CATÓLICA LISBON
BUSINESS & ECONOMICS

Corporate Bankruptcy: Can Machine Learning Methods Enhance the Prediction of Failure?

Eva Ferreira

Dissertation written under the supervision of Professor Dan Tran

Dissertation submitted in partial fulfilment of requirements for the MSc in Finance, at the Universidade Católica Portuguesa, January 2021.

Abstract

This dissertation aims to enhance the performance of traditional corporate bankruptcy prediction models through the application of machine learning techniques and models, and industry effects. The data used includes 3664 companies out of which 144 went bankrupt throughout the period of 2000 until 2019, and it was structured to emulate the design of the variables Campbell et al. (2008) used in their study. Evidence was found that implies the improvement of various metrics' results from the use of machine learning techniques and models. The model with the highest F1-score, meaning the most balanced, is the Logit with the application of hyperparameter tuning and industry effects. The model with the highest Recall, which means the percentage of bankruptcies correctly predicted, is the Logit with the application of the oversampling technique. Furthermore, both Support Vector Machines (SVM) and Artificial Neural Networks (NN) models delivered balanced and enhanced results compared with the two benchmarks (Altman Z-Score and Simple Logit models). The improvement techniques provided the models with distinct results. Oversampling led mostly to a higher percentage of bankruptcies predicted, while hyperparameter tuning and industry effects provided the models with more precise results. The variable importance in each type of model was also analysed. Overall, the Campbell et al. (2008) market variables (SIGMA, RSIZE, EXRET and PRICE) are highly significant for the positive results of all three types of models studied.

Keywords: Corporate Bankruptcy Prediction; Machine Learning; Logit Models; Support Vector Machines; Artificial Neural Networks.

Title: Corporate bankruptcy: Can Machine Learning Methods Enhance the Prediction of Failure?

Author: Eva Maria Ribeiro Silva Ferreira

Resumo

Esta dissertação tem como objetivo melhorar o desempenho de modelos tradicionais de previsão de falência corporativa através da aplicação de técnicas e modelos de *machine learning*, e efeitos de indústria. A data usada incluí 3664 empresas das quais 144 faliram no período de 2000 até 2019, e foi estruturada para emular o desenho de variáveis que Campbell et al. (2008) utilizaram no seu estudo. Evidências foram encontradas que indicam a melhoria dos resultados de várias métricas através do uso de técnicas e modelos de *machine learning*. O modelo com maior *F1-Score*, isto é, o mais equilibrado, é o Logit com a aplicação de ajuste de hiperparâmetros e efeitos de indústria. O modelo com maior *Recall*, ou seja, a percentagem de falências corretamente previstas, é o Logit com a aplicação da técnica de sobreamostragem. Ademais, ambos os modelos de Máquina de Vetores de Suporte (MVS) e Redes Neurais Artificiais (RN) forneceram resultados equilibrados e melhorados em comparação com os modelos de referência (Atlman Z-Score e Logit Simples). As técnicas de melhoramento geraram resultados diferentes aos modelos. Sobreamostragem levou a maiores percentagens de falência prevista, enquanto o ajuste de hiperparâmetros e os efeitos de indústria forneceram aos modelos resultados mais precisos. A importância de variáveis nos diferentes tipos de modelo também foi analisada. Em geral, as variáveis de mercado de Campbell et al. (2008) (SIGMA, RSIZE, EXRET e PRICE) são bastante significativas para os resultados positivos dos três diferentes tipos de modelo estudados.

Palavras-Chave: Previsão de Falência Corporativa; Machine Learning; Modelos Logit; Máquina de Vetores de Suporte; Redes Neurais Artificiais.

Título: Falência Corporativa: Podem os Métodos de *Machine Learning* Melhorar a Previsão de Falência?

Autora: Eva Maria Ribeiro Silva Ferreira

Acknowledgements

This dissertation marks the end of a tough journey but nonetheless pleasant and rewarding. During this degree and throughout my academic life I learnt deeply and broadly about topics I am passionate about, and I had the pleasure to do it in a nurturing but challenging environment where team spirit was put above any competition.

Firstly, I would like to thank my supervisor, Dan Tran, for creating the space for growth and learning since day one. He was a constant driving force of my work and challenged me like no other professor has, and that I can only appreciate.

Secondly, I want to thank the three key persons that pushed me when I could not find the motivation to continue my work, my partner, André Soares, and the two best friends this university could have given me, Catalina Rosero and Christine Neufeld.

Thirdly, I have to thank my parents for not only providing me with the opportunity to study what I wanted and where I believed was best for me, but also for always offering me support when I most needed it.

Last but not least, I would like to thank all my professors and colleagues that accompanied me on this journey and helped build the woman I am today.

Table of Contents

List of Figures

List of Tables

List of Appendices

1. Introduction

1.1. Motivation and Contextual Analysis

Prediction of events or variables in finance is one of the most popular topics since it is through these forecasting models that firms and banks can conduct, for example, risk management processes. In specific, prediction of corporate bankruptcy is significant for the whole economy to help prevent major default events such as crises. If prediction models are robust for these periods of economic downturn, financial institutions are more protected against the consequences of corporate bankruptcy which are more likely to happen in higher volumes when the economy is failing.

In the corporate bankruptcy prediction literature and in industry, models like the Altman Z-score created by Altman (1968) are still taken to high regard even though they do not take into account matters such as time-dependent variables that encompass trends throughout time. This means that, for instance, banks still use models that might not be the best at predicting a large number of bankruptcies with the precision necessary. Credit risk is defined as the chance of loss stemming from the counterparty's (in this case the borrower) failure to repay a loan provided by the lender (i.e., the bank). Banks not only need to predict default but also its probability, this way they can apply fair rates on their lending deals more efficiently. Thus, not only prediction of failure is important to understand which firms are viable counterparties, but also the probability of default is highly significant for the application of fair rates in the lending process. Another critical factor in models for banks to forecast failures is not only the percentage of bankruptcies they correctly predict but also the precision with which they do so, since if the precision is low banks will be overestimating rates for firms that will not default in the future.

Various models have been introduced throughout the decades, from the Altman Z-score to Merton's option-based model in 1974, to time-dependent logit models such as Shumway's in 2001 and finally machine learning models, ranging from Support Vector Machines to Artificial Neural Networks. Despite the extensive research on this topic, some faults can be found in most of these models and how they have been applied and evaluated. Firstly, and as aforementioned, most of these models do not take into account time-dependent variables which might capture trends that are important to understand whether or not a firm will default. Secondly, most of the literature in corporate bankruptcy prediction has evaluated the models in terms of their accuracy. This metric is reasonable for when the datasets are balanced, which does not happen if one has a dataset that correctly represents reality, meaning in fact bankruptcy is not equally as likely as non-bankruptcy in the real world, and if the dataset is imbalanced

accuracy is misleading. This happens because if the data is imbalanced, e.g. with a 90% majority class of non-bankrupt companies and 10% minority class of bankruptcies, if the model predicts all companies to be non-bankrupt it will be 90% accurate, which might seem good but the factor that one, in fact, wanted to predict (bankruptcy) is not forecasted, thus the model has no use. Research such as in Altman (1968) overcomes this by undersampling the dataset. However, this is first, not representative of reality, and second, leads to an immense loss of information.

This research will be focused on machine learning models, namely Support Vector Machines and Artificial Neural Networks, since literature has found that these typically outperform popular models such as the Altman Z-score. The research will not focus on emulating discriminant analysis with machine learning techniques, instead, it will be focused on the improvement of an already proven-to-be-effective model, the time-dependent logit model firstly created by Shumway (2001) and later improved with market variables by Campbell et al. (2008), and the way the data was structured for such models. Further detailed objectives of this research are explained in the next section, "1.2. Objective and Structure".

1.2. Objective and Structure

The main objective of this research is to outperform traditional models in the corporate bankruptcy literature with the use of machine learning techniques, including Neural Networks (NN) and Support Vector Machines (SVM). Thus, multiple research questions are to be answered:

1. Do NN and SVM display high accuracy as well as high scores for metrics such as the F1-score, which not only evaluate the efficiency of the model but take Type I and Type II errors into account? If not, what techniques improve the simple models?
2. Do market variables, such as the ones used by Campbell et al. (2008), lead to a high performance of machine learning models?
3. Does industry analysis improve the metrics of machine learning models?

Regarding the structure of this paper, the next chapter will focus on the literature review, providing an extensive overview of the different methods explored for the topic at hand. Furthermore, in Chapter 3, the dataset, data treatment and data exploration will be reviewed. In Chapter 4, the methodology for both the benchmark models and machine learning models will be explained. In Chapter 5, the model evaluation metrics will be presented alongside the variable importance metrics. The results of the analysis will be displayed and discussed in Chapter 6. Finally, in Chapter 7 the research will be concluded.

2. Literature Review

2.1. Early Empirical Approaches

Within the corporate bankruptcy literature, there are three main routes of research in terms of statistical models: traditional accounting-based models (i.e., discriminant analysis), contingency claims-based models and, finally, hazard/logit models.

The research in this field started with Beaver (1966) and Altman (1968) from whom traditional accounting-based models were born. These are structured fundamental analysis using publicly published financial statements and are based on the linear combination of financial ratios that are claimed to best differentiate between failed and non-failed companies (Bauer & Agarwal, 2014). Further research has been conducted in order to find the best ratios to predict bankruptcy and improve the existing model. In spite of the wide use of these models, both on academic and industry level, these are often criticised for their lack of theoretical basis. They are also known for their high misclassification rates (Begley et al., 1996). Agarwal & Taffler (2008) mention several pitfalls of these models, including the fact that financial ratios computed based on financial statements represent the past thus, may not be informative to predict the future, and that accounting numbers might be subject to manipulation by management. Another downside from this type of model is that it is static, which might produce biased and inconsistent bankruptcy probability estimates (Shumway, 2001). Furthermore, Begley et al. (1996) and Hillegeist et al. (2004) argue that updating the model's coefficients does not increase performance thus, the whole model has to be renewed on a periodical basis.

