FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

FAILURE PREDICTION MODELLING: A LITERATURE REVIEW

Word count: 11 726

Michiel Mahieu
Student number: 01803107

Promotor/ Supervisor: Prof. Dr. Heidi Vander Bauwhede

Master's Dissertation submitted to obtain the degree of:

Master of Science in Business Economics: Corporate Finance

Academic year: 2022 – 2023

UNIVERSITEIT
GENT

Confidentiality & Scientific integrity

Confidentiality of the Master's Dissertation

The author and the promotor give permission to use this master's dissertation for consultation and to copy parts of it for personal use. Every other use is subject to the copyright laws, more specifically the source must be extensively specified when using results from this master's dissertation.

Scientific Integrity

I declare that the research was conducted in accordance with the rules governing scientific and academic integrity. I have read, and acted in accordance with the Code of Ethics of the Faculty.

Abstract

Firms keep failing, and models keeps predicting their failure. Failure prediction (FP) modelling is a common research topic in literature since the sixties. Ranging from statistical models to artificial intelligence models, and models of theoretical importance. This study reviews the literature of failure prediction modelling for the period 2005-2022. The research starts with a bibliometric analysis to gain insight into the immense dataset of various failure prediction models. Afterwards, an in-depth search reviews the main statistical and artificial intelligence models. Although a part of the literature suggest a total shift to the artificial intelligence models, the statistical models continue to have a persistent value. The purely, accounting input data models are still used but few new models appear in this domain. Other input data as market-information, topological pattern recognition is used in the newer models. NNs, deep learning models, multi-criteria decision aid and other artificial intelligence models with their enhancements were all incorporated in this research. The improved flexibility in input data, prediction accuracy, and dataset requirements as normality contrasts with the complexity of model composition. This negatively affects the comprehensibility of a model and its universal use. The trade-off between accuracy and comprehensibility will appear to be a challenging issue.

Foreword

This Master's dissertation is written to obtain the degree of Master of Science in Business Economics, graduation subject Corporate Finance. During this research, I came across some previously seen concepts. However, the credo of Ghent University: "dare to think" was certainly applicable throughout my research. Many new topics crossed my path during this process. For instance, I was introduced to artificial intelligence models. This unfamiliar world brought many challenges with it. While writing this research, my analytical and critical thinking skills were constantly triggered. Throughout the process, several people helped me in every possible way. Therefore, I would like to thank thank some of them in particular for their unconditional support and helpful feedback.

First, I would like to thank Prof. Dr. Vander Bauwhede for all the constructive feedback, her guidance and useful insights. Furthermore, I would like to thank my girlfriend, friends, and family for all the support during this period.

Michiel Mahieu
Ghent, January 2023

Table of content

List of tables

List of figures

List of abbreviations

FP= Failure Prediction

MDA= Multiple Discriminant Analysis

SVM= Support Vector Machine

NN= Neural Network

MCDA= Multi Criteria Discriminant Analysis

1. Introduction

In June 2022, newspaper *De Tijd* publishes: “Signal jumps to orange” (Michielsen, 2022, p.1). The signal represents the number of companies going bankrupt per month. Bankruptcies are a structural problem and a never-ending story. So, it is important to predict which companies could fail.

With the appearance of failure prediction (FP), stakeholders became able to make predictions about bankruptcy. These predictions guide stakeholders in questions like; will this company be resilient enough to cope with certain shocks and survive in the current economic climate? Although exceptional situations cannot be predicted, and many external factors play an active role in the financial health of companies, FP can be a basis for decision-making. A considerable number of classification models for FP have been developed. These models were checked, rechecked, updated and combined with new models. There is no one-size-fits-all analysis, but all types have the same initial aim. Making the accuracy of prediction as high as possible. The most recent models are reported and critically assessed but, to the best of my knowledge, overview papers covering the period 2005-2022 do not yet exist.

This research extends on this relatively new period. As described in the methodology, the overview starts from 2005. The reasoning behind this starting year is based on two rationales. This research represents a follow-up on the models for FP from Financiële analyse van de onderneming (Vander Bauwhede et al., 2017). Additionally, 2005 has a practical relevance due to the credit crisis that followed shortly afterwards. This is discussed in more detail in the methodology. The research question will therefore be: what are the developments in FP modelling from 2005 onwards? This research question can be divided into two sub-questions which separate the models in both the literature review and the research results. First, which artificial intelligence models have been developed after 2005? Subsequently, to obtain a full picture of the trends at play in FP, the changes made in earlier developed models after 2005 will be incorporated. Therefore, the second sub-question will be: what enhancements in statistical models were developed after 2005?

2. Literature Review

2.1 Definition of failure

Balcaen & Ooghe (2006) pointed out the issue of the arbitrary definition of failure. The word failure is used in different contexts depending on the authors' interpretation. According to Balcaen & Ooghe (2006), bankruptcy, financial distress, cash insolvency, loan default… have all been linked to failure in literature.

To clarify what is meant with failure in this research, the criteria for failure are based on a legal definition. Bruloot (2021) describes this definition, which is in accordance with the Belgian criteria, as 1. being a trader, 2. having sustainably ceased to pay, and 3. having a shaken (unstable) credit. The latter implies that credit solutions have been exhausted, including the impossibility of postponing payments. Instead of a legal perspective, it is also possible to rely on an economic or financial perspective (Zhou, 2013).

2.2 Historical Evolution

After defining failure, one could look back in the evolution of FP to get an insight into the various models. As mentioned before, recent developments do include updated pre-existing models next to the new models. Therefore, this literature review section gives a clear overview of the evolution in FP modelling. From the very beginning with the univariate analysis of Beaver (1966) until models developed to approximately 2005, the starting point of the research itself. The research results section gives an overview of the recently developed models and re-evaluation of already long-standing ones.

The earliest statistical methods cannot be seen as a standalone type. These methods influenced the literature and development of new types. The real start of FP models dates to the thirties (Bellovary et al., 2007). However, these authors describe Beaver's work (1966) as the first real breakthrough in FP. Therefore, in the following, a concise overview of the first period, starting in the '60s will indicate the start of FP. In this period, FP is also called FP analysis because of the non-active use of ratio analysis in that period.

The literature of FP modelling can be split into distinct groups of models. Various segmentation methods have been used by different authors. According to the research of Balcaen and Ooghe (2006), the models for FP can be divided into four categories.1. univariate models, 2. risk index models, 3. Multiple Discriminant Analysis (MDA) and 4. conditional probability models. The work of Jackson and Wood (2013) applied an alternative categorization within FP modelling. These authors listed the models in decreasing order of popularity as follows: 1. MDA, 2. logit models, 3. Neural Networks (NN), 4. contingent claims, 5. univariate analysis. Another categorization from Aziz and Dar (2006) is more overarching. This research divides the models into 1. classical statistical models, 2. artificially intelligent expert system models, and a third and last section for the theoretical models.

In this literature review section, the categorization will be as in Aziz and Dar (2006) split up into classical statistical models, artificial intelligence models, and a theoretical model. The subdivision of the classical statistical models will be highly equal to the work of Balcaen and Ooghe (2006). After the classical statistical models, the review will progress to the artificial intelligence models. However, the NN already gained attention in the nineties. So after the section on classical statistical models, this research will go back in time to some extent.

2.2.1 Classical statistical models

2.2.1.1 Univariate models

To the best of my knowledge, the univariate analysis is the only used model in the starting period of FP modelling. After the period 1930-1965, the revolutionary works of Beaver (1966) and Altman (1968) mark the beginning of an era of new models. The different models can be seen in defined time ranges. From 1930-1965/1968 the main analysis methods were all univariate methods for ratio analysis. The combination of different univariate ratios into one model leads to the launch of the MDA.

Bellovary et al. (2007) indicate one ratio to be central in all studies in this time frame. The Working Capital to Total Assets ratio was unanimously indicated as a failing ratio in bankrupt firms. This research also indicated the Current Ratio as an indicator. However, the effect was seen as subordinate to the working capital to total assets ratio. Although there were differences in the number of analysed ratios, the method in this period was always the same. Based on a sample of non-failing firms, a standardized ratio value was indicated. A distinction between failing and non-failing firms was made based on a comparison between the obtained value and standardised value.

In 1966, Beaver used the Univariate analysis for 30 ratios. According to Bellovary, the main contribution of Beaver's work were the predictive abilities of individual ratios to classify a firm as bankrupt of non-bankrupt. The highest predictive value was obtained by the Net Income to Total Debt ratio (Beaver, 1966). Although Altman (1968) is seen as the founding father of MDA, Beaver already indicated the importance of different ratios in his analysis. With this claim, he proposed future research to compare multiple ratios in one analysis at the same time to improve the predictive ability of the model.

2.2.1.2 Multiple Discriminant analysis

What follows will become the most well-known work in the category of FP modelling. With his research, Altman (1968) would show a technique to attempt “the quality of ratio analysis as an analytical technique” (p.2). The implied MDA creates a Z-score. This score returns an overall viability value for a certain company. The formula combines discriminant coefficients with independent variables. The discriminant coefficients form a linear combination of characteristics (read: financial ratios) that best discriminates the groups of firms (bankrupt vs. non-bankrupt). The independent variables represent the actual values for the specific researched enterprise. The final discriminant function uses five different ratios, also called the five financial forces. 1. Working capital/ Total Assets (previously already indicated as relevant), 2. Retained Earnings/ Total Assets, 3. Earnings Before Interest and Taxes/ Total Assets, 4. Market Value of Equity/ Book Value of Total Debt, and 5. Sales/ Total Assets. The values for these ratios are combined with the discriminant coefficients to retrieve a company-specific Z-score. This method showed high predictability for the one-year before failure.

In 1977, Altman released a new article that includes his updated version of the Z-score. Furthermore, during the eighties, MDA remained a generally accepted analysis but its importance declined slightly (Balcaen & Ooghe, 2006). The focus changes to the logit analysis.

In 1982 Ooghe- Verbaere (Vander Bauwhede et al., 2017) invented some static and empirical models to conduct the viability of an enterprise. The discriminant score, equal to the Z-score obtained in the Altman model, is compared with a predefined critical value to categorize a specific firm as failing or non-failing (Ooghe & Verbaere, 1985). The independent variables were selected by a stepwise method, selecting the next best discriminator at each step. Another distinction with the original Altman model (1968) is the constant term in the function. According to the authors, the reliability of the model can be examined by the misclassification of observations. Besides, the authors also shared some critical thoughts about the model. This research has a retrospective character. This simply means that the model is not proven to be effective in predicting failure. However, it does work for known cases in the past*. Known cases* is not randomly chosen here. Not all annual accounts are immediately available at the end of the financial year. Especially enterprises in difficulties might postpone their annual accounts publication (Ooghe & Verbaere, 1985).

