to make informed, strategic decisions in competitive economic environment.

## 8. Appendix

### Appendix 1: Explanation of Key Financial Indicators Signalling Corporate Bankruptcy

Many financial indicators warn individuals of corporate bankruptcy. These indicators include decreasing cash flows, excessive interest payments and the burden of high fixed costs, all of which can be measured by the different ratios present in our dataset, as explained below.

**A. Decreasing Cash Flows**

Cash flows reflect a company's ability to generate sufficient cash to cover its operating expenses, debt obligations and financial needs. It is typically measured by ratios such as the Operating Cash Flow to Sales Ratio.

$$Operating\ Cash\ Flow\ to\ Sales\ Ratio = \frac{Operating\ Cash\ Flow}{Net\ Sales}$$

A lower Operating Cash Flow to Sales Ratio highlights a company's inefficiency in generating sales into cash, possibly due to issues like cash flow management or the need for working capital.

The Cash Flow to Liability Ratio is also another metric used for analysing cash flows.

$$Cash\ Flow\ to\ Liability\ Ratio = \frac{Operating\ Cash\ Flow}{Total\ Liabilites}$$

A lower Cash Flow to Liability Ratio suggests that the company may take a longer time to cover its total liabilities, assuming that all of the cash flow was dedicated to liability payment.

Hence, when a company is experiencing negative or reduced cash flows, it implies that either the company is not generating enough funds from its provision of goods and services or that the business is spending more than what it earns, both of which may lead to difficulties in repaying liabilities.

**B. Excessive Interest Payments**

When evaluating interest payments, the Interest Coverage Ratio (ICR) is often used to compare a company's ability to cover its interest expense with its operating income.

$$Interest\ Coverage\ Ratio = \frac{EBIT}{Interest\ Expense}$$

A lower ICR of less than 1 indicates that the company may be struggling to repay its debts, shedding light on the company's poor financial health, while an ICR of below 2 or 3 still highlights a company's riskiness in not meeting its debt obligations.

**C. High Fixed Costs**

Fixed costs are costs incurred by the firm that does not vary with the level of production. This means that these costs remain constant for most of the time and must be paid regardless of the firm's performance. Thus, if the firm perceives the fixed costs to be high, it could mean that they are experiencing reduced revenue, highlighting potential financial strain. The Fixed Cost Coverage Ratio is commonly used to analyse how sensitive a firm's operating income is to changes in sales, which is influenced by the level of fixed costs.

$$Degree\ of\ Financial\ Leverage\ (DOL) = \frac{Percentage\ Change\ in\ Operating\ Income}{Percentage\ Change\ in\ Sales}$$

Therefore, a firm with a lower DOL means that it has a higher proportion of fixed costs, making it more sensitive to changes in sales.

## Appendix 2: Models Used in Bankruptcy Prediction Models

**(a) Altman's Z-Score Model**

The model uses only five key financial ratios as its inputs uses them in a discriminate equation (Rashid et al., 2024).

The Z-score is calculated using the following formula:

$$Z\ Score = 1.2\ X1 + 1.4\ X2 + 3.3\ X3 + 0.6\ X4 + 1.0\ X5$$

where

X1 = Working Capital/Total Assets

X2 = Retained Earnings/Total Assets

X3 = Earnings Before Interest and Taxes/Total Assets

X4 = Market Value Equity/Book Value of Total Debt

X5 = Sales/Total Assets

- If the value of Z < 1.11 = Zone "distress"
    - The company is experiencing financial difficulties with a high likelihood of going bankrupt.
- If the value of 1.11 < Z < 2.6 = "gray" zone
    - The company is suffering from bad finanical health and requires management. If no timely intervention is administered, the company could go bankrupt. On the flip side, if effective measures are implemented, the company can suc successfully navigate through its financial difficulties.
- If the value of Z > 2.6 = "safe" zone
    - The company is financially healthy with a low likelihood of bankruptcy.

**(b) The Springate Model**

The model relies on four financial ratios which are used in a Multiple Discriminant Analysis (Rahayu, 2017).

The S-score is calculated using the following formula:

$$S\ Score = 1.03\ X1 + 3.07\ X2 + 0.66\ X3 + 0.4X4$$

where

X1 =Working Capital/Total Assets

X2 =Earnings before interest and taxes (EBIT) / Total assets

X3 =Earnings before tax (EBT) / Current Debt

X4 =Sales/Total Assets

- If the value of S < 0.862 = "distress" zone
    - The company is experiencing financial difficulties with a high likelihood of going bankrupt
- If the value is 0.862 < S < 1.062 = "grey" zone
    - The company is suffering from bad finanical health and requires management. If no timely intervention is administered, the company could go bankrupt. On the flip side, if effective measures are implemented, the company can successfully navigate through its financial difficulties.
- If the value of S > 1.062 = "safe" zone
    - The company is financially healthy with a low likelihood of bankruptcy.

**(c) The Zmijewski Model**

The model uses three ratios as inputs in probit regression to evaluate the company's performance, leverage, and liquidity. For this model, more emphasis is placed on the significance of the level of debt as the most influential factor in bankruptcy (Marsenne et al. 2023).