A different methodology to improve on the traditional models' performance was developed on the foundation of the Black and Scholes (1973) and Merton (1974) framework, which is based on the assumption of equity being a call option on the firm's assets. This model on contingency claims comes to reduce the hurdles of the previous accounting-based research. Firstly, prices represent, in efficient markets, both historical financial information and market sentiment on a firm. This comes to overcome the distrust in the veracity of the accounting based financial ratios. Moreover, since it is based on the renowned frameworks aforementioned it does not include the theoretical uncertainty that other traditional models offer. While this model is considered better compared to the discriminant analysis ratio combination it still poses difficulties in terms of application. Its biggest hurdle is possibly the fact that it is based on variables that are not retrievable from the market thus, they need to be estimated which could produce large errors. Finally, it is argued that market prices might not reflect the information from a firm's accounting data (Sloan, 1996), which although representative of the past is still important information to be taken into account.

The logistic regression, or simply logit, is a statistical model that uses the logistic function to determine conditional probability typically used to model a binary dependent variable. It is argued by Ohlson (1980), that this model avoids the statistical obstacles of the multivariate discriminant analysis introduced in the 1960s, such as the requirement of normally distributed predictors, thus being another tool that improves on the traditional model. Recently, models such as the ones from Shumway (2001) and Campbell et al. (2008) have emerged, that use both accounting and market data, though without the implementation barriers the contingency claim-based models impose. Additionally, these improve on one of the biggest problems from the traditional approach, which is its static nature. Shumway (2001) argues that not only does the hazard model produce more consistent and, in some cases, unbiased results, by including market data, it also overcomes the setback that accounting ratios are found to be statistically unrelated to bankruptcy probability. Despite the arguably fewer interpretability of the discrete-time hazard model, the critical significance of corporate bankruptcy prediction models is their ability to forecast potential failures, thus "(…) the true worth of different approaches should be measured by how good they are empirically rather than how sound they are theoretically" (Bauer & Agarwal, 2014). In their research, Bauer & Agarwal (2014) claim that the hazard/logit model dominates both previous approaches in terms of predictive ability.

2.2. Machine Learning Approaches

The machine learning methodologies differ from statistical approaches mostly in the fact that, for statistical techniques, the researcher needs to impose a structure to the model, such as linear or logistic regressions, to then estimate the parameters to fit the data at hand, while the machine learning techniques automatically derive the structure for the model representation of the dataset (Huang et al., 2004). This flexibility can allow for higher accuracy rates in the classification of bankruptcy which, in the end, is the critical deciding factor to using a model for this matter. Compared to a simple statistical model, this high level of adaptability comes with lower interpretability. Machine learning methodologies are normally split into two categories, supervised and unsupervised. Supervised models are trained using "labelled" data, meaning each observation in the training dataset comes with its correct answer for the problem in study. Contrarily, unsupervised models are mainly trained with "unlabelled" data, meaning it is not trained under supervision and it needs to discover information on its own.

Out of the panoply of methodologies machine learning offers, two main models have been applied and studied in the corporate bankruptcy forecasting matter, these are Support Vector Machines (SVM) and Artificial Neural Networks or, most commonly known as, simply

Neural Networks (NN). Both of these models can be applied to classification problems such as the topic at hand. As Min and Lee (2005) mention, SVM have been applied in several finance topics such as credit rating, time series prediction, insurance claim fraud detection and, the topic of this research, corporate bankruptcy prediction. A similar situation is encountered for NN. Tsai and Wu (2008) argue that NN results are superior to that of traditional models in matters such as corporate bankruptcy and credit scoring, being the most popular tool used in financial decision-making problems.

Developed by Vapnik (1998), SVM is a binary classification exercise which finds hyperplanes that maximize the distance from the hyperplane to the data points, and it is equivalent to solving a quadratic optimization problem (Shin et al., 2005). This way, SVM finds the maximum margin hyperplane which gives the maximum separation or distance between the decision classes. This methodology has mainly been used in corporate bankruptcy literature as a comparison to the discriminant analysis models, or other machine learning techniques. In most of the cases, SVM outperforms the traditional models but there is no clear consensus as to being the best machine learning methodology. For instance, Min and Lee (2005) use SVM and a 5-fold cross validation application to the corporate bankruptcy problem and find that compared to multiple discriminant analysis, logistic regression analysis and three-layer back-propagation neural networks, it outperforms the other methods in their study. More recently, Barboza et al. (2017) find that although machine learning techniques show a 10% higher accuracy, SVM does not lead to higher accuracy than other machine learning models.

Another well implemented methodology in the issue of corporate bankruptcy forecasting is (Artificial) Neural Networks (NN), which can be supervised or unsupervised. This technique is inspired by the biological nervous systems and the way they process information. Since, for example, the human brain learns by experience, the same happens with NN, that by taking previous examples can learn patterns to predict and make decisions (classify information) on new data. The most common type of NN in financial decision-making is, as argued by Tsai and Wu (2008), the multilayer perceptron (MLP) trained by the backpropagation learning algorithm. This consists of a layer of "input" units fully connected to a layer of "hidden" units (that can be connected to more hidden layers) and finally connected to a layer of "output" units (Tsai & Wu, 2008). The data is fed to the input layer, which then moves through the subsequent layers until reaching the output layer – this process is called the forward pass. Then the machine goes through the process of a backward pass in order to decrease the loss function calculated at the end of the forward pass until this function reaches a minimum, hence the name "training by the backpropagation learning algorithm". In the corporate

bankruptcy literature this technique has mostly been used in comparison with a multivariate discriminant analysis, which it outperforms most of the time, and to other machine learning methods, to which there is not a consensus on the best model to use.

Two of the first authors to use NN in the matter at hand were Odom & Sharda (1990) which used the same variables from the Altman (1968) accounting-based model and showed that the machine learning technique provided better results than the simple multivariate discriminant analysis. The same result has been proven in further research, including studies in different geographies (Lee & Choi, 2013; Tsai & Wu, 2008; Barboza et al., 2017; Tam & Kiang, 1992).

It is of importance to mention that no research has been found on the question of machine learning techniques in corporate bankruptcy prediction with a time-dependent logit model as a benchmark. Thus, this research is of the utmost value to fill this gap within the corporate bankruptcy prediction literature.

3. Data

3.1. Data Retrieval

The main purpose of this research is to conduct an analysis on the usefulness of machine learning techniques for the topic of corporate bankruptcy prediction. Thus firstly, one has to retrieve the data necessary for the computation of the accounting and market variables, as well as the labels of bankrupt or non-bankrupt company. These include data for the Campbell et al. (2008) variables and the Altman (1968) variables, the latter used only for the benchmark estimation.

The data were retrieved from the CRSP and CRSP/Compustat Merged databases available through the Wharton Research Data Services as well as from the Thomson Reuters Datastream database. The timeline chosen is from 2000 to 2019, to include enough data to conduct a thorough analysis. The method to label the data as bankrupt or non-bankrupt was through the use of the "Research Company Reason for Deletion" variable in the CRSP/Compustat Merged database, as this contains a set of characteristics classified by codes, within which are two that can be considered for the bankruptcy case. These are codes 2 (Bankruptcy) and 3 (Liquidation). The remainder of the companies which have been deleted from the database for other reasons such as Acquisition, was removed as to not interfere with the efficacy of the models. Moreover, the bankrupt companies were not considered bankrupt until the year where the last information is available and considered failed afterwards, where zeros were input in the variables for the years after bankruptcy, for which one did not have information. Lastly, only companies that were in the dataset for at least two years were included, in order to have enough data per company for the analysis to be conducted. Thus, within the time frame chosen, the number of companies to be analysed will be 3664, of those 144 are classified as bankrupt.

3.2. Data Treatment and Variable Construction

The variables were constructed according to the study by Campbell et al. (2008), though on an annual basis. As such, variable construction for the market variables was slightly altered. The variables and their calculation are explained in Table 3.1:

Variable	Calculation[1]
NIMTA	$Net\ Income_i / Market\ Value\ of\ Assets_i$
TLMTA	$Total\ Liabilities_i / Market\ Value\ of\ Assets_i$

[1] Subscript i refers to each company while subscript SP500 refers to the S&P500 index

Variable	Calculation[1]
CASHMTA	$Cash\ \&\ Cash\ Equivalents_i / Market\ Value\ of\ Assets_i$
MB	$Market\ Value\ of\ Equity_i / Book\ Value\ of Equity_i$
EXRET	$Monthly\ Log\ Return_i - Monthly\ Log\ Return_{SP500}$
SIGMA	$One\ year\ standard\ deviation\ of\ the\ monthly\ return_i$
RSIZE	$Log(Market\ Capitalization_i / Market\ Capitalization_{SP500})$
PRICE	$\text{Log(Monthly Price}_i\text{)}$ [Truncated above at 15$]

Table 3.1 | Variables used and their computation
In this table the construction of the variable is described alongside their names.

Variable standardisation is a common practice in machine learning, it is normally done by subtracting the mean and dividing by the unit variance, though this is highly prone to outlier influence. Thus, all variables suffered a process named Robust Standardisation using the Scikit-Learn RobustScaler method. It is preferred against the StandardScaler method as the former uses statistics which are robust for outliers, hence the name. According to the Scikit-Learn documentation, this scaler removes the median and scales the data according to the Interquartile Range (IQR), which is between the first quartile (25th percentile) and the third quartile (75th percentile). This will affect variables such as the MB (market-to-book ratio) tremendously as these include very extreme outliers and, thus, should be treated. For this Campbell et al. (2008) winsorised his variables at the 5th and 95th percentile, though in this analysis the predefined scaling of the RobustScaler class will be used.

The analysis for all models with the exception of the Altman Z-Score model was conducted on a five-year rolling window basis and, for each window, the train set consists of the first four years and the test set is the last year. With this, 15 overlapping windows will be analysed, and the metrics' results will be averaged for each model. One concern might appear, as after the companies went bankrupt and left the dataset zeros were input in the variables for years after bankruptcy, thus there would be rolling windows where a company was considered bankrupt and all its variable values were zeros. Though a company was only considered for each rolling window if at least for one of the year-company observations, there was real information available (not a full line of zeros), which corresponds to data from one or two years prior to bankruptcy. For instance, a company that bankrupted in 2005 and only has financial data until 2004 will only be analysed until the rolling window starting in 2004, thereafter it will not be included in each subset of the data.