As defined above, the misclassification of observations is a critical issue in FP. Two types of errors might arise when classifying firms. A company that has been classified as viable might be in bankruptcy after X years. This type of mistake is called a type I-error or credit error. The other situation in which a non-viable classified company would have been viable after X years is called a type II-error or commercial error. Watts and Zimmerman (1986) already indicated the expensiveness of a Type I-error. Later, the review of Jayasekera (2018) also mentioned that the Type I-error is the most expensive one. This review also looked at Beaver's study (1966) and added that this model classifies non-failed firms better than failed ones. This implies a higher Type I- error which must be considered when using the model. The question to ask when categorizing edge observations is all about the opportunity cost for society by indicating a particular firm as failing vs non-failing (Jayasekera, 2018).

2.2.1.3 Conditional probability models

Logit analysis

Altman's breakthrough triggered an era of new models. In 1980, Ohlson established the logit analysis. Ohlson used the shortcomings of the MDA to establish his model. The urge for normally distributed predictors is one of the drawbacks according to Ohlson. Another drawback is related to the outcome. The Z-score implies an ordinal ranking, and is therefore limited in terms of interpretation. The last drawback is the use of variables and predictors for the matching of firms. It would be more relevant to use them in a non-arbitrary way (Ohlson, 1980). The author concluded that the predictive value of MDA was overestimated, and the error rates were higher in practice than mentioned in the prior literature. Ohlson (1980) indicated four factors that significantly influenced the default probability. 1. The size of the company, 2. measures of the financial structure of the firm, 3. performance measures, and 4. liquidity measures. The implied probabilistic model includes random effects to the logarithm in combination with a vector of predictors. The result is a multivariate probability score between zero and one. Together with the cut-off-point, firms are indicated as failing vs. non-failing.

Ohlson used the simple Binary Logit model for his analysis. Besides this model, the multinomial logit model and its extension, the nested logit models gained more literature attention in the sixties and seventies (Jones & Hensher, 2004). The multinomial logit model improves the binary Logit model by allowing independent variables to have more than two categories (Kwak & Clayton-Matthews, 2002). The binary logit model and multinomial logit model are just weight functions of fixed parameters without incorporating behavioural information into the parameters. The mixed logit model maximizes the use of this behavioural information.

The inclusion of additional behavioural information might improve the predictability of the model (Hensher & Greene, 2003).

Probit analysis

As the logit model assumes a logistic distribution (Balcaen & Ooghe, 2006), the probit model requires a cumulative distribution. The main distinction between the two models is the distribution of the observations (Klieštik et al., 2015). The logit model has flatter tails than the probit model. In the case where many extreme observations occur, the outcomes of the two models can differ significantly.

2.2.2.2 Survival Analysis

Another model that obtained increasing attention was the survival analysis. This model is a dynamic statistical tool to analyse the time until a certain event. The survival function represents the possibility that the business will survive past a certain time. The hazard function represents the rate of failure at a certain time. (Gepp & Kumar, 2008). Essentially, survival analysis is a useful technique in examining the effects of variables on the timing of events (Parker et al., 2002). The low awareness of the time dimension in FP is early indicated as a general issue in FP, especially in the statistical models (Balcaen & Ooghe, 2006).

2.2.2. Artificial intelligence models

As previously stated, the models are divided into three categories. The following will enhance on the artificial intelligence models. However, it must be clear that there is an overlap in periods between the two categories.

2.2.2.1 Neural Networks

The nineties could be described as a disruptive period in the landscape of FP modelling. This period includes the start-up of the Neural Network (NN). These models are artificial intelligence-based predictions used to indicate the viability of firms. Odom & Sharda (1993) described the distinctive nature of these models. NNS made it possible to e.g., analyse imbalanced data sets. ANN’s are self-improving streams. By letting it analyse more data sets, it will improve itself and analyse new data sets with previously acquired knowledge.

Balcaen & Ooghe (2006) made a clear overview of problematic topics related to statistical classification models. The earlier mentioned timing issue is one of the most prevailing issues in FP. Schumway (1999) emphasises the disturbed prediction ability of only one “snapshot” of the firm. The urge for a time-series-based model is certainly visible. The previously mentioned retrospective character of statistical models is also highlighted by these authors as an unsolved problem. Besides these timing related issues, some data-related problems also seem to occur in FP. The biggest twist in NNs is the ability to cope with certain levels of data instability and stationarity problems. The stationarity problem means the constant relationship between independent and dependent variables, which is not existing in reality. Another key contribution of NNs is their capacities for non-linear modelling (Van Gestel et al., 2009). Other problems as variable selection, and the use of annual account information are subordinate to the first two but also highly relevant. However, they are not further specified in this research.

In 1990, Odom & Sharda were the first to use a NN model to test company failure and compare it with the output of traditional methods. Simply stated, this method uses a training sample (with a percentage of bankrupt and non-bankrupt firms) to train the network. Afterwards the data set that one wants to examine is inserted into the network to make predictions. Odom & Sharda (1990) concluded that this research has a better predictive value in all scenarios compared to the MDA. The predictive accuracy was also more consistent with different distributions in the training sample.

2.2.2.3 Decision Trees

Unlike logit analysis and discriminant analysis, decision trees are non-parametric. Decision tree modelling constitute a breakthrough in simplifying FP modelling (Gepp & Kumar, 2010). decision tree modelling is one of the more recent techniques. This model is normally used to distinguish two groups based on a predetermined variable (Gepp & Kumar, 2010). Within the field of bankruptcy prediction, this binary separation creates a group of failing and non-failing elements. This recursive process exists of diverse levels and uses cut-off values to indicate two, non-overlapping groups. The main advantage of this technique is the user-friendly interpretation of the graphical model. In this model, it is possible to incorporate different misclassification costs for Type I and Type II error as inputs. The first similar model was applied in 1985 under the recursive partitioning algorithm (Frydman et al.,1985). Around 2000, Joost et al. (1998) introduced the decision tree model in FP. The research of Frydman established a clear step-by-step plan to set up such an algorithm and compared the results with a discriminant analysis. Without neglecting the disadvantages of this algorithm, this research demonstrated the superiority over discriminant analysis. In the following years, variants of this recursive partitioning algorithm were established, and the techniques seem still pertinent today. This model will be further examined in the following section.

2.2.2.4 Support Vector Machines

Machine learning models, as previously mentioned, have the capacity to deal with difficult, imbalanced data sets. The Support Vector Machine (SVM) is a machine learning model that has a high degree of generalization (Cortes & Vapnik, 1995). This made it possible to map non-linear data sets in a linear decision surface. Literature found a higher performance for SVM compared to the classic NNs (Li & Sun, 2009).

SVM and decision trees are discussed briefly in this section. These models are among the newer ones before 2005 so it makes sense that most of the literature for these models can be found in the research results section.

2.2.3 Theoretical models

The above-discussed models can be seen as practical methods for FP. Other models like the option pricing model, also called contingent claims analysis, extend the category of FP from a more theoretical point of view. As described by Charitou A. (2000), the standard option pricing model is easily determined by five variables. The underlying variables in this model are 1. the book value of total liabilities due at maturity, 2. the current market value of the firm's assets, 3. the standard deviation of % firm value changes, 4. the average time to the debt's maturity, 5. the difference between the riskless return and the firm's pay-out yield. The intuitive reasoning based on these variables is as follows. If the firm's value of the assets falls below the amount of debt outstanding to creditors, equity holders will arrange for bankruptcy so that the firm's assets are transferred to the creditors. Because of their limited liability right, the equity holders will not be charged for anything because there are no liabilities anymore.

2.2.4 Causes of failure and the failure processes

So far, all discussed articles have addressed predictions for failure. However, it later turned out to be crucial to have a broader view of all factors which could cause failure. Ooghe & De Prijcker (2008) describe bankruptcy causes and failure processes. With this research, Ooghe initiated factors, both financial and non-financial, which meet the inclusion of the time dimension of failure. This research is concentrated on four distinct types of failure processes. First, the failure process of an unsuccessful start-up. Secondly, the failure process of an ambitious growth company. As a third option, he included the failure process of a dazzled growth company. And the last one is the failure process of an established company. According to the author, it would be very absurd to use the same ratios to all companies in different life stages. As an example, one of the main reasons for failure in start-up firms is the inexperienced management and the deficit of obtained advice by experts. Contrarily, in more established firms, the management is becoming expert but might become to over-optimistic because the firm is doing so well. Taking ungrounded risks might lead to failure in these cases. External factors as wrong estimated market size, and restricted change in customer behaviour towards the new strategy might also be important here. Argenti (1976) excludes the latter, external causes, as drivers for failure in this stage. This author indicates that failure is in this stage mainly related to the policy of the company, and management. As the company is reaching a certain level of maturity, management might become less committed and motivated to keep up to date. A loss in competitive advantage and following increasing expenses might also influence the bankruptcy probability. This section aside to ensure the multifactorial view on FP. It might be relevant to not use a unilateral view, only focused on financial data.

A brief summary of all categories has now been established. From the above, it might be important to retain a broad view on FP. The research results section will have to reveal whether this broader view is incorporated in the post-2005 enhancements.

3. Methodology

To examine the developments in FP, a dataset of relevant articles has been composed. To clarify the selection, the source database and search query will be explained. Then, a clear timeframe will be established. In this way, all relevant articles will be incorporated in the research.

3.1 Data set

On 18/09/2022, the final dataset was retrieved from Web of Science, as opposed to the search query of Zambrano et al. (2021) which was conducted on Scopus. This research is a suffix to the literature in reviewing failure prediction modelling. Web of Science has, in contrast to Scopus, a completer coverage of book reviews (Scopus versus Web of Science, 2022). The review in the research results was therefore executed with Web of Science. By this, no degrade is meant towards Scopus.

