Master's Thesis in Accounting, Valuation and Financial Management
Stockholm School of Economics
Autumn 2023

The Tale of Two Techniques:

The comparative accuracy of machine learning and statistical techniques in predicting corporate bankruptcy for Swedish industrial firms

Edvin Lindhout (50592) Hampus Ljunglöf (42042)

Supervisor: Stina Skogsvik
Date: 2023-12-04

Abstract

Bankruptcy prediction has long been an important area of study, yet the evolution of these predictive models in the context of modern machine learning techniques remains underexplored. Our thesis addresses this by comparing the effectiveness of probit analysis – a time-tested statistical approach – with XGBoost – a new-era machine learning technique – in predicting corporate bankruptcy among Swedish firms. Utilizing a dataset of Swedish industrial firms, we meticulously assess the accuracy of each model, looking also at their capacity to select and leverage relevant independent variables. Our findings reveal notable differences in the performance of these models, providing valuable insights for researchers and practitioners. While the probit model offers a reliable, well-established framework, XGBoost demonstrates superior adaptability and performance, marking a significant advancement in bankruptcy prediction methodologies. The machine learning technique also proves better at extracting useful information through feature selection and appears more generalizable when tested on firms of different industries and sizes. We perform several robustness checks to ensure the viability of these conclusions and end by discussing our findings, limitations and potential future research directions.

Keywords: Bankruptcy prediction, Machine learning, XGBoost, Probit analysis, Swedish firms

A tremendous thank you to our supervisor Stina Skogsvik. Her immense knowledge of the subject and impeccable ability to steer this thesis in the right direction made it more rewarding to write and hopefully more interesting to read. A special thank you also to Artin Mirzaian for his guidance in applying XGBoost to the bankruptcy prediction problem, to André Brogärd for his superb suggestions on our research design, and to Erik Larsson and William Löthén for their encouragement and advice.

Table of Contents

1. Introduction

No one can have missed the recent boom in artificial intelligence (AI). A recurring topic of news reporting, included in the strategies of most Fortune 500 companies and even the topic of casual conversation – the recent advancements in AI has had a pervasive impact on our lives. Today, we can witness how AI programs power autonomous vehicles, detect signs of disease on medical images that are otherwise invisible to the human eye, and enable you to unlock your smartphone with face recognition. They paint portraits, prepare tax returns, respond to emails, and debug code. The list can be made very long. This transition represents the most important technological breakthrough since social media (Time, 2023).

The current boom in AI is driven by breakthroughs in an area known as machine learning – specifically generative AI – with McKinsey recently dubbing 2023 "Generative AI's breakout year" in their global survey on the current state of AI (McKinsey, 2023). Rather than depend on human programming, computers are "trained" to execute tasks based on examples. While the models are trained by humans, their abilities can often extend far beyond those of any human – a famous example being when AI defeated the world champion of a complex board game called Go in 2016 (Wired, 2023). McKinsey states in the survey that AI historically has been a topic for tech employees but has risen as a focus of managers amid recent advances (McKinsey, 2023). The two tech-giants, Microsoft, and Alphabet reported complete shifts to their corporate strategies to capture what they see as a new infrastructure layer of the economy. For example, Microsoft is investing USD 10 billion in OpenAI (the organization behind ChatGPT and Dall-E) and is planning on updating its Office software and search engine, Bing, with generative AI integrations. As a response, Google was quick to launch their own chatbot, Bard, onto the market (Time, 2023). While there are tons of advocators of AI, there are just as many critics. Criticism is often directed towards generative AI and centered around concerns such as ethics, sustainability, disinformation, bias and even copyright (Wired, 2023). Yet beyond all the hype of generative AI, there is an area where machine learning has been applied ever since the late 1990s: bankruptcy prediction.

Corporate failure remains a ubiquitous topic due to its negative effects on stakeholders and society, including financial distress, job losses, and a reduced trust in the affected business. Counterparty risk has always been a central part in doing business and it has been in the interest of both governments, lenders, financial institutions, fund managers, and financial market participants alike to assess the risk in order to manage it. As such, researchers started exploring methods of corporate bankruptcy prediction as far back as the 1930s. An early discovery was that capital market information can be utilized to develop bankruptcy prediction models, despite the stochastic nature of default events (Altman et al., 2017). Bankruptcy prediction research has evolved since it first emerged nearly a century ago, with various approaches employed, frequently relying on quantitative analysis of financial ratios and accounting metrics (Bellovary et al., 2007).

As the access to more computational power exploded near the turn of the millennia, researchers began experimenting with machine learning techniques to see if corporate failure could be predicted more accurately than with traditional statistical techniques. And yet, nearly thirty years later, there is still an ongoing debate whether machine learning models are better (Clement, 2020). The initial critique largely concerned the lack of any noteworthy performance gains (Altman et al., 1994), but this was likely due to constraints in the computational capacity of widely available computers. Today, the critique revolves around the lack of transparency (Kim et al., 2022) and overfitting tendencies (Ptak-Chmielewska, 2019). Machine learning models are continuously being improved upon with efforts being made to make them both more transparent and resistant to overfitting. We will refer to these models as new-era machine learning models, as they are distinct from their predecessors and do not come with the same flaws. We ask ourselves, can the recent advancements in AI put an end to a thirty-year-old debate?

1.1 Purpose & research questions

Our purpose with this study is to evaluate the comparative accuracy of an established statistical technique compared to a new-era machine learning model for the task of predicting corporate bankruptcy. An important aspect of maximizing accuracy is to extract useful information from the independent variables. Our purpose results in the following primary research questions:

(i) *How does the comparative accuracy of a new-era machine learning model differ from an established statistical technique in predicting corporate bankruptcy?*

(ii) *How effective are the techniques at identifying useful independent variables from a larger set of potentially useful variables?*

Another critical aspect to consider is the generalizability of the models. While comparative accuracy may indicate superior performance in a test setting resembling the training data conditions, it could be susceptible to overfitting, leading to greater performance drops on settings different from the training data. Two pivotal dimensions for exploration are industry and size, aiming to discern whether the relative performance of statistical and machine learning techniques remains consistent when asked to predict bankruptcies in firms distinct from their training samples. To address these considerations, we delve into the secondary research questions:

(iii) *How does the comparative accuracy of a new-era machine learning model differ from an established statistical technique in predicting corporate bankruptcy for firms in other industries than the training sample?*

(iv) *How does the comparative accuracy of a new-era machine learning model differ from an established statistical technique in predicting corporate bankruptcy for firms of other sizes than the training sample?*

2. Review of literature

Since our study aims to compare an established statistical technique to a new-era machine learning model in predicting corporate bankruptcy, the literature review is divided into six sections, each covering information relevant to the study. The first section presents guidance from theory on capital investment risk and its importance in business. The second section explores how researchers started using statistics to measure bankruptcy risk and presents the most prominent traditional models in the field. The third section gives an overview of the most used machine learning techniques in bankruptcy prediction. The fourth section covers selected studies comparing different techniques. The fifth section explores the field of bankruptcy prediction in a Swedish context and discusses a potential research gap in current literature. Finally, the sixth section summarizes the key takeaways from previous literature and ends with a proposal on how our study can contribute.

2.1 Capital investment risk

Risk is a crucial component in business. Understanding, managing, and sometimes embracing risk is key for business success (Moore, 1983). A manager faces many types of risks in their role, from supply chain risks to compliance risk and counterparty risk. A capital investor looking to invest in a business by acquiring company shares needs to consider both variability risk and bankruptcy risk. Variability risk refers to the "return variability risk of a capital investment" whereas bankruptcy risk is the "drop-dead risk of a capital investment" (Skogsvik, S., p.3, 2021). With regards to the latter, Skogsvik highlights a component to valuation, denoted *Pfail*. It is an estimate to account for the probability of a firm going bankrupt and it can be assessed through: (i) employing statistical models like probit and logit analysis based on financial data, (ii) considering bond ratings, or (iii) implied by market prices, i.e. reverse engineering.[1] While (ii) and (iii) might be useful for large public firms, the majority of firms are SMEs. Furthermore, a study conducted by Dichev (1998) found evidence that suggested a higher bankruptcy risk is not necessarily compensated by higher returns, casting doubt on the suitability of deriving risk from market implications.

Damodaran (2006) discusses the implications of bankruptcy risk in traditional valuation techniques, arguing that its effects on value are often short-changed, if not completely ignored. For example, in Discounted Cash Flow analyses (DCF) and relative valuation, firms are implicitly assumed to be going concerns and that any exposure to financial distress is temporary. A large part of the value in a DCF is captured in the terminal value, often far into the future – but what if there is a real chance that the company will not make it to the terminal value and the financial distress is not temporary? Damodaran argues that we tend to overvalue such firms using traditional valuation models primarily because the effect of financial distress is difficult to capture in the discount rate and expected cash flows (Damodaran, 2006).

Empirical observations indicate that many companies – primarily smaller and higher growth – will fail (Damodaran, 2006). Some companies fail because they borrow to fund their operations and default on these debt payments, while others lack sufficient cash to cover operating needs. The number of bankruptcies will vary depending on the definition of distress. For example, distress is much more common when defining it as failing to make interest payments or meet contractual commitments rather than the number of firms entering chapter 11 (Damodaran, 2006).[2]

[1] Starting with the known market price of e.g. a bond, and working backward, or reverse-engineering the bond valuation from its market price, we can work out the implied probability of failure

[2] A case filed under chapter 11 of the United States Bankruptcy Code is frequently referred to as a "reorganization" bankruptcy. Usually, the debtor remains "in possession," has the powers and duties of a trustee, may continue to operate its business, and may, with court approval, borrow new money (US Courts).

From a valuation perspective, failing to consider bankruptcy risk is problematic as it will prevent us from reaching the true market value of a business. As a company defaults on its debt, it is often forced to liquidate assets at bargain prices.[3] However, the consequences of financial distress go far beyond the direct costs of bankruptcy. The perception of financial distress can harm a company significantly as it affects employees, customers, suppliers, and lenders. Distressed firms experience customer loss, increased employee turnover, and face stricter supplier terms. These indirect bankruptcy costs can be devastating, effectively turning the perception of distress into a reality for many companies (Damodaran, 2006). The widespread interest in the ability to predict corporate failure becomes evident and will be further explored and validated in the forthcoming section as we review the existing literature to date.

2.2 Pioneering bankruptcy prediction models

Existing literature on bankruptcy prediction consists mostly of empirical studies, comparing data of bankrupt and non-bankrupt firms rather than applying theory (Brenes, 2022). The author discusses why there is an apparent lack of theoretical models and points to the complexity of bankruptcy as a potential explanation. While bankruptcy is driven by several deterministic factors such as the company's financial health, market conditions and industry trends, there are instances of bankruptcies that cannot be adequately explained by these factors alone. Sudden economic downturns, shifts in consumer behavior and even natural disasters can all drive a company bankrupt and are difficult to predict with a single theory (Bradley, 2004). In a review paper, Scott (1981) examines the alignment of empirical and theoretical models for bankruptcy prediction. While fundamental theoretical models, such as the Gambler's Ruin model, explain some of the empirical findings, they fail to account for it fully.[4] This suggested that an empirically based model could successfully predict causes of bankruptcy that bankruptcy theory at the time failed to explain. Empirical models are practical – they make sense. If we think of financial data as a proxy for a company's true health, looking at the financial data of companies that went bankrupt and those that remained solvent should help us predict how well a company might do in the future based on its financial information.

The modern use of financial ratios to predict bankruptcy was pioneered in the 1960s, in papers by Beaver (1966) and Altman (1968). Previous research consisted mainly of univariate ratio analysis, comparing the ratios of healthy and bankrupt firms, and identifying indicators of oncoming financial distress. With Beaver and Altman, the hypothesis that financial ratios might be employed to statistically predict bankruptcy was introduced, first with Beaver's univariate approach followed by Altman's multivariate (Bellovary et al., 2007). The aim of Beaver's study was to empirically verify the usefulness (predictive ability) of accounting data (financial statements). In doing so, the author investigated 79 failed and 79 non-failed firms across 38 industries by comparing the mean values of 30 ratios. The study was conducted during a period between 1954 to 1964 and the data was based on publicly listed industrial firms in the US with an average asset size of USD 7.4 million. Based on a paired-sample design, a non-failed firm of the same asset size and industry was selected for each failed firm in the sample. "Failure" was used as the dependent variable, defined as:

> "*The inability of a firm to pay its financial obligations as they mature. Operationally, a firm is said to have failed when any of the following events have occurred: bankruptcy, bond default, an overdrawn bank account, or nonpayment of a preferred stock dividend*" (Beaver, p. 71, 1966).

[3] Found by, for example, Shleifer & Vishny (1992), Aghion et al. (1994), Berkovitch & Israel (1995), White (1994), and Pulvino (1998). Eckbo & Thorburn (2000) fail to find support for asset fire-sales in Swedish bankruptcy auctions.
[4] The Gambler's Ruin model is a construct in probability theory, exploring the probable outcomes of the gambler either attaining a predetermined fortune or succumbing to financial ruin (Coolidge, 1909). A key concept is that a game with unfavorable odds, i.e. of negative value to the gambler, will always result in the gambler's ruin. For bankruptcy prediction, this should imply that firms who fail to create value will eventually go bankrupt.

The author examined multiple forecast horizons, including one-, two-, three-, four-, and five years prior to failure, achieving average prediction errors as low as 13%, 20%, and 23% for the first three years, respectively. The author then computed 30 ratios from every available set of financial statements selected based on: (i) the frequency of the ratio in previous literature, (ii) ratios with high performance in recent studies (to be able to analyze the consistency), and (iii) ratios defined based on a "cash-flow" concept. Certain ratios with obvious overlaps were excluded to ensure each single ratio extracts as much additional information as possible. Unlike previous studies in the field, the author tested predictive abilities of the individual ratios in classifying firms as either bankrupt or non-bankrupt (Beaver, 1966). His results indicated that the ratio of net income to total debt achieved the highest predictive ability one year prior to failure with an accuracy of 92%, followed by net income to sales at 91%. Net income to net worth, cash flow to total debt, and cash flow to total assets showed accuracies of 90% (Bellovary et al., 2007). Beaver's univariate study paved the way for multivariate bankruptcy prediction models (Bellovary et al., 2007). However, several other univariate studies have been developed (e.g. Pinches et al., 1975 and Chen & Shimerda, 1981)

Pointing to the shortcomings of univariate analysis, such as the inherent flaws of looking at a single ratio in isolation, Altman was the first to publish a multivariate study in 1968. He employed multivariate discriminant analysis (MDA) to formulate a five-factor model for the prediction of bankruptcy among manufacturing firms in the US. This model, known as the "Z-Score Model", indicated bankruptcy when a firm's score fell within a specific range. Using a paired-sample design, a total of 66 manufacturing corporations were divided into two groups. The first group consisted of 33 corporations ("failure" firms) that had filed for bankruptcy under Chapter 10 of the National Bankruptcy Act between any of the years 1946 to 1965. The asset size of the corporations ranged between USD 0.7 million and USD 25.9 million, with a mean of USD 6.4 million. The second group contained the remaining 33 corporations and were still in existence in 1966. The data for the second group was based on the same period as for the first group. With the sample of companies in place, the author compiled a set of 22 ratios for evaluation, each classified into one of five categories including liquidity, leverage, solvency, profitability, and activity ratios. Most ratios were selected based on popularity in previous studies and potential relevancy to the study, while a few new ratios were created (Altman, 1968). Altman analyzed the 22 ratios' relative contribution, inter-correlations, predictive accuracies, as well as applied his own judgement. This resulted in a five-variable model as the overall best predictor of bankruptcy.

The Z-Score Model exhibited a predictive accuracy in the initial sample of 95% one year prior to failure and 72% two years prior to failure. When tested out-of-sample, the model obtained a predictive accuracy of 79%.[5] However, its predictive performance diminished rapidly after year two. Corresponding accuracies three, four and five years prior to bankruptcy amounted to 48%, 29%, and 36%, respectively (Altman, 1968). In summary, the author's results suggest that bankruptcy can be accurately predicted at least one year prior to failure.

Over the years, researchers have investigated new quantitative approaches to improve bankruptcy prediction models. For example, as Ohlson (1980) developed the first logit-based bankruptcy prediction model, the next two decades saw a shift towards models based on logit and probit analysis (Bellovary et al., 2007). Unlike Altman's (1968) model, which assigns scores to classify observations as good or bad payers, Ohlson's (1980) model calculates the default probability of the prospective borrower (Altman et al., 2017). Although initially designed for industrial firms, the model is now considered an industry-agnostic model, similar to Altman's original 1968 model (Bellovary et al., 2007). In line with Altman, Ohlson applied a

[5] Out-of-sample generally refers to the performance or testing of a model on data that was not used during the model's training phase. It is important for assessing a model's ability to generalize to real-world situations beyond the data it was trained on.

legalistic definition of failure stating that the firms: "… *must have filed for bankruptcy in the sense of Chapter 10, Chapter 11, or some other notification indicating bankruptcy proceedings*" (Ohlson, p. 114, 1980). The sample consisted of 2,163 listed industrial firms, 105 bankrupt and 2,058 non-bankrupt, with data from the period 1970 to 1976. Exclusions were made to small and privately held corporations as well as utilities, financial services companies, and transportation companies. Three sets of estimates were computed for the logit model: (i) one year prior to bankruptcy, (ii) two years prior to bankruptcy, and (iii) one or two years prior to bankruptcy. Although Ohlson did not attempt to create any "new or exotic" ratios, his model stood out from its peers by incorporating unique independent variables (Ohlson, 1980). For example, it emphasized the significance of firm size, which was found to be one of the main drivers of failure, as well as the development of profitability over time. Furthermore, Ohlson's study differed from others at the time as he was able to consider the timing issue by assessing whether the company underwent bankruptcy before or after the release date of the financial statements. Several earlier studies have overlooked this issue by assuming the financial statements were available at the fiscal year-end date (Ohlson, 1980).