[1] Subscript i refers to each company while subscript SP500 refers to the S&P500 index

3.3. Data Exploration

For further understanding of the data at hand, data exploration techniques had to be applied and their results analysed. The first method is to conduct a summary statistics analysis which is displayed in Tables 3.2 and 3.3. The first table is referring to the whole of the dataset including bankrupt and non-bankrupt companies, while the second refers only to the companies that went bankrupt at any time in the time frame. While for the models the company observations are only considered positive (1) for the bankrupt variable when they leave the database, in the summary statistics analysis all observations for one company are considered positive (1) for the bankrupt variable, independently of when the company left the database because of bankruptcy reasons. This way it is made sure one analyses all the time frame for which bankrupt companies were alive and not just in the year they went bankrupt. The input zeros were not considered for the summary statistics analysis as they would skew the results. The summary statistics results are prior to the scaling of the variables so to show their real values.

As observable, the scaling of the data is highly important for some variables, as for example the MB variable shows extreme outliers on the positive end leading to high standard deviation whether for the whole dataset as for the bankrupt sample. Thus, it was reasonable to standardise the variables at the 25^{th} and 75^{th} percentiles. Comparing the two datasets, it is observable that variables such as NIMTA, MB, PRICE and EXRET display lower values for bankrupt companies than for the overall average as expected, while volatility represented by the SIGMA variable shows higher levels for bankrupt companies. Furthermore, most variables display positive values across the board and lower values for the bankrupt sample, with the exception of RSIZE and a large part of EXRET that display negative values, which is in line with the Campbell et al. (2008) results.

	NIMTA	TLMTA	CASHMTA	MB	SIGMA	PRICE	RSIZE	EXRET
Mean	-0.02	0.41	0.11	35.09	0.14	2.22	-10.01	-0.05
Standard Deviation	0.19	0.28	0.15	818.05	0.13	0.86	2.19	0.60
Minimum	-8.30	0.00	0.00	0.02	0.00	-3.06	-17.16	-5.70
25th Percentile	-0.01	0.18	0.02	1.23	0.07	2.01	-11.57	-0.26
Median	0.02	0.36	0.06	1.86	0.10	2.71	-9.98	-0.03
75th Percentile	0.04	0.63	0.13	2.93	0.16	2.71	-8.53	0.19
Maximum	4.70	1.00	3.07	103783.76	2.60	2.71	-2.96	6.47
Excess Kurtosis	297.61	-1.02	30.75	6546.53	26.56	3.87	-0.30	10.19
Skewness	-10.71	0.42	4.16	65.86	4.05	-2.03	0.06	0.35

Table 3.2 \| Summary Statistics for entire dataset
This table represents the summary statistics of the eight variables in the entire dataset. Values are displayed in units.

	NIMTA	TLMTA	CASHMTA	MB	SIGMA	PRICE	RSIZE	EXRET
Mean	-0.12	0.47	0.17	12.41	0.23	1.49	-11.64	-0.31
Standard Deviation	0.35	0.31	0.26	95.44	0.19	1.14	1.49	0.85
Minimum	-4.08	0.00	0.00	0.05	0.00	-2.67	-16.86	-4.73
25th Percentile	-0.16	0.17	0.02	0.92	0.11	0.72	-12.59	-0.70
Median	-0.02	0.46	0.07	1.44	0.18	1.70	-11.62	-0.21
75th Percentile	0.02	0.77	0.20	2.52	0.28	2.52	-10.54	0.17
Maximum	1.50	1.00	2.15	1965.00	1.64	2.71	-7.05	4.22
Excess Kurtosis	18.71	-1.11	8.74	17.94	1.20	-0.59	0.22	2.37
Skewness	-3.61	-0.46	2.86	4.00	1.11	-0.23	-0.14	-0.61

Table 3.3 | Summary Statistics for bankrupt companies
This table represents the summary statistics of the eight variables in the dataset filtered for bankrupt companies. Values are displayed in units.

Another step to understanding the topic at hand is to assess the number of bankruptcies that happened each year, such is displayed in Figure 3.1 below. As observable the years of the highest number of bankruptcies are 2001 and 2008, which might be explained by the dotcom bubble and the financial crisis, respectively. After 2014 there are no bankruptcies reported in the dataset, this could have happened for a number of reasons and does not represent reality in these last years. Overall, the number of bankruptcies is variant throughout the years and thus one should take into account this factor when analysing results from the models, as different models might be better at predicting bankruptcies in different time periods with different economic realities.

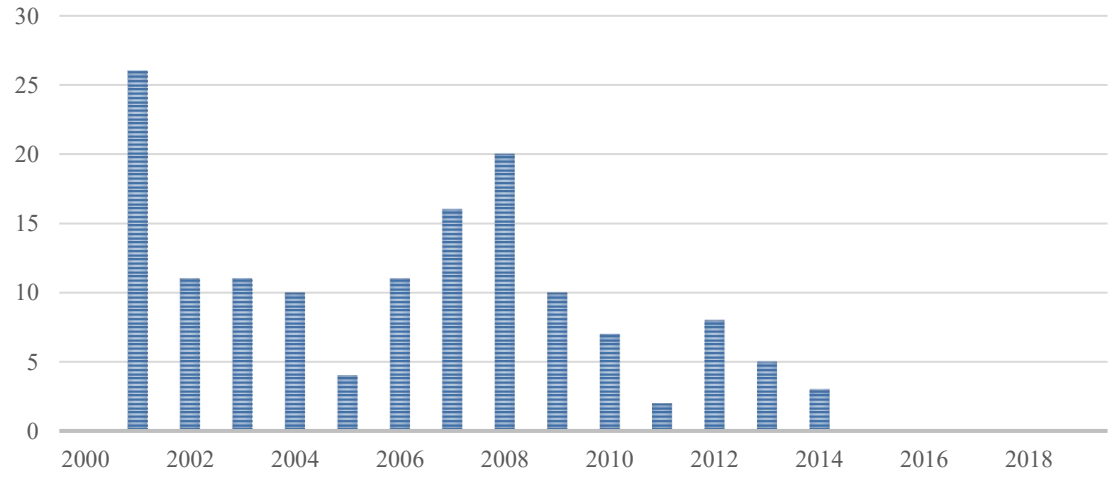


Figure 3.1 | Number of Bankruptcies per year
This graph represents the distribution of bankruptcies per year in the dataset throughout the time period studied.

Finally, one other factor important for the analysis is to understand how bankruptcy varies in different industries, specifically to understand which industries have the greatest number of bankruptcies within the time frame analysed, and such is shown in Figure 3.2 below. The industry classification used is the Global Industry Classification Standard (GICS), thus the dataset is divided into eleven industries. The imbalance of the two labels is clear in this chart, with bankruptcy being the minority in all industries. The industries with the highest number of bankruptcies within the time frame are Consumer Discretionary, Information Technology and Financials, though the industry with the second highest proportion of bankruptcies to the total number of companies is Consumer Staples (5%) following Consumer Discretionary (7%). Important to point out are also the most common industries in the dataset, which are Health Care, Financials and Information Technology.

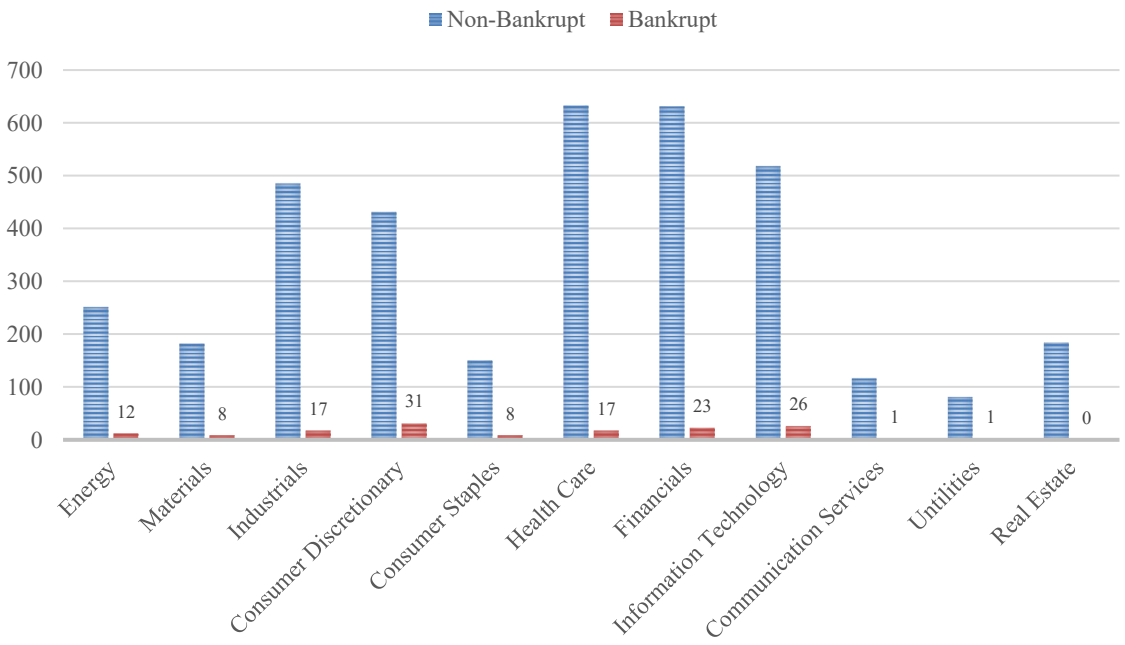


Figure 3.2 | Number of Bankrupt and Non-Bankrupt Companies per Industry
This graph represents the distribution of bankruptcies and non-bankruptcies in the dataset for the GICS industries.

Finally, and as previously mentioned, the dataset is highly imbalanced having only 4% of the companies leaving the database for bankruptcy reasons. Additionally, the average life of a company in the dataset is twelve years and the variables with the highest correlation with the bankruptcy variable are PRICE, SIGMA, RSIZE and NIMTA.

4. Methodology

All the following models (except for the Altman Z-Score) will be conducted on a 5-year rolling window analysis with a robust standardisation of the variables. All data treatment processes are explained in the previous chapter (Chapter 3 – Data) under the subsection "3.2. Data Treatment and Variable Construction".