The keywords of this search were established based on the research of Shi & Li (2019) but were somewhat different. "Topic" was chosen as field search for the keywords. It became already clear that different articles in the dataset did mention the keywords in the abstract but not in the title. The keywords "failure prediction" OR "insolvency prediction" OR "bankruptcy prediction" OR "financial distress prediction" OR "default prediction" OR "early warning" were used for the search. This was combined with an AND search for "business". There was a check search made with synonyms for "business", but this did not include any extra relevant articles. The keywords were kept general for a specific reason. Predefining the search query with model names in combination with "failure prediction" limits the search engine to find new models. As the purpose of this research is to look for different, updated models, predefining them does not seem suitable for this work.

The term FP might sound overarching. Failure can be related to company failure, specific machines, stock models, or even heart disease. All categories in the dataset were checked to make sure that the group of available papers only includes the relevant ones. All relevant categories were summarized in the table 3 in the appendix. Selecting the relevant articles from a dataset can be done by e.g. citations, by impact factor or by selecting certain journals. In this research, the relevant articles were selected by reviewing the quality of an article by the Australian Business Deans Council (ABDC) list (Tattersall, 2022). This list classifies journals by a rating, starting from A*. To ensure the quality of the dataset used in this paper, the list of relevant articles was limited to the A* and A journals. The dataset was computed based on 213 different journals. These were all checked one-by-one manually with the quality list from the ABDC. This results in a list of 31 different relevant journals.

3.2 Time determination

After specifying the relevant journals and categories, the timeframe has been established. In the very beginning, the number of articles was limited. Figure 3 in appendix gives an overview of the released papers since 1954 until now without the prespecified categories and journals as described above. In 1960 the two researches include Altman's research and Beaver's work.

The biggest twist in the number of articles can be found around 1980 and in the period 2005-2015. The latter period is relevant for the starting point of this research and explicable. Starting in 2007, reaching the worst in 2008, Belgium and the entire world suffered a credit crisis until 2011 (Crouhy et al., 2008). This crisis found its origin in the United States real estate market. This disaster had an influence on all sectors and so business failure became a hot topic. Until then, bankruptcy prediction modelling had not received sufficient attention. This crisis made it a prominent theme (Barboza et al., 2017). The changing interest is also remarkable in the number of released papers. With this reasoning in mind, the period of 2007 has been determined as an interesting starting period for this reviewing paper. However, this paper is a follow-up to the chapter of failure prediction modelling in the book Financiële Analyse van de Vennootschap (Ooghe et al., 2021). This chapter assesses various models until 2005.

Since these two time periods fall in the same range, 2005 will be taken as the starting point of the review. Again, the research question therefore is: what are the developments in FP modelling from 2005 onwards? This period of 17 years covers the interest of both statistical and artificial intelligence models. Adjusting the dataset with this timeframe, a total of 338 articles were determined relevant for the period 2005-2022.

4. Research Results

4.1 Bibliometric analysis

It is impossible and of no added value to describe each of the 338 articles of the dataset. Therefore, a bibliometric analysis might help to give an insight into the retrieved dataset before going into more detail. Bibliometric analysis is used in part to identify and visualize relations between different journals or authors (Henninger, 2012). It also allows to indicate related keywords. As mentioned by Shi & Li, (2019), bibliometric analysis does not only look for links between articles but can also identify trends in enormous amounts of quantitative data.

Before using the bibliometric analysis on the dataset, a summarizing table of the most cited works in the dataset is shared. In appendix table 4, sixteen researches with their respective publication year, author and journal can be found. The cut-off value for this table was set at more than 150 citations. The journal Expert Systems with Applications is most frequently cited in these researches. This journal is specified in domains related to artificial intelligence, computer science application and engineering (Resurchify, 2022). This might indicate a major interest of artificial intelligence models in failure prediction modelling over the statistical ones. However, a more profound search will be necessary to extract conclusions from the dataset.

To gain insight into the dataset, a bibliometric analysis was conducted. Starting from the Web of Science dataset, the 338 relevant articles were inserted into VOSviewer. This is a visualization tool for bibliometric networks developed by Van Eck & Waltman from Leiden University, Netherlands. First, the important authors and interrelatedness between authors was examined. VOSviewer makes it possible to search for this interrelatedness. The figure below shows the interrelatedness between authors with a minimal of three articles within the period 2005-2022. Combined with the Web of Science list of most mentioned authors, an enumeration of regularly active authors in the timeframe can be established.

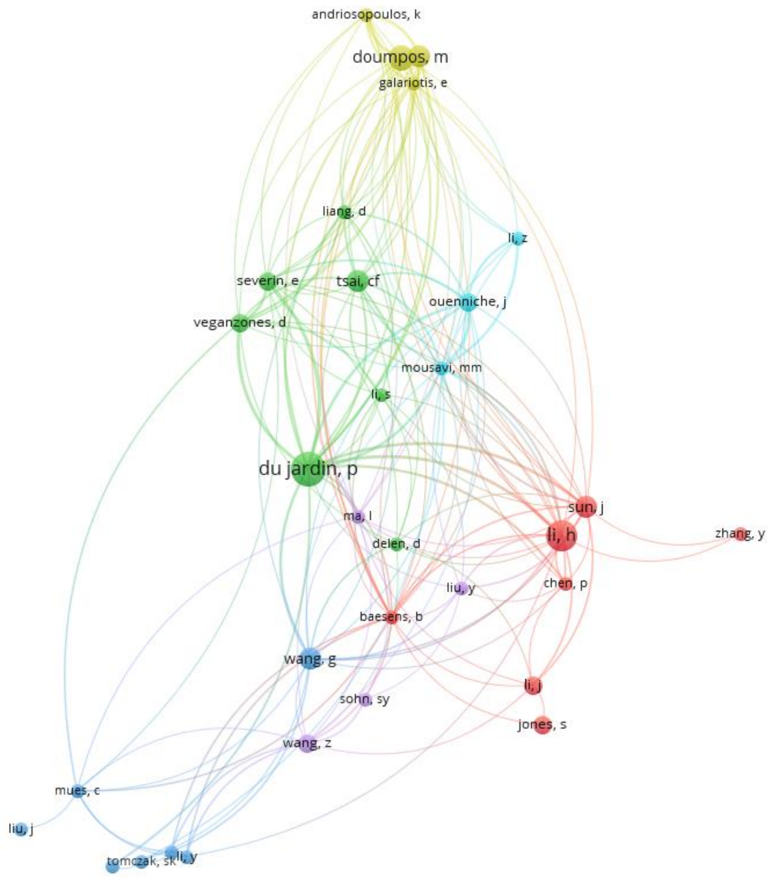

Figure 1: Interrelatedness authors VOSviewer

Du Jardin, Li, Sun, Doumpos and Quenniche are authors worth mentioning. The researches of these authors are certainly relevant for further examination. However, plenty authors have a certain research domain so they publish several articles of one or two models. This research has the purpose to give an overview of all relevant models and enhancements. It is therefore possible that the appearance of the authors is less prominent in the discussed articles below.

The keyword search, another visualization possibility, might be the most important related to this research. This tool searches for common keywords in the articles. This visualization does not only indicate which keywords are used intensively, it also shows which keywords occur frequently together. The figure below shows all the author keywords that meet the minimum occurrence of five. These author keywords are keywords that are, according to the author, most appropriate in accordance with the article. The figure's colours range from blue to yellow to sort them by period. Bankruptcy prediction and financial distress are obviously the most used ones. Terms as SVM and logistic regression are also standard practices in FP modelling and appear frequently according to the visual. It is interesting to see that some of the statistical models also preserve their value. As such, logistic regression, and financial ratios are still indicated as keywords in some authors' work. NN and ensemble learning are also frequently mentioned and well-known in this research area. The relatively newer techniques as Adaboost, genetic algorithms and deep learning were also all indicated to occur frequently. Without extracting direct conclusions from this, this might prescribe a shift in models. It is possible that smaller, less-known, newer models are less visible in this search because of the minimum occurrence requirement of five. Therefore, more in-depth research is needed.

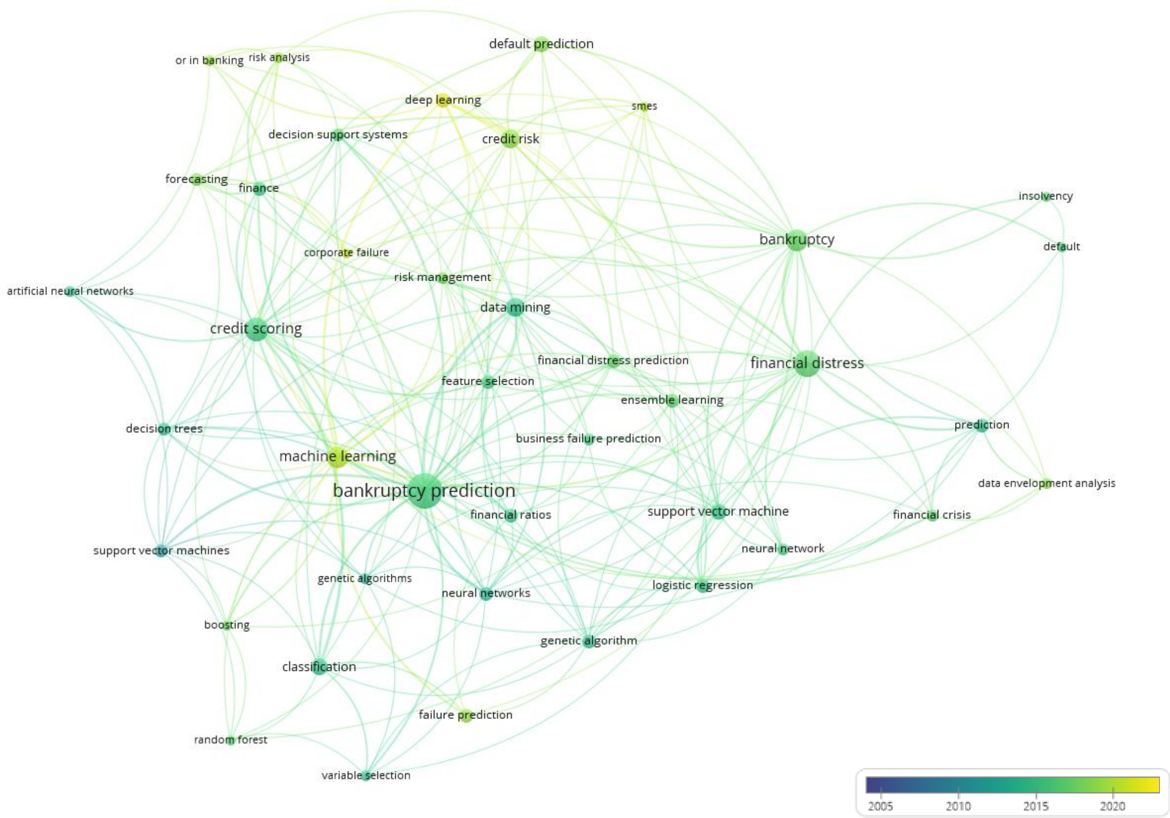

Figure 2: Author keywords VOSviewer

The last feature of the bibliometric analysis that is relevant for this research is the geographical distribution. In appendix figure 4, the geographical distribution is shown by number of citations per country. The density of the country in the figure illustrates the magnitude of citations. It is clear that China, England, and USA are the main countries in this visual. However, Belgium also seems to appear in the top ten equal to Germany and France.