The X-score is calculated using the following formula:

$$X\ Score = -4.3 - 4.5\ X1 + 5.7\ X2 - 0.004\ X3$$

where

X1 = Earning after Taxes/Total Assets (Profitability)

X2 =Total Debt/Total Assets (Leverage)

X3 =Current Assets/Currrent Liabilities (Liquidity)

- The higher the X-Score, the greater the likelihood of a company going bankrupt.
    - If X-Score > 0, the company is highly likely to go bankrupt
    - If X-Score < 0, the company is unlikely to go bankrupt

## Appendix 3: Evaluation of the Different Models Used in Bankruptcy Prediction

**Table 1.** Advantages and limitations of the bankruptcy prediction models.

| Group of Bankruptcy Prediction Models | | Advantages | Limitations |
|---|---|---|---|
| **Traditional statistical** | **Models of discriminant analysis**<br>(Altman 1968)<br>(Edmister 1972 (Edmister 1972))<br>(Taffler and Tisshaw 1977)<br>(Springate 1978)<br>(Fulmer et al. 1984)<br>(Sandin and Porporato 2007)<br>(Slefendorfas 2016)<br>Bonity index (Sedlacek 2001) | + simple calculation methodology;<br>+ high reliability-accuracy;<br>+ broad application and level of investigation practice. | - evaluates dependency of only linear financial indicators and bankruptcy likelihood; impact of macroeconomic changes is not evaluated;<br>- likelihood is most precisely calculated a year prior to the bankruptcy;<br>- abundance of derivative models from Altman practice caused the creation of similar models;<br>- often models are applied irrespective of industry branches, scope of companies; therefore, reliability declines;<br>- company's financial state, internal/external environment of activity, and trends of development are not thoroughly assessed. |
| | **Logistic regression model**<br>(Chesser 1974)<br>(Zavgren 1985)<br>(Grigaravicius 2003)<br>(Stoskus et al. 2007; Butkus et al. 2014)<br>(Ohlson 1980) | + evaluates rapidly changing, competitive, economic environment. | - similar to models of discriminant analysis;<br>- precision when more than 1 year to bankruptcy remains is lower than models based on discriminant analysis. |
| **Alternative** | **Models of neural networks**<br>(Iturriaga and Sanz 2015)<br>(Chung et al. 2016)<br>(Du Jardin 2019)<br>(Slavici et al. 2016) | + high reliability-accuracy;<br>+ integrated and applicable to a specific company-critical bankruptcy features are clearly and precisely identified;<br>+ able to model especially complex non-linear functions;<br>+ possible to apply for the analysis of combined, broad sets of data. | - new, little studied when compared to traditional models;<br>- rarely used in practice;<br>- difficult to identify the most precise neural network for evaluating problem characteristics due to their complexity, difficulty; require broad data base and is time-consuming;<br>- require computer software resources-additional costs for companies. |

*Figure A: Advantages and Disadvantages of Models Used in Bankruptcy Prediction (Giriuniene et al., 2019)*

## Appendix 4: Density Distribution of Top Five Important Variables

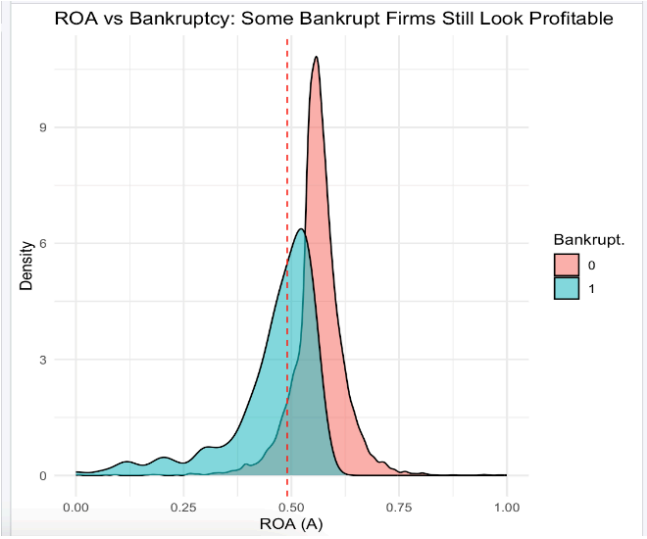


*Figure B: Density plot of ROA vs Bankruptcy*

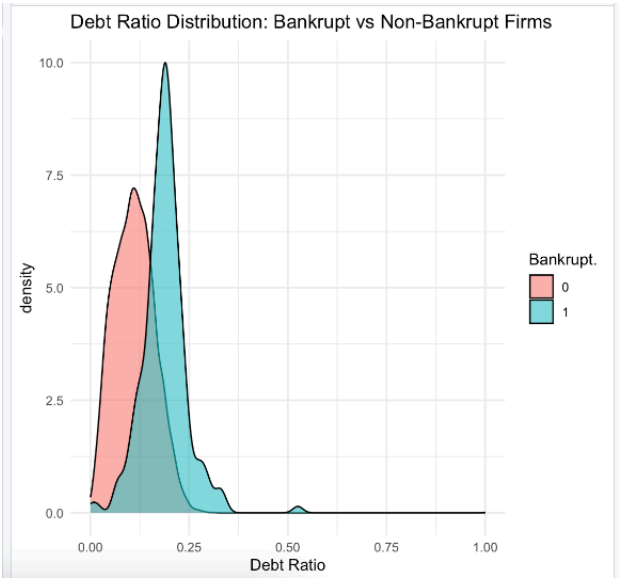


*Figure C: Density Plot of Debt Ratio Distribution*

**Appendix 5: Final Features After Data Preprocessing Using Logistic Regression and Correlation**

```
['Borrowing.dependency',
 'Total.assets.to.GNP.price',
 'ROA.A..before.interest.and...after.tax',
 'Accounts.Receivable.Turnover',
 'Pre.tax.net.Interest.Rate',
 'Net.Value.Per.Share..B.',
 'Non.industry.income.and.expenditure.revenue',
 'Net.worth.Assets',
 'Total.Asset.Turnover',
 'Cash.flow.rate',
 'Revenue.per.person',
 'Realized.Sales.Gross.Margin',
 'Revenue.Per.Share..Yuan...',
 'Cash.Total.Assets',
 'Quick.Assets.Total.Assets',
 'Liability.Assets.Flag',
 'Cash.Turnover.Rate',
 'Fixed.Assets.to.Assets',
 'Cash.Flow.to.Liability',
 'Equity.to.Liability',
 'Inventory.Turnover.Rate..times.']
```