4.1. Altman Z-Score - Benchmark

The first benchmark of this research is the established methodology in corporate bankruptcy known as the Altman Z-Score pioneered in 1968. Altman initiated the use of accounting ratios to predict default, creating a static one-period multivariate analysis based on financial ratios, defining a score to determine the probability of default. For the sake of this analysis and to turn the result of the equation binary, company-year observations with a Z-Score below 1.8 are considered as bankrupt (1), and all other scores refer to non-bankruptcy (0). The model will be conducted using the following equation and applied to each company-year observation:

$$Z - Score_t = 1.2\ \frac{Working\ Capital_{t-1}}{Assets_{t-1}} + 1.4\ \frac{Retained\ Earnings_{t-1}}{Assets_{t-1}} + 3.3\ \frac{EBITA_{t-1}}{Assets_{t-1}} + 0.6\ \frac{MV(Equity)_{t-1}}{BV(Assets)_{t-1}} + 0.99\ \frac{Sales_{t-1}}{Assets_{t-1}} \quad (4.1)$$

where MV(Equity) and BV(Assets) refer to the market values and book values of the variables, respectively.

4.2. Logistic Regression (Simple Logit) - Benchmark

The second model used is also to be considered a benchmark although it is an adaptation of the long-horizon forecasting methodology from Campbell et al. (2008). This model differs on some aspects to the original: (1) the original is conducted on a monthly basis and this analysis will be done on an annual basis thus, as mentioned in Chapter 3, the variables suffered some alterations; (2) on the original article the authors tested a lag of zero, six, twelve, 24 and 36 months and in this analysis, only a lag of two years (24 months) will be used; (3) the authors had only explanatory variables for one period (month) per company and in this analysis, another period per company will be added; (4) this analysis will be based on a 5-year rolling window. The model will be conducted to mimic equation 4.2 below.

$$Bankruptcy_{i,t+2} = \frac{1}{1 + exp(-\alpha - \beta_t\, x_{i,t} - \beta_{t+1}\, x_{i,t+1})} \quad (4.2)$$

where $Bankrutpcy_{i,t+2}$ is an indicator that equals one if the firm is predicted to go bankrupt or fail in year t+2, and $x_{i,t}$ and $x_{i,t+1}$ are vectors of explanatory variables known at the end of the previous year and two years, respectively. A higher level of $\alpha + \beta_t x_{i,t} + \beta_{t+1} x_{i,t+1}$ implies a higher probability of bankruptcy or failure.

The tool used for the estimation of this model is the Scikit-Learn Logistic Regression class. Logistic regression is a classifier commonly used in several research areas. It can be used to estimate the probability of an observation being in one class rather than the second, which then is converted to positive (1) or negative (0) through a threshold, which is normally defaulted as 50%. The way it works is similar to a typical linear regression where the weighted sum of a vector of variables with a vector of weights is calculated, though in the logistic regression case it is then put through a sigmoid function, represented below in equation 4.3, that outputs a result between zero and one.

$$\sigma(t) = \frac{1}{1 + exp(-t)} \tag{4.3}$$

Then the output is put through a threshold function that decides whether the result will be a positive or negative observation. The way this model is trained is through a log loss function that emulates the output of system of equations 4.4 below.

$$c(\theta) = \begin{cases} -\log(\hat{p}), & \text{if } y = 1 \\ -\log 1(1 - \hat{p}), & \text{if } y = 0 \end{cases} \tag{4.4}$$

This system represents the behaviour of the log loss function in logistic regressions, for example when t equals zero the function -log(t) grows to infinity which means the cost will be large if the model estimates a probability close to zero when the real observation is positive (1); similarly, the cost will be large when the model estimates a probability close to one when the real observation is negative (0). Contrarily, the cost will be low when the model correctly identifies the observations as positive or negative, as for example if t is close to one the function -log(t) will be close to zero, thus if the model estimates a probability close to zero or one when the real observations are zero or one, respectively, the cost will be very low. The overall cost function can be written in one equation which represents the average of all observations' costs and is represented in equation 4.5 below.

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log\left(\hat{p}^{(i)}\right) + \left(1 - y^{(i)}\right)\log\left(1 - \hat{p}^{(i)}\right)\right] \tag{4.5}$$

As the log loss function is convex, an optimisation algorithm can be applied to find its minimum. In the Scikit-Learn Logistic Regression class the solver is set to default as the 'lbgs', which stands for Limited-memory Broyden–Fletcher–Goldfarb–Shanno. This solver

approximates each second derivative matrix update with gradient evaluations and it also stores only the last few updates, meaning it saves some memory. This hyperparameter value will be set as default for all variations of the Logit model.

4.3. Shared Machine Learning Techniques and Concepts

For the following three models, several machine learning techniques will be used and are common to most of them. Thus, the following section will be used to explain them and provide a basis for all the further model setup and estimation.

Firstly, a common characteristic of machine learning models is hyperparameters and their respective tuning. A hyperparameter is a parameter that is used to structure the model and whose value cannot be determined via training. These are also used to control the learning process, thus they have a high impact on the results of each model. This means that the selection of the values for these parameters has to be done with some care and to do that one can use hyperparameter tuning or optimisation. Tuning is the process used to find the best group of parameter values that yield an optimal model, minimising the predefined loss function and/or maximising the metrics chosen in advance. The GridSearchCV tool from the Scikit-Learn library will be used for this tuning process. This tool conducts an exhaustive search of pre-specified hyperparameter values for a certain model to find the one that results in the best score for a certain pre-determined metric. The metric used for every following hyperparameter tuning is the F1-score.

Within the previous class of characteristics of machine learning models there is an important hyperparameter called regularisation. This is a technique that ensures that the model avoids overfitting the training data by discouraging learning a model with higher complexity or flexibility. There are three prominent types of regularisation, these are L1, L2 and Elastic Net. The basic premise of these three methods is to add a component in the loss function that penalises it, so it does not overfit the data with the objective of creating more sparse models that perform better in data not previously seen, which is the goal in prediction. The main difference between L1 and L2 regularisation is the type of norm they use to compute the added parameter in the loss function, the former uses an L1 norm while the latter uses an L2 norm. Both norms are displayed in the equations 4.6 and 4.7 below.

$$L1 = \sum_{i=1}^{n} |w_i| \tag{4.6}$$

$$L2 = \sum_{i=1}^{n} w_i^2 \tag{4.7}$$

The Elastic Net, or sometimes called L1-L2 regularisation, is simply a linear combination of the previous two, which yields the benefits of both types. If not mentioned within the following model descriptions in this chapter, the default values for the regularisation parameters were kept.

Secondly, a technique used as an improvement attempt to all models will be described. As this dataset is highly imbalanced, one could use methods such as undersampling of the non-bankrupt data, as previously done by Altman (1968) and Barboza et al. (2017), though with such technique the loss of useful data is immense. Thus, a relatively recent technique of oversampling will be used. That is the Synthetic Minority Oversampling Technique (SMOTE) developed by Chawla et al. (2002), highly used for the problem of imbalanced data in several research areas. The basis for this technique is to oversample the minority class by creating synthetic observations between the existing (real) ones. The way it is conducted is based on another machine learning technique called K-Nearest Neighbours, where between each minority observation it draws imaginary lines and creates a new synthetic observation within that line. More specifically, firstly a minority class instance (i.e. x_1) is found, then from that example the K-nearest observations are found, usually five, and finally one of the neighbours (i.e. x_2) is randomly selected and a synthetic example is produced at a randomly selected point within the pre-calculated distance between the two points (x_1 and x_2) in the feature space. SMOTE was conducted following the train and test split to not spoil the test sets. In other words, the test sets are still imbalanced and were directly extracted from the original dataset, and the oversampling technique was conducted on the training sets.

4.4. Logistic Regression (Logit) with Machine Learning

On an attempt to improve the Simple Logit model's predictability, two machine learning methods were applied: oversampling and class-weight tuning. Oversampling is a methodology typically applied to imbalanced datasets such as the one at hand and was previously explained in section 4.3. Class-weight tuning is another technique used in imbalanced datasets and it is done in accordance with class distributions in the dataset, meaning in this example the distribution of bankrupt and non-bankrupt companies. Class-weight is a hyperparameter available for logistic regression models and represents how far the model is punished for every incorrect prediction of a class. In a base setting, the class-weight hyperparameter is set to penalise the incorrect predictions equally, with a weight of one for each class, but for

imbalanced data it can be set to punish more the incorrect predictions of the minority class (in this case, bankruptcy). It is typically set as the opposite weights of the distribution, meaning in the dataset at hand it can be set as 96 for the bankrupt class (1) and 4 for the non-bankrupt class (0). Furthermore, a more advanced technique will be used for hyperparameter optimization purposes. That is the GridSearchCV tool from the Scikit-Learn library explained in the previous section.

4.5. Support Vector Machines (SVM)

Support Vector Machines are a form of supervised learning, meaning it infers a function from labelled data (input-output pairs) to then deduce an estimate for an unlabelled data point. It can be used for both regression and classification problems, though for what concerns this research only the classification methodology will be explained. The SVM classification differs from other classifiers in the way that it tries to find the hyperplane that maximises the margin that separates the classes at hand. This allows for better results out of sample, since other classifiers might be able to separate the classes in the training process but have such small margins dividing these that they might fail in the test process. SVM classification can be linear or non-linear and either of these two types can be applied depending on the problem's data and efficiency requirements. Linear SVM might be more efficient and work well in certain types of problem but it does not fit problems where data is not linearly separable. In either case, SVM methodology will seek the optimal parameters associated with a kernel function. A kernel is a function capable of computing the dot product $\phi(\mathrm{a})^{\mathrm{T}} \cdot \phi(\mathrm{b})$ based only on the original vectors a and b, without having to compute (or even know) the transformation ϕ (Géron, 2019). Multiple types of kernel functions are available but according to Min & Lee (2005), the one that produces the best result for this problem is the Radial Basis Function (RBF), thus it will be used in this research as well. Nonetheless, the four kernel functions available in the Scikit-Learn library for Support Vector Machines models are Linear, Polynomial, RBF and Sigmoid, being the RBF a non-linear kernel for SVM classification. The RBF kernel is represented by the equation 4.8 below.

$$K(a, b) = \exp\left(-\gamma \|a - b\|^2\right) \tag{4.8}$$

When training an SVM with the RBF kernel, one must take into account two parameters: C and γ. Also called the regularisation parameter, C is common to all kernel functions and it characterises a trade-off between the misclassification of training examples with the simplicity of the decision boundary or surface. A low C leads to a smooth decision surface while a high C will fit the training examples as much as possible, sometimes leading to overfitting. As for γ

(gamma), it defines how significant a single training example is and it also acts like a regularisation hyperparameter. A large γ will lead to low variance but also high bias models, meaning each instance's range of influence is smaller which leads to the decision boundary being more irregular, worming around individual observations. Contrarily, a small γ will lead to higher variance but lower bias models, meaning each instance has a larger range of influence leading to a smoother decision boundary. In Figure 4.1 below one can observe the effect of different values for C and γ using an RBF Kernel taken from Géron (2019).