Zambrano et al. (2021) also used a bibliometric analysis tool to analyse their data. However, this research did not do an in-depth search in the data. The basic models that were indicated in studies of Zambrano et al. (2021) and Shi & Li (2018) were together summarized in table 5 and table 6 in appendix. These will not or briefly be explained in this research, as there is no added value in re-explaining the basic findings of these models. However, this research does look at certain general trends in NNs, and further enhancements on models that are worth mentioning. The way this research categorizes a model or enhancement as worth mentioning is discussed below. The figure of VOSviewer with the author keywords gives a general trend implication. However, in the following section, an in-depth search was done.

4.2 In-depth search

To the best of my knowledge, there is no program or algorithm that sorts all articles one by one in the right category. So, every article of the 338 articles was first analysed based on the title and abstract to find which model was used. However, for various articles, it was still not clear what method was exactly used for the research. In these articles, a more in-depth search was needed.

Based on this research, a total list of models and enhancements was established. The models with an occurrence of five different researches were directly classified in the table. The ones that did not match the five-time occurrence were added based on relevance or high number of citations in relation with the time passed since publishing. It might be possible that models were already existing before 2005 but did not receive the same attention then. Because the research question is about developments in failure prediction modelling since 2005, a reuse or more intensive use of a pre-existing model is also an important development. Together with this rationale, the table below was created.

As previously mentioned in the literature review, the classifying technique is equal to Aziz and Dar (2006). The only difference is the absence of theoretical models as the Option Theory in the literature review. The table below shows a summary of the further discussed models. In other words, it constitutes a summary of all developments in failure prediction modelling from 2005 based on the retrieved dataset.

Statistical models	**Enhancement one**
Discriminant analysis	Markov for discriminant analysis
	Magic square
Logistic regression	Random effects logistic regression
	Tax arrears
	FiTo-meter
Hazard model	Improved discrete time hazard model
	The frailty model

Table 1: Statistical models and enhancements from dataset

Artificial Intelligence models	Enhancement one	Enhancement two
Neural Network	General Regression NN	Fruit Fly Optimization algorithm
	Self-Organizing Maps	Growing Hierarchical Self-Organizing Maps
		Topological Patterns
	Quantile Regression	
Support Vector Machine		
Genetic Algorithm	Hybrid Genetic algorithm	
	Real-valued Genetic algorithm	
Bayesian Network		
Case-based reasoning		
Ensemble models	Bagging	Random Forest
	Boosting	Adaptive Boosting
		Gradient Boosting
		Extreme Gradient Boosting
	Stacking	
Deep learning	Convolutional NNs	
	Natural Language Processing	
MCDA	PROMETHEE	
	Data Envelopment Analysis	Graph Theory

Table 2: Artificial Intelligence models and enhancements from dataset

The following section will go more in depth into new, updated, or combined models. Categories for which no new or relevant information was found in the dataset are not discussed in the text below. To limit the table to the essentials, applications on certain models were not listed and explained if they were not significantly different. E.g., the Adaboost was executed on the SVM in the dataset, but this research limits itself to AdaBoost in general.

4.2.1 Classical Statistical models

4.2.1.1 Discriminant analysis

The oldest models can be found within the statistical models, and more precisely in discriminant analysis. In the literature review, considerable attention was already given to this part. These are easy-to-interpret models and proved to be established practices today. Discriminant analysis is mentioned in almost every literature review of researches in the dataset and is discussed in more detail or used as a basis for comparison. The authors still indicate discriminant analysis as a keyword of their research in 81 of the 338 articles. With this, it was not said that the predictive value is equal to more recently developed models. Very few renewing works were published in this research domain due to an increased attention for more dynamic and flexible models (Volkov et al, 2017). This author used Markov for discrimination on a Belgian database. This model (partly) fulfils the dynamic requirement which most statistical models failed to achieve. The latter mostly use a moment analysis. Using multiple time periods in the non-ensemble models showed a better classification accuracy than the traditional one-time evaluations. This Markov analysis also shows remarkable results for the logistic regression.

Besides, another innovative scoring method, called the magic square, has been implemented in macro-economics. Bemš J. et al (2015) introduced this in the field of corporate failure cases. Based on the financial indicators as in Altman's research (1968), a polygon area was indicated for each company. Although the method is not outperforming artificial intelligence models, it is an improvement on simplification of interpretation (Bemš J. et al, 2015). These two enhancements demonstrate that the statistic models are not just used to compare and show the performance superiority of certain artificial intelligence models. Furthermore, artificial intelligence models will also prove more difficult on implementation and comprehensibility. However, Du Jardin (2017) concluded that the FP accuracy declines with financial ratios when the time horizon enlarges.

4.2.1.2 Logistic regression model

Ooghe is a widely cited author in this research. He is a non-negligible author in the Belgian literature of FP. In 2005, Ooghe & Spanjers created the FiTo-meter which is the first model in the Simple-Intuitive models, as they call it. The model is constructed based on eight different ratios, assembled from the four dimensions of a company's financial health: liquidity, solvability, profitability, and added value. By using a logit transformation, all ratios get a value between zero and one. Based on this, the FiTo-score is calculated as an average of these logit scores. This model does not enhance much on misclassification errors. Although, the model seems to perform well one and three years before failure which enhances on the number of times a model needs to be calculated. Besides, the FiTo-compas improves the interpretability by assigning the ratio performance of a certain company in a visual. Further, by excluding the coefficients in discriminant analysis, the model is certainly an improvement when it comes to stability and robustness.

Alongside the discriminant analysis, the logit model was also previously mentioned in the literature review. These models already made a shift from the original binary model to the nested logit model (Klieštik et al. 2015). Sohn & Kim (2007) applied the random effects logistic regression model for bankruptcy prediction. They identified the problem that until then, existing models could not explain the varying default probability under same conditions. Their model also incorporates uncertainty that cannot be explained by company's characteristics. The results show the superior classification accuracy of the model over the fixed effects logistic regression model.

Lukason & Andresson (2019) focused on tax arrears in FP. The authors start from the rationale that frequently postponing payments lead to an increase in the likelihood of failure. Adapting this hypothesis on the tax arrears of a company leads to a logistic model that accounts for postponed tax payments in addition to financial ratios. The combination of these model increases the prediction accuracy above the levels of both models separately. Moreover, the tax arrears are particularly useful when the bankruptcy date is approaching. This is in accordance with the prediction accuracy of financial ratios. It also increments when the time horizon narrows down (Du Jardin, 2017).

4.2.1.3 Hazard model

Hwang & Chu (2014) used an approved model of the discrete time hazard model. Before this research, the model assumed constant firm-specific predictors. By using this more flexible variant, it is possible to incorporate the evolution of macroeconomic dynamics e.g., economic downturns in crises. The possibility for these coefficients to fluctuate enhanced the predictive power of the model.

Another evolution in the hazard models is the use of the frailty model. The basic idea behind the frailty model is that a firms default probability correlates high with its industry default probability. Hertzelf & Officer (2012) already prove that default clustering is intra-industry related. The research of Chen & Wu (2014) goes even further, they use a frailty model with a hazard rate function. These researchers state that accounting for sectoral frailties decreases the bias and improves the predictive power of FP forecasting.

4.2.2 Artificial Intelligence models

4.2.2.1 Neural Networks

Alongside the hazard model, the NNS are also previously described in the literature review section. The broad scope of NNs and their various forms of application are both theoretical and practical. Shi and Li (2019) reported that NNs combined with the logit models are among the most frequently used ones.

The general regression NN as described by Specht & Romsdahl (1994) is “based on finding the expected value of a dependent variable given a set of input measures” (p.1). This model differs with the previous NNs in estimating continuous variables. The continuous variables occur when time-variation is considered. Furthermore, this model improves a lot on speed in training data processing. An important application of this model is in the Fruit Fly Optimization algorithm. Pan’s (2011) highly influential work, according to the number of citations, is an algorithm based on the food finding behaviour of the fruit fly. This gives a minimum and maximum value to the firms. Pan used four financial ratio variables. Being the Revenue Growth Rate, Fixed-Asset Growth Rate, Operating Profit Margin, and Profit Margin. The dependent variable in this research are the companies in risk. The main advantage of this algorithm is the continuing quality improvement of the result during the iterative process. The general regression NN indicates an improved prediction accuracy by implementing this algorithm.

Also, in the category of NNs, the Kohonen Map was the first in the self-organizing maps. This new machine learning technique is typified by unsupervised learning. Unsupervised learning can essentially be described as searching for patterns and similarities based on input data without predefining a certain output or pattern. In machine learning, this output will be represented as a two-dimensional map (Lee et al., 2005). The output is mostly represented on a rectangular grid. This type of visual representation often ends up in a reduction of dimensionality. Although this method was already invented in the nineties, the application of Kohonen maps to bankruptcy prediction was around 2000. Lee, K. et al (2005) also described the change in interest from back propagation (an algorithm for data training) to Kohonen maps, mainly because of their unsupervised nature. It is prominent that in the period 2005-2012, almost no researches about self-organizing maps were published in the retrieved dataset. The subsequent period does include considerable researches on this topic. These researches include e.g. Kernel-based fuzzy self-organizing maps and growing hierarchical self-organizing maps. The latter is an extension on the self-organizing map and is used by Huang, S. et al. (2014) for topological pattern discovery. By using this clustering model, the self-organizing map serves again as fraud detection model. This enables other sources than purely accounting data in failure prediction modelling. The section of deep learning will incorporate some of these alternative sources and go into more detail.