*Figure D: List of 21 Variables Used in Train Set*

## Appendix 6: Calculations of Metrics Used in Model Evaluation and Their Significance

### (a) Metrics from Confusion Matrix

#### (i) Sensitivity Value

Sensitivity measures how well the model identifies actual bankrupt companies. A High sensitivity means the model detects most companies that are truly at risk of bankruptcy. It is computed using the following formula:

$$Sensitivity = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Negatives\ (FN)}$$

#### (ii) Precision Value

Precision tells us how many of the companies predicted as bankrupt are actually bankrupt. A high precision means fewer "false alarms". It can be computed using the following formula:

$$\Pr Pr\ e\ cision = \frac{True\ Positives\ (TP)}{True\ Positives\ (TP) + False\ Positives\ (FP)}$$

### (b) Balanced Accuracy

Balanced accuracy corrects for imbalanced datasets and ensures both classes are evaluated equally. It is computed using the following formula:

$$Balanced\ Accuracy = \frac{1}{2}\left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP}\right)$$

### (c) AUC-ROC Score

AUC-ROC score measures how well the model distinguishes between the two classes (bankrupt vs non-bankrupt) across all thresholds, with AUC ranging from 0.5 (random guessing) to 1.0 (perfect classification). In our case, a high AUC-ROC means the model is good at ranking companies by their risk of bankruptcy. Even if the exact threshold is off, the model still assigns higher probabilities to truly bankrupt companies.

### (d) F1 Score

F1 Score balances precision and senstivity, allowing for the assessment of trade-offs and can be calculated in the following manner:

$$F1\ Score = 2\left(\frac{precision \cdot sensitvity}{precision + sensitivity}\right)$$