Table 1. Overview of pioneering bankruptcy prediction models

Period	Data	Model	Dependent variable	Independent variable	Accuracy
Beaver (1966)					
1954-1964	79 failed 79 non-failed	UDA	Failure: (i) bankruptcy, (ii) bond default, (iii) an overdrawn bank account, or (iv) nonpayment of a preferred stock dividend	30 financial-statement ratios (incl. cash-flow, profitability, capital structure, liquidity, turnover)	50% to 92%
Altman (1968)					
1946-1965	33 failed 33 non-failed	MDA	Filing for bankruptcy under Chapter X of the National Bankruptcy Act	5 ratios (incl. liquidity, leverage, solvency, profitability, activity)	Year before failure: (1) - 95% (2) - 72% (3) - 48% (4) - 29% (5) - 36% Hold-out - 79%
Ohlson (1980)					
1970-1976	105 failed 2,058 non-failed	Logit	Filing for bankruptcy under Chapter X, Chapter XI, or other notification indicating bankruptcy proceedings	9 ratios, (incl. size, changes over time, profitability, capital structure, liquidity)	Year before failure (1) - 96% (2) - 96 % (1 or 2) - 93% 1* - 96.3%
Skogsvik (1990)					
1966-1979	51 failed 328 non-failed	Probit	Failure: (i) Bankruptcy or composition agreement, (ii) Voluntary shut-down of primary production activity or (iii) Receipt of a substantial state subsidy	17 HCA** ratios from 71 initially (incl. profitability, cost structure, capital turnover, liquidity, asset structure, financial structure, growth)	Year before failure: (1) - 83.3% (2) - 78.4% (3) - 74.7% (4) - 73.9% (5) - 74.6% (6) - 73.3%

***** Including two additional factors. ****** The author also tested current cost accounting (CCA) ratios.

With regards to the predictive ability one year prior to bankruptcy, the model achieved an accuracy of 96% and Ohlson identified four factors influencing the probability of failure: "*(i) the size of the company; (ii) a* *measure(s) of the financial structure; (iii) a measure(s) of performance; (iv) a measure(s) of current liquidity (the evidence regarding this factor is not as clear as compared to cases (i)- (iii))"* (Ohlson, 1980, s. 110).

In 1984, Zmijewski proposed a new logit model (similar to Ohlson's), although with two important contributions pertaining to the data selection phase. The author argued that since bankruptcy prediction models are typically estimated on non-random samples, it can result in biased parameter and probability estimates unless appropriate estimation techniques are used. Zmijewski identified two biases: (i) the bias of using a matched-pair research design, since the real population of bankrupt vs. non-bankrupt firms is not balanced, and (ii) the bias of excluding firms with missing data, arguing that incomplete accounting data is correlated with a higher risk of bankruptcy, which makes for a non-representative sample. By examining 17 bankruptcy prediction studies, the author found that 12 of the studies used a matched-pair design to collect the non-bankrupt sample. Furthermore, all 17 studies used a share of bankrupt firms far above the actual population, with only three of the studies using shares lower than 40%. Therefore, it is likely that these studies constructed models using non-random samples with compositions significantly different from that of the overall population, without proper adjustments (Zmijewski, 1984).

Numerous, subsequent studies have attempted to replicate Altman's and Ohlson's models in new settings (e.g., Hillegeist et al., 2004 and Griffin & Lemmon, 2002). Hillegeist et al. (2004) examined whether publicly available information regarding the probability of failure is effectively captured in accounting-based measures, particularly Altman's (1968) Z-Score Model and Ohlson's (1980) O-Score Model. The authors developed a market-based measure of the probability of bankruptcy using the Black-Scholes-Merton option-pricing model (referred to as BSM-Prob), which is then compared to the two scores in terms of their relative information content. The results indicate that significantly more information is provided by the BSM-Prob compared to the two accounting-based measures. The authors argue that accounting-based bankruptcy prediction models often lack measures of asset volatility. Without accounting for asset volatility, these models may not accurately assess the risk of firms defaulting on their debt, even when they have identical leverage ratios. According to the authors, both Altman's (1968) and Ohlson's (1980) prediction models fail to include volatility as a variable (Hillegeist et al., 2004).

Furthermore, the higher information content of the BSM-prob also holds true after making several modifications to the Z-Score Model and O-Score Model, for example by revising the coefficients and making industry adjustments. The authors conclude that many of the coefficients from Altman's and Ohlson's models are outdated. Their findings are in line with Begley et al. (1996), who contended that the widely used models rooted in Altman (1968) and Ohlson (1980) had lost accuracy and proposed improvements in the modelling of default risk. Around this time, following advancements in computer technology, a new family of models was made available for the task of bankruptcy prediction. AI models, such as neural networks, emerged and quickly became popular choices for researchers (Bellovary et al., 2007). Capitalizing on more recent technological progress, scholars and practitioners began exploring new tools to assess corporate failure in ways that were less susceptible to the limitations of the once prominent model (Altman et al., 2017).

2.3 Bankruptcy prediction using machine learning models

Despite having been the most popular approach for predicting corporate bankruptcy throughout the second half of the 20th century, statistical techniques such as logistic regression have been shown to have certain limitations (Kruppa et al., 2013). They come with several, potentially unrealistic, underlying assumptions (Tay & Shen, 2002) and can cause model misspecification if not dealt with correctly (Malley et al., 2012). For instance, Skogsvik (1990) performs several statistical tests to conclude that MDA is not an ideal approach for his sample of firms despite having been used by popular studies, with similar firm samples, prior to his. Consequently, when the access to more computational power for the masses introduced machine learning models as potentially useful techniques, researchers have increasingly turned away from statistical techniques in recent years (Clement, 2020). Machine learning models, being non-parametric, are not subject to the same set of underlying assumptions and are therefore less restrictive (Altman, 2017). Additionally, since credit risk assessment resembles pattern-recognition problems – a specialty of machine learning models – some researchers have emphasized the potential of their use to solve the bankruptcy prediction problem (Kruppa et al., 2013).

Broadly speaking, the key difference between the two approaches is that a traditional statistical model makes assumptions about data distribution while a machine learning model finds generalizable patterns (Bzdok et al., 2018). One approach cannot be hailed superior to the other. Instead, the optimal choice depends on the specific circumstances and perhaps most crucially – the data. To showcase how the approaches solve the same problem differently, let us use a concrete example. Consider a binary classification problem, where the task is to predict whether a passenger on the Titanic survived or not based on their demographic information. A formulation of the probability using logistic regression might look like this:

$$P(Survival) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 * Age + \beta_2 * Sex + \beta_3 * Siblings\&Spouses)}}$$

Where β_0, β_1, β_2 and β_3 are the independent variables estimated during training and P(Survive) is the probability of the passenger surviving. A decision tree-based machine learning model instead creates several decision trees to derive a prediction. The first decision tree might have the following information:

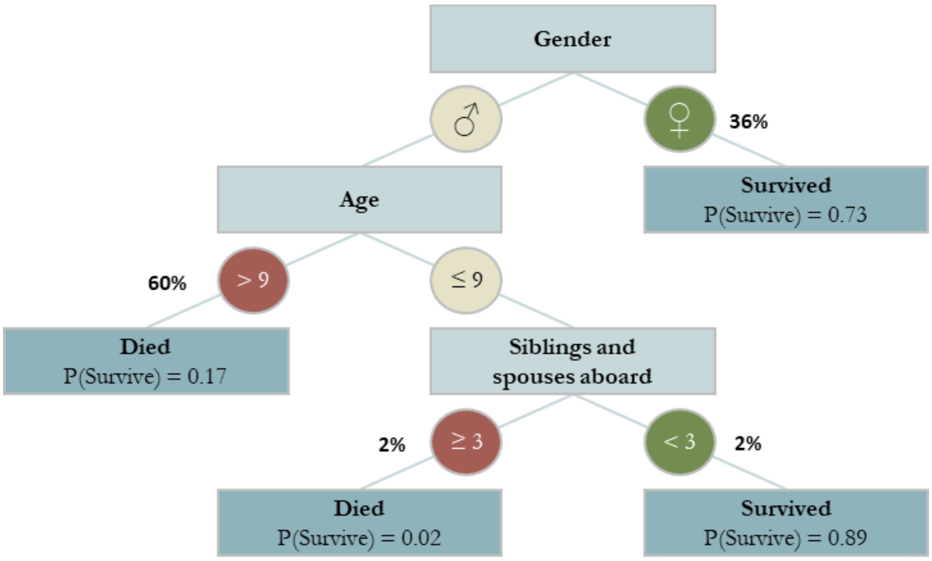

Figure 1a. Illustrative example of a potential decision tree classifying passengers on the Titanic as either survivors or deceased, based on gender, age and the number of siblings or spouses aboard. The number next to a node denotes the share of the sample following that path in the decision tree and the number in each end-leaf denotes the probability of survival. There is an implied cut-off of P(Survive), where a passenger is either classified as deceased or survived, between 0.17 and 0.73. The reason for the male-part of the tree having more layers is because the model could not effectively categorize men as dead or alive, as it could with women. A woman was likely to survive, solely based on gender, while a man was likely to die or survive based on both age and the number of siblings and spouses aboard.

Naturally, creating a single deep decision tree runs the risk of overfitting to the training data.[6] Evidence of this, in the context of bankruptcy, was found by Wang et al. (2012). The goal of training a model is not to

[6] Deep refers to how many layers a decision tree has, the more layers, the deeper the tree.

perfectly capture the differences between survivors and deceased in the training data, it is to find generalizable patterns from which it can make accurate and robust out-of-sample predictions. A solution to overfitting is to use an ensemble model, that combines multiple estimators to a combined prediction (Wang et al., 2011). Random Forest is a popular and well-known ensemble model that combines multiple decision trees, each trained on different parts of the data, to derive a single prediction.

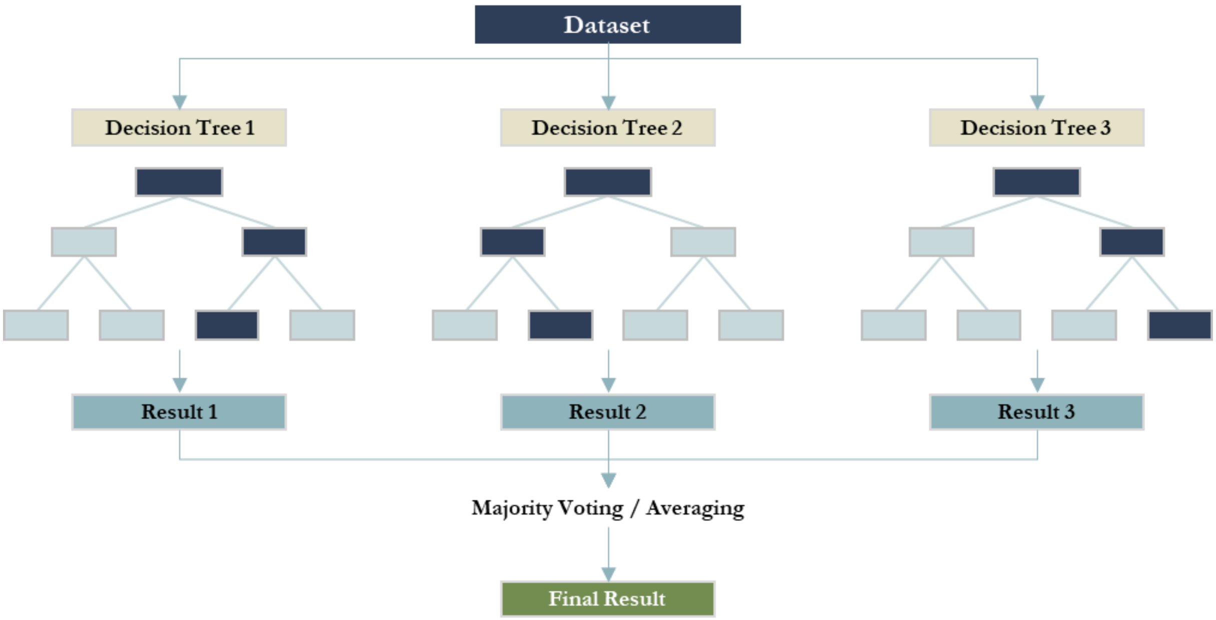

Figure 1b. Illustrative example of a potential Random Forest model. Each tree is given a random subsample of the dataset, causing it to construct different routes to classify the data. The results of all decision trees are then combined to a final result, typically by majority voting for classification tasks and averaging for regression tasks.

Once a model stops improving, the predictions of the trees are combined, which for an ensemble model typically means to simply average the probabilities in order to arrive at a final probability:

$$P(Survival) = \frac{Probability\ from\ Tree\ 1 + Probability\ from\ Tree\ 2 + Probability\ from\ Tree\ N}{N}$$

This method varies rather drastically from the statistical approach. The use of algorithms to find patterns makes machine learning models better at handling certain data structures (Bzdok, 2018). One such structure is "wide data", where the proportion of independent variables is high relative to the number of observations. Another structure is non-linear data, where the independent variables and the dependent variable do not show a linear relationship. While statistical models like logistic or polynomial regression can effectively capture nonlinear shapes such as curves or waves, an algorithm does not infer any probability distribution and is therefore even less sensitive to complex, nonlinear relationships (Kumar & Ravi, 2007). However, corporate failure is a complex occurrence and it was not certain at the time that machine learning models would be better equipped to predict it. As such, numerous studies were dedicated to exploring the efficacy of machine learning techniques for bankruptcy prediction.

Tam & Kiang (1992) employed Artificial Neural Networks (ANN) to forecast bankruptcy risk in the banking industry and subsequently conducted a comparative analysis with a linear discriminant model, a logistic regression model, decision tree (DT), and K-Nearest Neighbor algorithm (KNN). The authors describe a neural net as: "*a nonlinear discriminant function as a pattern of connections between its processing units*" (Tam & Kiang, p. 926, 1992). The study is based on data from Texas banks one and two years prior to failure over the period 1985 to 1987. A training sample of 118 banks were selected for each period, of which 59 failed and 59 non-failed. Failed banks were matched with non-failed banks based on their respective asset size, number of branches, age, and charter status. The authors used 19 financial ratios to describe each bank (Tam & Kiang, 1992). With regards to predictive accuracy, adaptability and robustness, the authors conclude that neural nets can be used to evaluate the risks in the banking industry. However, several limitations of using ANN in forecasting bankruptcy are discussed. First, when the number of network layers is large relative to the training sample, the ANN runs the risk of overfitting the network. Second, while neural networks can tell you the predictive accuracy of the training sample, they lack a prescribed method for deriving the relative importance of each input based on the weights. Assessing the individual inputs' significance can therefore become a constraint.

As one of the domain-defining researchers, Altman continued his research by examining how AI models could be used to predict corporate failure. He corroborates the observations of Tam & Kiang in his 1994 study (Altman et al., 1994). The study compared Linear Discriminant Analysis (LDA) and neural networks (NNs) for evaluating the financial data of approximately 1,000 companies registered in the Italian database Centrale dei Bilanci over the 1982-1992 period. While NNs showed strong performance, concerns were raised regarding their black-box nature and occasional production of seemingly illogical conclusions. LDA was ultimately deemed superior due to its transparency and interpretability. The study recommended further exploration of NNs, suggesting they could be a valuable technique when their limitations, particularly related to transparency, are addressed.

Although multiple machine learning models have been developed and applied to the bankruptcy prediction problem, Wang et al. (2014) argued that there was still no consensus around the overall best technique used for bankruptcy prediction. The authors concluded that the performance of prediction depends on multiple factors, including specific details of the classification problem, data structure, characteristics used, the degree to which the classes can be segregated using those characteristics, as well as the objective of classification (Wang et al., 2014). Altman co-authored a study commemorating the 50-year anniversary of his 1968 paper where he contributes to the discussion of an overall best technique (Altman et al., 2017). The authors focused on four promising machine learning techniques, namely support vector machines, bagging, boosting and Random Forest. Their analysis covered over 10,000 firm-year observations on North American companies over the period 1985 to 2013. Certain findings are interesting to highlight. First, unlike the 1994 paper by Altman et al., the authors now find that the machine learning models led to a notably higher prediction accuracy, amounting to a 10-percentage increase on average when compared to the traditional models. The difference in performance is enhanced further when the authors include several complementary financial indicators as predictive variables (e.g. change in price-to-book, change in return-on-equity, operating margin, and growth in sales, assets and number of employees). Second, the results suggest that bagging, boosting and Random Forest are the best predictors of bankruptcy one year prior to default, for firms with the same characteristics as in their sample (Altman et al., 2017).

Apart from the highlighted studies, there have been numerous researchers contributing to the literature. Several studies have dealt with various types of non-parametric models such as Extreme Learning Machine (Yu et al., 2014), Fuzzy-set Qualitative Comparative Analysis (Boratyńska & Grzegorzewska, 2018), various types of Neural Networks including Feed-forward, General Regression, Multilayer, Multilayer Perceptron and Recurrent Neural Networks (e.g. De Andrés et al., 2011; Song, Cao & Zhang, 2018; Korol, 2019; Tsai, Hsu & Yen, 2014; Ozbayoglu, Gudelek & Sezer, 2020), Gaussian Processes (Antunes, Ribeiro & Pereira, 2017), Classification and Regression Tree (Durica, Frnda & Svabova, 2019), Decision Rule Inducer (Parsania, Jani & Bhalodiya, 2014) and k-Nearest Neighbor (Kruppa et al., 2013). In 2020, Clement conducted a systematic review of recent machine learning techniques, analyzing 32 texts published between 2016 and 2020. The findings of the review align with Wang et al.'s (2014) conclusions that there is a lack of definitive evidence supporting any single model as the superior choice (Clement, 2020). The author underlines the need for further research and exploration.

2.4 Review papers comparing techniques

While the overall objective of the bankruptcy prediction models has remained relatively unchanged since Beaver's 1966 study, the techniques employed have been numerous and the subject of much debate. A recurring observation is that the optimal technique depends very much on the specific circumstances and data structure (Kumar & Ravi, 2007). The best model among several might be different for one researcher than for another, because they tweak the models differently or use different data. To investigate this issue, numerous articles have been published that compare techniques and offer methodological guidance.

In Dimitras et al.'s 1996 survey of business failures, the primary focus was to compare a comprehensive array of viable prediction techniques spanning from 1932 to 1994, predominantly encompassing statistical methods but also incorporating a few, popular at the time, AI approaches. Their selection method involved the examination of 47 journal articles featuring models or industrial/retail applications related to business failure prediction. The survey revealed that many of the prediction methods emerged primarily after the 1980s, aimed at addressing the limitations of discriminant analysis (DA) that had been identified at this point in time. It was observed that the most crucial financial ratios for prediction were from the solvency category (e.g. WC/TA, TD/TA), with profitability ratios also playing a significant role, signifying the vital role of a firm's ability to generate and retain profits.

O'Leary (1998) reviewed studies on predicting corporate failure using neural networks, which had become popular at the time. The primary comparison revolved around the performance of neural networks in comparison to traditional statistical techniques. The selection method involved an analysis of 15 articles that presented research using neural networks for bankruptcy prediction, covering the 1980s to the 1990s. Notably, the study found that in certain cases, neural networks performed on par with or even better than statistical methods, but in some specific settings, were still found to be inferior. The study suggested that there is potential for further development and refinement of neural network models in the context of corporate failure prediction, indicating that with time, the new techniques would consistently outperform the old ones.