The last enhancement in the NNs is the quantiles regression function. Most of the regression functions have the purpose of minimizing the squared error loss function. Recently, Krüger & Rösch (2017) implemented the quantile regression function which is based on minimizing the loss in a quantile. Even more recently, Kellner et al. (2022) enhanced this function by using the quantile regression with a NN. This must enhance the flexibility of the model and allows non-linearity in quantiles. This research has the title “ Opening the black box -Quantiles neural networks for loss given default prediction”. The purpose is, other than the majority of the dataset researches, to look beyond the prediction accuracy improvement. It tries to give an understandable processing of the data which is a problem artificial intelligence models still struggle with.

4.2.2.2 Support vector machine

In accordance with the starting data of the timeframe (2005), the support vector machine learning model began to arise (Min & Lee, 2005). This model was originally invented for pattern recognition and found its way as an unsupervised machine learning technique around 2003. Van Gestel et al. (2003) were among the first to use the Least square SVM in comparison to a NN. Throughout the years, SVM continues to appear on a regular basis in the dataset. 30% of the authors in this dataset indicated SVM as an author keyword. According to Veganzones & Séverin (2018), the intensive use of this model over time is mainly related to imbalanced data sets. These authors indicate the processing of imbalanced datasets as one of the main weaknesses of the classic prediction models (logit analysis, random forests) For the SVM model, there is an almost unchanging prediction accuracy by using balanced or unbalanced data inputs. The vast majority of researches in the dataset use SVM as a point of comparison with other models (Kumar, P. & Ravi, V.; 2007), (Lessman et al., 2015) or as an ensemble processing model for a particular case (Wu et al., 2007), (Sun et al., 2017). More detailed research to all these ensemble models would lead this study too far. Exploring the evolution in SVM ensemble models (read: combinations with SVM) would be an interesting path for follow-up research. This is further discussed in the future research section.

4.2.2.3 Genetic algorithm

Genetic algorithms have frequently been combined with other artificial intelligence models. This hybrid algorithm is mainly done in combination with a NN and the later discussed case-based reasoning. The idea behind this algorithm is the survival of the fittest. This algorithm is used to optimize the selection of indicators for the NN. By selecting the “fittest” indicators, a string is formed which keeps optimizing itself until no more value adding indicators can be found (Shin, & Lee, 2002). The evaluation function in genetic algorithms is called the fitness function. According to Kozeny (2015), the importance of the fitness function was neglected too often. This author also indicates the existence of similar functions as the fitness function. An example is a function based on a bitmask. However, these alternative functions will not be further elaborated.

The genetic algorithm knows a much broader application range in recent years. E.g., the study of Min et al. (2006) applies the GA in a hybrid form with the SVM to optimize the feature subset and parameters of SVM. Beside the hybrid form to optimize the SVM's parameters, the real-valued genetic algorithm also aims to optimize the parameters (Wu et al., 2007). The real-valued genetic algorithm is more straightforward, faster and more efficient than the standard binary genetic algorithm.

4.2.2.4 Bayesian Network models

The use of the bayesian model in the domain of failure prediction was first examined in 2001. Because the major theoretical contributions lay behind 2005, this model is not yet covered in the literature review section. This model mainly differs in requirements for the underlying distribution of variables (Sun, I. & Shenoy, P.; 2006). This model is very adaptive and can model complex relationships and observations with missing values.

4.2.2.5 Case-based reasoning

The main difference with case-based reasoning is that it uses a similar enterprise to assess the bankruptcy classification instead of searching for a pattern as done in unsupervised learning (Alaka et al., 2018). The basic for this technique is the nearest neighbour algorithm (Kumar & Ravi; 2007). Based on the research in the dataset, case-based reasoning is still used recently but the usage is declining. Fabio et al (2016) used a general-purpose case-based reasoning. This technique uses an overall similarity in accordance with the case base and a local similarity to retrieve the most similar cases. Based on the dataset, other articles use case-based reasoning in combination with e.g., the genetic algorithm. Overall, the use of case-based reasoning is limited in the dataset.

4.2.2.6 Deep Learning

Deep learning is relatively new within FP. These models use multiple layers of NNs, but also extract features automatically. So even less human intervention is necessary than in NNs (Qu et al., 2019). This can be helpful in unstructured data inputs as images and texts (Mai et al, 2019). However, these models counter with an even larger black box towards statistical models. Vis-à-vis the NNs, the deep learning models also require more training data. These models are more accommodating for large data sets.

According to Du Jardin (2022) the number of researches to deep learning models increased significantly in the latest years. convolutional NNs is an example of such a model. It received much theoretical attention but had very little practical application. Although it is a NN, it suits more within the category of deep learning models. As explained in Hosaka (2019), this network can predict bankruptcy by inputting financial ratios as an image. According to Du Jardin, the limited use is related to the unadjusted accounting input data. Du Jardin (2022) proposes a model that transforms data into topological data. This gives geographical designs to analyse the dataset. In this case, the results are presented as an image, based on a self-organizing NN. Using this data in combination with the convolutional NN model shows a major predictive ability than traditional models as MDA, SVM, and ensemble models.

Du Jardin is not the only researcher that refers to the problem of input data. In 2015, Lang & Stice-Lawrence also indicate a problem in quantifying textual disclosures. Besides market-based and accounting-based variables, the interest for textual disclosures in financial reporting is also gaining notice (Mai et al. 2019). By using Natural Language Processing, the textual component as the management discussion and analysis is transferred into numerical units. The outputs from the word embedding layer, described by Milkov et al. (2013) are used in two deep learning model architectures. The average embedding model, and the convolutional N as already described above. The research results of Mai et al. (2019) confirm the incremental value of textual disclosures in combination with accounting/ market-based information. In 2019, Matin et al. conducted similar research to the predictive value of text segments on predicting distress accuracy. This article shows the difference in added value between textual segments from managers analysis vs. auditor analysis. As could be expected, the prediction accuracy increased more in the model that incorporated audit reports.

4.2.2.6 Ensemble modelling

All models in the literature review are homogenous methods. In contrast, heterogenous ensemble models are gaining much attention in the research results. Those heterogeneous models, combinations of homogenous models, mainly improve the field of FP by improving the prediction accuracy in almost all models (Kim et al., 2010) (Xia et al., 2018).

Within the ensemble modelling, bagging, boosting and stacking are frequently used methods (Wang et al., 2011). Although they already exist for a long time, the occurrence in the dataset is undeniable. Both Xia et al. (2018) & Wang et al. (2011) refer to logistic regression analysis, decision tree, artificial NN, and SVM as the four most common used base models, also called weak learners. The bagging, boosting and stacking approaches are, simply put, ways to combine these weak learners. It is important that the chosen aggregation method is in occurrence with the weak learners. e.g., models with a high bias and low variance should have an aggregation method that has the purpose of reducing bias (Rocca J, 2019).

Bagging

The first one, bagging (read: bootstrap aggregating) came up with the highest predictive value in the researches above (Xia, et al, 2018) & (Wang et al. 2011). The concept behind bagging is to combine the predictions of several base learners to create a more accurate output. Each training data subset is used to train a different base learner of the same type.

In 2001, Breiman implemented the random forests based on the bagging method. More awareness came especially after 2010. Since then, the model has been frequently used in works in the dataset of this research. The algorithm processes parallel, in contrast to the AdaBoost model in the next section, and improves especially in terms of overfitting. Overfitting is the problem where a model incorporated every aspect of the training data too well. It uses all the negative noisy data and details on new samples of the database which makes it inconsistent (Dietterich, T.1995). The main point of divergence between bagging and the updated version random forests is the forced choice to only some of the features. This choice is random, so the different decision trees use differing features on their sample of the dataset. As a result, the different outcome will not only be based on differences in the data sample, but also on differences in the features. In this way, the prediction accuracy raises. The main drawback of random forest is that it cannot be used in time-varying data assessments. (Nikulski, 2020). Brown & Mues (2012) pointed out that both Random Forests and the further discussed gradient boosting have good prediction accuracies and work well when coping with imbalanced data sets.

Boosting

The second ensemble model is boosting. The basic idea of boosting is to repeatedly apply a base learning to modified versions of the training dataset, thereby producing a sequence of base learners for a predefined number of iterations. From the dataset, it became clear that the Boosting method is a popular technique, especially its enhancements are widely used. Moreover, the AdaBoost (read: Adaptive Boosting) technique emerged as a commonly used method. Alfaro et al. (2008) describes AdaBoost as “based on building consecutive classifiers on modified versions of the training set” (p.11). This method has a sequential pattern and penalizes wrong predicted samples from a dataset by giving it more weight in the next round. In this way, these errors will be avoided, and the prediction accuracy will keep increasing. The weighting given to this adapted training set was indicated as the biggest difference between bagging and boosting. Alfaro's research also highlights the impressive usefulness of AdaBoost based on the decision tree models.

Extreme gradient boosting (XGBoost) is an improvement and extension to GBDT (gradient boosted decision trees). The XGBoost is an ensemble of boosted trees (Chen, 2015). XGBoost can better avoid the overfitting problem and optimizes the objective function (Xia Y. et al; 2017). According to Zieba et al (2016), the main motivation for the use of the XGBoost method is explained by an example. The estimators of economic indicators can have high variances due to huge changes in a small sample of companies. The poor prediction coming from Gradient-Based models like NNs and logistic regression might be overcome by using the XGBoost. Alongside deep learning, XGBoost is one of the most successful methods for large scale data classification.

Stacking

Although the stacking method was already developed many years ago, around 1990, it is included in the result section and not in the literature review. It is just recently getting more attention in FP (Wang et al., 2011). "stacking is an ensemble learning and general method of using a high-level base learner to combine lower-level base learners to achieve greater predictive accuracy" (p.5) Wolpert (1992). The main difference with bagging and boosting is the heterogeneousness of the underlying models. Besides the research of Wolpert (2011), Zhang, W. et al (2021), Yao J. (2022) and many other recently presented researches based on ensemble modelling using the stacking approach. These authors declare using the method to remove e.g., noisy data, parameter optimization and outliers.

The stacking method forms together with the bagging, boosting, and other derivative methods most of the ensemble models. The aforementioned better prediction accuracy and, especially with bagging, relative easiness in use highlights the most positive contributions of the ensemble models.