**Appendix 7: Plot for Model Selection for Basic MARS model**

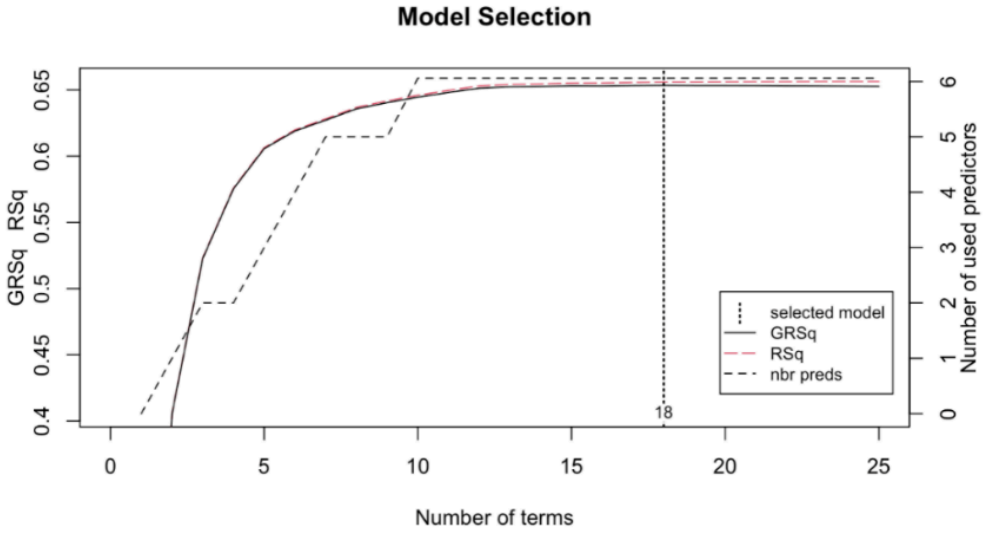


*Figure E: Plot for Model Selection*

## Appendix 8: Backward Elimination for Comparing GCV and RSS

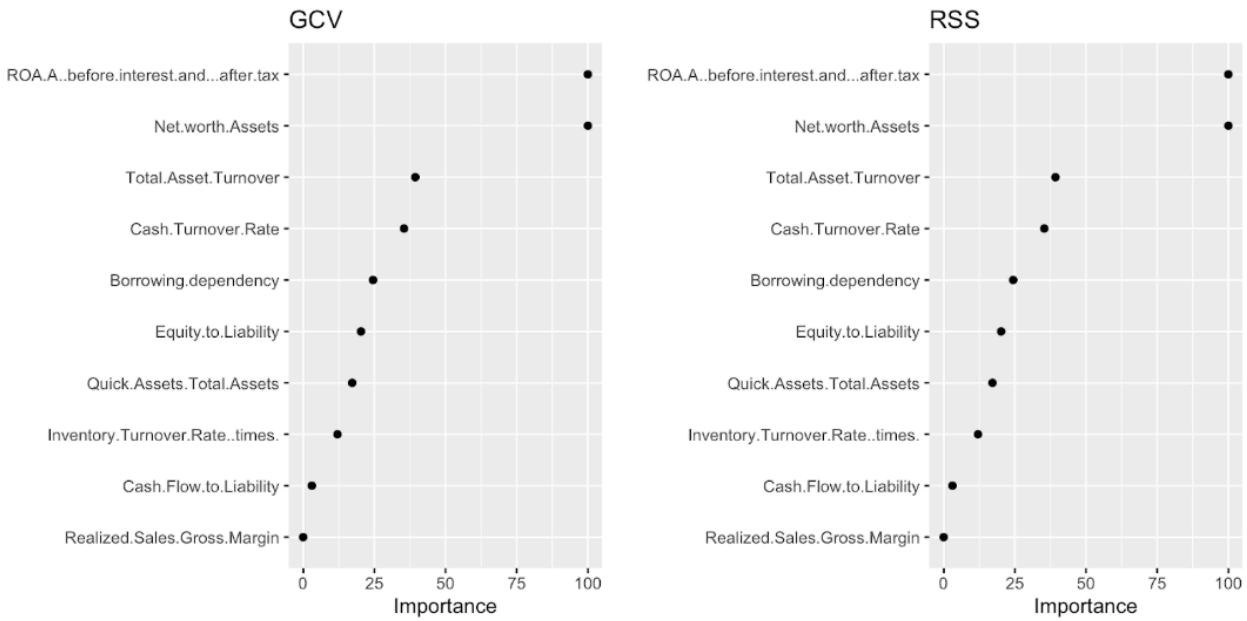


*Figure F: Comparison of GCV and RSS Values*

# Appendix 9: Hinge Terms from Optimised MARS Model

```
> named_coefficients
                        h(Borrowing.dependency-0.39106) * h(Net.worth.Assets-0.778951)
                                                                          -7.325961e+02
                       h(Borrowing.dependency-0.372537) * h(Net.worth.Assets-0.778951)
                                                                           6.448824e+02
                       h(Net.worth.Assets-0.778951) * h(0.0334497-Equity.to.Liability)
                                                                          -2.156536e+02
    h(ROA.A..before.interest.and...after.tax-0.534507) * h(Net.worth.Assets-0.778951)
                                                                           1.057181e+02
    h(ROA.A..before.interest.and...after.tax-0.570377) * h(Net.worth.Assets-0.778951)
                                                                          -8.928232e+01
h(ROA.A..before.interest.and...after.tax-0.431862) * h(0.0522486-Equity.to.Liability)
                                                                          -7.930961e+01
    h(ROA.A..before.interest.and...after.tax-0.353212) * h(Net.worth.Assets-0.778951)
                                                                          -3.696100e+01
                                   h(ROA.A..before.interest.and...after.tax-0.568142)
                                                                           1.508561e+01
    h(0.570377-ROA.A..before.interest.and...after.tax) * h(Net.worth.Assets-0.778951)
                                                                          -1.234614e+01
                                   h(ROA.A..before.interest.and...after.tax-0.531474)
                                                                          -1.156048e+01
                                             h(0.536287-Realized.Sales.Gross.Margin)
                                                                          -9.967176e+00
                 h(Net.worth.Assets-0.778951) * h(0.288516-Quick.Assets.Total.Assets)
                                                                          -6.653425e+00
h(0.515953-ROA.A..before.interest.and...after.tax) * h(0.323838-Total.Asset.Turnover)
                                                                          -6.132046e+00
                                                   h(0.323838-Total.Asset.Turnover)
                                                                           5.357330e+00
h(ROA.A..before.interest.and...after.tax-0.515953) * h(0.323838-Total.Asset.Turnover)
                                                                          -4.697113e+00
h(ROA.A..before.interest.and...after.tax-0.431862) * h(Equity.to.Liability-0.0522486)
                                                                           4.419984e+00
                       h(Net.worth.Assets-0.778951) * h(Equity.to.Liability-0.0334497)
                                                                           2.385545e+00
                                   h(ROA.A..before.interest.and...after.tax-0.431862)
                                                                           2.173073e+00
```

*Figure G: Significant Hinge Terms Explained in Section 5.3.2*

| | names | x |
|---|---|---|
| | <chr> | <dbl> |
| 1 | h(0.778951-Net.worth.Assets) | -6.83e+ 0 |
| 2 | h(ROA.A..before.interest.and...after.tax-0.431862) | 3.65e+ 1 |
| 3 | h(ROA.A..before.interest.and...after.tax-0.570377) * h(Net.worth.Assets-0.778951) | -7.30e+ 3 |
| 4 | h(0.570377-ROA.A..before.interest.and...after.tax) * h(Net.worth.Assets-0.778951) | -1.13e+ 2 |
| 5 | h(0.323838-Total.Asset.Turnover) | 6.27e+ 1 |
| 6 | h(0.323838-Total.Asset.Turnover) * h(Cash.Turnover.Rate-7.07e+08) | 7.12e- 8 |
| 7 | h(0.323838-Total.Asset.Turnover) * h(7.07e+08-Cash.Turnover.Rate) | -7.11e- 8 |
| 8 | h(ROA.A..before.interest.and...after.tax-0.515953) * h(0.323838-Total.Asset.Turnove… | 5.23e+ 1 |
| 9 | h(0.515953-ROA.A..before.interest.and...after.tax) * h(0.323838-Total.Asset.Turnove… | -8.85e+ 1 |
| 10 | h(ROA.A..before.interest.and...after.tax-0.431862) * h(Equity.to.Liability-0.052248… | -3.51e+ 0 |
| 11 | h(ROA.A..before.interest.and...after.tax-0.431862) * h(0.0522486-Equity.to.Liabilit… | -1.30e+ 3 |
| 12 | h(Borrowing.dependency-0.372537) * h(Net.worth.Assets-0.778951) | 8.72e+ 3 |
| 13 | h(Net.worth.Assets-0.778951) * h(Quick.Assets.Total.Assets-0.288516) | 1.94e+ 1 |
| 14 | h(Net.worth.Assets-0.778951) * h(0.288516-Quick.Assets.Total.Assets) | -1.26e+ 2 |
| 15 | h(Borrowing.dependency-0.39106) * h(Net.worth.Assets-0.778951) | -2.07e+ 3 |
| 16 | h(Net.worth.Assets-0.778951) * h(Equity.to.Liability-0.0334497) | 1.43e+ 1 |
| 17 | h(Net.worth.Assets-0.778951) * h(0.0334497-Equity.to.Liability) | -3.20e+ 3 |
| 18 | h(0.323838-Total.Asset.Turnover) * h(Inventory.Turnover.Rate..times.-1.74757e+09) | -1.52e- 9 |
| 19 | h(0.323838-Total.Asset.Turnover) * h(1.74757e+09-Inventory.Turnover.Rate..times.) | -2.31e- 9 |
| 20 | h(Cash.Flow.to.Liability-0.12872) | -1.45e+ 1 |
| 21 | h(ROA.A..before.interest.and...after.tax-0.568142) | -4.92e+ 2 |
| 22 | h(ROA.A..before.interest.and...after.tax-0.534507) * h(Net.worth.Assets-0.778951) | 4.87e+ 2 |
| 23 | h(ROA.A..before.interest.and...after.tax-0.531474) | -6.11e+ 1 |
| 24 | h(ROA.A..before.interest.and...after.tax-0.353212) * h(Net.worth.Assets-0.778951) | -3.59e+ 2 |
| 25 | h(0.323838-Total.Asset.Turnover) * h(Cash.Turnover.Rate-1.93262e+08) | -7.26e- 8 |
| 26 | h(ROA.A..before.interest.and...after.tax-0.0929089) | -2.16e+ 1 |
| 27 | h(Realized.Sales.Gross.Margin-0.536287) | -2.55e+ 1 |
| 28 | h(0.536287-Realized.Sales.Gross.Margin) | -1.93e+ 2 |
| 29 | h(Net.worth.Assets-0.778951) * h(Cash.Turnover.Rate-2.66e+09) | 2.88e-10 |
| 30 | h(Net.worth.Assets-0.778951) * h(2.66e+09-Cash.Turnover.Rate) | 3.63e- 9 |