In their review titled "Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques", Kumar & Ravi (2007) comprehensively examined bankruptcy prediction methods spanning from 1968 to 2005, encompassing an extensive range of techniques. Their selection method involved scrutinizing papers published in peer-reviewed journals, international conferences, and edited volumes across various domains, including accounts, finance, management, operational research, neural networks, expert systems, and decision support systems. The review yielded several noteworthy conclusions: bankruptcy research had evolved, and now considered a greater number of independent variables; the focus of predictions had shifted towards corporates over banks; data between 1980 and 2003 was most commonly analyzed; neural networks tended to outperform statistical techniques, as neural networks can be likened to a combination of parallel logistic regressions, thereby contradicting the findings of O'Leary (1998); decision trees were a favored machine learning model due to their transparency, aiding researchers in understanding the rationale behind predictions; and both neural networks and decision trees exhibited superior performance in different datasets, leading to a reduced preference for statistical techniques due to their lower accuracy in bankruptcy prediction.

In recent years, focus has shifted from trying to find the superior technique to realizing that each model comes with its benefits and drawbacks (Clement, 2020). In their 2018 systematic review Alaka et al. conducted an extensive analysis of bankruptcy prediction techniques, focusing on the eight most used methods (comprising two statistical and six intelligent techniques) and their performance across 13 criteria. They selected 49 studies, drawing from the papers of three scientific journals published between 2010 and 2015. A motivation for the study, as with other review papers comparing techniques, was to provide an understanding of the attributes of the techniques used to develop bankruptcy prediction models as well as their shortcomings. The review showed that many of the techniques are used in the wrong situation or with the wrong data conditions. No single technique outperformed the others consistently across the 13 criteria, highlighting the need to choose a technique based on the specific circumstances and objectives rather than pursuing a one-size-fits-all approach.

2.5 Bankruptcy prediction in Sweden

Most bankruptcy prediction research up until 2005 had been conducted on listed US firms (Bellovary et al., 2007). Since then, more attention has been paid to private firms from other regions (Alaka et al., 2018). There is still an ongoing debate about the generalizability of models with regards to the performance loss when the models are used on firms in different countries, industries and sizes than what they were trained on (Altman et al., 2017). While one might equate an untested sample of firms to a research gap, one must first reasonably ascertain some novel characteristic of the sample firms. Some differences are more apparent. For instance, Altman's original model includes the market value of equity, which poses a practical challenge to apply on private firms. Models developed for specific industries could also be expected to perform worse on other industries where the normal ratio levels vary drastically (Smith & Liou, 2007). With other differences, such as the country of origin, it is more difficult to conclude that a model cannot be generalized. A previous study by Ooghe & Balcaen (2007) found that with re-estimated coefficients, some models were widely usable in other countries while other models performed far worse. Altman (2017) found that bankruptcy year and size were two factors that strongly impaired the performance of the Z-Score Model, while country and industry did not. These mixed findings warrant a further exploration of the generalizability of models, especially with regards to the comparative performance of statistical and machine learning techniques.

Although data availability in Sweden is very good, especially on private firms, the body of bankruptcy prediction research remains relatively limited. Some studies have been conducted, exploring both a Swedish setting holistically (e.g. Yazdanfar, 2008) and specific hypotheses (e.g. Skogsvik, 1990 or Ivanova, 2023). Using probit analysis, Skogsvik (1990) was the first to explore bankruptcy prediction in a Swedish setting. The article's focus was to empirically test how current cost accounting information (CCA) could predict bankruptcy – as opposed to historical cost accounting (HCA). Skogsvik conducted an empirical test of the ability of current cost accounting ratios to predict business failure on a sample of 51 failed and 328 non-failed Swedish industrial firms. Failure companies were identified using the following three criteria: (i) bankruptcy and/or a composition agreement, (ii) voluntary shutdown of the primary production activity, and (iii) receipt of a substantial subsidy provided by the state. In groups of six or seven companies, the non-failure firms were matched with the failure firms. Using probit analysis, prediction models were then estimated for each year leading up to bankruptcy, covering all the way back to six years prior to failure. Based on Swedish data from the period 1966 to 1979, the author used principal component analysis to analyze a total of 79 CCA ratios and 71 HCA ratios. The analysis resulted in 20 CCA- and 17 HCA-components being generated, covering everything from profitability, cost structure and capital turnover to liquidity, asset structure, financial structure, and growth. The comparison of predictive performance between CCA and HCA was evaluated based on estimated error rates. A holdout sample was used, in which the companies were classified based on calculated probabilities each year prior to failure as either "Error type I: Prediction of non-failure for a failure company" or "Error type II: Prediction of failure for a non-failure company" (Skogsvik, p. 146, 1990).[7] The empirical results indicate that CCA ratios are rather similar to HCA ratios with regards to predictive performance, although to a slight advantage of CCA except for five and six years prior to failure (Skogsvik, 1990).

The most recent contribution to the bankruptcy prediction research in Sweden is Ivanova et al.'s (2023) study, which adds a qualitative variable to popular models to investigate the relationship between directors' and CEOs' prior corporate bankruptcy experiences and financial risk. The authors used a large sample of more than one million firm-year observations covering Swedish private firms from 1998 to 2014. By modifying Altman's Z-Score Model and Ohlson's O-Score Model to include executives' corporate bankruptcy experience measures, the study investigates if prediction accuracy can be improved. The authors find that companies with directors and CEOs that have prior experience from bankruptcies tend to have a higher risk of failure, riskier financial policies, and higher cost of debt (Ivanova et al., 2023).

Additional research within a Swedish context primarily consists of various bachelor's and master's theses. These studies explore the specifics of Swedish samples and the application of established models. For instance, a recent study investigated private retail firms in Sweden (Larsson & Lindhout, 2021), investigating the applicability of Altman's Z-Score Model and Ohlson's O-Score Model on Swedish private retailers. Similarly, Charraud & Garcia Saez (2021) tested the applicability of Altman's, Ohlson's, and Zmijewski's models across approximately 350,000 active Swedish companies between 2017 and 2018. Surprisingly, neither they nor any other study has explored the effectiveness of machine learning techniques in a Swedish setting. The only exception is Seidu's (2015) master's thesis, which examined the classification capabilities of a Gaussian processes model. Yet, given that only one model was tested on a sample of 2,000 firms, it is evident that there is much more to discover in the area of machine learning techniques to predict bankruptcy on Swedish firms.

2.6 Summary of findings

Concluding the literary review of bankruptcy prediction models, several key insights have emerged that illuminate our understanding of this complex field. Firstly, we drew on theory to determine the necessity of these models in assessing the capital investment risks accurately. Traditional valuation methods, notably the Discounted Cash Flow (DCF) analysis, have been consistently found to fall short in adequately accounting for the nuances of bankruptcy risk. This gap underscores the importance of integrating more sophisticated – but also more accessible – risk assessment methodologies into financial analyses. A retrospective look at the development of bankruptcy prediction models reveals a significant evolution from the early empirical models introduced by Beaver and Altman. These foundational models, leveraging financial ratios to predict bankruptcy, have paved the way for more advanced approaches. It is evident from their results that later models, such as Ohlson's and Skogsvik's, provided a more layered and accurate assessment of bankruptcy risk with methodological improvements.

Literature points to a pivotal advancement in the field – the emergence of machine learning models. The combined literature seems to indicate a superiority of these models for bankruptcy prediction, attributed to their flexibility and pattern recognition capabilities. However, numerous studies across a long period have pointed to flaws in their applicability and usefulness, showing that there still is no consensus on their superiority. A more harmonized finding from the comparative analyses is the lack of a one-size-fits-all technique. The effectiveness and applicability of a model are highly context-dependent, influenced by the specific economic environment, the nature of the business under consideration, and the availability and

[7] An important clarification is that for our study, we define a type 1 error as an incorrect prediction of bankruptcy (false positive) and a type 2 error as an incorrect prediction of non-bankruptcy (false negative). Skogsvik (1990) uses the opposite definition.

structure of data. The literary review also sheds light on the relevance of these models within specific contexts, such as the Swedish market. The adaptability of bankruptcy prediction models to various regulatory and economic environments is emphasized as a critical factor. A useful piece of the puzzle is therefore knowing how a machine learning technique compares to a statistical technique when it comes to not only prediction accuracy, but also the ability to identify useful information to train on and be generalizable across different types of firms.

3. Research design

The research design is structured as follows: first, we present the sample and data collection process, followed by the description of the dependent and independent variables used, including key decisions we have made. Then, we present and motivate our choice of models for the comparison of techniques, as well as the settings in which they were trained and tested. Finally, we conclude the chapter by disclosing how the models were evaluated before presenting initial data analysis, showing key characteristics and differences between the bankrupt and non-bankrupt firms. The aim of the research design is to create a setting in which we can investigate our four research questions accurately and reliably. To do this, we drew inspiration from Skogsvik (1990), replicating his study but comparing a machine learning model to a statistical model instead of a CCA-model to an HCA-model. Utilizing the method of an existing study, as opposed to creating our own, makes for a more efficient mode of comparison as we can use the study to guide our own research design. While other bankruptcy prediction studies have often been the subject of a comparative study and would be suitable for a comparison, we find it especially fitting to use Skogsvik's model. Firstly, because it allowed us to conduct the study on a sample of Swedish firms without risking that the potentially unique characteristics of Swedish firms play a too large part in the prediction results. Secondly, because Skogsvik identified a large set of potentially useful independent variables, we could use these to investigate our second research question, pertaining to the technique's ability to extract useful information. Thirdly, because no study has investigated Skogvik's study to the same extent as in this study, we add novelty to our contribution, which might be diminished had the study been conducted on Altman's or Ohlson's model.

While our research design attempts to replicate Skogsvik's as closely as possible, some adjustments had to be made, which we carefully disclose. Skogsvik's methodology will not be accounted for in full, so we refer readers to his original dissertation for all details.

3.1 Data and collection process

Data availability for firms, especially private ones, is much better today than it was at the time of Skogsvik's data collection, resulting in our study having a larger sample. Skogsvik's sample consisted of 379 companies that met the following criteria:

(i) Joint-stock limited company ("Aktiebolag"),
(ii) Classified as either mining or manufacturing business,
(iii) During the period 1966-1971 had either more than 200 employees, or
(iv) During the period 1966-1971 had assets of at least 20 million SEK[8], and
(v) Not subject to acquisition, fusion or any similar business combination.

Our data was collected from the Serrano Database – an extensive database on Swedish private (non-listed) firms, containing financial statement information and information on bankruptcy filings. It is based on financial statement data from the Swedish Companies Registration Office (Bolagsverket), including companies' income statement and balance sheet items. Additionally, general company data is sourced from Statistics Sweden (SCB), bankruptcy information from the Swedish Companies Registration Office, and group data from Bisnode. The Serrano Database covers most legal forms in the Swedish business community. As of October 20th, 2023, there were more than 1,200,000 firms in the total dataset, with data spanning from 1932 up to 2021. By gathering bankruptcy data from 2022 and 2023, we were able to use the Serrano data all the way up to 2021. This additional data was sourced from *Kreditrapporten*, a publicly accessible compilation of bankrupt firms. After applying the same criterion as Skogsvik, we end up with a

[8] 20 million SEK at price levels prevailing in the 1970's. Therefore, we inflation-adjusted all firm-year observations before applying the criteria.

sample of 192,668 firm-year observations, spanning from 2000 to 2021. All observations preceding the year 2000 were excluded to prevent the models from being trained on outdated business trends. The selection of a specific year is somewhat arbitrary, and it would be equally justifiable to exclude firms from 2002 and prior.

To answer the research questions pertaining to the accuracy of models on firms different from the training sample, data for two additional samples was collected, to be used in Setting B and Setting C. For Setting B, we inverted criteria (ii), effectively creating a sample of all non-industrial firms otherwise fulfilling the criterion of size and legal form. For Setting B, we instead inverted criterion (iii) and (iv), thereby creating a sample of small and medium sized industrial firms.

3.2 Variables and techniques

3.2.1 Dependent variable

Our study focuses on one specific economic event to serve as an indicator of financial distress: bankruptcy filings. In Sweden, a bankruptcy filing marks a critical juncture where management and shareholders promptly relinquish control of the company, leading to it either being liquidated or integrated into a bankruptcy auction buyer's receiving company, effectively culminating its operations (Thorburn, 2000). Another type of firm failure that could have been relevant to include in the definition of failure is composition (Sw. Företagsrekonstruktion). Apart from being an alternative to bankruptcy filings and therefore a sign of failure, a new law on corporate composition took effect in Sweden in 2022.[9] The law both lowers the requirements for a company to enter composition and allows the composed company to terminate lengthy contracts, such as rental contracts. Surprisingly, the number of compositions remains at a low level since the law took effect, despite the number of bankruptcy filings being at its highest in twenty years (Karlsson-Tuula, 2023). Ultimately, because composition data was not possible to collect from *Kreditrapporten* or elsewhere, we used bankruptcy filings as the sole indicator of failure. Skogsvik used a broader scope of bankruptcy in his study, with a definition that includes compositions, a board decision to shut down production activity as well as various types of government support beyond a simple bankruptcy. Yet again, due to the sheer volume of data, we were limited to using readily available variables, preventing us from incorporating such a broad definition.

With the publication of financial accounts often lagging several months after the books have been closed, there is a risk that a company will file for bankruptcy before their numbers are published for the prior fiscal year. This warrants a methodological decision on how to adjust for the chance of a bankruptcy filing happening before the financial statements for the previous year are published. There are three options: (1) validating that the annual report was publicly available before the bankruptcy for each company in the dataset, (2) not making any adjustments or (3) adjusting so that the model can only look at data up to two years before the bankruptcy, thereby creating a “safety margin”. Option one is undoubtedly the preferable option, and what Skogsvik did in his study, but practically impossible given the large amount of data. Serrano does not include the publication date of the annual reports, meaning they would have to be collected manually. Option two and three both have pros and cons, respectively.

Option two involves utilizing all available data, including post-bankruptcy information. This approach offers several advantages, notably the capacity to leverage a comprehensive dataset, potentially enhancing the predictive accuracy of the model. Moreover, it mirrors the practical scenario in which stakeholders, such as creditors and investors, may possess access to non-public information, thereby enabling more informed decision-making. Nevertheless, a significant drawback is that it may render the model less

[9] Lag (2022:964) om företagsrekonstruktion

effective for real-time bankruptcy event prediction or for forecasting future periods, as it incorporates data that was not publicly available at the time of the bankruptcy event. Consequently, the alignment with the objective of creating a model that accurately reflects the information available at a specific historical point may be compromised.

Conversely, option three restricts the data used for modeling to pre-bankruptcy information exclusively. This choice maintains congruence with the information available up to the moment of bankruptcy filing, rendering the model more attuned to predicting bankruptcies as they would have been projected at that precise juncture. Notably, it upholds the temporal validity of the model, making it more applicable for the anticipation of future bankruptcy events. Nevertheless, the trade-off lies in the potential exclusion of valuable information that could otherwise augment the model's predictive accuracy, possibly resulting in reduced predictive performance.

The decision between option two and three hinges on the research objectives, reflecting the trade-off between comprehensiveness and temporal alignment of the data used in bankruptcy prediction modeling. In line with previous research (e.g. Ivanova et al., 2023 and Ohlson, 1980), and because this study has not stated a specific purpose of developing a model for stakeholders with insider information, we opted for option three. Another potential benefit of option three is its potential to enhance the model's sensitivity to early warning signs of financial distress. By excluding immediate pre-bankruptcy financial data, this approach compels the model to seek out other, potentially earlier, indicators of deteriorating financial health. Rather than relying solely on the conventional income statement and balance sheet metrics, the model is driven to consider a broader range of financial and non-financial factors that may serve as leading indicators of distress. These could encompass shifts in working capital, declining profitability, changes in management or business strategies, and other factors that precede financial distress. This approach offers the advantage of potentially providing more advanced warnings of financial troubles, aligning more closely with the decision-making needs of stakeholders who require early signals to make informed choices. It underscores the trade-off between data comprehensiveness and the potential for timelier, more sensitive predictions in the realm of bankruptcy prediction modeling.

This study aims to contribute to bankruptcy prediction literature by leveraging the latest data available. The Serrano Database, current as of October 20th, 2023, provides financial data up to 2021. With our choice to include a safety margin for our prediction horizon, the models would be able to analyze financial performance up to 2019, before making a prediction for up to 2021. However, by including bankruptcy data from *Kreditrapporten* from 2022 and 2023, the models can look at data up to 2021, before making predictions up to 2023.

3.2.2 Independent variables

For our independent variables, we turn to Skogsvik's prior research. In his doctoral dissertation, Skogsvik (1987) outlined a large set of key ratios from accounting components that were originally identified by Pinches et al. (1973). The components were seven in total and pertained to (1) profitability, (2) cost structure, (3) capital turnover, (4) liquidity, (5) asset structure, (6) capital structure and (7) growth. Altogether, the components contained 71 ratios, of which 69 were financial ratios and 2 absolute measures of size. These constituted Skogsvik's primary set of independent variables and our starting point, from which we assembled four models.

Due to constraints in the data, several of Skogvik's original independent variables required modifications while a few ratios had to be excluded altogether. In most of these cases, a specific entry from the balance sheet or income statement was absent in the dataset. Upon our assessment, this absence was deemed unlikely to significantly impair the variable's informational content. Consequently, we opted to retain the feature in its modified form. The only data limitation with a considerable impact was the absence of detailed categorization of the various untaxed reserve items – only the aggregated sum was available. Skogsvik utilized different components of untaxed reserves to compute an alternative measure of a firm's implied tax liability. Among the 71 independent variables, several functioned as pairs or triplets, differing only in their measurement of implied tax liability. As such, all pairs and triplets containing these alternative measures were consolidated into single variables. Subsequently, Skogsvik introduced 12 new variables in a later phase of the study. These variables were normalized versions of existing ones, assessing deviations from a firm's historical benchmarks in metrics such as profit and solvency. Following our modifications and exclusions, our core set comprised 77 independent variables. For a comprehensive view of all variables and the modifications, please refer to Appendix 1 and Appendix 2.