4.2.2.7 Multi Criteria Decision Aid

An overarching model in FP modelling is the Multi-Criteria Decision Aid (MCDA). This method is already used in a many finance fields as portfolio selection and investment appraisal (Mousavi & Lin, 2020). According to these authors, the model helps to solve the conflicting ranking problem of different performance criteria in mono-criterion evaluation. They propose PORMETHEE II, a multi-criteria evaluation of different prediction models, that meets the multidimensional requirements of financial decisions. A combination of statistical and artificial intelligence models can be used with this method. The research of Mousavi & Lin (2020) concludes that incorporating corporate governance indicators outperforms the statistical models that do not include these indicators. The study of Mousavi & Lin (2019) on MCDA refers to the corporate governance indicators as board composition and director characteristics in FP modelling. A further improvement of the model performance can be achieved by incorporating Ensemble methods (Mousavi & Lin, 2019). Like Deep Learning models, the MCDA can also process enormous amounts of data. According to the authors, MCDA is used too little compared to the mono-criterion model. The ease of using mono-criterion is in contradiction with the usefulness of conflicting rankings and the restricted number of criteria.

Data Envelopment Analysis

A subcategory in MCDA is the data envelopment analysis (Li et al., 2017). Premanchandra, I. et al (2011) described the nature of data envelopment analysis “to assess the efficiency of decision-making units that have multiple inputs and outputs” (p.2). This technique differs as it does not apply for any requirements regarding relationships between inputs and outputs. Neither does it requires a large sample size, and it evaluates each decision-making unit individually (Premanchandra L. et al. 2011). These features caused data envelopment analysis to be highly relevant in literature in general (Paradi, J. et al;2014) and in the dataset of this research.

As data envelopment analysis began to appear in literature in the nineties, the application to bankruptcy prediction was only from 2000-2004 onwards. In 2004, Cielen et al. compared the predictive accuracy of data envelopment with more traditional techniques, thereby focusing on the Type-I and Type-II forecast errors. These authors concluded that the prediction accuracy of data envelopment analysis outperforms the universally used decision trees. Furthermore, the model shows the highest Type-I accuracy of all the models. Type-I, as previously elaborated in the literature review, is the most important and costly error that must be minimized. Due to its popularity in the dataset and great prediction accuracy, data envelopment analysis should be added to the list of influential and accurate models.

In the dataset, several types of models within data envelopment analysis began to appear. The Malmquist- and Slacks-based data envelopment analysis are two examples of submodels. All submodels showed little appearance in number of articles and citations. This can partly be explained by the only recent appearance of these models (2018-2022). A thorough analysis would lead this research too far and provides little added value. A fairly complete overview of the submodels has been presented in Mousavi et al. (2022).

Graph theory

This research already showed some models that incorporate the importance of non-accounting data. Yildirim et al. (2021) highlight the importance of the less accessible invoice data to analyse the relational trading information among companies. The clients/suppliers of your clients/suppliers are important for your own default prediction. A company in default will first affect the companies which are incorporated in its value chain, but also the related companies. This innovative default prediction model is based on the graph theory in big data analytics. This research works with a weighted customer score for each client in order to obtain a total score for failure prediction. The perfect information assumption in this model is less obvious. Other factors as knowing all your customers and time expenditure are not realistic for some companies. Although this technique is part of data envelopment analysis, the practical implementation might be limited. Therefore, this model floats between the data envelopment analysis and the theoretical models.

The more theoretical graph theory is the last model in the overview of the artificial intelligence models. Hereby, all trends arising from the dataset and Bibliometric Analysis are discussed. For models like the SVM, a more general trend is covered. In turn, other models received a more detailed analysis. In following, a general conclusion will be drawn. First, a discussion section will follow with limitations of this research and future research directions.

5. Discussion

Limitations

Although this research is intended to be an added value to the already existing literature and review articles on FP, some limitations need to be highlighted. First of all, it should be clear that the strict quality restriction on the data collection might have an impact on the research results and conclusion. This research uses the ABDC list to indicate quality journals. As a result, influential articles from lower than indicated quality journals are possibly not included in the research results. Furthermore, newer models developed in 2021 and 2022 that received few citations may also not be included in the results.

Regarding the database, Web of Science has been used in this study. In Zambrano et al (2021), Scopus was used as database. By choosing Web of Science, all relevant reviewing articles were certainly incorporated but it is possible that some articles from the Scopus database have not been included. The researches from this dataset were categorized into statistical, artificial and theoretical models. The latter domain was only narrowly developed. In this study, the choice was made to prioritize to the models used in practice. Furthermore, in the research results section a country level analysis of the models was made. This clarifies which countries, by number of citations, are the most important in literature. However, there was no breakdown by popularity of certain models within countries.

Future research directions

This latter constraint directly develops the first opportunity for future research directions. An in-depth search to country-specific popularity in FP models could clarify country-level differences. It might for instance be possible that a particular country has a general preference to the statistical models with accounting input data vs. countries preferring deep learning models with a broader input data requirement.

As mentioned with the SVM models, numerous variations were implemented in the dataset. Especially the ensemble methods that use SVM know a considerable number of variations. These were not further specified in this research but might be an interesting starting point for in-depth research to this model. This study concisely discussed the main developments in FP modelling. A more comprehensive analysis with further comparison between the different models would be a worthwhile extension in this study area.

Firms fail in all divergent phases of the firm life cycle. An introductory reflection of these researches was reported in the literature review. This might be an important fallacy that current models do not take into consideration. Should statistical vs/and artificial intelligence models be more adapted regarding the life stage a firm is in? This could be the research question of a next study.

6.Conclusion

What are the developments in failure prediction modelling from 2005 onwards? This was the research question that has been covered by this study.

Generally, the number and importance of artificial intelligence models have been increased. However, as discussed in the bibliometric analysis, the statistical models reach a persistent appearance in literature. Fewer new models were developed in this category, but the already existing models retain their use in practice. First, it is important to mention that a one-fits-all model did not exist before 2005, and still does not exist today. Differences in available input data: balanced vs. imbalanced, accounting vs. market vs. corporate governance data influence the type of model required/ possible. Ease of interpretability is also a distinctive feature in the discussed models. A linear model as the Z-score model is more useful and explainable than a deep learning technique. However, the quantile regression function is an example which tries to enhance the interpretability of the artificial intelligence models.

The main development can be described in one word: multidisciplinarity. The ensemble models know a great prediction accuracy improvement because of the combination of models. The MCDA also uses a multiview technique. This combination of models and data enhances the accuracy but requires more abstract thinking skills, time to calculate. This also effects the difficulty for others to recheck a calculation.

When choosing a suitable model, it will be important to distinguish what is important for the user and the situation in which they want to use the model. In my opinion, most of the artificial intelligence models fail to enhance the interpretability of the model to the level of the statistical models. The main question before using a FP model should be: Is interpretability in this particular case more important than a difference of 1,2,3… percent accuracy? Although, most of the artificial intelligence models are relevant if you want to reach a higher accuracy and you are familiar with e.g. analysing models of textual patterns in financial statements. Besides the user of the model, the other main element for some artificial intelligence model is the availability of the necessary data. If you e.g. can include corporate governance indicators in the model, then MCDA might be a suitable model for the analysis. In short, the persistence of statistical models in the literature proves that the more dynamic models bring several challenges in addition to their potential.

7.Reference list

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. https://doi.org/10.1016/j.eswa.2017.10.040

Alfaro, E., García, N., Gámez, M., & Elizondo, D. (2008). Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. *Decision Support Systems*, 45(1), 110-122. https://doi.org/10.1016/j.dss.2007.12.002

Altman, E. I. (1968). Financial Ratios, Discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

Altman, E., I., & Hotchkiss, E. (2005). *Corporate Financial Distress and Bankruptcy: Predict and Avoid Bankruptcy, Analyze and Invest in Distressed Debt*, (3rd ed.). Wiley.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2014). Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2536340

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2016). 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. https://doi.org/10.1111/jifm.12053

Argenti, J. (1976). Corporate planning and corporate collapse. *Long range planning*, 9(6), 12-17. https://doi.org/10.1016/0024-6301(76)90006-6

Aziz, M. A., & Dar, H. A. (2006). Predicting corporate bankruptcy: where we stand? Corporate Governance. *The international journal of business in society*. https://doi.org/10.1108/14720700610649436

Balcaen, S., & Ooghe, H. (2006). 35 years of studies on business failure: an overview of the classic statistical methodologies and their related problems. *The British Accounting Review*, 38(1), 63–93. https://doi.org/10.1016/j.bar.2005.09.001

Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405-417. https://doi.org/10.1016/j.eswa.2017.04.006

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

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.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Brown, I., & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, 39(3), 3446-3453. https://doi.org/10.1016/j.eswa.2011.09.033

Bruloot, D. (2021,26,02). II FEB-HAWE insolventierecht 2021 [Powerpoint-slides]. Faculty of Law, Ghent University. Consulted on 28 June 2022.

Charitou, A., & Trigeorgis, L. (2000). Option-based bankruptcy prediction. *SSRN 248709*.

Chen, T., & He, T. (2015). Higgs boson discovery with boosted trees. *Proceedings of Machine Learning Research*. (pp. 69-80).

Cielen, A., Peeters, L., & Vanhoof, K. (2004). Bankruptcy prediction using a data envelopment analysis. *European Journal of Operational Research*, *154*(2), 526-532. https://doi.org/10.1016/S0377-2217(03)00186-3

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

Crouhy, M. G., Jarrow, R. A., & Turnbull, S. M. (2008). The subprime credit crisis of 2007. *The Journal of Derivatives*, 16(1), 81-110. https://doi.org/10.3905/jod.2008.710899

Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert systems with applications*, 40(10), 3970-3983. https://doi.org/10.1016/j.eswa.2013.01.012

Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM computing surveys,* 27(3), 326-327.

Du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications*, *75*, 25-43. https://doi.org/10.1016/j.eswa.2017.01.016

Du Jardin, P. (2022). Designing topological data to forecast bankruptcy using convolutional neural networks. Annuals of Operations Research, 1-42. https://doi.org/10.1007/s10479-022-04780-7

Frydman, H., Altman, E. I., & Kao, D. L. (1985). Introducing recursive partitioning for financial classification: the case of financial distress. *The journal of finance*, 40(1), 269-291. https://doi.org/10.1111/j.1540-6261.1985.tb04949.x

Gepp, A., Kumar, K., & Bhattacharya, S. (2010). Business failure prediction using decision trees. *Journal of forecasting*, 29(6), 536-555. https://doi.org/10.1002/for.1153

Gepp, A., & Kumar, K. (2008). The role of survival analysis in financial distress prediction. *International research journal of finance and economics*, 16(16), 13-34.