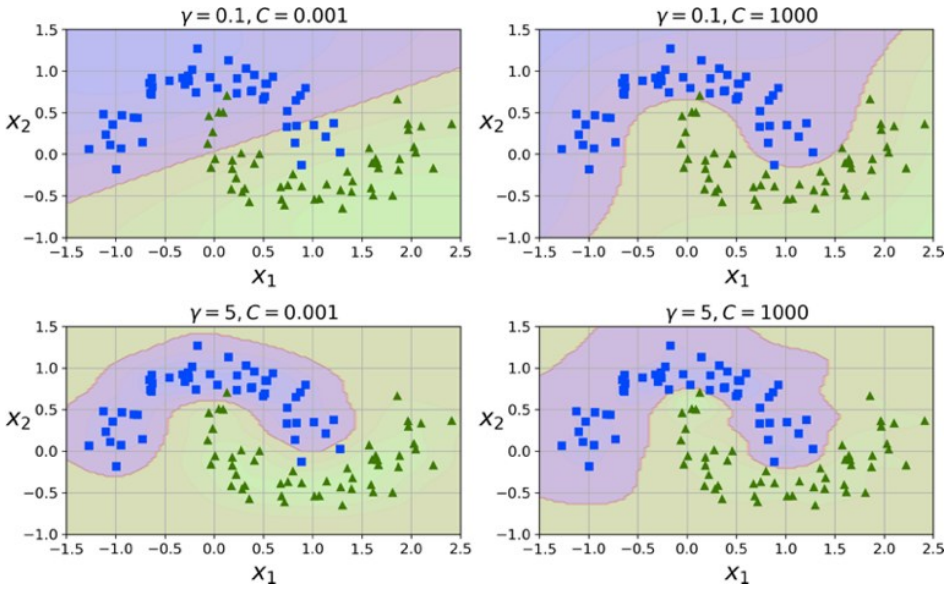


Figure 4.1 | SVM classifiers using an RBF kernel with different hyperparameters (Géron, 2019)
This figure represents the different fits of an SVM model depending on the different values assigned to the hyperparameters γ and C.

Initially, a simple Support Vector Classification (SVC) model with an RBF kernel function from the library Scikit-Learn will be used and the C and γ hyperparameters will be set as the default. The data input in the SVM models is structured in the same way as for the Logit models. Moreover, and on an attempt to improve the initial results, oversampling and hyperparameter tuning will be employed. For the tuning technique, the hyperparameters to be adjusted are C, γ and the class-weight, which is also available for the SVM model, and the tool used is, as aforementioned and explained, the GridSearchCV from the Scikit-Learn library.

4.6. Backpropagation Neural Networks (BPNN)

Neural Networks can be used for several different types of problems, both for regression and classification. As mentioned previously, the multi-layer perceptron (MLP) architecture trained by the backpropagation learning algorithm is the most commonly used in the topic at hand.

Figure 3.1 shows an example of a three-layer neural network, with one input, one output and one hidden layer.

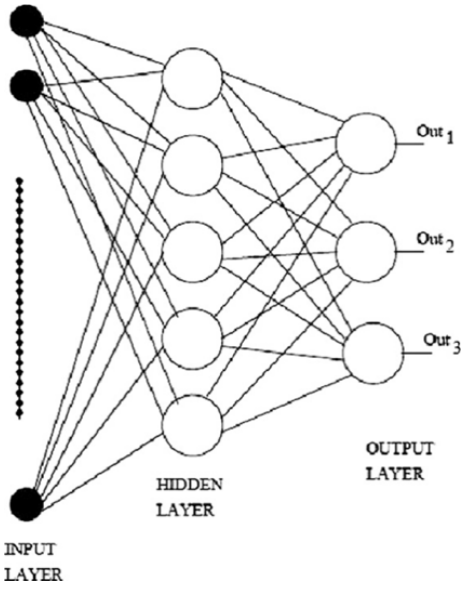


Figure 4.2 | Three-layer Neural Network (Tsai & Wu, 2008)
This figure represents the structure of a three-layer Neural Network, with an input layer, a hidden layer and an output layer similar to the structure of the BPNN used in this research.

The backpropagation learning algorithm is used to efficiently compute the gradients automatically, meaning in just two steps through the network (one forward, one backward), the algorithm computes the gradient of the network's error in regard to every parameter of the model. Thus, it can find how each weight and bias term within the network should be altered to reduce the error of the model. The steps it takes to conduct this process are the following: (1) it processes one mini-batch at a time (the size of the batch is something to be defined beforehand) and it goes through the training set in batches multiple times, each time it processes the training set is called an epoch (the number of epochs is also to be set beforehand); then (2) each mini-batch is sent through the first hidden layer for the algorithm to compute the output of this layer which then passes through the next hidden layer(s) until it reaches the output layer, this is called the *forward pass* (all the intermediate results are saved to be used for the *backward pass*); next (3) the algorithm measures the output error (using a loss function that is to be defined previously); then (4) it computes how much each hidden-output layer connection contributed to the error until reaching the input layer, this is done analytically by applying the *chain rule* (this reverse pass efficiently measures the error gradient throughout all the connection weights in the network by backpropagating the error gradient through the network); finally, (5) the algorithm performs a Gradient Descent step to alter all the connection weights, using the error gradients computed (Géron, 2019).

For the creation of a NN, some hyperparameters have to be set, these include the aforementioned, batch size, epochs, and loss function, plus the following parameters: activation functions for each layer, number of neurons in the hidden layer(s), the optimiser or solver, and metrics for the model to use. From the latter group, activation layers are needed to add a degree of non-linearity between layers, since if there is only a chain of several linear transformations from layer to layer that is equivalent to having a single linear layer which does not allow for the solving of complex problems. For what concerns neurons, these are determined by the input and output the problem requires, it used to be standard procedure to form an inverted pyramid along the network with fewer neurons at each subsequent layer. Nowadays simply using the same number of neurons in all hidden layers is the common practice as it not only performs equally well in most cases, it is easier and more efficient to tune since there is only one hyperparameter to optimise instead of one per layer. (Géron, 2019)

The data input in the BPNN models is structured in the same way as for the Logit and SVM models. Initially, a simple Sequential model with three Dense (fully connected) layers from the Keras library will be used, and the following parameters will be set as such:

- Input and hidden layers' activation functions as the Rectified Linear Unit (ReLu) which is the most widely used function in designing networks;
- Output layer activation function as the sigmoid activation function since this is a binary classification problem;
- The number of neurons in the hidden layer is an arbitrary decision and is usually tested to find its optimal value, though the value for hidden layer neurons will be set as 32 which is one of the values used by Tsai & Wu (2008);
- The loss function to be used is the "binary_crossentropy" as it is the standard for binary classification and is only to be altered for specific situations;
- The optimizer used is the Stochastic Gradient Descent ("sgd") which is also a standard solver used in machine learning;
- The number of batch units is also arbitrary, thus the starting point will be 100 for the batch size, a lower value can produce better results but will also be more computationally exhaustive and therefore slower.
- The number of epochs will be set as 50 which is the minimum value used by Tsai & Wu (2008) and can be increased to produce better results if needed.

– Finally, accuracy and recall are to be set as a metric to be used by the algorithm, as accuracy is not the best metric for imbalanced datasets but still important while recall is very significant for this type of analysis.

Additionally, for result improvement purposes, oversampling will be employed. Hyperparameter tuning could also be conducted as further research, though one must keep in mind that such technique applied to Neural Networks models takes a great level of computational power and as such should be used with caution.

4.7. Industry Fixed Effects and Other Improvements

Fixed effects methodologies have been used in econometrics to control for individualistic or idiosyncratic characteristics of the observations at study, in the case at hand industry effects are the logical option to control for industry-specific characteristics. For instance, one industry might have a lower financial ratio by definition and that, in said industry, does not typically translate into bankruptcy as it would in another industry. The fixed effect coefficients absorb all the across-group action, meaning what is left is the within-group action, which is what one would be interested in when solving a classification problem of this sort. This methodology also assists in reducing the threat of omitted variable biases. Hence, as an attempt to improve on the previous methodologies, industry dummies for the Global Industry Classification Standard were added with code 10 (Energy) as the benchmark, meaning this dummy was removed from the variables. One might also explore, in further research, individual fixed effects methodologies which take dummies (or equivalent methodology) of each of the companies and add such as variables in the models. The industry fixed effects technique is to be applied in all models except the Altman Z-Score.

Another improvement applied was the use of the probabilities extracted from the Logistic Regression, SVM and BPNN, applying a different threshold than the default 50% to determine if an observation is positive (1) or negative (0). The alteration of the threshold was conducted as follows: in models where Recall is high and Precision low (metrics further explained in Chapter 5) a higher threshold, meaning above 50%, was applied. This, in theory, would improve the Precision of the model without hurting Recall too much. For models where the inverse happens a lower threshold, meaning below 50%, was applied. Ideally, this would encompass a larger number of firms to be considered bankrupt leading to a higher Recall without affecting Precision significantly.

5. Evaluation

5.1. Evaluation Metrics

The metrics used for a classification problem are based on a matrix that divides the results according to the real solutions for the problem. This is called a confusion matrix and an example for corporate bankruptcy is represented in Figure 5.1. As one can observe, the optimal results lie in the True Positive (TP) and True Negative (TN) sections of the figure, though as this is an estimation analysis it is bound to have some False Positive (FP) or False Negative (FN) results, which correspond to Type I and Type II errors, respectively. This is what one wants to minimise.

		Predicted Values	
		No Bankruptcy	Bankruptcy
Real Values	No Bankruptcy	True Negative (TN)	False Positive (FP)
	Bankruptcy	False Negative (FN)	True Positive (TP)

Figure 5.1 | Confusion Matrix for the Corporate Bankruptcy Problem
This figure represents a confusion matrix for the corporate bankruptcy problem with the positive event being bankruptcy.