3.2.3 Statistical technique

Probit regression is a type of regression where the dependent variable can only take two values. In the context of financial crisis forecasting, probit and logit analysis assumes that a certain (unobservable) financial crisis index exists and that this can be determined as a linear function of certain independent variables (Skogsvik, 1990). Probit regression is statistically rigorous, built on a solid foundation with assumptions like a normal distribution for the error term, and is robust to outliers, making it suitable for financial data that may exhibit variability over time. Probit regression allows for the comparison of variable coefficients, offering insights into the relative importance of each predictor. It assumes linearity between independent variables and the latent variable, and its interpretability can be challenging for non-statisticians. As highlighted in our literature review, Skogsvik (1990) justifies the choice of probit by noting that the data did not meet all the necessary conditions for the proper application of MDA, the prevailing method for bankruptcy prediction at that time.[10] Consequently, he opted for probit analysis due to its better alignment with the data characteristics, acknowledging that there were no essential grounds for choosing probit over logit. Our selection of probit as the statistical technique aligns with Skogsvik's methodology in his research.

3.2.4 Machine learning technique

XGBoost, an ensemble learning algorithm, is a popular choice for predicting corporate bankruptcy and various other classification tasks. XGBoost has received attention in the latest years for its exceptional predictive performance and versatility (e.g. Nielsen, 2016). Similar to Random Forest, XGBoost is an ensemble model that combines multiple decision trees. A key difference, which is also a reason behind its superior performance, is how it combines the trees. While Random Forest simultaneously trains individual trees and combines their predictions, XGBoost sequentially trains trees, enabling each subsequent tree to learn from its predecessors' errors and thereby enhancing the model's predictive accuracy. This is achieved by employing a loss function, where each new tree aims to reduce the overall loss of the ensemble model. For an in-depth explanation of the XGBoost algorithm and the math behind it, we refer the readers to Chen & Guestrin (2016).

Relevant to this study is that XGBoost has proven robustness to overfitting, otherwise a common concern in predictive machine learning modeling. It is also able to produce feature importance scores that, albeit not as transparent and intuitive as the statistical significance of statistical models, can aid in the identification of key variables influencing bankruptcy prediction. We chose XGBoost as the machine learning technique for these reasons. However, it is important to note that other machine learning techniques, like support

[10] Multivariate Discriminant Analysis assumes homoscedasticity, which is the equality of variance-covariance for the independent variables between groups. In his study, Skogsvik could not reject the null hypothesis that there was heteroscedasticity.

vector machines, various bagging techniques or LightGBM would also make for a valid comparison to the statistical techniques.

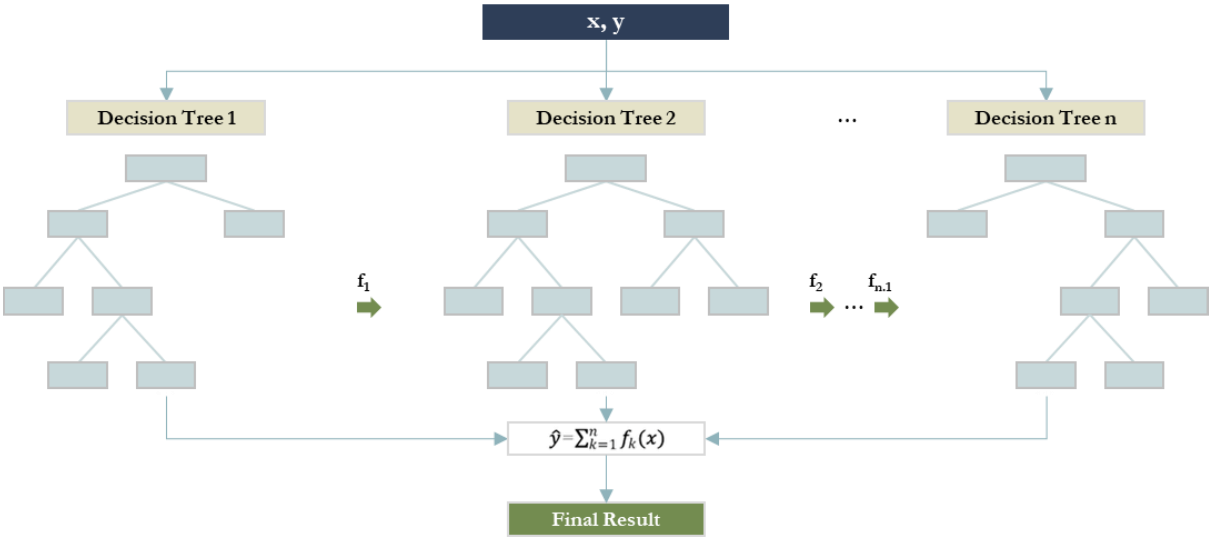

Figure 1c. Simplified illustration of the XGBoost architecture. A key difference for tree boosted models like XGBoost is that trees are created sequentially and not in parallel as with Random Forest (Figure 1b). This effectively allows for each new tree to learn from the previous trees' mistakes by sequentially combining the trees so that the new tree decreases the error term of the ensemble model.

3.3 Models and settings

This section explores the bankruptcy prediction models we developed and the settings in which they were evaluated. Four models were developed, two of which are based on probit analysis (referred to as Base Probit and Modified Probit) and two using XGBoost (referred to as Base XGBoost and Modified XGBoost). The base models use a limited number of predetermined independent variables, while the modified models start with a larger set, from which the techniques use different ways to select the most useful variables for their final model.

3.3.1 Model 1: Base Probit

Relying on previous research, domain knowledge and statistical techniques such as principal component analysis (PCA) and univariate tests, Skogsvik narrowed down his features from the original 71. Using PCA, he identified 17 variables that explained approximately 80 percent of the total variation in the original set of ratios. Among these 17 variables, 5 captured changes over time while the remaining 12 measures were snapshots of the firm's performance at one point in time. Skogsvik argued that additional useful information could be extracted by creating normalized versions of the 12 ratios that measure deviations from a historical average. As a result, 12 new ratios were added to the set of candidate variables. Using a statistical selection method, mainly based on backward selection, Skogsvik narrowed it down to six independent variables for his 1-year forecast model. These were *Return on Assets R(1)T*, *Interest Expense R(1)Sk*, *Inventory Turnover Ratio TVL(1)*, *Solvency SD(1)*, *Change in Equity* E(1)' and *Normalized Interest Expense N.R(1)Sk*. Replicating Skogsvik, our Base Probit model is therefore:

$$
\begin{aligned}
probit\big(P(Bankruptcy_{it+2} = 1 | X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)\big) \\
= \Phi(\alpha_0 + \beta_1 R(1)T_{it} + \beta_2 R(1)Sk_{it} + \beta_3 TVL(1)_{it} + \beta_4 SD(1)_{it} + \beta_5 E'_{it} \\
+ \beta_6 N.(R(1)Sk)_{it})
\end{aligned}
$$

3.3.2 Model 2: Base XGBoost

The Base XGBoost model uses the same six independent variables as the Base Probit model, but with XGBoost as the classification technique. A comparison of the Base Probit model and the Base XGBoost model shows the difference in how probit and XGBoost can predict bankruptcy for the sample. They were estimated with the same dataset, using the same variables and evaluation metrics.

3.3.3 Model 3: Modified Probit

The next pair of models include a recalibration of independent variables. The Modified Probit is a recalibration of Skogsvik's original probit model, starting from the 17 + 12 variables he identified using PCA. From there, we repeat the backward selection process in an attempt to identify variables potentially more useful for the new data, since the original model was estimated on company data from the 1960s-1980s. Another option would be to start with the 77 variables in the primary set and repeat the principal component analysis. This is most likely a more methodologically sound approach, as the Modified Probit model would be a full re-estimation starting from the primary set. However, performing principal component analysis falls outside the scope of this study. Instead, we rely on Skogvik's PCA. We thereby assume that the 17 + 12 variables are still the most useful variables. This is not an unreasonable assumption, given that both samples consist of large Swedish industrial firms. The one altered characteristic is the time component, as our sample contains more recent firm-year observations. After backward selection, we end up with the following variables for our Modified Probit model:

$$\begin{aligned} probit\big(P(Bankruptcy_{it+2} = 1 | X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)\big) \\ = \Phi(\alpha_0 + \beta_1 R(1)T_{it} + \beta_2 R(1)Sk_{it} + \beta_3 SD(1)_{it} + \beta_4 OHT(1)_{it} + \beta_5 ATA_{it} \\ + \beta_6 MATA_{it} + \beta_7 LNT(1)_{it} + \beta_8 N.(R(5)E, es)_{it}) \end{aligned}$$

3.3.4 Model 4: Modified XGBoost

Finally, the Modified XGBoost model includes all the 77 primary variables and allows XGBoost to perform its own feature selection. This model shows not only a machine learning model's classification ability, but also its ability to find the most useful independent variables without the aid of a researcher. A comparison of the Modified Probit model and the Modified XGBoost model shows how the techniques differ in feature selection. Given a larger set of variables to choose from, the models were free to identify the most important variables to solve the problem. While the comparison lacks the same level of purity as the comparison of the Base Probit model and the Base XGBoost model, because the Modified Probit model does not start with as many variables to choose from as the Modified XGBoost, it still gives a sense of how effective the feature selection processes are. This could offer a valuable contribution to researchers curious to know the implications of feature selection using different techniques.

3.3.5 Setting A: Large industrials

Skogsvik (1990) analyzed Swedish industrial companies of the legal form joint-stock limited company, with either more than 200 employees or total assets of at least SEK 20 million (1970 price levels). We applied the same selection criteria to both our training sample (firm-year observations 2000-2015) and our main test setting (firm-year observations 2016-2021).

3.3.6 Setting B: Large non-industrials

The second setting explores research question (iii) by comparing the four different models' ability to predict bankruptcy for firms of different industries from the training sample. Ceteris paribus, the industry test examines how accurately the models can predict bankruptcy in industries they have not been estimated on.

It gives an indication of whether one technique is more susceptible to overfitting. The models' performance is disclosed for all industries, categorized by SNI-group.[11]

3.3.7 Setting C: SME industrials

The third setting explores research question (iv) by introducing firms of smaller sizes than the ones the models were trained on. As with the second setting, the models were asked to make predictions on firms with differing characteristics, this time in terms of size rather than industries. The normal levels of ratios for smaller firms are likely to vary, presenting a potential challenge for the models. The models' performance is disclosed for different firm sizes, using the number of employees as an indicator of size. The groups are 1-4 employees, 5-9 employees, 10-19 employees, 20-49 employees, 50-99 employees and 100-199 employees.[12]

3.4 Model evaluation

To evaluate the models' performance across the different settings, we primarily look at the Area Under the Curve (AUC), as it is a commonly used evaluation metric in current bankruptcy prediction literature and shows the trade-off between sensitivity and specificity for every possible cut-off. In addition, a decision rule is introduced to enhance the visibility of the models' classification performance. For each setting, the firms with a predicted probability of default in the top 3% of the distribution are classified as bankrupt while the rest are classified as healthy firms. With this rule introduced, confusion matrices are presented showing the models' predictions, accuracy, sensitivity and specificity. 3% should be a good level for illustrative purposes with regards to the specific samples, since the share of bankrupt firms is slightly smaller (0.5-1%) and the predictions will therefore likely yield a number of false positives and false negatives alongside the accurate predictions.

3.4.1 Descriptive statistics

In this section, we explore the potential differences in financial ratios between a bankrupt and a solvent firm using a simple mean analysis. Table 2 displays our results from the six independent variables employed in Skogsvik's (1990) model. However, a more detailed review of all the 77 independent variables identified by Skogsvik is presented in Appendix 1, along with our calculated mean results for each ratio.

[11] The Swedish Standard Industrial Classification (SNI) is used to classify enterprises and workplaces according to the activity carried out.

[12] The group with 100-199 employees was subsequently excluded due to a limited number of bankruptcy observations.

Table 2. Characteristics of bankrupt and non-bankrupt firms in their last reported financial statements

Variable		Bankrupt firms Mean	Non-bankrupt firms Mean
BANKRUPTCY		1	0
R(1)T	Return on total capital	-0.08	0.10
R(1)Sk	Interest expense	-0.04	-0.03
TVL(1)	Inventory turnover rate	0.22	0.18
SD(1)	Solvency	0.05	0.41
E(1)	Growth in equity	0.69	1.15
N.(R(1)Sk	Normalized interest expense	-1.17	-0.90
N = 124,279		1,128	123,151

The difference in means were significant beyond the 0.05 level for all variables tested.

Conducting a simple mean analysis in a bankruptcy prediction study helps identify average values of financial metrics over a certain period. These trends can highlight deviations or anomalies that might indicate financial distress. According to the information presented in Table 2, solvent firms tend to show higher profitability, healthier balance sheets as well as positive trends. One striking difference lies in the solvency levels and growth patterns of equity. On average, bankrupt firms display a solvency ratio of 5%, experiencing a substantial decrease in equity of over 30%. In contrast, solvent firms exhibit an average solvency of 41%, paired with a 15% increase in equity. Furthermore, return on capital is significantly negative for bankrupt firms as opposed to solvent firms.

4. Empirical results & analysis

This section explores the empirical findings, organized into two primary sections. First, we present the results from the training phase, then we present the actual results observed during the testing phase. Table 4 provides an overview of each model's performance across different settings. As outlined in the research design, four models underwent testing across three settings, resulting in a total of 12 tests conducted. Results for all four models are initially presented within a single setting before addressing their performance in subsequent settings.

4.1 Training phase

Our training set, after removing observations with missing values, consists of 124,279 firm-year observations of industrial firm annual reports in the years 2000-2015. The probit models were trained with the *statsmodels* Python package using the standard parameters, whereas the XGBoost models were trained using the *XGBoost* Python package. Since machine learning models tend to be more sensitive to overfitting due to erroneous hyperparameter configuration, we iteratively adjusted the settings in an effort to find the best configuration. In our final configuration, the XGBoost models both had a max depth of three in their decision trees and 300 training iterations. Each training iteration was only given a subsample of 70% of the training set.

As shown in Table 3, the Modified Probit model, using backward selection to construct the final model from Skogsvik's primary set of variables, ends up with 8 independent variables. Three of these were shared with the Base Probit model: *Return on Assets* R(1)T, *Interest expense* R(1)Sk and *Solvency* SD(1). In addition, the model selects: *Asset Turnover* OHT(1), *Proportion of Fixed Assets* (ATA), *Proportion of Tangible Assets* (MATA), *Size* LNT(1) and *Normalized Return on Equity* N.R(5)E,es.

To our surprise, the Modified XGBoost model does not exclude any of the 77 available independent variables. This indicates that all variables, to some extent, contain useful information. However, the difference in feature importance spans several orders of magnitude, with the *Return on Assets* variable R(3)T having a feature importance of 0.8% and the *Solvency* variable SD(2) having a feature importance of 7.2%.

The training phase indicates that the independent variables were of varying usefulness to the different models. For the Base Probit model, only two variables, *Return on Assets* R(1)T and *Interest Expense* R(1)Sk, were significant beyond the 0.05 level. The Modified Probit model indicated that more variables were useful, with six of the eight being significant. This, combined with the higher log likelihood score, which is a measure of the goodness of fit of the model in which a less negative value is better, would suggest that the additional five variables contain useful information for predicting bankruptcy. The difference in log likelihood between the probit models faded in comparison with the log likelihood scores of the XGBoost models, however. The Base and Modified Probit models have log likelihood scores of -6158 and -6054, respectively, while the Base and Modified XGBoost models have log likelihood scores of -3832 and -2393, respectively.

A few observations pertaining to the non-linearity of the data can be made. The *Solvency* variable SD(1), used by all four models, is statistically insignificant beyond the 0.05 level for the probit models, but the most useful variable for the Base XGBoost model and third most useful for the Modified XGBoost model (outperformed only by the *Flow of Capital Ratio* Kap(1) and the other *Solvency* metric SD(2)).[13] A potential conclusion for the seemingly low importance of solvency for the probit models is that while the linear effect

[13] The difference between SD(1) and SD(2), also disclosed in Appendix 1, is that the former excludes the implied tax liability from the untaxed reserves in the solvency ratio.

of solvency is not statistically significant, the XGBoost models' ability to capture non-linear relationships make the solvency variable more useful. For example, higher solvency might not always be better, but certain thresholds might indicate a lower bankruptcy risk which the XGBoost technique can use to accurately split the data in a decision tree. However, the difference in variable usefulness can have other explanations, such as potential overfitting.

Table 3. Training results of in-sample estimation using the four bankruptcy prediction models

Model / Variable	Base Probit	Base XGBoost	Modified Probit	Modified XGBoost
	Coefficient (Z-stat)	Feature importance	Coefficient (Z-stat)	Feature importance
Constant	-2.328 (-199.392)*	-	-2.384 (-28.661)*	-
R(1)T	-0.881 (-24.348)*	0.125	-0.902 (-24.007)*	0.011
R(1)Sk	-0.099 (-2.480)*	0.105	-0.120 (-2.994)*	0.013
TVL(1)	-0.003 (-0.335)	0.098	- -	0.013
SD(1)	-0.003 (-1.760)	0.242	-0.002 (-1.440)	0.032
E(1)	-0.003 (-1.614)	0.104	- -	0.014
N.R(1)Sk	0.000 0.302	0.089	- -	0.011
OHT(1)	- -	-	0.079 (8.168)*	0.010
ATA	- -	-	-0.152 (-2.780)*	0.010
MATA	- -	-	0.581 (11.097)*	0.010
LNT(1)	- -	-	-0.035 (-4.204)*	0.012
N.R(5)E,es	- -	-	-0.001 (-1.115)	0.010
N	124,279	124,279	124,279	124,279
Log likelihood	-6,158	-3,832	-6,054	-2,393

* Significant beyond the 0.05 level. **Feature importance** refers to an independent variable's ability to correctly split the data into bankrupt and healthy firms. The higher the number, the more useful the variable is in distinguishing bankrupt firms from healthy. With regards to the Modified XGBoost model, all 77 ratios are not presented in this table as that would be cumbersome.

4.2 Testing phase: AUC and other performance indicators

Overall, Table 4 shows how the two XGBoost models outperform the probit models across all three test settings when looking at the AUC-score. As mentioned in our choice of evaluation metric, the AUC captures the models' ability to trade-off specificity and sensitivity at any cut-off point. We applied a cut-off point at the 97th percentile, classifying the 3% of firms with the highest probability of default as bankrupt. However, depending on the objective of the test, the cut-off can be lowered in order to classify more bankruptcies correctly (increased sensitivity) at the cost of classifying more non-bankruptcies incorrectly (decreased specificity). The higher the AUC-score, the more a model can increase its sensitivity without sacrificing specificity. The Base Probit model is the only probit model to achieve an AUC-score as high as 80% (79.94% to be exact). However, this is only the case when it is applied to large industrial firms as part of Setting A. In contrast, both the Base and Modified XGBoost models consistently achieve an AUC-score of 80% or higher, regardless of the test setting.