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, 30(2), 133-176. https://doi.org/10.1023/A:1022558715350

Henninger, M. (2012). Bibliometric analysis- an overview. *Social Media for Academics*. Retrieved from https://www.sciencedirect.com/topics/computer-science/bibliometric-analysis

Hertzel, M. G., & Officer, M. S. (2012). Industry contagion in loan spreads. *Journal of Financial Economics*, 103(3), 493-506. https://doi.org/10.1016/j.jfineco.2011.10.012

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2018.09.039

Huang, S. Y., Tsaih, R. H., & Yu, F. (2014). Topological pattern discovery and feature extraction for fraudulent financial reporting. *Expert systems with applications*, *41*(9), 4360-4372. https://doi.org/10.1016/j.eswa.2014.01.012

Hwang, R. C., & Chu, C. K. (2014). Forecasting forward defaults with the discrete-time hazard model. *Journal of Forecasting*, *33*(2), 108-123. https://doi.org/10.1002/for.2278

Jackson, R. H., & Wood, A. (2013). The performance of insolvency prediction and credit risk models in the UK: A comparative study. *The British Accounting Review*, 45(3), 183–202. https://doi.org/10.1016/j.bar.2013.06.009

Jayasekera, R. (2018). Prediction of company failure: Past, present and promising directions for the future. *International Review of Financial Analysis*, 55, 196–208. https://doi.org/10.1016/j.irfa.2017.08.009

Jones, S., & Hensher, D. A. (2004). Predicting Firm Financial Distress: A Mixed Logit Model. *The Accounting Review*, 79(4), 1011–1038. https://doi.org/10.2308/accr.2004.79.4.1011

Kim, M. J., & Kang, D. K. (2010). Ensemble with neural networks for bankruptcy prediction. *Expert systems with applications*, 37(4), 3373-3379. https://doi.org/10.1016/j.eswa.2009.10.012

Kellner, R., Nagl, M., & Rösch, D. (2022). Opening the black box–Quantile neural networks for loss given default prediction. *Journal of Banking & Finance*, 134, 106334. https://doi.org/10.1016/j.jbankfin.2021.106334

Klieštik, T., Kočišová, K., & Mišanková, M. (2015). Logit and Probit Model used for Prediction of Financial Health of Company. *Procedia Economics and Finance*, 23, 850–855. https://doi.org/10.1016/s2212-5671(15)00485-2

Kozeny, V. (2015). Genetic algorithms for credit scoring: Alternative fitness function performance comparison. *Expert Systems with applications*, 42(6), 2998-3004. https://doi.org/10.1016/j.eswa.2014.11.028

Krüger, S., & Rösch, D. (2017). Downturn LGD modeling using quantile regression. *Journal of Banking & Finance*, 79, 42-56. https://doi.org/10.1016/j.jbankfin.2017.03.001

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. https://doi.org/10.1016/j.ejor.2006.08.043

Kwak, C., & Clayton-Matthews, A. (2002). Multinomial logistic regression. *Nursing research*, 51(6), 404-410.

Lee, K., Booth, D., & Alam, P. (2005). A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms. *Expert Systems with Applications*, 29(1), 1-16. https://doi.org/10.1016/j.eswa.2005.01.004

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, 247(1), 124-136. https://doi.org/10.1016/j.ejor.2015.05.030

Li, H., & Sun, J. (2009). Predicting business failure using multiple case-based reasoning combined with support vector machine. *Expert systems with applications*, 36(6), 10085-10096. https://doi.org/10.1016/j.eswa.2009.01.013

Li, Z., Crook, J., & Andreeva, G. (2017). Dynamic prediction of financial distress using Malmquist DEA. *Expert Systems with Applications*, 80, 94-106. https://doi.org/10.1016/j.eswa.2017.03.017

Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European journal of operational research*, 274(2), 743-758. https://doi.org/10.1016/j.ejor.2018.10.024

Matin, R., Hansen, C., Hansen, C., & Mølgaard, P. (2019). Predicting distresses using deep learning of text segments in annual reports. *Expert Systems with Applications*, 132, 199-208. https://doi.org/10.1016/j.eswa.2019.04.071

Michielsen, S. (2022). Signaal springt op oranje. *De Tijd.* Retrieved from https://www.tijd.be/opinie/commentaar/signaal-springt-op-oranje/10392958.html

Min, S. H., Lee, J., & Han, I. (2006). Hybrid genetic algorithms and support vector machines for bankruptcy prediction*. Expert systems with applications*, 31(3), 652-660. https://doi.org/10.1016/j.eswa.2005.09.070

Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert systems with applications*, 28(4), 603-614. https://doi.org/10.1016/j.eswa.2004.12.008

Mousavi, M. M., & Lin, J. (2020). The application of PROMETHEE multi-criteria decision aid in financial decision making: Case of distress prediction models evaluation. *Expert Systems with Applications*, 159, 113438. https://doi.org/10.1016/j.eswa.2020.113438

Mousavi, M. M., Ouenniche, J., & Tone, K. (2022). A dynamic performance evaluation of distress prediction models. *Journal of Forecasting*. https://doi.org/10.1002/for.2915

Nikulski, J. (2020). The Ultimate Guide to AdaBoost, random forests and xgboost. Retrieved from https://towardsdatascience.com/the-ultimate-guide-to-adaboost-random-forests-and-xgboost-7f9327061c4f

Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. In 1990 IJCNN International Joint Conference on neural networks (pp. 163-168). IEEE. DOI: 10.1109/IJCNN.1990.137710

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy*. Journal of Accounting Research,* 18(1), 109. https://doi.org/10.2307/2490395

(onder)zoektips. Scopus versus Web of Science - de verschillen | (onder)zoektips. (n.d.). Retrieved from https://researchtips.ugent.be/nl/tips/00001697/

Ooghe, H., Joos, P., & De Bourdeaudhuij, C. (1995). Financial distress models in Belgium: The results of a decade of empirical research. International Journal of Accounting, 30, 245-274.

Ooghe, H., & De Prijcker, S. (2008). Failure processes and causes of company bankruptcy: a typology. *Management decision*. https://doi.org/10.1108/00251740810854131

Ooghe, H., & Spaenjers, C. (2005). *De Financiële toestand van de Belgische ondernemingen*. Retrieved from https://www.researchgate.net/publication/293490258_De_FiToR-meter_een_nieuwe_eenvoudige_en_geintegreerde_maatstaf_voor_de_financiele_toestand_van_een_onderneming

Ooghe, H., Vander Bauwhede, H., & Van Wymeersch, C. (2021). *Financiële analyse van de vennootschap* (6th edition, e.d.). België: Intersentia.

Ooghe, H., Verbaere, E., (1985). Predicting business failure on the basis of accounting data: The Belgian experience. *The International Journal of Accounting* 9 (2), 19–44

Paradi, J. C., Asmild, M., & Simak, P. C. (2004). Using DEA and worst practice DEA in credit risk evaluation. *Journal of productivity analysis*, 21(2), 153-165. https://doi.org/10.1023/B:PROD.0000016870.47060.0b

Parker, S., Peters, G. F., & Turetsky, H. F. (2002). Corporate governance and corporate failure: a survival analysis. *Corporate Governance*. https://doi.org/10.1108/14720700210430298

Premachandra, I. M., Chen, Y., & Watson, J. (2011). DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment. *Omega*, 39(6), 620-626. https://doi.org/10.1016/j.omega.2011.01.002

Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science*, 162, 895-899. https://doi.org/10.1016/j.procs.2019.12.065

Resurchify. (n.d.). Impact score, overall ranking, h-index, SJR, rating, publisher, ISSN, and other important metrics. *Expert systems with applications*. Retrieved from https://www.resurchify.com/impact/details/24201

Rocca J. (2022) Ensemble methods: Bagging, boosting and stacking. Retrieved from https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The journal of business*, 74(1), 101-124. https://doi.org/10.1086/209665

Shi, Y., & Li, X. (2019). An overview of bankruptcy prediction models for corporate firms: A Systematic literature review. *Intangible Capital*, 15(2), 114. https://doi.org/10.3926/ic.1354

Shin, K. S., & Lee, Y. J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert systems with applications*, *23*(3), 321-328. https://doi.org/10.1016/S0957-4174(02)00051-9

Sohn, S. Y., & Kim, H. S. (2007). Random effects logistic regression model for default prediction of technology credit guarantee fund. *European Journal of Operational Research*, 183(1), 472-478. https://doi.org/10.1016/j.ejor.2006.10.006

Specht, D., & Romsdahl, H. (1994). Experience with adaptive probabilistic neural networks and adaptive general regression neural networks: Semantic scholar. *IEEE*. Retrieved from https://www.semanticscholar.org/paper/Experience-with-adaptive-probabilistic-neural-and-Specht-Romsdahl/2767b9a211857c8bdb361054c5989086a76478f8

Sun, J., Fujita, H., Chen, P., & Li, H. (2017). Dynamic financial distress prediction with concept drift based on time weighting combined with Adaboost support vector machine ensemble. *Knowledge-Based Systems*, 120, 4-14. https://doi.org/10.1016/j.knosys.2016.12.019

Tattersall, T. (2022). Research. Australian Business Deans Council. Retrieved from https://abdc.edu.au/research/

Vander Bauwhede, H., Van Wymeersch, C., & Ooghe, H. (2017). *Financiële analyse van de onderneming: theorie en toepassing op de jaarrekening volgens Belgian GAAP en IFRS*. Intersentia.