The four metrics to be used are, as aforementioned, all based on the confusion matrix. These are Accuracy, Recall, Precision and the F1-score which will be described in the following section.

Accuracy is the typical measure used in the corporate bankruptcy prediction research though, it might not be the best measure when one uses an imbalanced test set. Nevertheless, accuracy represents the proportion of correct predictions (TP or TN) within the whole number of examples analysed. The formula for its calculation is the following:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.1}$$

Recall is another metric that can be calculated from the confusion matrix and it refers to the proportion of positives that are correctly identified. It is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{5.2}$$

Precision is a measure of the proportion of true positives among the predicted positive values and it is calculated as in the following formula:

$$Precision = \frac{TP}{TP + FP} \tag{5.3}$$

Finally, the F1-score is a combination of equations (5.2) and (5.3), the harmonic mean of recall and precision and it measures the overall model accuracy. It is the preferred metric for this type of problem, and as one is not so concerned with the proportion of True Negatives (correctly predicted non-bankruptcies) no other measure will be used. The F1-score is computed as follows:

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5.4}$$

For all of the metrics, the closer to one the measure the better, though as found in previous research the accuracy for machine learning methods in the corporate bankruptcy prediction problem is always high and close to 1, thus one must not give as much importance to this measure as it is skewed by the True Negatives in an imbalanced dataset of this topic.

Additionally, confusion matrices are also going to be used to further understand the aforementioned metrics. These confusion matrices will be represented as percentages of real positives and negatives. Therefore, the perfect confusion matrix would have 1 or 100% in the top left corner, meaning 100% of the negative cases were correctly predicted and 0 or 0% on the top right corner, conversely it would have 1 or 100% in the bottom right corner, meaning 100% of the positive cases were correctly predicted and 0 or 0% on the bottom left corner.

5.2. Variable Predictive Power

For further understanding of the predictive power of each variable, different methods were applied to assess such information depending on the model at hand. This evaluation is highly significant not only to measure which variables have a high influence on each model's results, but also to demystify some of the behaviour of the machine learning models. For each of the three types of model (Logit, SVM and BPNN) only one evaluation was conducted, as each model variations are rather similar. This was done on the best predicting model (or model improvement) based on the metric F1-score afore explained. The methodologies explained below were applied on each rolling window to assess if the variable importance changed throughout the periods.

For the Logit model picked, the methodology applied was based on the weights available for extraction. The most predictive variable was then chosen based on the maximum absolute weight of each variable. As the data is standardised it is possible to use this methodology.

For both machine learning model types, the methodology applied was one similar to the feature selection application. After the training process of the model, each column of variable

test data was shuffled separately. Then based on the F1-score metric, the variable data shuffling that hurt the model result the most is said to be the variable with the most predictive power for that model.

6. Results

A summary of the results is presented in Table 6.1 where the models with and without modifications are shown, taking into account the four metrics used. As previously mentioned, the most important metric for imbalanced datasets is the F1-score, and as one is trying to correctly predict positive classifications of the corporate bankruptcy problem Recall is the metric that evaluates this, thus it will also be important for the assessment of the models. Nonetheless, Accuracy and Precision are important to address the overall performance of the model and one should attempt to also keep the values for these metrics high. One factor important to mention is that the values in Table 6.1 are the averages of the rolling windows' metrics for each model (except for the Altman Z-Score).

For Table 6.1. and hereafter the following abbreviations will be used: "WC" for weighted-class with sample class weights, "SM" for oversampling (SMOTE), "TUN" for hyperparameter tuning, and an "I" at the end of the previous abbreviations or "IND" for industry fixed-effects.

Model/Metric	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Altman Z-Score	*83.17%*	*0.97%*	*46.53%*	*1.99%*
Simple Logit	*97.97%*	*14.46%*	*25.28%*	*13.53%*
Logit (WC)	85.05%	4.04%	59.39%	7.20%
Logit (TUN)	88.84%	5.32%	51.70%	8.73%
Logit (SM)	30.59%	1.06%	**78.13%**	2.08%
Logit (IND)	81.23%	2.89%	67.52%	5.50%
Logit (WCI)	92.97%	6.48%	54.96%	11.38%
Logit (TUNI)	96.60%	10.67%	34.43%	**15.60%**
Logit (SMI)	40.14%	1.14%	76.97%	2.24%
SVM	99.13%	0.00%	0.00%	0.00%
SVM (TUN)	93.55%	6.15%	46.12%	10.50%
SVM (SM)	84.97%	3.66%	55.41%	6.72%
SVM (IND)	99.13%	0.00%	0.00%	0.00%
SVM (TUNI)	95.84%	7.67%	36.49%	**12.26%**
SVM (SMI)	90.83%	4.18%	35.92%	7.32%
BPNN	99.12%	6.11%	0.65%	1.14%
BPNN (SM)	86.80%	4.06%	52.27%	7.31%
BPNN (IND)	99.10%	0.83%	0.18%	0.30%
BPNN (SMI)	93.37%	7.30%	47.36%	**11.76%**

Table 6.1 \| Evaluation metrics´ results for all models
This table represents the results for all models. In italic are the results for the two benchmarks. In bold is the model with the highest Recall and the best models based on their F1-scores. Abbreviations used: WC (Class-Weight with sample weighs), SM (Oversampling), TUN (Hyperparameter Tuning), I or IND (Industry effects).

6.1. Altman Z-Score

As expected, the results for the Altman Z-Score are unsatisfactory, though the Recall (rate of true positives among all possible positives) is high. Its accuracy is around 83% which is rather low compared to other corporate bankruptcy models. The factor that is the least satisfactory is its precision. If one looks at the confusion matrix displayed on the left side of Figure 6.1 compared to the perfect confusion matrix on the right side it is clear that, even though almost half of the positives are correctly predicted, that comes at the expense of losing precision, meaning 17% of the negative cases are predicted as positive, leading to a low F1-score and lower Accuracy.[1]

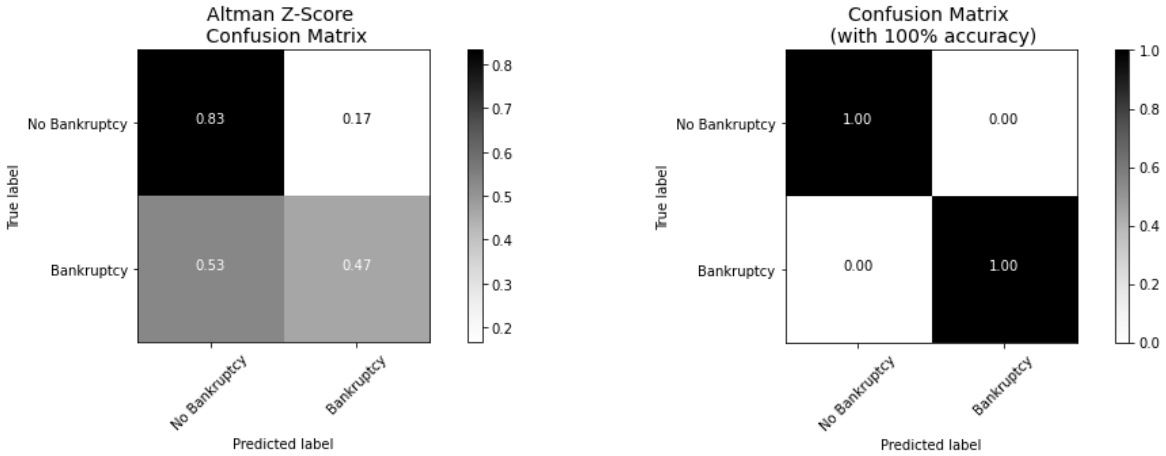


Figure 6.1 | Confusion matrix for the Altman Z-Score model vs. perfect confusion matrix
In this figure it is shown that 47% of the bankruptcies are correctly predicted while 17% of the of the non-bankrupt companies are incorrectly predicted as bankrupt (Total Non-Bankrupt: 40898, Total Bankrupt: 144)

6.2. Logistic Regression (Simple Logit and Logit with Machine Learning)

The Simple Logit model with no improvements led to somewhat satisfactory results, it is one of the models with the highest F1-Score and the model with the highest Precision (true positives among all positives predicted) though it is not better in terms of Recall compared with the machine learning models. This means this model, although precise in its bankruptcy predictions, does not predict the most bankruptcies among all models (on average, only 25% of the bankruptcies were predicted). All improvement attempts were successful in terms of correctly predicting more bankruptcies, though the improvement with best F1-score is the hyperparameter tuned with industry fixed effects (TUNI), which although it is not the variation with the most bankruptcies predicted, it is the most precise one. Improvements with oversampling (SM and SMI) led to an extremely low Accuracy and Precision though they are the models that identified the most bankruptcies in the dataset. Displayed in Figure 6.2 is the Recall metric for all Logit models (with the exception of the WC variations) throughout the

[1] The confusion matrix from the Altman Z-Score differs from the ones from other models as the analysis is not conducted on a rolling window basis

predicted time period. It is important to note that all metrics after 2016 are significantly decreased since during this period there are no bankruptcies detected in the dataset, thus being impossible to have true positives above zero.

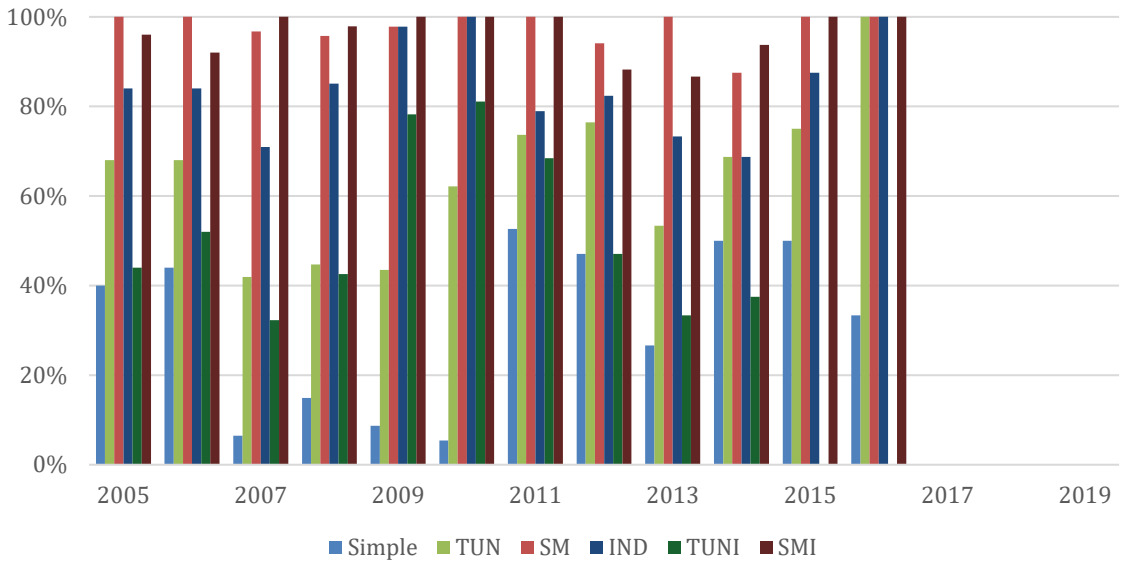


Figure 6.2 | Recall throughout the predicted time period for Logit models
This figure represents the distribution of different Logit models' Recalls throughout the predicted time period.