Table 4. Overview of the four developed models and their three test settings

Setting / Model	**A. Large industrials**	**B. Large non-industrials**	**C. SME industrials**
	>200 FTEs or >SEK 20m total assets*		<200 FTEs and <SEK 20m total assets*
1. Base Probit	**A1** AUC: 80%	**B1** AUC: 72%	**C1** AUC: 75%
2. Base XGBoost	**A2** AUC: 86%	**B2** AUC: 80%	**C2** AUC: 82%
3. Modified Probit	**A3** AUC: 78%	**B3** AUC: 72%	**C3** AUC: 77%
4. Modified XGBoost	**A4** AUC: 87%	**B4** AUC: 85%	**C4** AUC: 84%

*1970 price levels

The most notable difference in performance arises when comparing the Modified XGBoost model with the Modified Probit model. In the case of large industrial firms (Setting A), the Modified XGBoost model achieves an AUC score that is 9.3 percentage points higher than its probit counterpart. This gap widens when applied to large non-industrial firms (Setting B), showing a substantial 13.8% advantage for the Modified XGBoost model. Interestingly, the closest instance where a probit model approached the performance of an XGBoost model was observed when applying the Base Probit and Base XGBoost models to large industrial firms, revealing a 6.5 percentage point superiority for the latter.

4.2.1 Setting A: Large industrials

Setting A, the study's primary test, displays the models' ability to predict bankruptcy for firms similar to what they were trained on. The only difference is time, with the training set consisting of firm-year observations from 2000-2015 and the test set of Setting A consisting of firm-year observations from 2016 to 2021.

The Receiver Operating Characteristics (ROC) graphs, plotting the true positive rate against the false positive rate, graphically represent the AUC-score, i.e. the trade-off between correctly predicting a bankruptcy versus erroneously predicting bankruptcy for a non-bankrupt firm. As can be observed in Figures 2a-2d, the Modified XGBoost achieved the highest performance in Setting A with an AUC-score of 87%, followed closely by the Base XGBoost model at 86%. Subsequently, there is a substantial gap in performance to the Base Probit model which, in turn, is closely followed by the Modified Probit model.

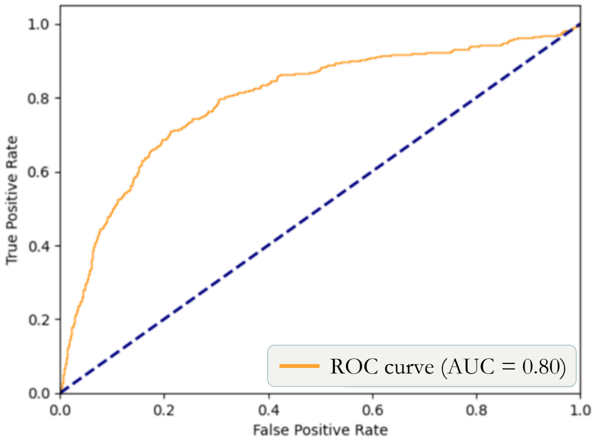

Figure 2a. ROC for the Base Probit model

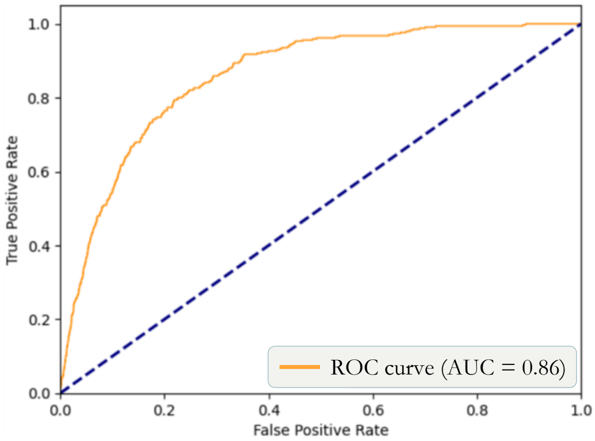

Figure 2b. ROC for the Base XGBoost model

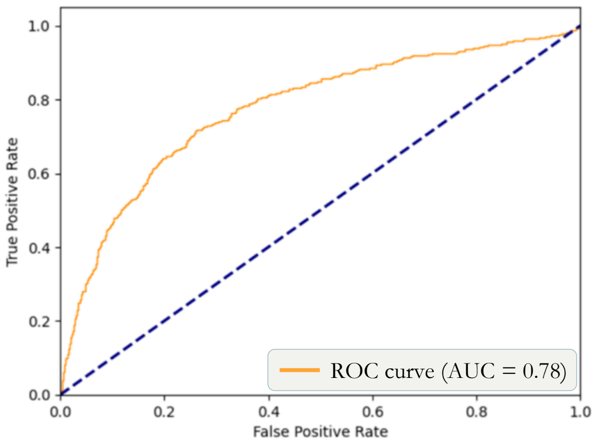

Figure 2c. ROC for the Modified Probit model

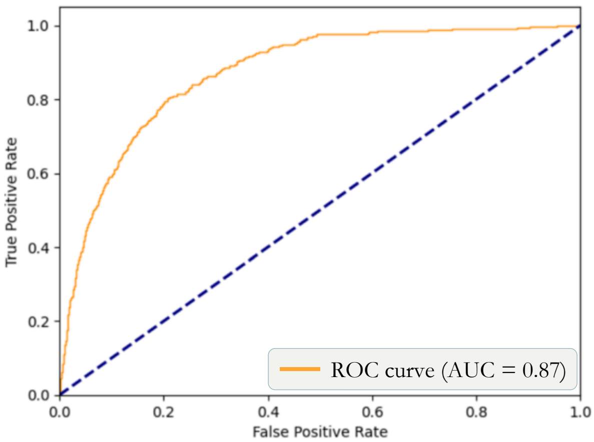

Figure 2d. ROC for the Modified XGBoost model

The performance gap between the two techniques is another sign of the difference in their inherent ability to extract useful information from the data. The Modified XGBoost performs the best, showing that access to more information improves performance rather than subduing the model to overfitting. However, the performance gain is relatively small, with an improvement of only 1 percentage point compared to the Base XGBoost model. The Modified Probit model performs worse than the Base Probit model out-of-sample, despite being subject to backward selection which resulted in better in-sample performance.

Looking at the confusion matrices, where the outcome of our classification rule is reported, the four models are all closely matched in both accuracy (overall rate of correct predictions) as well as specificity (rate of correctly identified non-bankrupt firms). This is driven by the overwhelming proportion of non-bankruptcies in the dataset (99.5%). The metric with more varying results is the sensitivity (rate of correctly identified bankrupt firms). For both probit models, the sensitivity is slightly below 21%, whereas the Base XGBoost model correctly identifies 25% of the bankrupt firms. The Modified XGBoost shows the strongest performance at this cut-off level, with a sensitivity of almost 30%.

Table 5. Classification results for the four models when applied to large industrials

N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Base Probit								
68,389	71	66,068	1,981	269	96.71%	20.88%	97.09%	79.94%
Base XGBoost								
68,389	85	66,082	1,967	255	96.75%	25.00%	97.11%	86.47%
Modified Probit								
68,389	70	66,067	1,982	270	96.71%	20.59%	97.09%	78.15%
Modified XGBoost								
68,389	100	66,097	1,952	240	96.79%	29.41%	97.13%	87.45%

Table 5 reports classification results using the decision rule in which firms with a score above the cut-off of 97% are predicted as likely going bankrupt. **"True Positive"** represents accurately predicted bankruptcies, **"True Negative"** represents accurately predicted non-bankruptcies, **"Type I Error"** represents non-bankrupt firms incorrectly predicted as bankrupt, **"Type II Error"** represents bankrupt firms incorrectly predicted as non-bankrupt, **"Accuracy"** is the share of accurate predictions out of all observations, **"Sensitivity"** is the share of predicted bankruptcies out of all bankruptcies and **"Specificity"** is the share of predicted non-bankruptcies out of all non-bankruptcies.

4.2.2 Setting B: Large non-industrials

With Setting B, we are asking the models to predict bankruptcies for firms of different industries than the sample they were trained on. As one might expect, the overall performance is worse than in Setting A. However, the performance drop differs greatly between the four models. As can be seen in Figures 3a-3d, the Base Probit model showed the highest drop, with an AUC that was eight percentage points lower than in Setting A. The Modified XGBoost model, however, proved more robust and only had a decrease in AUC of two percentage points compared to Setting A.

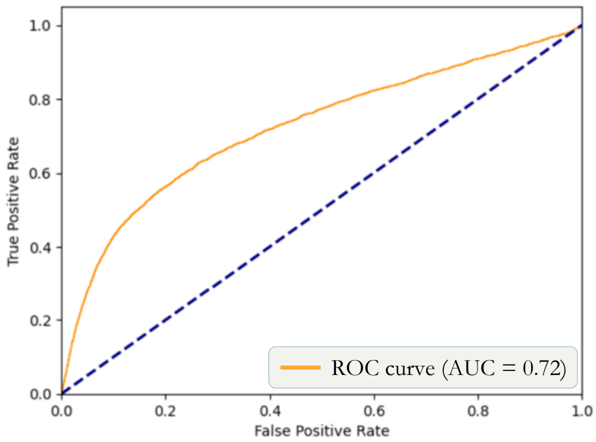

Figure 3a. ROC for the Base Probit model

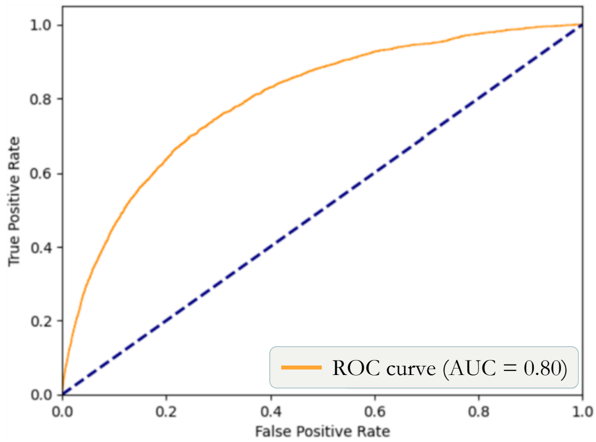

Figure 3b. ROC for the Base XGBoost model

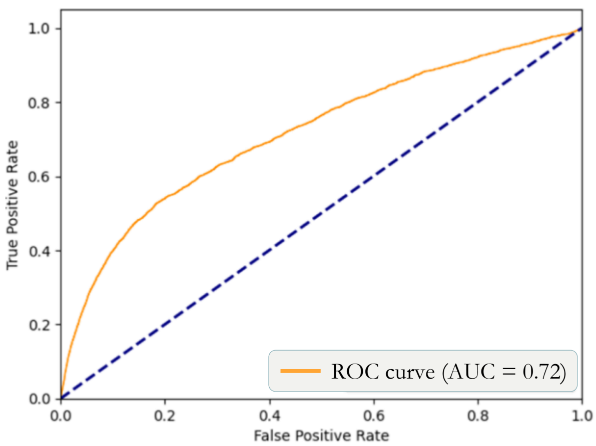

Figure 3c. ROC for the Modified Probit model

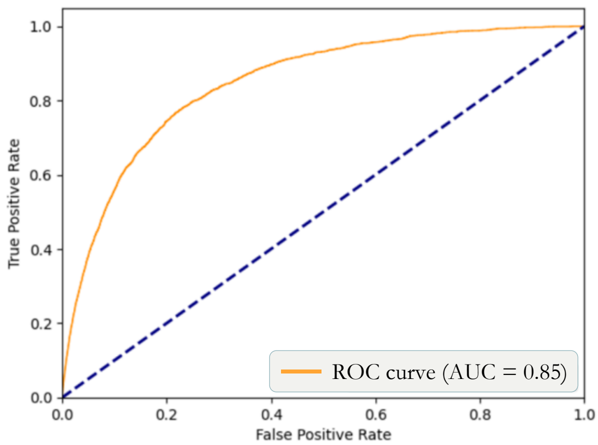

Figure 3d. ROC for the Modified XGBoost model

Following up on the observation from Setting A, with the Modified Probit model performing worse than its base version out-of-sample, they are more closely matched in Setting B. However, the Base model still has a slight advantage.

The confusion matrices indicate that, for the given cut-off level of 3%, bankruptcy prediction was more difficult in some industries compared to others. For example, real estate and financial firms are often studied specifically and subject to exclusion from industry-agnostic bankruptcy prediction studies due to their different regulatory requirements and capital structures (see for example Ivanova et al., 2023). As expected, our models encountered challenges in analyzing data from these specific industries, particularly noting the significant struggle with real estate companies, which consistently exhibited the poorest performance across all four models.

The variance in performance yields several noteworthy observations. As evidenced by their closely comparable AUC-scores, the probit models demonstrated a near match in their performance. Specifically, the Base Probit model outperformed the Modified Probit model in eight out of the fifteen industry subsamples. The Modified Probit model displayed an overall lower performance compared to the other three models. Notably, it also exhibited the most inconsistent performance trends, with an AUC of only 47% for real estate firms – worse than simply tossing a coin – yet with a slight redemption as it outperformed the Base XGBoost model in the *Water supply* industry subsample. This discrepancy ultimately boiled down to just two firms being predicted differently.

The XGBoost models showed more robust performances. Apart from the *Water supply*, the Base XGBoost model outperformed both probit models across all subsamples. The Modified XGBoost outperformed its base version across all industries except education and other service activities. It is worth noting that *Water supply*, *Education* and *Other service activities* – the three industries with unexpected trend breaks in performance – are the three subsamples with the fewest observations. Evidently, the smaller number of observations could have implications for the robustness of the models' predictions.

Consistent with the findings in Setting A, the predominant proportion of non-bankrupt observations in Setting B yields minimal variations in accuracy and specificity among the models and their respective subsamples. However, the sensitivity measure provides more nuanced insights. Notably, the lowest sensitivity can be observed for the Base Probit model in the *Water supply; sewerage, waste management & remediation services* subsample, in which it failed to identify any bankruptcies and therefore registered a sensitivity score of 0. Conversely, the highest sensitivity was obtained by the Modified XGBoost model in the *Agriculture, forestry and fishing* subsample, with a score of 47%.

Table 6a. Classification results for the Base Probit model for large non-industrials by industry

N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Base Probit								
Agriculture, forestry & fishing								
24,340	7	23,586	724	23	96.93%	23.33%	97.02%	75.84%
Water supply; sewerage, waste management & remediation services								
2,579	0	2,488	78	13	96.47%	0.00%	96.96%	68.40%
Construction								
95,344	134	91,955	2,727	528	96.59%	20.24%	97.12%	72.61%
Wholesale & retail trade; repair of motor vehicles & motorcycles								
133,779	127	129,194	3,887	571	96.67%	18.19%	97.08%	77.94%
Transportation & storage								
39,405	41	38,040	1,142	182	96.64%	18.39%	97.09%	72.10%
Accommodation & food service activities								
25,506	20	24,595	746	145	96.51%	12.12%	97.06%	69.50%
Information & communication								
35,463	22	34,308	1,042	91	96.81%	19.47%	97.05%	76.61%
Financial & insurance activities								
13,383	3	12,956	399	25	96.83%	10.71%	97.01%	63.76%
Real estate activities								
115,985	15	112,432	3,465	73	96.95%	17.05%	97.01%	61.31%
Professional, scientific & technical activities								
97,999	44	94,839	2,896	220	96.82%	16.67%	97.04%	71.88%
Administrative & support service activities								
26,563	24	25,632	773	134	96.59%	15.19%	97.07%	72.51%
Education								
8,192	6	7,921	240	25	96.77%	19.35%	97.06%	69.78%
Human health & social work activities								
18,233	15	17,614	532	72	96.69%	17.24%	97.07%	70.96%
Arts, entertainment & recreation								
10,160	4	9,822	301	33	96.71%	10.81%	97.03%	67.53%
Other service activities								
5,187	4	5,010	152	21	96.66%	16.00%	97.06%	73.52%
Total								
652,118	466	630,392	19,104	2,156	96.74%	17.77%	97.06%	71.72%

Table 6a reports classification results based on the same decision rule as in **Table 5.** The results in this table are split into multiple industries to evaluate how the model performs out-of-sample. Five industries are not shown in the table as the subsample contained too few bankruptcies. These are: (1) *Mining and quarrying*, (2) *Electricity, gas, steam and air conditioning supply*, (3) *Public administration and defence; compulsory social security*, (4) *Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use* and (5) *Activities of extraterritorial organisations and bodies.*

Table 6b. Classification results for the Base XGBoost model for large non-industrials by industry

N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Base XGBoost								
Agriculture, forestry & fishing								
24,340	8	23,587	723	22	96.94%	26.67%	97.03%	89.86%
Water supply; sewerage, waste management & remediation services								
2,579	3	2,491	75	10	96.70%	23.08%	97.08%	71.15%
Construction								
95,344	135	91,956	2,726	527	96.59%	20.39%	97.12%	81.98%
Wholesale & retail trade; repair of motor vehicles & motorcycles								
133,779	164	129,231	3,850	534	96.72%	23.50%	97.11%	84.02%
Transportation & storage								
39,405	43	38,042	1,140	180	96.65%	19.28%	97.09%	80.74%
Accommodation & food service activities								
25,506	29	24,604	737	136	96.58%	17.58%	97.09%	74.84%
Information & communication								
35,463	24	34,310	1,040	89	96.82%	21.24%	97.06%	80.32%
Financial & insurance activities								
13,383	5	12,958	397	23	96.86%	17.86%	97.03%	76.59%
Real estate activities								
115,985	10	112,427	3,470	78	96.94%	11.36%	97.01%	63.20%
Professional, scientific & technical activities								
97,999	66	94,861	2,874	198	96.87%	25.00%	97.06%	83.71%
Administrative & support service activities								
26,563	28	25,636	769	130	96.62%	17.72%	97.09%	76.68%
Education								
8,192	7	7,922	239	24	96.79%	22.58%	97.07%	84.23%
Human health & social work activities								
18,233	22	17,621	525	65	96.76%	25.29%	97.11%	77.30%
Arts, entertainment & recreation								
10,160	4	9,822	301	33	96.71%	10.81%	97.03%	75.00%
Other service activities								
5,187	6	5,012	150	19	96.74%	24.00%	97.09%	85.50%
Total								
652,118	554	630,480	19,016	2,068	96.77%	21.13%	97.07%	80.34%