*Figure H: Additional Analysis on Hinge Interactions*

```
                                                                  (Intercept)
                                                                 1.577205e+00
                                h(ROA.A..before.interest.and...after.tax-0.0929089)
                                                                -1.356721e+00
                                           h(Realized.Sales.Gross.Margin-0.536287)
                                                                -1.216860e+00
                h(Net.worth.Assets-0.778951) * h(Quick.Assets.Total.Assets-0.288516)
                                                                 1.027028e+00
                                                 h(Cash.Flow.to.Liability-0.12872)
                                                                -8.515469e-01
                                                      h(0.778951-Net.worth.Assets)
                                                                -5.424086e-01
                   h(0.323838-Total.Asset.Turnover) * h(7.07e+08-Cash.Turnover.Rate)
                                                                -6.175599e-09
                h(0.323838-Total.Asset.Turnover) * h(Cash.Turnover.Rate-1.93262e+08)
                                                                -6.130586e-09
                   h(0.323838-Total.Asset.Turnover) * h(Cash.Turnover.Rate-7.07e+08)
                                                                 5.977533e-09
                       h(Net.worth.Assets-0.778951) * h(2.66e+09-Cash.Turnover.Rate)
                                                                 3.748721e-10
h(0.323838-Total.Asset.Turnover) * h(1.74757e+09-Inventory.Turnover.Rate..times.)
                                                                -1.749904e-10
                       h(Net.worth.Assets-0.778951) * h(Cash.Turnover.Rate-2.66e+09)
                                                                 1.475278e-10
h(0.323838-Total.Asset.Turnover) * h(Inventory.Turnover.Rate..times.-1.74757e+09)
                                                                -1.168780e-10
```

*Figure I: Hinge Terms and their Intercepts*

## Appendix 10: Interpretations of Confusion Matrix

In a confusion matrix, the different values hold different meanings in the context of bankruptcy prediction, which is explained as follows:

(a) **True Positives (TP):** These are companies that were actually bankrupt and were correctly identified by the model.

(b) **False Negatives (FN)**: These are bankrupt companies that the model incorrectly predicted as not bankrupt.

(c) **True Negatives (TN):** These are solvent companies correctly classified as not bankrupt.

(d) **False Positives (FP):** These are solvent companies that were mistakenly flagged as bankrupt.