As observable, all improvement models outperform the Simple Logit significantly in terms of true positives predicted among all possible positives, meaning Recall. This metric is highly important for the prediction in imbalanced datasets and therefore displayed throughout time in Figure 6.2. Although this metric is one of the most important if the goal is to predict the highest number of bankruptcies, it is not the only one. According to this metric, the two best models to use would be both oversampling models that display Recall above 80% in all periods where bankruptcies existed. Though, if one observes Figure 6.3 that displays the F1-scores for all Logit models (with the exception of the WC variations) throughout the predicted period the conclusion would be different. This is because Precision is also important in any prediction problem, and F1-score calculates a harmonic mean of Recall and Precision. As observable, the F1-score is not as stable as the Recall throughout time, with some models having peaks and valleys. The F1-scores for the TUNI model outperform most models in various periods followed by the Simple Logit and TUN which seem to outperform the TUNI variation in the remaining periods.

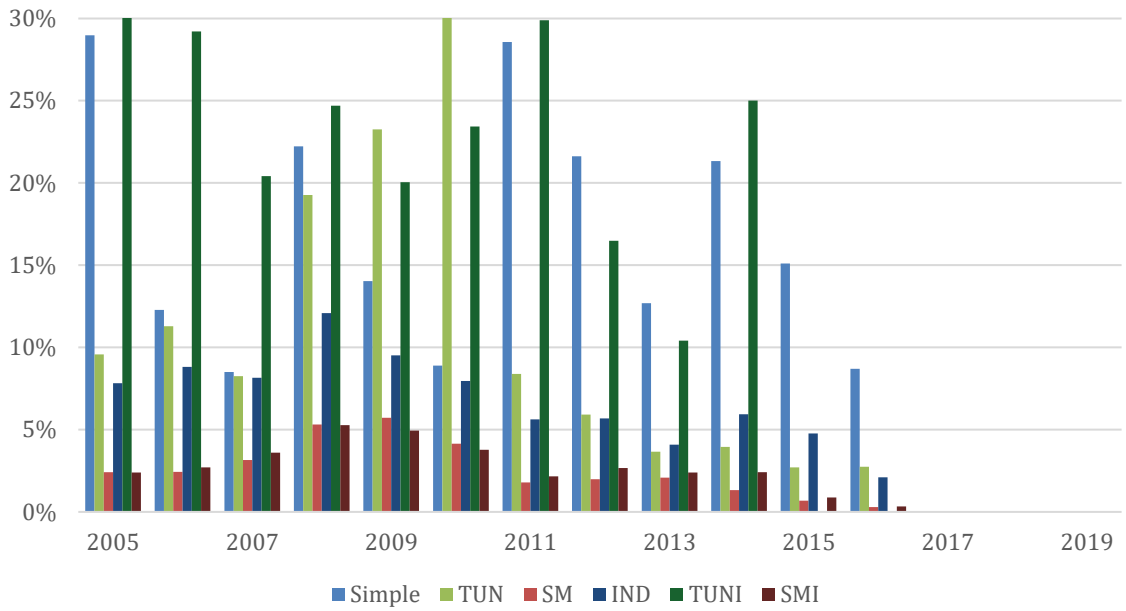


Figure 6.3 | F1-score throughout the predicted time period for Logit models
This figure represents the distribution of different Logit models' F1-scores throughout the predicted time period.

A frequently used technique to understand better the previous evaluation metrics is the confusion matrix. In Figure 6.4 the confusion matrices for the simple model, Logit SM, and Logit TUNI are displayed for the year of 2008, where the dataset had the highest number of bankruptcies following 2001[1], plus a perfect confusion matrix for that year. From the confusion matrices shown, one can recognise that if the most important characteristic in a model is detecting the most bankruptcies the best choice would be to oversample the training dataset, though this puts a toll on the Accuracy of the model possibly leading to an overestimation of lending rates applied to firms if the model would be used by banks. If instead Recall is not the most important factor but still a significant one in the equation, the best model to pick is the tuned hyperparameter variation with industry effects which led to predicting approximately 43% of the bankruptcies without lacking accuracy and with a more satisfactory precision level.

[1] Note that the number of bankruptcies in the test set of 2008 does not correspond to the number of bankruptcies in said year, but to the sum of the bankruptcies in 2006, 2007 and 2008 because of the way the data was structured.

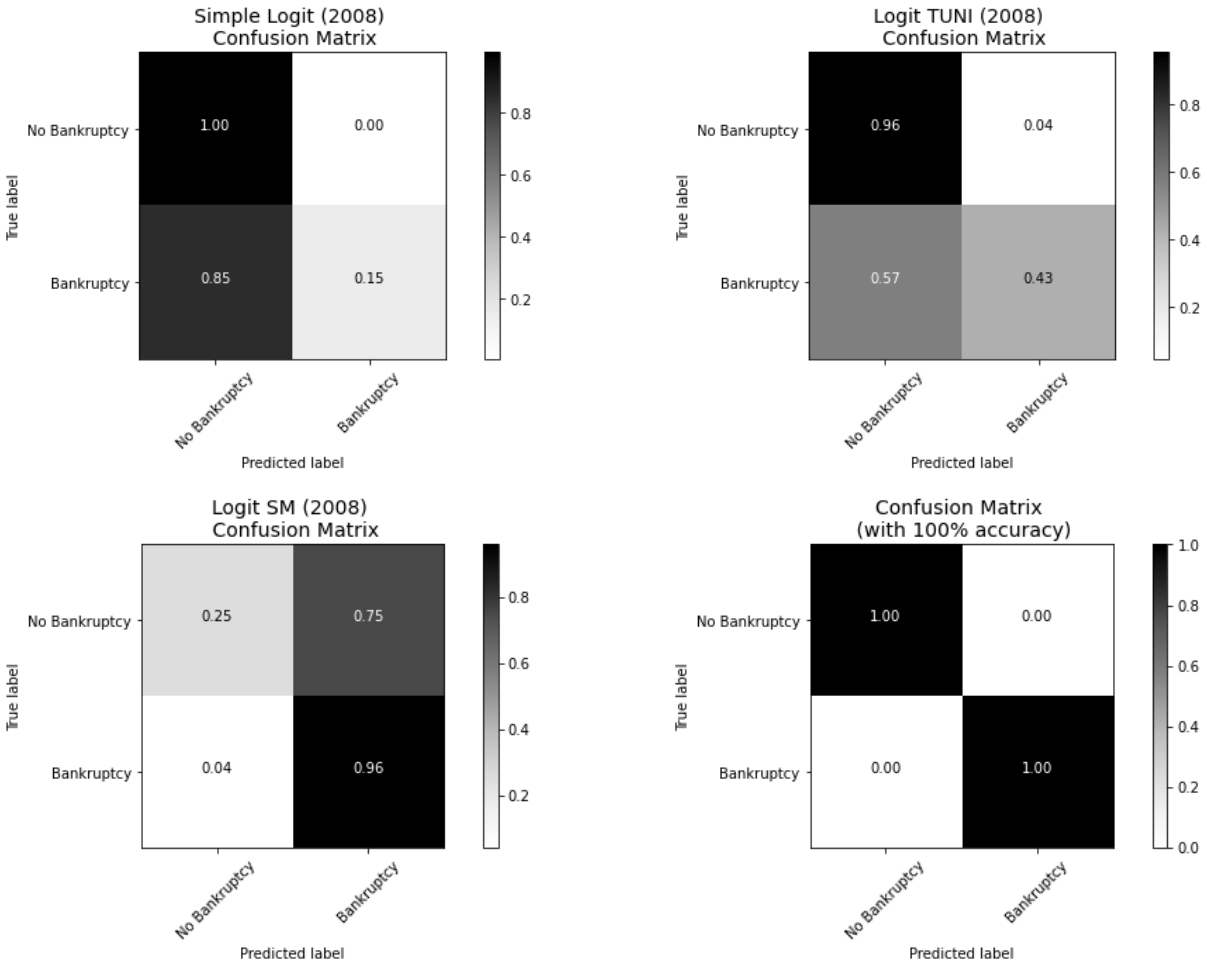


Figure 6.4 | Confusion matrices for Logit models vs. perfect confusion matrix
In this figure the confusion matrices for three different Logit models in 2008 are shown. It is observable that the Logit SM is the model with highest Recall (96%). (Total Non-Bankrupt: 2135, Total Bankrupt: 47)

6.3. Support Vector Machines (SVM)

The SVM model is completely ineffective (0% F1-score) without some sort of tuning or alteration, thus the only relevant results are the ones where an improvement was applied (with the exception of industry fixed effects alone which also had unsatisfactory results). Out of all SVM improvements, the tuning (which includes a weighted-class hyperparameter adjustment) with industry fixed effects was the variation with the highest F1-score and Precision. Though this model improvement was not the one with the highest Recall, meaning the highest number of true positives out of all possible positive cases. For the highest Recall, the oversampling (SMOTE) SVM without industry fixed effects was the best model. The Recall metric for all improved SVM models (except SVM IND) throughout the predicted time period is shown in Figure 6.5 compared to the Simple Logit model (benchmark).