Table 6b reports classification results based on the same decision rule as in **Table 5.** The results in this table are split into multiple industries to evaluate how the model performs out-of-sample. These are: (1) *Mining and quarrying*, (2) *Electricity, gas, steam and air conditioning supply*, (3) *Public administration and defense; compulsory social security*, (4) *Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use* and (5) *Activities of extraterritorial organizations and bodies.*

Table 6c. Classification results for the Modified Probit model for large non-industrials by industry

N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Modified Probit								
Agriculture, forestry & fishing								
24,340	4	23,583	727	26	96.91%	13.33%	97.01%	77.29%
Water supply; sewerage, waste management & remediation services								
2,579	5	2,493	73	8	96.86%	38.46%	97.16%	72.36%
Construction								
95,344	138	91,959	2,723	524	96.59%	20.85%	97.12%	72.69%
Wholesale & retail trade; repair of motor vehicles & motorcycles								
133,779	119	129,186	3,895	579	96.66%	17.05%	97.07%	74.58%
Transportation & storage								
39,405	46	38,045	1,137	177	96.67%	20.63%	97.10%	74.79%
Accommodation & food service activities								
25,506	15	24,590	751	150	96.47%	9.09%	97.04%	66.10%
Information & communication								
35,463	19	34,305	1,045	94	96.79%	16.81%	97.04%	72.21%
Financial & insurance activities								
13,383	1	12,954	401	27	96.80%	3.57%	97.00%	62.06%
Real estate activities								
115,985	12	112,429	3,468	76	96.94%	13.64%	97.01%	47.05%
Professional, scientific & technical activities								
97,999	46	94,841	2,894	218	96.82%	17.42%	97.04%	73.65%
Administrative & support service activities								
26,563	26	25,634	771	132	96.60%	16.46%	97.08%	69.93%
Education								
8,192	3	7,918	243	28	96.69%	9.68%	97.02%	64.80%
Human health & social work activities								
18,233	17	17,616	530	70	96.71%	19.54%	97.08%	73.95%
Arts, entertainment & recreation								
10,160	3	9,821	302	34	96.69%	8.11%	97.02%	64.71%
Other service activities								
5,187	6	5,012	150	19	96.74%	24.00%	97.09%	74.81%
Total								
652,118	460	630,386	19,110	2,162	96.74%	17.54%	97.06%	71.60%

Table 6c reports classification results based on the same decision rule as in **Table 5.** The results in this table are split into multiple industries to evaluate how the model performs out-of-sample. These are: (1) *Mining and quarrying*, (2) *Electricity, gas, steam and air conditioning supply*, (3) *Public administration and defense; compulsory social security*, (4) *Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use* and (5) *Activities of extraterritorial organizations and bodies.*

Table 6d. Classification results for the Modified XGBoost model for non-industrials by industry

N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Modified XGBoost								
Agriculture, forestry & fishing								
24,340	14	23,593	717	16	96.99%	46.67%	97.05%	91.51%
Water supply; sewerage, waste management & remediation services								
2,579	3	2,491	75	10	96.70%	23.08%	97.08%	84.54%
Construction								
95,344	168	91,989	2,693	494	96.66%	25.38%	97.16%	85.43%
Wholesale & retail trade; repair of motor vehicles & motorcycles								
133,779	177	129,244	3,837	521	96.74%	25.36%	97.12%	84.64%
Transportation & storage								
39,405	58	38,057	1,125	165	96.73%	26.01%	97.13%	84.28%
Accommodation & food service activities								
25,506	22	24,597	744	143	96.52%	13.33%	97.06%	79.90%
Information & communication								
35,463	27	34,313	1,037	86	96.83%	23.89%	97.07%	83.00%
Financial & insurance activities								
13,383	2	12,955	400	26	96.82%	7.14%	97.00%	81.39%
Real estate activities								
115,985	20	112,437	3,460	68	96.96%	22.73%	97.01%	73.93%
Professional, scientific & technical activities								
97,999	61	94,856	2,879	203	96.86%	23.11%	97.05%	86.27%
Administrative & support service activities								
26,563	35	25,643	762	123	96.67%	22.15%	97.11%	81.19%
Education								
8,192	7	7,922	239	24	96.79%	22.58%	97.07%	80.87%
Human health & social work activities								
18,233	19	17,618	528	68	96.73%	21.84%	97.09%	82.19%
Arts, entertainment & recreation								
10,160	7	9,825	298	30	96.77%	18.92%	97.06%	80.04%
Other service activities								
5,187	9	5,015	147	16	96.86%	36.00%	97.15%	81.22%
Total								
652,118	629	630,555	18,941	1,993	96.79%	23.99%	97.08%	85.43%

Table 6d reports classification results based on the same decision rule as in **Table 5.** The results in this table are split into multiple industries to evaluate how the model performs out-of-sample. These are: (1) *Mining and quarrying*, (2) *Electricity, gas, steam and air conditioning supply*, (3) *Public administration and defense; compulsory social security*, (4) *Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use* and (5) *Activities of extraterritorial organizations and bodies.*

4.2.3 Setting C: SME industrials

In our last setting, Setting C, the models are tasked with predicting bankruptcy for industrial firms of a smaller size compared to their training data. Similar to Setting B, the firms in the test set are expected to exhibit varying capital structures and normal ratio levels.

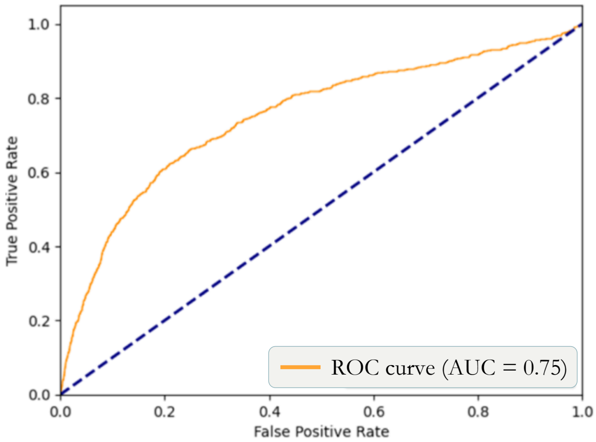

Figure 4a. ROC for the Base Probit model

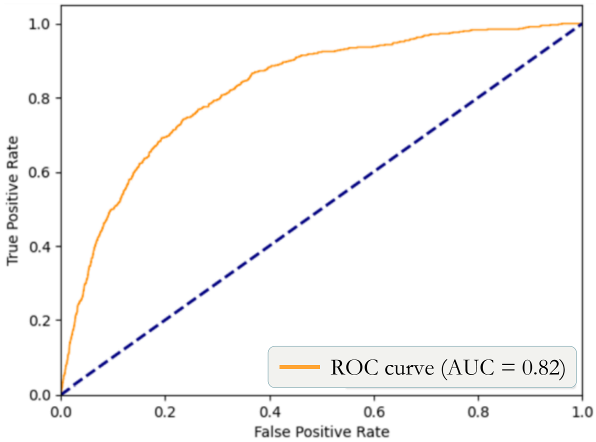

Figure 4b. ROC for the Base XGBoost model

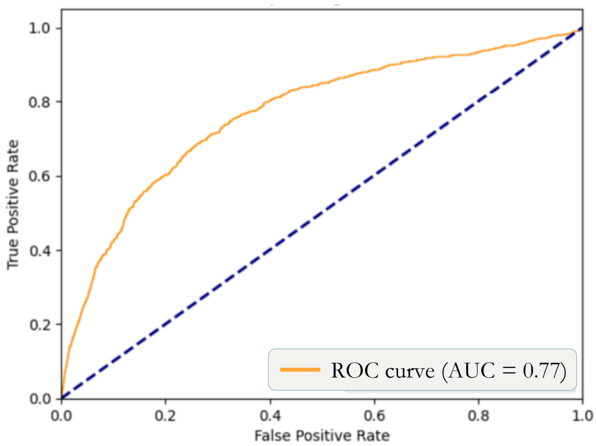

Figure 4c. ROC for the Modified Probit model

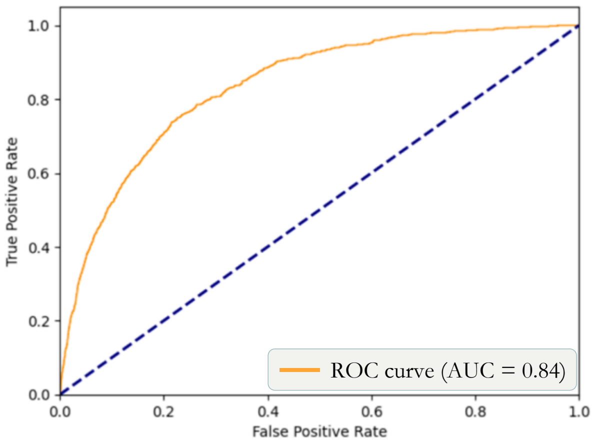

Figure 4d. ROC for the Modified XGBoost model

Figures 4a-4d allows for an observation to be made in line with Setting B: all models displayed an inferior performance compared to Setting A. However, in contrast to Setting B, three models exhibited an improvement in performance while the Modified XGBoost model, with its 77 independent variables, demonstrated an increased difficulty when predicting bankruptcy for smaller industrial firms as opposed to larger non-industrial ones, suggesting that size has a larger effect than industry on performance. Despite this, it maintained the highest performance across all subsamples, except for the *5-9 FTEs* subsample, where its base version excelled. In Setting C, the Modified Probit model performed nearly as well as in Setting A, with only a one-percentage-point difference in AUC. After lagging behind its Base version in Settings A and B, it showcased improved performance in this setting.

Observing the subsamples categorized by the number of FTEs, the performance closely aligns with firm size. As the firm size decreases, all models face escalating challenges in accurately predicting bankruptcy. An exception arises with the Modified Probit model, which performed best on the 20-49 subsample, followed by 5-9, 50-99, 10-19 and 1-4.

Table 7a. Classification results for the Base Probit model for SME industrials by size (number of FTEs)

Size (FTEs)	N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Base Probit									
1-4	32,717	36	31,564	946	171	96.59%	17.39%	97.09%	70.21%
5-9	15,065	21	14,527	431	86	96.57%	19.63%	97.12%	79.33%
10-19	12,417	18	11,976	355	68	96.59%	20.93%	97.12%	79.92%
20-49	10,109	12	9,761	292	44	96.68%	21.43%	97.10%	83.12%
50-99	3,948	3	3,818	116	11	96.78%	21.43%	97.05%	87.72%

Table 7a reports classification results based on the same decision rule as in **Table 5.** The results in this table are grouped into different size categories based on the number of employees.

Table 7b. Classification results for the Base XGBoost for SME industrials by size (number of FTEs)

Size (FTEs)	N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Base XGBoost									
1-4	32,717	45	31,573	937	162	96.64%	21.74%	97.12%	80.72%
5-9	15,065	26	14,532	426	81	96.63%	24.30%	97.15%	84.13%
10-19	12,417	18	11,976	355	68	96.59%	20.93%	97.12%	85.43%
20-49	10,109	13	9,762	291	43	96.70%	23.21%	97.11%	86.88%
50-99	3,948	3	3,818	116	11	96.78%	21.43%	97.05%	88.52%

Table 7b reports classification results based on the same decision rule as in **Table 5.** The results in this table are grouped into different size categories based on the number of employees.

Table 7c. Classification results for the Modified Probit for SME industrials by size (number of FTEs)

Size (FTEs)	N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Modified Probit									
1-4	32,717	35	31,563	947	172	96.58%	16.91%	97.09%	74.45%
5-9	15,065	21	14,527	431	86	96.57%	19.63%	97.12%	79.21%
10-19	12,417	20	11,978	353	66	96.63%	23.26%	97.14%	77.54%
20-49	10,109	14	9,763	290	42	96.72%	25.00%	97.12%	83.63%
50-99	3,948	3	3,818	116	11	96.78%	21.43%	97.05%	79.09%

Table 7c reports classification results based on the same decision rule as in **Table 5.** The results in this table are grouped into different size categories based on the number of employees.

Table 7d. Classification results for the Modified XGBoost for SME industrials by size (number of FTEs)

Size (FTEs)	N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Modified XGBoost									
1-4	32,717	48	31,576	934	159	96.66%	23.19%	97.13%	81.81%
5-9	15,065	25	14,531	427	82	96.62%	23.36%	97.15%	83.29%
10-19	12,417	25	11,983	348	61	96.71%	29.07%	97.18%	87.13%
20-49	10,109	16	9,765	288	40	96.76%	28.57%	97.14%	88.23%
50-99	3,948	7	3,822	112	7	96.99%	50.00%	97.15%	89.03%

Table 7d reports classification results based on the same decision rule as in **Table 5.** The results in this table are grouped into different size categories based on the number of employees.

5. Additional analysis & robustness checks

To demonstrate the reliability and stability of our results, we have performed several robustness checks. Our main results indicate that the comparative accuracy of a new-era machine learning technique is greater than a traditional statistical technique. However, the difference could be circumstantial, pertaining to the specific models, our choice of training and testing data, respectively, as well as the metrics we used to evaluate the models' accuracy.

To investigate if our results are consistent with other machine learning and statistical techniques, we repeated the test in Setting A, but used logit instead of probit for the base statistical model and Random Forest instead of XGBoost for the base machine learning model. They are both popular techniques, not only in general but also specifically within bankruptcy prediction (as referred to in the literature review). The main difference between logit and probit is a different assumption about the distribution of the error term (logistic versus normal distribution). The main difference between Random Forest and XGBoost is how the decision trees are trained (parallel versus sequentially). With our new techniques, the performance in Setting A decreased slightly for both models. The Base Logit model showed an AUC of 78% (compared to 80% for the Base Probit) while the Base Random Forest model had an AUC of 86% (compared to 86% for the Base XGBoost). These findings illustrate that although the selection of a particular technique can influence outcomes, the consistent performance gap between traditional statistical models and modern machine learning models persists.

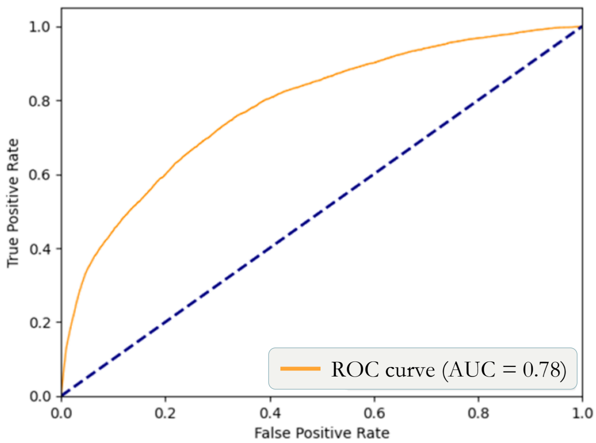

Figure 5a. ROC for the Base Logit model

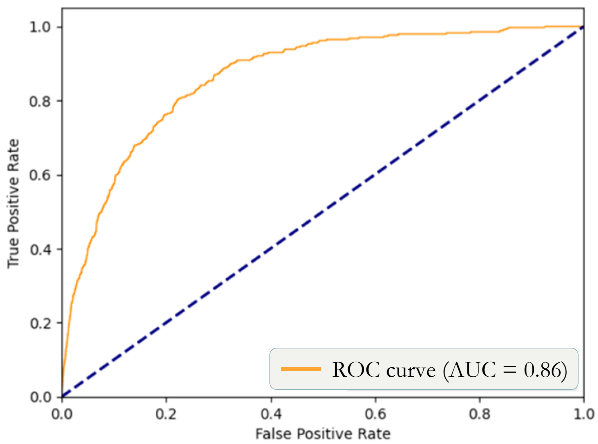

Figure 5b. ROC for the Base Random Forest model

Table 8. Classification results for the base models with alternative techniques

N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Base Probit								
68,389	71	66,068	1,981	269	96.71%	20.88%	97.09%	79.94%
Base XGBoost								
68,389	85	66,082	1,967	255	96.75%	25.00%	97.11%	86.47%
Base Logit								
68,389	69	66,066	1,983	271	96.70%	20.29%	97.09%	78.15%
Base Random Forest								
68,389	104	66,101	1,948	236	96.81%	30.59%	97.14%	86.08%

Table 8 reports classification results based on the same decision rule as in **Table 5.** The results from the original techniques (probit & XGBoost) are compared to our alternative techniques (logit & Random Forest).

We further investigated the robustness of our results by altering the training and testing data. A recurring limitation of bankruptcy prediction studies is that the results of a study do not generalize well outside the specific setting, as discussed by Ooghe and Balcaen (2007). This comes as a result of "over-modelling", which means the models are optimized to solve the problem (classifying the data) as opposed to creating a stable and useful model for bankruptcy prediction. By altering which firm-year observations are included in the training and testing set, we can investigate whether the performance gap is stable over time. For our main results, we limited the training data to observations from 2000-2015 and the testing data to 2016-2021. Experimenting with different combinations of years – ranging from shorter to longer training periods and even an inversion (training data from 2016-2021, testing data from 2000-2015) still resulted in a relatively higher performance from the machine learning models.

Finally, to ensure the findings were not specific to our choice of evaluation metrics, we tested other alternatives. While the ROC-curve and AUC-score is a particularly useful metric for a binary classification task because it demonstrates the relative performance – and trade-off between specificity and sensitivity – across all potential cut-off points, it rewards sensitivity and specificity equally. Although two models may share the same AUC, when you examine their ROC-curves, you might find that one model achieves higher sensitivity but at the cost of lower specificity, while the other model achieves higher specificity but with lower sensitivity. Arguably, because bankruptcies are so costly, a model that finds more bankruptcies at the cost of lower specificity (more false positives) is the more useful model. The F2-score is a metric that rewards sensitivity twice as much as specificity, making it a useful metric when capturing bankruptcies is the priority. When evaluated with the F2-score, both machine learning models still outperform the statistical models in Setting A.

$$F2 = \frac{(1 + 2^2) \times Precision \times Recall}{(2^2 \times \text{Precision}) + \text{Recall}}$$

Table 9. Classification results for all models with additional evaluation metrics

N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC	Precision	Recall	F2 Score
Base Probit											
68,389	71	66,068	1,981	269	96.71%	20.88%	97.09%	79.94%	3.46%	20.88%	10.40%
Base XGBoost											
68,389	85	66,082	1,967	255	96.75%	25.00%	97.11%	86.47%	4.14%	25.00%	12.46%
Modified Probit											
68,389	70	66,067	1,982	270	96.71%	20.59%	97.09%	78.15%	3.41%	20.59%	10.26%
Modified XGBoost											
68,389	100	66,097	1,952	240	96.79%	29.41%	97.13%	87.45%	4.87%	29.41%	14.65%

Table 9 reports classification results based on the same decision rule as in **Table 5.** The results from Setting A are reported with additional evaluation metrics: precision, recall (which is sensitivity) and the F2-score.