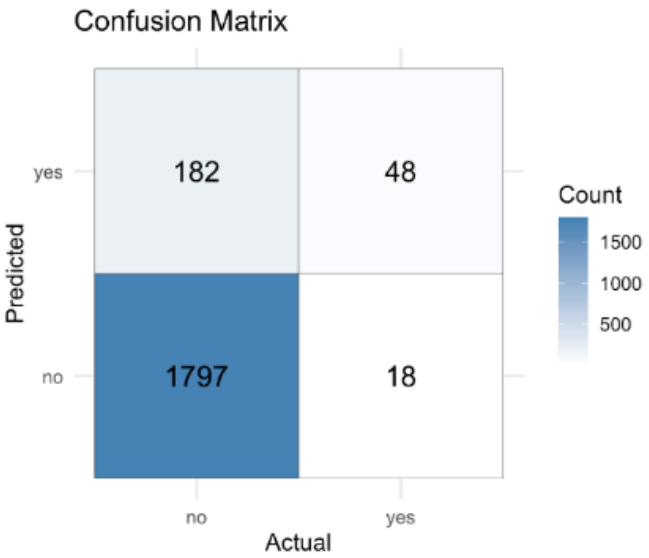


*Figure J: Confusion Matrix of Optimum MARS model*

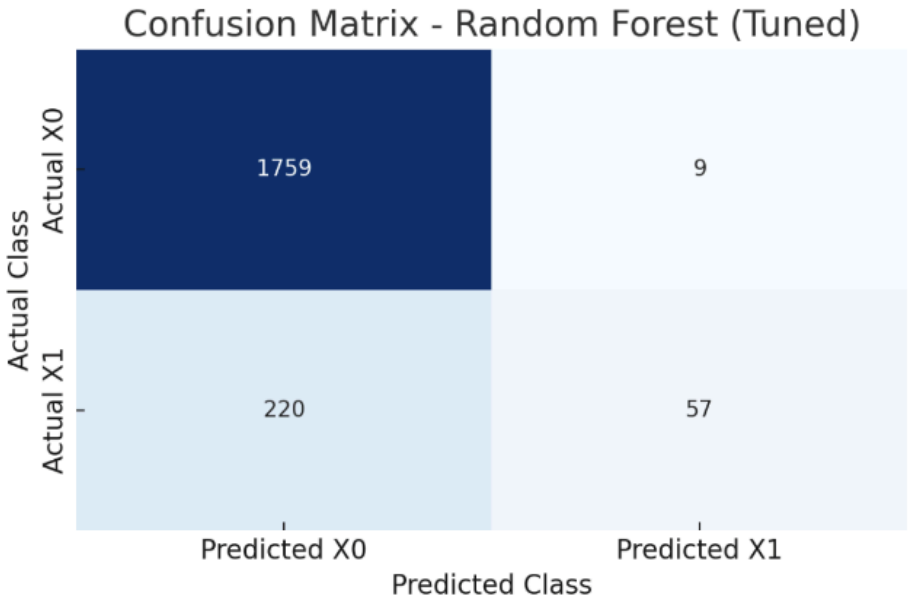


*Figure K: Confusion Matrix of Random Forest*

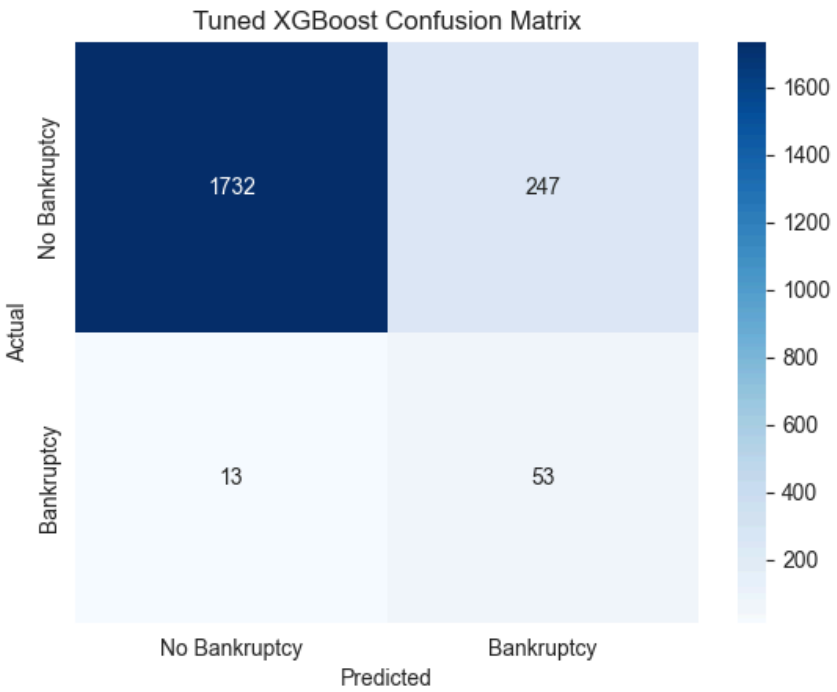


*Figure L: Confusion matrix of XGBoost*

## Appendix 11: Plot of SHAP Analysis

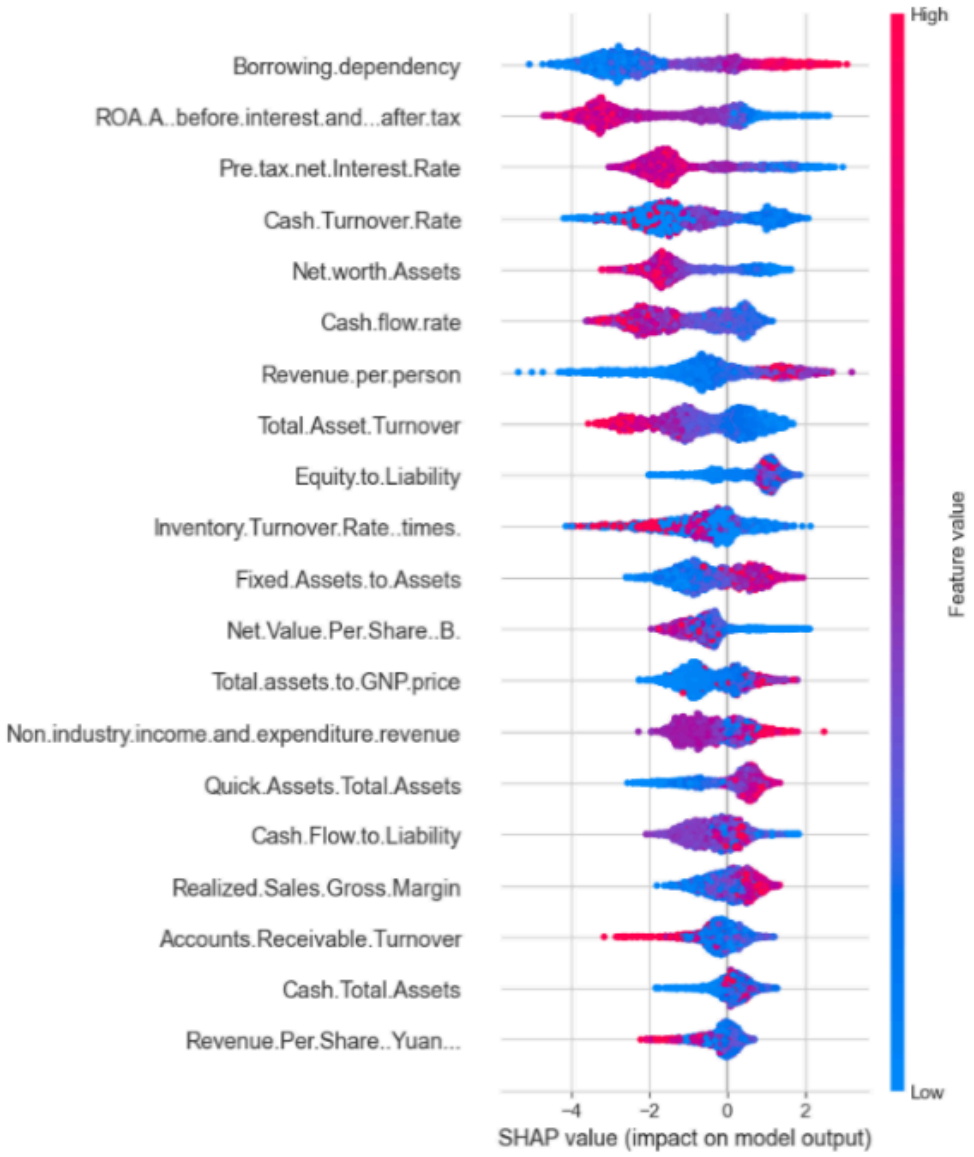


*Figure M: Variables and their associated SHAP values*

## Appendix 12: Summary of Benefits, Limitations and Future Improvements of Each Model

| | MARS | XGBoost | Random Forest |
|---|---|---|---|
| **Benefits** | - Simple and easy to interpret<br>- Customizable thresholds<br>- Quick implementation | - High predictive accuracy<br>- Handles complex data<br>- Scalable and adaptable | - Robust to missing data<br>- Feature importance<br>- Flexible and scalable |
| **Limitations** | - Requires manual updates<br>- Limited In handling complex data | - Requires computational resources<br>- Difficult to interpret without tools | - Risk of overfitting<br>- Difficult to interpret interactions between trees |
| **Future Improvements** | - Automatic threshold updates<br>- Better representation of interactions | - Improved explainability<br>- Optimize for imbalanced datasets | - Mitigate overfitting<br>- Better visualization of interactions |

*Figure N: Summarised Evaluation of Our Predictive Models*

## 9. References


Alfaro, L., Larrain, M., Vilches, C., & Jeong, S. (2020, August). *Latam Airlines and COVID-19: Seeking Bankruptcy Protection in the United States*. Harvard Business School. https://www.hbs.edu/faculty/Pages/item.aspx?num=58535

*All You Need To Know About Corporate Bankruptcy In Singapore*. (2023, January 9). Smart Towkay. Retrieved April 5, 2025, from https://www.smart-towkay.com/blog/view/313-all-you-need-to-know-about-corporate-bankruptcy-in-singapore

Barboza, F., Kimura, H., & Altman, E. (2017). Expert Systems with Applications. *83*(0957-4174), 405-417. https://doi.org/10.1016/j.eswa.2017.04.006.

Basu, S. (2021, August). Personalized product recommendations and firm performance. *Electronic Commerce Research and Applications*, *48*. https://doi.org/10.1016/j.elerap.2021.101074

Brownlee, J. (2021, April 27). *Multivariate Adaptive Regression Splines (MARS) in Python - MachineLearningMastery.com*. Machine Learning Mastery. Retrieved April 6, 2025, from https://machinelearningmastery.com/multivariate-adaptive-regression-splines-mars-in-python/