Van Gestel, T., Baesens, B., & Martens, D. (2010). From linear to non-linear kernel based classifiers for bankruptcy prediction. *Neurocomputing*, 73(16-18), 2955-2970. https://doi.org/10.1016/j.neucom.2010.07.002

Van Gestel, T., Baesens, B., Suykens, J., Espinoza, M., Baestaens, D. E., Vanthienen, J., & De Moor, B. (2003). Bankruptcy prediction with least squares support vector machine classifiers. *IEEE*. DOI: 10.1109/CIFER.2003.1196234

Veganzones, D., & Séverin, E. (2018). An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems*, 112, 111-124. https://doi.org/10.1016/j.dss.2018.06.011

Volkov, A., Benoit, D. F., & Van den Poel, D. (2017). Incorporating sequential information in bankruptcy prediction with predictors based on Markov for discrimination. *Decision Support Systems*, 98, 59-68. https://doi.org/10.1016/j.dss.2017.04.008

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. https://doi.org/10.1016/j.eswa.2010.06.048

Wolpert, D. H. (1992). Stacked generalization. *Neural networks*, *5*(2), 241-259. https://doi.org/10.1016/S0893-6080(05)80023-1

Wu, C. H., Tzeng, G. H., Goo, Y. J., & Fang, W. C. (2007). A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy. *Expert systems with applications*, 32(2), 397-408. https://doi.org/10.1016/j.eswa.2005.12.008

Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert systems with applications*, 78, 225-241. https://doi.org/10.1016/j.eswa.2017.02.017

Yao, J., Wang, Z., Wang, L., Liu, M., Jiang, H., & Chen, Y. (2022). Novel hybrid ensemble credit scoring model with stacking-based noise detection and weight assignment. *Expert Systems with Applications*, 198, 116913. https://doi.org/10.1016/j.eswa.2022.116913

Zambrano Farias, F., Valls Martínez, M. D. C., & Martín-Cervantes, P. A. (2021). Explanatory Factors of Business Failure: Literature Review and Global Trends. *Sustainability*, 13(18), 10154. https://doi.org/10.3390/su131810154

Zhang, W., Yang, D., Zhang, S., Ablanedo-Rosas, J. H., Wu, X., & Lou, Y. (2021). A novel multi-stage ensemble model with enhanced outlier adaptation for credit scoring. *Expert Systems with Applications*, 165, 113872. https://doi.org/10.1016/j.eswa.2020.113872

Zhou, L. (2013). Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems*, 41, 16–25. https://doi.org/10.1016/j.knosys.2012.12.007

6. Appendix

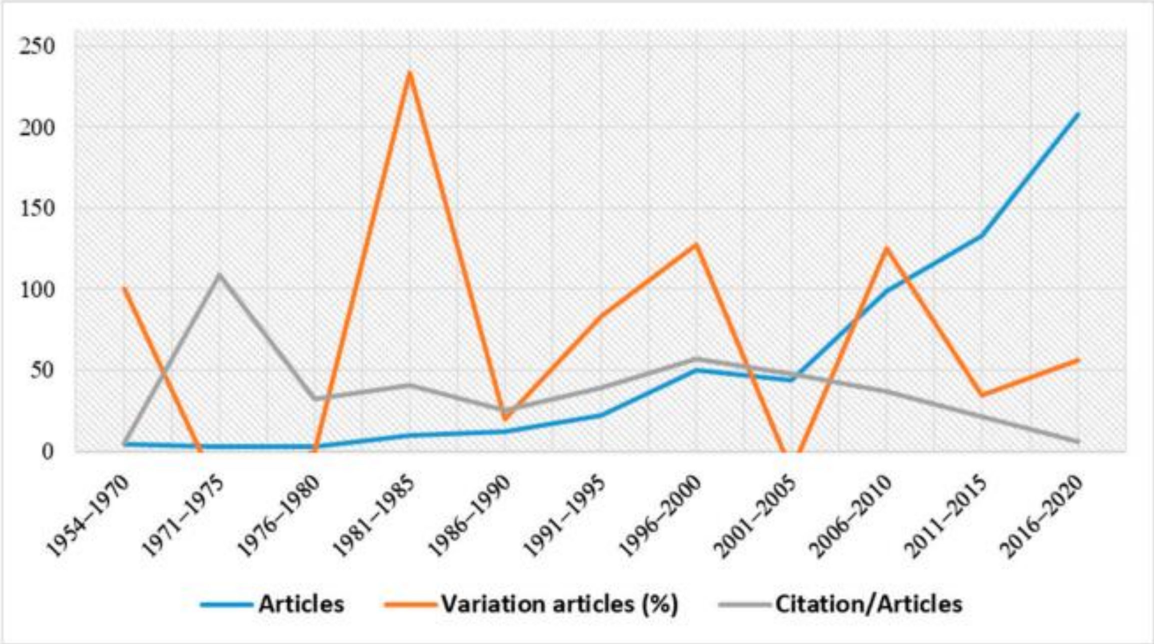

Figure 3: Number of published articles in FP modelling (Zambrano et al., 2021)

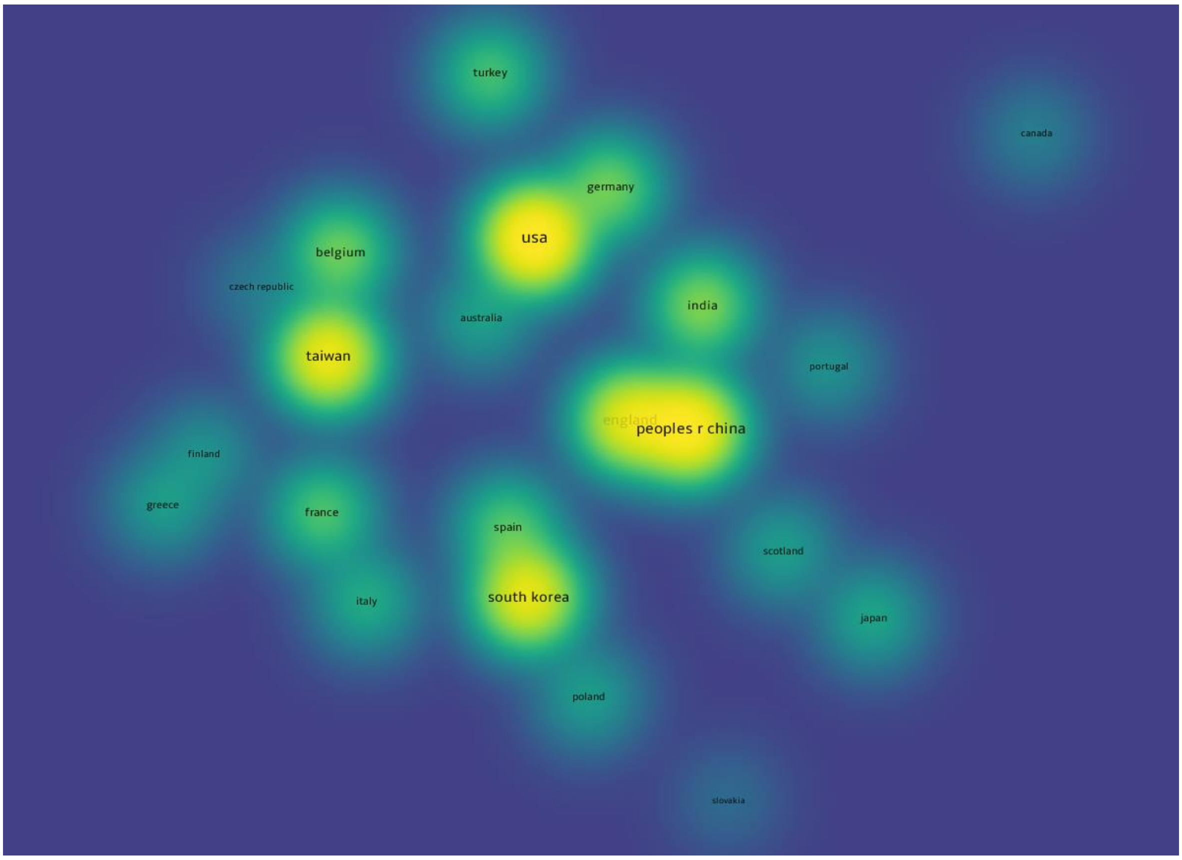

Figure 4: Geographical distribution of the dataset, visualized by VOSviewer

Relevant Web of Science categories in the dataset
Operations Research Management Science
Computer Science Artificial Intelligence
Management
Business Finance
Economics
Computer Science Information Systems
Business
Information Science library
Computer Science Interdisciplinary Applications
Science Software Engineering
Computer Science Theory Methods

Table 3: Relevant WoS categories in dataset

1	Pan, WT	2012	A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example	Knowledge-Based Systems
2	Kumar, PR & Ravi, V.	2007	Bankruptcy prediction in banks and firms via statistical and intelligent techniques- A review	European Journal of Operational Research
3	Min, JH. & Lee, YC	2005	Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters	Expert Systems with Applications
4	Shin, KS, Lee, TS, Kim, HJ	2005	An application of Support vector machines in bankruptcy prediction model	Expert Systems with Applications
5	Lessmann et al.	2015	Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research	European Journal of Operational Research
6	Duffie, D. et al.	2007	Multi-period default prediction with stochastic covariates	Journal of Financial Economics
7	Brown & Mues	2012	An experimental comparison of classification algorithms for imbalanced credit scoring data sets	Expert Systems with Applications
8	Xia, YF et al.	2017	A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring	Expert Systems with Applications
9	Min et al.	2006	Hybrid genetic algorithms and support vector machines for bankruptcy prediction	Expert Systems with Applications
10	Wu et al.	2007	A real- valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy	Expert Systems with Applications
11	Barboza et al.	2017	Machine learning models and bankruptcy prediction	Expert Systems with Applications
12	Geng et al.	2015	Prediction of financial distress: An empirical study of listed Chinese companies using data mining	European Journal of Operational Research
13	Sun et al.	2014	Predicting financial distress and corporate failure: A review from the state-of-the-art definitions, modelling, sampling, and featuring approaches	Knowledge-based systems
14	Ziebe et al.	2016	Ensemble boosted trees with synthetic features generation in application to bankruptcy	Expert Systems with Applications
15	Olson et al.	2012	Comparative analysis of data mining methods for bankruptcy prediction	Decision Support Systems
16	Delen et al.	2013	Measuring firm performance using financial ratios: A decision tree approach	Expert Systems with Applications

Table 4: Most cited researches

Statistical models
Discriminant analysis
Logistic regression/logit
Probit
Hazard model
Partial least squares

Table 5: Statistical models from Zambrano et al. (2021) and Shi & Li (2018)

Artificial Intelligent techniques
Neural Network
Support Vector Machine
Data Mining
Decision Tree
Genetic Algorithm
Rough set
Fuzzy logic
Case-based reasoning
Data development analysis
Adaboost
K-nearest neighbours
Bayesian network

Table 6: artificial intelligence models from Zambrano et al. (2021) and Shi & Li (2018)