Another key finding of our main result was that the machine learning model is more effective at identifying useful variables from a larger set of potential useful variables. But there are many feature selection methods that can be used for probit analysis, and it could be that backward selection did not work optimally for the specific circumstances. Using forward selection, the Modified Probit model selected the same set of independent variables. With LASSO regression as the selection method, another selection method used in bankruptcy prediction studies (Tian et al., 2015), different variables were selected. However, performance for the Modified Probit model was even worse using these variables. Our findings, along with the results from these robustness checks, suggest that the selected variables do not demonstrate strong generalization capabilities. While the training results surpass those of the Base Probit model, the out-of-sample classification performance is notably inferior.

Forward selection:

$$probit\big(P(Bankruptcy_{it+2} = 1|X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)\big) = \Phi(\alpha_0 + \beta_1 R(1)T_{it} + \beta_2 R(1)Sk_{it} + \beta_3 SD(1)_{it} + \beta_4 OHT(1)_{it} + \beta_5 ATA_{it} + \beta_6 MATA_{it} + \beta_7 LNT(1)_{it} + \beta_8 N.(R(5)E, es)_{it})$$

LASSO:

$$probit\big(P(Bankruptcy_{it+2} = 1|X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11},)\big) = \Phi(\alpha_0 + \beta_1 ATA_{it} + \beta_2 OR'_{it} + \beta_3 RBS'_{it} + \beta_4 LI(1)III_{it} + \beta_5 LI(3)I_{it} + \beta_6 OHT(1)_{it} + \beta_7 R(1)T_{it} + \beta_8 R(5)E, es_{it} + \beta_9 ATA_{it} + \beta_{10} N.(LI(3)I)_{it} + \beta_{11} N.(SD(1))_{it})$$

Another potential limitation lies in the construction of the Modified Probit model. The omission of principal component analysis constrained the model to a more restricted set of variables, albeit likely containing the most pertinent ones. To explore the efficacy of the variables employed by the Modified XGBoost model, mirroring the selection process of the Modified Probit, we confined the Modified XGBoost model to the 29 independent variables identified through backward selection by the Modified Probit. While this yielded an equivalent AUC of 87%, the classification performance at the cut-off point was marginally inferior. Taking an additional step, we assessed the Modified XGBoost's performance with only the 8 variables chosen by the Modified Probit model following feature selection. The AUC remained at 87%, but there was a slight, further decrease in classification performance at the cut-off point. These results imply that while the access to more information through additional variables yielded higher performance, the performance gap remains relatively large when the access to variables is constrained.

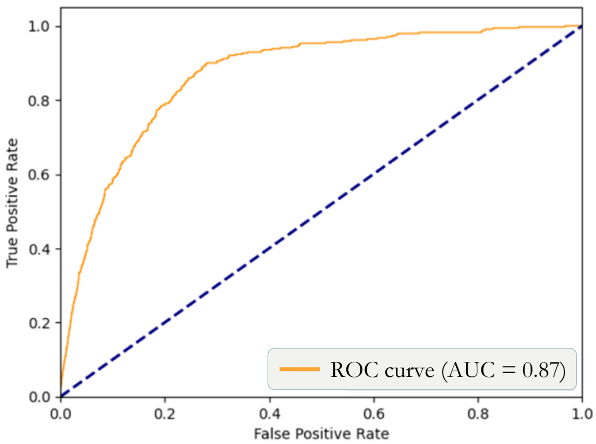

Figure 6a. ROC for the Modified XGBoost with 29 independent variables

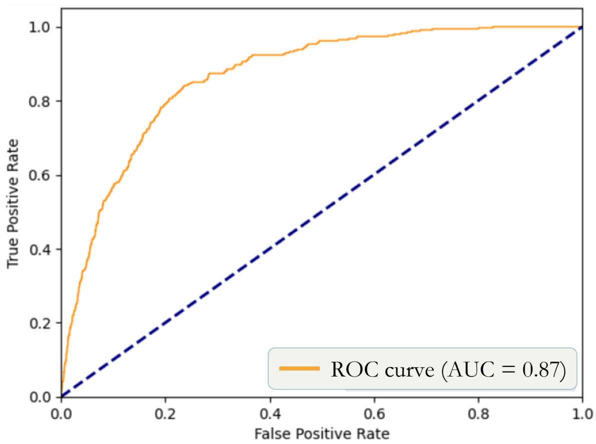

Figure 6b. ROC for the Modified XGBoost with 8 independent variables

Table 10. Classification results for the modified models with alternative variable constrictions

N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Modified Probit								
68,389	70	66,067	1,982	270	96.71%	20.59%	97.09%	78.15%
Modified XGBoost								
68,389	100	66,097	1,952	240	96.79%	29.41%	97.13%	87.45%
Modified XGBoost (29 independent variables)								
68,389	92	66,089	1,960	248	96.77%	27.06%	97.12%	86.93%
Modified XGBoost (8 independent variables)								
68,389	85	66,082	1,967	255	96.75%	25.00%	97.11%	86.66%

Table 10 reports classification results based on the same decision rule as in **Table 5.** The modified models are compared to alternative configurations of the Modified XGBoost model, using both 29 variables and 8 variables.

Our final key finding in the main results was that machine learning techniques seem to perform better than statistical techniques when asked to predict bankruptcy on firms different from what they were trained on. However, this might only hold true when the training sample is industrial firms, potentially because the decision rules it creates generalize well on other firms. To ensure the superior performance still holds true when trained on another industry, we replaced the training set with retail firms instead.

Tested on Setting B, all models performed similarly when trained on retail firms compared to industrial firms. However, the results are still consistent with our main results, as both XGBoost models achieve higher AUC-scores than the probit models. As a final test, we repeated the same procedure for Setting C, investigating potential changes if we train the models on small firms instead of large, and test them on large firms as opposed to small firms. In other words, using firms from Setting C as the training data and firms from Setting A as the testing data. The models all performed worse, but the XGBoost models were consistent in their superior performance. This indicates that large firms contain information that generalize better on small firms than vice versa.

Table 11. Classification results for all models using retail firms as the training sample

N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Base Probit								
586,728	419	567,181	17,188	1,940	96.74%	17.77%	97.06%	71.66%
Base XGBoost								
583,728	499	567,261	14,108	1,860	97.26%	21.15%	97.57%	80.05%
Modified Probit								
586,728	418	567,179	17,190	1,941	96.74%	17.72%	97.06%	72.23%
Modified XGBoost								
586,728	567	567,328	17,041	1,792	96.79%	24.04%	97.08%	85.45%

Table 11 reports classification results based on the same decision rule as in **Table 5.** The results show the models' performance when trained on retail firms and tested on all other firms (including industrial).

Table 12. Classification results for all models using SMEs as the training sample

N	True Positive	True Negative	Type I Error	Type II Error	Accuracy	Sensitivity	Specificity	AUC
Base Probit								
68,389	76	65,953	1,999	360	96.55%	17.42%	97.06%	73.82%
Base XGBoost								
68,389	91	65,993	1,960	345	96.63%	20.94%	97.12%	81.63%
Modified Probit								
68,389	77	65,968	1,985	359	96.57%	17.76%	97.08%	76.52%
Modified XGBoost								
68,389	104	66,002	1,950	333	96.66%	23.79%	97.13%	83.77%

Table 12 reports classification results based on the same decision rule as in **Table 5.** The results show the models' performance when trained on SME industrial firms and tested large industrial firms.

In summary, our additional checks confirmed the robustness of our main findings, indicating that machine learning techniques consistently outperform traditional statistical methods in bankruptcy prediction. Through various checks, we validated the stability of the performance gap between these approaches, considering alternative techniques, variable selection methods and evaluation metrics. Notably, even when adjusting for different training and testing data periods, our machine learning models exhibited superior predictive accuracy. Furthermore, our exploration of industry and firm-size variations upheld the superiority of XGBoost models. Whether trained on retail or industrial firms, or on large or small firms, and tested accordingly, the performance gap endured.

6. Discussion

The results of our study make a compelling case for machine learning techniques with regards to the thirty-year-old debate. Machine learning techniques, in their current state, outperform traditional statistical techniques decidedly and reliably. With our four research questions, we sought to thoroughly explore the comparative performance of the two categories of techniques, not only in terms of accuracy but also in terms of extracting useful information. The results overwhelmingly point to the conclusion that new-era machine learning models are better at predicting corporate bankruptcy. We find support for this conclusion all throughout our results, with both superior goodness of fit in the training phase and more accurate predictions across all three test settings. Our findings contribute to the ongoing discussion in the bankruptcy prediction field and corroborate the findings of Altman (2017) and Kumar & Ravi (2007). In Setting A, our main test setting, both XGBoost models display a convincingly higher AUC, indicating that for any cut-off point, they will be able to classify bankrupt and non-bankrupt firms more accurately than their probit counterparts. To put things into perspective, Skogsvik's ambitious current cost accounting approach, where he meticulously derived current cost values in a manner that arguably required more effort than feeding raw data to the XGBoost algorithm, resulted in largely equal performances compared to historical cost accounting information. As shown by comparing the Base Probit to the Base XGBoost model, by simply substituting the probit technique for XGBoost, all else equal, the AUC increased by 6.5 percentage points.

We also find evidence of machine learning models being better at identifying variables with useful information through feature selection. However, this conclusion is not immune to objections. Our approach of modifying the probit model can be argued to be too simplistic and not the optimal way to modify the probit model to incorporate useful information from new independent variables. Our intention, however, was to evaluate just what effect an uncomplicated feature selection approach would have for the probit model and the XGBoost model, respectively. The Modified XGBoost model clearly improved its performance when all we did was to give it more information and leave the algorithm to make use of the information. The Modified Probit model did not demonstrate a clear improvement in performance. Furthermore, as demonstrated by our robustness checks, even with alternative feature selection methods for the Modified Probit model and heavy restrictions on the Modified XGBoost model, the results were consistent.

Contributing to the ongoing debate of bankruptcy prediction models' generalizability in general, and machine learning techniques in particular, we formulated research questions (iii) and (iv). The results of Settings B and C demonstrate that the machine learning models were more robust outside their "comfort zone" than the statistical models. This discovery introduces nuance to the conclusions of Hillegeist et al. (2004) and Begley et al. (1996), both of whom identified poor generalizability in bankruptcy prediction models. Our findings suggest that this issue can be mitigated by leveraging machine learning techniques.

While the XGBoost algorithm showed superior performance, we do not necessarily suggest it will be the optimal technique for any bankruptcy classification study. As the literature review revealed, recent papers by Alaka et al. (2018) and Clement (2020) argue that the search for a one-size-fits-all is futile. Substituting XGBoost for Random Forest in our robustness check yielded a nearly identical performance. Instead, our results demonstrate the robustness of the new-era machine learning techniques. The results also indicate that they are more "forgiving" than their statistical counterparts. Consider the performance drop of the Modified XGBoost in Setting B and C, compared to Setting A. It is intuitive to think that a model with 77 independent variables will overfit and perform poorly when faced with firms of different financial structure than those it was trained on. Nevertheless, the largest performance drop for the Modified XGBoost was a three-percentage point decrease in AUC, compared to six percentage points for the Modified Probit and eight for the Base Probit.

6.1 Limitations and suggestions for further research

Naturally, our study has a number of limitations, most of them relating to the research design. Among other things, our choice to follow Skogsvik's method – while having several advantages – limited our exploration of potentially useful variables to the 71 (77) in his primary set. While there are a considerable number of relevant variables, an XGBoost algorithm is capable of handling many more variables, which would have been interesting to explore. Further research should collect an even larger set of variables, say the 500 most commonly used financial and non-financial ratios in financial statements, to see if performance could be improved further. Another limitation with the study is our narrow definition of failure. While it was motivated by both data collection procedures and the fact that non-legalistic definitions of failure are fuzzy by nature, the fact remains that many firms would be considered failed without ever reaching bankruptcy. Consider Samhällsbyggnadsbolaget (SBB), a Swedish real estate firm that has been the subject of intense media coverage during 2023 due to their declining financial situation. Many would agree the firm failed in spirit, despite not formally declaring bankruptcy. Regardless, further research could more carefully review literature and create a research design that more sophisticatedly deals with failure to better capture real-world failures exhaustively. Also related to our research design is the choice to only look at different industries and sizes when assessing the generalizability of our models. As found in the literature, country of origin and age, among other factors, might pose different challenges to the models and should be investigated by further research. As a final limitation, our comparison of AUC scores and other metrics could be deemed as rather subjective. We find differences in the models' respective results but lack a clear framework or test with which to compare them. As such, we can definitively say that a higher score is better, but any assessment of the magnitude of the difference comes with a degree of subjectivity. Further research should seek ways to enable an objective assessment of the differences.

7. Appendix

Appendix 1. Overview of Skogsvik's independent variables

#	Variable	Symbol	Formula	Bankrupt firms Mean	Non-bankrupt firms Mean
	Profitability				
1	Profit margin	VM	RR(2)/RI	-0.14	-0.03
2	Return on total capital	R(1)T	(RE(2)+FK-EO(3)-EO(4))/T	-0.08	0.10
3	Return on total capital	R(2)T	(RE(2)+FK-EO(3)-EO(4))/(T-KS+KS(4))	-0.04	0.18
4	Return on total capital	R(3)T	(RE(2)+FK-EO(3)-EO(4))/(T-KS+KS(4)-Lask(2))	-0.04	0.19
5	Return on total capital	R(4)T	(RE(2)+FK-EO(3)-EO(4))/(T-KS)	-0.07	0.19
6	Return on total capital	R(5)T	(RE(1)-FK)/(T-KS)	-0.05	0.17
7	Return on total capital	R(6)T	(RE(1)+RK(3)+RK(4)+RK(5)+FK+EO(5))/(T-KS)	0.09	0.26
8	Return on equity	R(1)E,es	(RE(2)-SKT-Ch(Lask(2))-(1-ss)*(EO(3)+EO(4))/(ER+OR+BR)	-0.28	0.20
9	Return on equity	R(2)E,es	(RE(2)-SKT-(1-ss)*(EO(3)+EO(4))/(ER+OR+BR)	-0.15	0.18
10	Return on equity*	R(3)E,es	(RE(2)-SKT-Ch(Lask(1))-(1-ss)*(EO(3)+EO(4))/(ER+BR+OR-Lask(1))	-	-
11	Return on equity	R(4)E,es	(RE(1)-SKT-Ch(Lask(2))+ss*((EO(3)+EO(4))/(ER+BR+OR-Lask(1))	-0.20	0.12
12	Return on equity	R(5)E,es	(RE(1)+RK(3)+RK(4)+RK(5)+EO(5)-SKT)/(ER+BR+OR)	-0.01	0.34
13	Value added ratio	FVK(1)	(RE(2)+RK-FI(1)+FK-EO(3)-EO(4))/(T-AT(1))	-2.15	-1.58
14	Value added ratio	FVK(2)	(RE(2)+RK-FI+FK-EO(3)-EO(4))/(T-OT(2)-AT(1)-AT(2)-AT(3))	-2.19	-1.61
	Cost				
15	Proportion of salaries	LÖA	RK(1)/(RK+ED(2)-min(EO(5),0))	-0.33	-0.31
16	Proportion of depreciation	AVA	(RK(3)+RK(4)+RK(5))/(RK+EO(2)-min(EO(5),0))	-0.04	-0.04
17	Interest expense	R(1)Sk	FK/(KS+LS+Lask(2))	-0.04	-0.03
18	Interest expense	R(2)Sk	FK/(KS+LS)	-0.04	-0.03
19	Interest expense*	R(3)Sk	FK/(KS+LS+Lask(1))	-	-
20	Interest expense	R(4)Sk	FK/(KS(4)+LS)	-0.19	-1.06
21	Proportion of taxes	S(1)	(SKT+Ch(Lask(2)))/RE(2)	-0.40	-0.19
22	Proportion of taxes	S(2)	SKT/RE(2)	-0.43	-0.20
23	Proportion of taxes*	S(3)	(SKT+Ch(Lask(1)))/RE(2)	-	-
24	Operating allowance*	DBA	EO(3)/(RK+FK+EO(2)-min(EO(5),0))	-	-
	Capital turnover				
25	Asset turnover	OHT(1)	RI/T	1.93	1.72
26	Asset turnover	OHT(2)	RI/(T-KS+KS(4))	2.63	3.43
27	Asset turnover	OHT(3)	RI/(T-KS+KS(4)-Lask(2))	2.73	3.65
28	Asset turnover	OHT(4)	RI/(T-KS)	4.17	4.01
29	Accounts receivables turnover rate	TKF	(OT(3)+OT(5))/RI	0.31	0.56
30	Inventory turnover rate	TVL(1)	OT(6)/RI	0.22	0.18
31	Inventory turnover rate	TVL(2)	OT(6)/RK	0.19	0.13
32	Cash asset turnover rate	TCH	(OT(1)+OT(2))/RI	0.10	0.46
33	Liquid current asset turnover rate	TLI	(OT-OT(6))/RI	0.54	1.31
34	Current asset turnover rate	TOT	OT/RI	0.76	1.49
35	Working capital turnover rate	TRK	(OT-KS)/RI	0.14	0.82
36	Working capital size	NRK	(OT-KS)/OT(6)	1.00	8.34
	Liquidity				
37	Liquidity ratio I	LI(1)I	(OT(1)+OT(2))/T	0.03	0.14
38	Liquidity ratio I	LI(2)I	(OT(1)+OT(2))/KS	0.07	0.99
39	Liquidity ratio I	LI(3)I	(OT(1)+OT(2))/(KS-KS(2))	0.33	12.78
40	Liquidity ratio II	LI(1)II	(OT-OT(6))/T	0.41	0.48
41	Liquidity ratio II	LI(2)II	(OT-OT(6))/KS	0.83	2.43
42	Liquidity ratio II	LI(3)II	(OT-OT(6))/(KS-KS(2))	3.09	25.55
43	Liquidity ratio III	LI(1)III	OT/KS	1.40	3.15
44	Liquidity ratio III	LI(2)III	(OT)/(KS-KS(2))	4.50	29.45
45	Flow of capital ratio	KAP(1)	(RE(1)+RK(3)+RK(4)+RK(5)+FK+EO(5))/KS	-3.75	-4.23
46	Flow of capital ratio	KAP(2)	(RE(1)+RK(3)+RK(4)+RK(5)+FK+EO(5))/FK	0.31	-162.89
47	Flow of capital ratio	KAP(3)	(RE(1)+RK(3)+RK(4)+RK(5)+FK+EO(5))/(KS+LS)	-0.03	0.40