Buchholz, P., Schumacher, A., & Al Barazi, S. (2022, September 12). Big data analyses for real-time tracking of risks in the mineral raw material markets: implications for improved supply chain risk management. *Mineral Economics*, *35*, 701-744. https://link.springer.com/article/10.1007/s13563-022-00337-z

Buckley, J. (2023, February 11). *How the pandemic killed off 64 airlines*. CNN. Retrieved April 6, 2025, from https://edition.cnn.com/travel/article/pandemic-airline-bankruptcies/index.html

Calomiris, C. W., Longhofer, S. D., & Jaffee, D. M. (2008). Credit Rationing. *The New Palgrave Dictionary of Economics*, 1-10. Springer Nature Link. https://doi.org/10.1057/978-1-349-95121-5_479-2

Cross, G. (2025, February 13). *JCPenney is closing 8 locations by mid-2025. See the list*. USA Today. Retrieved April 5, 2025, from https://www.usatoday.com/story/money/2025/02/13/jcpenney-store-closings/78536216007/

Elemes, A., & Filip, A. (2022). Financial Reporting Quality and Private Firms' Access to Trade Credit Capital. *The International Journal of Accounting*, *57*(2). https://doi-org.remotexs.ntu.edu.sg/10.1142/S109440602250010X

Gibson, J., & Petrick, B. (2025, March 14). *Chapter 11 Bankruptcy: What's Involved, Pros & Cons of Filing*. Investopedia. Retrieved April 5, 2025, from https://www.investopedia.com/terms/c/chapter11.asp

Giriuniene, G., Giriunas, L., Morkunas, M., & Brucaite, L. (2019, August). A Comparison on Leading Methodologies for Bankruptcy Prediction: The Case of the Construction Sector in Lithuania. *Economies*, *7*(3), 82. 10.3390/economies7030082

Goldstein, I., Xiong, Y., & Yang, L. (2024). Information sharing in financial markets. *Journal of Financial Economics*, *163*, 103967. https://doi.org/10.1016/j.jfineco.2024.103967

Hernandez, J. (2023, September 15). *How Lehman's collapse 15 years ago changed the U.S. mortgage industry*. NPR. Retrieved April 5, 2025, from https://www.npr.org/2023/09/15/1199321274/lehman-brothers-collapse-2008-mortgages

ICAEW. (n.d.). *Six signs that a business is in distress*. ICAEW.com. Retrieved April 5, 2025, from https://www.icaew.com/regulation/insolvency/understanding-business-restructuring-and-insolvency/six-signs-that-a-business-is-in-distress

Jones, J. S., Lee, W. Y., & Yeager, T. J. (2012). Opaque banks, price discovery, and financial instability. *Journal of Financial Intermediation*, *21*(3), 383–408. https://doi.org/10.1016/j.jfi.2012.01.004

Kavalakglu, E., & Russi, E. (2024, May 9). *What is XGBoost?* IBM. Retrieved April 6, 2025, from https://www.ibm.com/think/topics/xgboost

Kearney. (2025, February 11). *More corporate bankruptcies expected in 2025: challenges and developments in distressed M&A*. Kearney. https://www.kearney.com/industry/private-equity/article/more-corporate-bankruptcies-expected-in-2025-challenges-and-developments-in-distressed-m-a

Kenge, R. (2020, October). Machine Learning, Its Limitations, and Solutions Over IT. *International Journal of Information Technology Modeling and Computing*, *11*(2), 73-83. 10.5958/0975-8089.2020.00009.3 Authors:

Knauth, D., & Blackwell, H. (2025, January 2). *Bankruptcy trends to watch in 2025*. Reuters. Retrieved April 5, 2025, from https://www.reuters.com/legal/litigation/bankruptcy-trends-watch-2025-2025-01-02/

Lazzaro, N. (2025, January 7). *US corporate bankruptcies soar to 14-year high in 2024; 61 filings in December*. S&P Global. Retrieved April 5, 2025, from https://www.spglobal.com/market-intelligence/en/news-insights/articles/2025/1/us-corporate-bankruptcies-soar-to-14-year-high-in-2024-61-filings-in-december-87008718

Marsenne, M., Ismail, T., Taqi, M., & Hanifahd, I. A. (2023, September 30). Financial distress predictions with Altman, Springate, Zmijewski, Taffler and Grover models. *Decision Science Letters*, *13*, 181-190. https://www.growingscience.com/dsl/Vol13/dsl_2023_51.pdf

Máté, D., Raza, H., & Ahmad, I. (2023). Comparative Analysis of Machine Learning Models for Bankruptcy Prediction in the Context of Pakistani Companies. *11*(10), 176. https://doi.org/10.3390/risks11100176

Ministry of Law, Singapore. (2020, July 23). *Insolvency, Restructuring and Dissolution Act to Commence on 30 July 2020*. Ministry of Law. Retrieved April 5, 2025, from https://www.mlaw.gov.sg/news/press-releases/2020-07-23-insolvency-restructuring-and-dissolution-act-commencement/

Nalley, S. (2024, July 9). *Five Top Causes of Business Bankruptcy and How to Avoid Them*. Kiplinger. https://www.kiplinger.com/kiplinger-advisor-collective/top-causes-of-business-bankruptcy-and-how-to-avoid-them

Oh, T. (2025, February 4). *Number of compulsory wind-ups rise more than 50% in 2024*. The Business Times. Retrieved April 5, 2025, from https://www.businesstimes.com.sg/companies-markets/number-compulsory-wind-ups-rise-more-50-2024

Ooi, K. L. (2020, July). Social Factors and Herd Behaviour in Developed Markets, Advanced Emerging Markets and Secondary Emerging Markets. *Journal of Contemporary Eastern Asia*, *19*(1), 97-122. ResearchGate. http://dx.doi.org/10.17477/jcea.2020.19.1.097

Patritti, M. (2024). *Leading Financial Indicators of Corporate Bankruptcy*. ScholarWorks at University of Montana. Retrieved April 5, 2025, from https://scholarworks.umt.edu/cgi/viewcontent.cgi?article=1516&context=utpp

Peh, A. (2024, March 11). *Commentary: Why are more people in Singapore going bankrupt?* CNA. Retrieved April 5, 2025, from https://www.channelnewsasia.com/commentary/singapore-bankruptcy-applications-debt-challenging-economy-money-4172186

Radovanovic, J., & Haas, C. (2023). The evaluation of bankruptcy prediction models based on socio-economic costs. *Expert Systems with Applications*, *227*. https://doi.org/10.1016/j.eswa.2023.120275

Rahayu, A. (n.d.). *Financial Distress Analysis to Detect Potential Bankruptcy Using the Altman (Z-Score)*. Springate. http://eprints.polsri.ac.id/4824/

Rashid, F., Khan, R. A., & Quresh, I. H. (2023, December 4). A Comprehensive Review of the Altman Z-Score Model Across Industries.

Stapleton, C., & Kenton, W. (2021, May 29). *Overleveraged: Meaning & Adverse Outcomes*. Investopedia. Retrieved April 6, 2025, from https://www.investopedia.com/terms/o/overleveraged.asp

The Straits Times. (2024, November 13). *Chinese EV start-up WM Motor files for bankruptcy*. The Straits Times. Retrieved April 6, 2025, from https://www.straitstimes.com/business/chinese-ev-start-up-wm-motor-files-for-bankruptcy

Valinsky, J. (2020, June 29). *Cirque du Soleil files for bankruptcy protection and cuts 3,500 jobs*. CNN Business. https://edition.cnn.com/2020/06/29/business/cirque-du-soleil-bankruptcy/index.html

Vukčević, M., Lakićević, M., Melović, B., Backović, T., & Dudić, B. (n.d.). *Modern models for predicting bankruptcy to detect early signals of business failure: Evidence from Montenegro*. PloS ONE. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0303793

Wang, Y. (2023, February 11). *The Impact of Information Asymmetry on Investment Behavior in the Stock Market | Highlights in Business, Economics and Management*. Darcy & Roy Press. Retrieved April 6, 2025, from https://drpress.org/ojs/index.php/HBEM/article/view/11867

World Bank Group. (n.d.). *Financial Access*. World Bank. Retrieved April 6, 2025, from https://www.worldbank.org/en/publication/gfdr/gfdr-2016/background/financial-access