Appendix 1 - continued. Overview of Skogsvik's independent variables

#	Variable	Symbol	Formula	Bankrupt firms Mean	Non-bankrupt firms Mean
	Asset structure				
48	Proportion of inventory	VLA	OT(6)/T	0.27	0.20
49	Proportion of working capital	RKA	(OT-KS)/T	0.05	0.28
50	Proportion of fixed assets	ATA	AT/T	0.32	0.32
51	Proportion of tangible assets	MATA	(OT(6)+AT(4)+AT(5)+AT(6))/T	0.51	0.44
52	Size (1)	LNT(1)	ln(T)	9.02	9.34
53	Size (2)	LNT(2)	ln(T-KS)	8.15	8.79
	Capital structure				
54	Solvency (1)	SD(1)	(ER+BR+OR-Lask(2))/T	0.05	0.41
55	Solvency (2)	SD(2)	(ER+BR+OR)/T	0.05	0.43
56	Solvency (3)*	SD(3)	(ER+BR+OR-Lask(1))/T	-	-
57	Proportion of untaxed reserves	ORA	OR/T	0.02	0.10
58	Proportion of short-term liabilities	KSA	KS/T	0.63	0.40
59	Proportion of accounts payable	LESA	(KS(1)+KS(3))/T	0.27	0.17
60	Proportion of interest-bearing liabilities	RBSA	(KS(4)+LS)/T	0.40	0.20
	Growth				
61	Growth in revenue	RI'	Ch(RI)/RI(t-1)	1.17	1.30
62	Growth in assets	T'	Ch(T)/T(t-1)	1.04	1.20
63	Growth in equity (1)	E(1)'	Ch(ER+BR+OR-Lask(2))/[(ER+BR+OR-Lask(2)](t-1)	0.69	1.15
64	Growth in equity (2)	E(2)'	Ch(ER+BR+OR)/[(ER+BR+OR)](t-1)	0.70	1.15
65	Growth in equity (3)*	E(3)'	Ch(ER+BR+OR-Lask(1))/[(ER+BR+OR-Lask(1)](t-1)	-	-
66	Growth in untaxed reserves	OR'	Ch(OR)/OR(t-1)	0.70	1.28
67	Growth in short-term liabilities	KS'	Ch(KS)/KS(t-1)	1.28	1.41
68	Growth in accounts payable	LES'	Ch(KS(1)+KS(3))/[(KS(1)+KS(3)](t-1)	2.35	2.99
69	Growth in interest-bearing liabilities	RBS'	Ch(KS(4)+LS)/[(KS(4)+LS)](t-1)	1.27	2.29
70	Growth in fixed assets	AT'	Ch(AT)/AT(t-1)	1.38	1.70
71	Growth in tangible assets	MAT'	Ch(OT(6)+AT(4)+AT(5)+AT(6))/[(OT(6)+AT(4)+AT(5)+AT(6))](t-1)	1.35	1.64
	Normalized ratios				
72	Normalized interest expense	N.R(1)Sk	R(1)Sk / R(1)Sk last four year average	-1.17	-0.90
73	Normalized asset turnover	N.OHT(1)	OHT(1) / OHT(1) last four year average	-0.08	-0.11
74	Normalized inventory turnover rate	N.TVL(1)	TVL(1) / TVL(1) last four year average	1.47	0.92
75	Normalized proportion of taxes	N.S(1)	S(1) / S(1) last four year average	-7.95	0.50
76	Normalized proportion of taxes	N.S(2)	S(2) / S(2) last four year average	-7.00	-4.86
77	Normalized return on total capital	N. R(1)T	R(1)T / R(1)T last four year average	-1.38	-0.05
78	Normalized return on total capital	N.R(5)E,es	R(5)E,es / R(5)E,es last four year average	-1.30	-0.25
79	Normalized liquidity ratio I	N.LI(3)I	Standard deviation of LI(3)I based on the last 4 years	136.62	36.07
80	Normalized liquidity ratio III	N.LI(1)III	Standard deviation of LI(1)III based on the last 4 years	0.93	1.18
81	Normalized proportion of fixed assets	N.ATA	Standard deviation of ATA based on the last 4 years	1.11	1.15
82	Normalized proportion of tangible assets	N.MATA	Standard deviation of Mata based on the last 4 years	1.22	1.05
83	Normalized solvency (1)	N.SD(1)	Standard deviation of SD(1) based on the last 4 years	0.66	1.04

Appendix 1 shows a gross list of all independent variables included in this study, as well as their respective results from the simple mean analysis. We excluded a total of six ratios, still presented in this table and marked with a ***** symbol, to reach a primary set of 77 independent variables. The following exclusions were made: **#10** Return on equity: R(3)E,es, **#19** Interest expense: R(3)Sk, **#23** Proportion of taxes: S(3), **#24** Operating Allowance: DBA, **#56** Solvency (3): SD(3) and **#65** Growth in equity (3): E(3)'. **#24** was excluded since a detailed item was missing from the financial statements. The other five exclusions were made due to data limitations, where a categorization of the different items related to untaxed reserves were missing.

Appendix 2. Overview of all financial statement items and related variables in Serrano

Financial statement item	Symbol	Serrano variable
Balance sheet		
Assets		
Cash and bank balances	OT(1-2)	KABASU
Accounts receivable	OT(3)	KUNDFORD
Other current receivables	OT(4)	KFORDOV
Current receivables - group companies and associates	OT(5)	KFORDKNC
Inventory	OT(6)	LAGERSU
Other current assets	OT(7)	OT less the sum of OT(1) through OT(6)
Current assets	OT	OMSTGSU
Participation in group companies and associates	AT(1)	ANDKNC
Long-term receivables	AT(2)	LFORDKNC
Loans to partners and related parties	AT(3)	LANDELAG
Machinery and equipment	AT(4)	MASKINV
Buildings and land	AT(5)	BYGGMARK
Intangible fixed assets	AT(7)	IMANLSU
Other fixed assets	AT(8)	AT less the sum of AT(1) through AT(7)
Total fixed assets	AT	ANLTSU
Total assets	T	TILLGSU
Liabilities		
Accounts payable	KS(1)	KSKLEV
Other current liabilities	KS(2)	KSKOV
Current liabilities - group companies and associates	KS(3)	KSKKNC
Current liabilities - credit institutions	KS(4)	KSKKRIN
Övriga korta skulder	KS(5)	KS less the sum of KS(1) through KS(4)
Total current liabilities	KS	KSKSU
Non-current liabilities - group companies and associates	LS(1)	LSKKNC
Non-current liabilities - credit institutions	LS(2)	OBLLAN+LSKKRIN
Contingent liabilities	LS(3)	ANSVFSU
Total non-current liabilities	LS	LSKSU
Untaxed reserves	OR	OBESKRES
Equity		
Share capital	ER(1)	AKTIEKAP
Other restricted equity	ER(2)	OVRGBKAP
Accumulated profit or loss	ER(3)	BRUTORES
Profit/loss for the year	ER(4)	RESARB
Total equity	ER	EKSU

Appendix 2 – continued. Overview of all financial statement items and related variables in Serrano

Financial statement item	Symbol	Serrano variable
Income statement		
Net sales	RI	NTOMS
Personnel expenses	RK(1)	PERSKOS
Other external expenses	RK(2)	EXTKOSOV
Operating profit/loss before depreciation and amortization	RR(1)	RORRESUL-AVSKRIV
Depreciation and amortization	RK(3-5)	AVSKRIV
Operating profit/loss	RR(2)	RORRESUL
Operating expenses	RK	NTOMS-RORRESUL
Profit/loss from group companies and associates	FI(1)	RESAND
Interest income	FI(2)	RTEINKNC+RTEINEXT+RTEINOV
Interest expense	FK	RTEKOKNC+RTEKOEXT+RTEKOOV
Profit/loss after net financial income	RE(1)	RESEFIN
Group contributions	EO(4)	KNCBDR
Other extraordinary income and expenses	EO(5)	EXTRAINT+EXTRAKOS
Profit/loss before appropriations and tax	RE(2)	RESAR-SKATTER-BSLDISP
Change in untaxed reserves	BO	BSLDISP
Profit/loss before tax	RE(3)	RESAR-SKATTER
Recorded tax	SKT	SKATTER
Profit/loss for the year	RE(4)	RESAR
Other items		
Implied tax liabilities	Lask(1)	OBESKRES * ss
Changes in implied tax liabilities	Ch(Lask(1))	Lask/Lask (t-1)
Tax rate	ss	-

References

Aghion, P., Hart, O. D., & Moore, J. (1992). The economics of bankruptcy reform.

Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. Expert Systems with Applications, 94, 164-184.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The journal of finance, 23(4), 589-609.

Altman, E. I. (2000). Predicting financial distress of companies: Revisiting the Z-Score and Zeta® models.

Altman, E. I. and Levallee, M. L. (1981) Business failure classification in Canada. Journal of Business Administration 12(I), 147-164.

Altman, E. I., & Saunders, A. (1997). Credit risk measurement: Developments over the last 20 years. Journal of banking & finance, 21(11-12), 1721-1742.

Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETATM analysis A new model to identify bankruptcy risk of corporations. Journal of banking & finance, 1(1), 29-54.

Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). Journal of banking & finance, 18(3), 505-529.

Altman, E., 1993. Corporate Financial Distress and Bankruptcy. Wiley, New York.

Altman, E. I., Iwanicz‐Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z‐score model. Journal of International Financial Management & Accounting, 28(2), 131-171.

Antunes, F., Ribeiro, B., & Pereira, F. (2017). Probabilistic modeling and visualization for bankruptcy prediction. Applied Soft Computing, 60, 831-843.

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. Expert Systems with Applications, 83, 405-417.

Beaver, W. H. (1966). Financial ratios as predictors of failure. Journal of accounting research, 71-111.

Begley, J., Ming, J., & Watts, S. (1996). Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models. Review of accounting Studies, 1(4), 267-284.

Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. Journal of Financial education, 1-42.

Berkovitch, E., & Israel, R. (1998). The bankruptcy decision and debt contract renegotiations. Review of Finance, 2(1), 1-27.

Boratyńska, K., & Grzegorzewska, E. (2018). Bankruptcy prediction in the agribusiness sector: Lessons from quantitative and qualitative approaches. Journal of Business Research, 89, 175-181.

Bradley, D. B., & Cowdery, C. (2004). Small business: Causes of bankruptcy. SBANC: Small Business Advancement National Center. University of Central Arkansas.

Brenes, R. F., Johannssen, A., & Chukhrova, N. (2022). An intelligent bankruptcy prediction model using a multilayer perceptron. Intelligent Systems with Applications, 200136.

Bzdok, D., Krzywinski, M., & Altman, N. (2018). Machine learning: supervised methods. Nature methods, 15(1), 5.

Charraud, J., & Garcia Saez, A. (2021). Bankruptcy prediction models on Swedish companies.

Chen, K. H., & Shimerda, T. A. (1981). An empirical analysis of useful financial ratios. Financial Management, 51-60.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).

Chow, A. R., & Perrigo, B. (2023, February 17). The AI Arms Race Is Changing Everything. TIME Magazine. https://time.com/6255952/ai-impact-chatgpt-microsoft-google

Chui, M., Yee, L., Hall, B., & Singla, A. (2023). The state of AI in 2023: Generative AI's breakout year. McKinsey & Company Global Survey. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year#

Clement, C. (2020). Machine Learning in Bankruptcy Prediction–a Review. Journal of Public Administration, Finance and Law, (17), 178-196.

Coolidge, J. L. (1909). The gambler's ruin. The Annals of Mathematics, 10(4), 181-192.

Damodaran, A. (2006). The cost of distress: Survival, truncation risk and valuation. Stern School of Business.

De Andrés, J., Lorca, P., de Cos Juez, F. J., & Sánchez-Lasheras, F. (2011). Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). Expert Systems with Applications, 38(3), 1866-1875.

Dichev, I. D. (1998). Is the risk of bankruptcy a systematic risk? The Journal of Finance, 53(3), 1131-1147.

Durica, M., Frnda, J., & Svabova, L. (2019). Decision tree based model of business failure prediction for Polish companies. Oeconomia Copernicana, 10(3), 453-469.

Eckbo, B. E., & Thorburn, K. S. (2000). Gains to bidder firms revisited: Domestic and foreign acquisitions in Canada. Journal of Financial and Quantitative analysis, 35(1), 1-25.

Griffin, J. M., & Lemmon, M. L. (2002). Book-to-Market equity, distress risk, and stock returns. The Journal of Finance, 57 (5), 2317–2336.

Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. Review of Accounting Studies, 9 (1), 5–34.

Ij, H. (2018). Statistics versus machine learning. Nat Methods, 15(4), 233.

Ivanova, M. N., Nilsson, H., & Tylaite, M. (2023). Yesterday is history, tomorrow is a mystery: Directors' and CEOs' prior bankruptcy experiences and the financial risk of their current firms. Journal of Business Finance & Accounting.

Karlsson-Tuula, M. (2023, November 9). *Det behövs ett alternativ till företagsrekonstruktion*. *Altinget*. https://www.altinget.se/artikel/det-behovs-ett-alternativ-till-foretagsrekonstruktion

Kim, H., Cho, H., & Ryu, D. (2022). Corporate bankruptcy prediction using machine learning methodologies with a focus on sequential data. Computational Economics, 59(3), 1231-1249.

Korol, T. (2019). Dynamic bankruptcy prediction models for European enterprises. Journal of Risk and Financial Management, 12(4), 185.

Kreditrapporten. (n.d.). Konkurser. Kreditrapporten. https://kreditrapporten.se/konkurser

Kruppa, J., Schwarz, A., Arminger, G., & Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. Expert systems with applications, 40(13), 5125-5131.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. European journal of operational research, 180(1), 1-28.

Lindhout, E., & Larsson, E. (2021). Who Saw It Coming? The Efficacy of Bankruptcy Prediction Models on Private Retail Firms.

Malley, J. D., Kruppa, J., Dasgupta, A., Malley, K. G., & Ziegler, A. (2012). Probability machines. Methods of information in medicine, 51(01), 74-81.

Moore, P. G. (1983). The business of risk. Cambridge university press.

Nielsen, D. (2016). Tree boosting with xgboost-why does xgboost win" every" machine learning competition? (Master's thesis, NTNU).

O'leary, D. E. (1998). Using neural networks to predict corporate failure. Intelligent Systems in Accounting, Finance & Management, 7(3), 187-197.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. Journal of accounting research, 109-131.

Ooghe, H., & Balcaen, S. (2007). Are failure prediction models widely usable? An empirical study using a Belgian dataset. Multinational Finance Journal, 11(1/2), 33-76.

Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. Applied Soft Computing, 93, 106384.

Parsania, V., Jani, N. N., & Bhalodiya, N. (2014). Applying Naïve bayes, BayesNet, PART, JRip and OneR algorithms on hypothyroid database for comparative analysis. Int. J. Darshan Inst. Eng. Res. Emerg. Technol, 3(1), 1-6.

Pinches, G. E., Eubank, A. A., Mingo, K. A., & Caruthers, J. K. (1975). The hierarchical classification of financial ratios. Journal of Business Research, 3(4), 295-310.

Ptak-Chmielewska, A. (2019). Predicting micro-enterprise failures using data mining techniques. Journal of Risk and Financial Management, 12(1), 30.

Pulvino, T. C. (1998). Do asset fire sales exist? An empirical investigation of commercial aircraft transactions. The Journal of Finance, 53(3), 939-978.

Seidu, M. N. (2015). Predicting bankruptcy risk: A gaussian process classifciation model.

Shleifer, A., & Vishny, R. W. (1992). Liquidation values and debt capacity: A market equilibrium approach. The journal of finance, 47(4), 1343-1366.

Simonite, T. (2023, February 8). Guide to Artificial Intelligence. The WIRED. https://www.wired.com/story/guide-artificial-intelligence/

Skogsvik, K. (1987). Prognos av finansiell kris med redovisningsmått: en jämförelse mellan traditionell och inflationsjusterad redovisning. Ekonomiska forskningsinstitutet vid Handelshögskolan. EFI.

Skogsvik, K. (1990). Current cost accounting ratios as predictors of business failure: The Swedish case. Journal of Business Finance & Accounting, 17(1), 137-160.

Skogsvik, S. (2021). Financial Analysis: Capital Investment Risk [Lecture two]. Stockholm School of Economics, Stockholm, Sweden.

Smith, M., & Liou, D. K. (2007). Industrial sector and financial distress. Managerial Auditing Journal, 22(4), 376-391.

Song, Y. G., Cao, Q. L., & Zhang, C. (2018). Towards a new approach to predict business performance using machine learning. Cognitive Systems Research, 52, 1004-1012.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. Management science, 38(7), 926-947.

Tay, F. E., & Shen, L. (2002). Economic and financial prediction using rough sets model. European Journal of Operational Research, 141(3), 641-659.

Thorburn, K. S. (2000). Bankruptcy auctions: costs, debt recovery, and firm survival. Journal of financial economics, 58(3), 337-368.

Tian, S., Yu, Y., & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. Journal of Banking & Finance, 52, 89-100.

Tsai, C. F., Hsu, Y. F., & Yen, D. C. (2014). A comparative study of classifier ensembles for bankruptcy prediction. Applied Soft Computing, 24, 977-984.

Wang, G., Hao, J., Ma, J., & Jiang, H. (2011). A comparative assessment of ensemble learning for credit scoring. Expert systems with applications, 38(1), 223-230.

Wang, G., Ma, J., & Yang, S. (2014). An improved boosting based on feature selection for corporate bankruptcy prediction. Expert Systems with Applications, 41(5), 2353-2361.

Wang, G., Ma, J., Huang, L., & Xu, K. (2012). Two credit scoring models based on dual strategy ensemble trees. Knowledge-Based Systems, 26, 61-68.

Weidenman, Per. "The Serrano Database for Analysis and Register-Based Statistics." at Swedish House of Finance Research Data Center. Accessed 2023-10-20. https://www.hhs.se/en/houseoffinance/data-center/

White, M. J. (1994). The costs of corporate bankruptcy: A US-European comparison. Available at SSRN 5552.

Yazdanfar, D., & Nilsson, M. (2008). The bankruptcy determinants of Swedish SMEs. In ISBE International Entrepreneurship Conference, Belfast, Ireland, November, 4-7, 2008.

Yu, Q., Miche, Y., Séverin, E., & Lendasse, A. (2014). Bankruptcy prediction using extreme learning machine and financial expertise. Neurocomputing, 128, 296-302.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting research, 59-82.