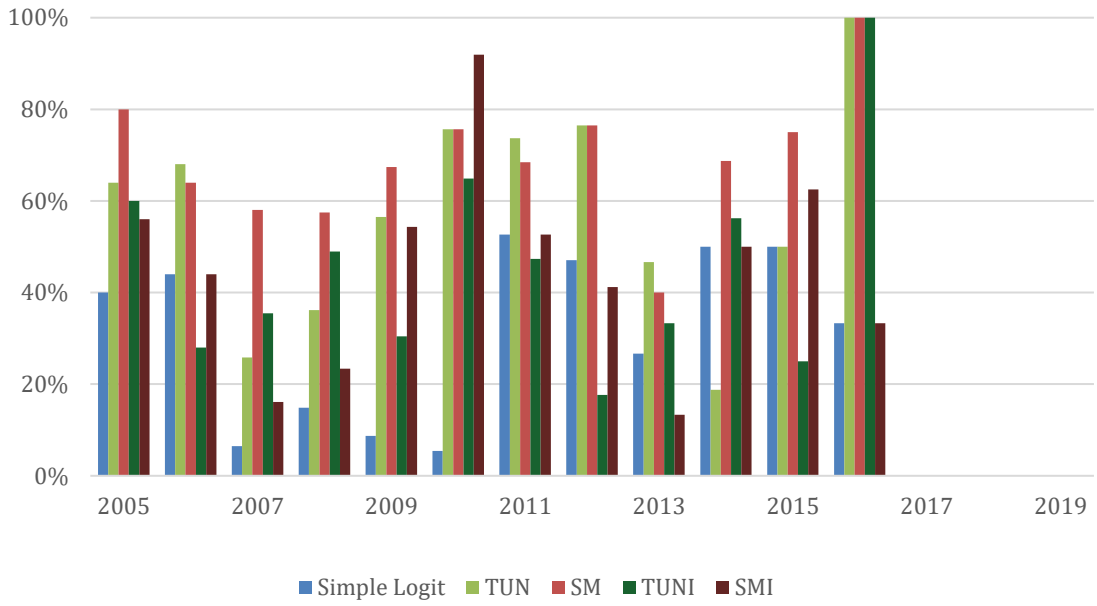


Figure 6.5 | Recall throughout the predicted time period for improved SVM models vs. Simple Logit
This figure represents the distribution of different SVM models' Recalls versus the Recall of the Simple Logit throughout the predicted time period.

As can be seen, all improved SVM models displayed outperform the Simple Logit until 2011 and the best model throughout the majority of the period in terms of true positives predicted among all possible positives, meaning Recall, is the oversampling improvement model without industry effects. Though, if one observes Figure 6.6, that displays the F1-scores for all improved SVM models (except SVM IND) throughout the predicted period, the conclusion again would be different. In terms of F1-score, the Simple Logit model outperforms the improved SVM models in the large majority of the periods though in some periods the hyperparameter tuned SVM models outperform the Simple Logit.

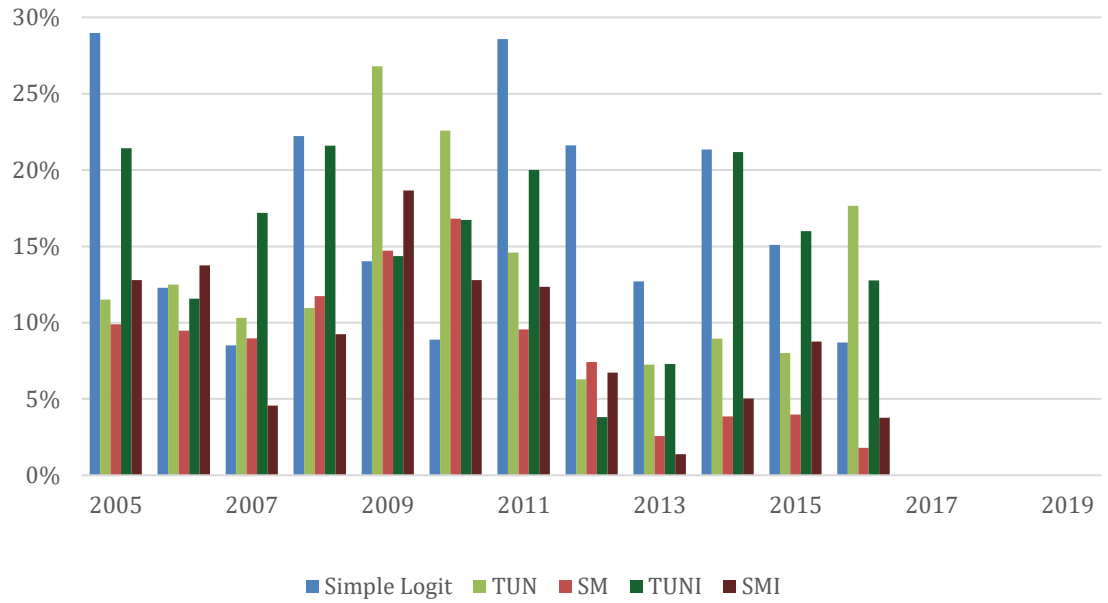


Figure 6.6 | F1-score throughout the predicted time period for improved SVM models vs. Simple Logit
This figure represents the distribution of different SVM models' F1-scores versus the F1-score of the Simple Logit throughout the predicted time period.

In Figure 6.7 the confusion matrices for the Simple Logit, the SVM TUNI and SVM SM for the year of 2008 are displayed alongside a perfect confusion matrix for the same year. From the matrices, one can observe that the two best SVM models outperform the Simple Logit model in terms of number of bankruptcies correctly predicted in the year of 2008. Furthermore, oversampling, as in the Logit models, is better at correctly predicting more bankruptcies while the tuned model is more precise. In terms of percentage of bankruptcies predicted out of all possible bankruptcies the SVM improvements do not outperform the Logit SM model for 2008.

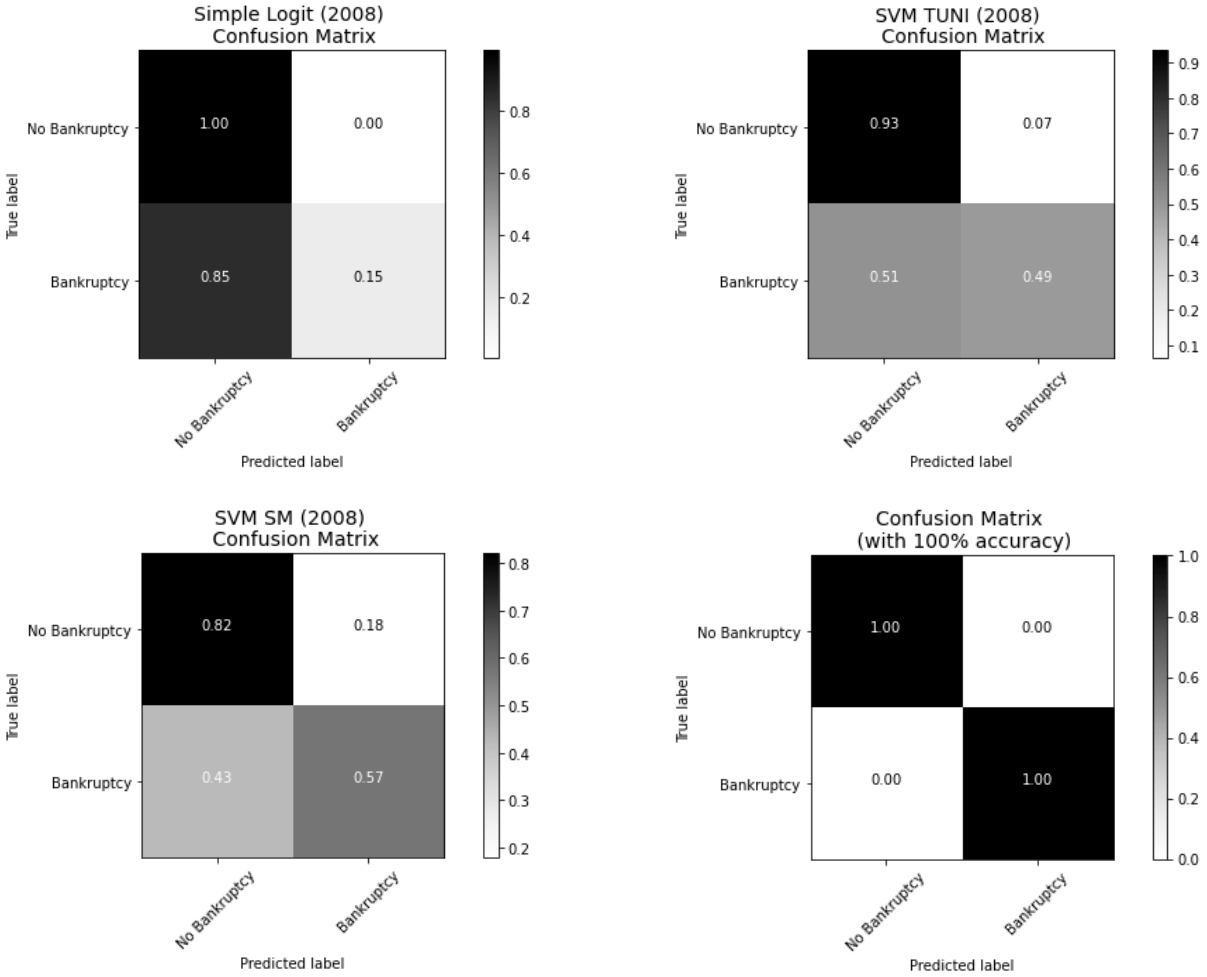


Figure 6.7 | Confusion matrices for improved SVM models vs. Simple Logit vs. perfect confusion matrix
In this figure the confusion matrices for two different SVM models in 2008 are shown. The SVM SM is the model with highest Recall (57%). (Total Non-Bankrupt: 2135, Total Bankrupt: 47)

6.4. Backpropagation Neural Networks (BPNN)

The simple BPNN model is unsatisfactory but not completely inefficient as the simple SVM model, though the model is significantly improved by applying the oversampling technique. Both with and without industry fixed effects this adjustment outperforms some of the previous models, though the best F1-score comes from applying the industry improvement. This, once more, does not imply a higher number of correctly predicted bankruptcies, it means in this case that the predictions are more accurate and precise. In Figure 6.8 the Recall metric for all improved BPNN models (except BPNN IND) throughout the predicted time period is shown compared to the Simple Logit model (benchmark).

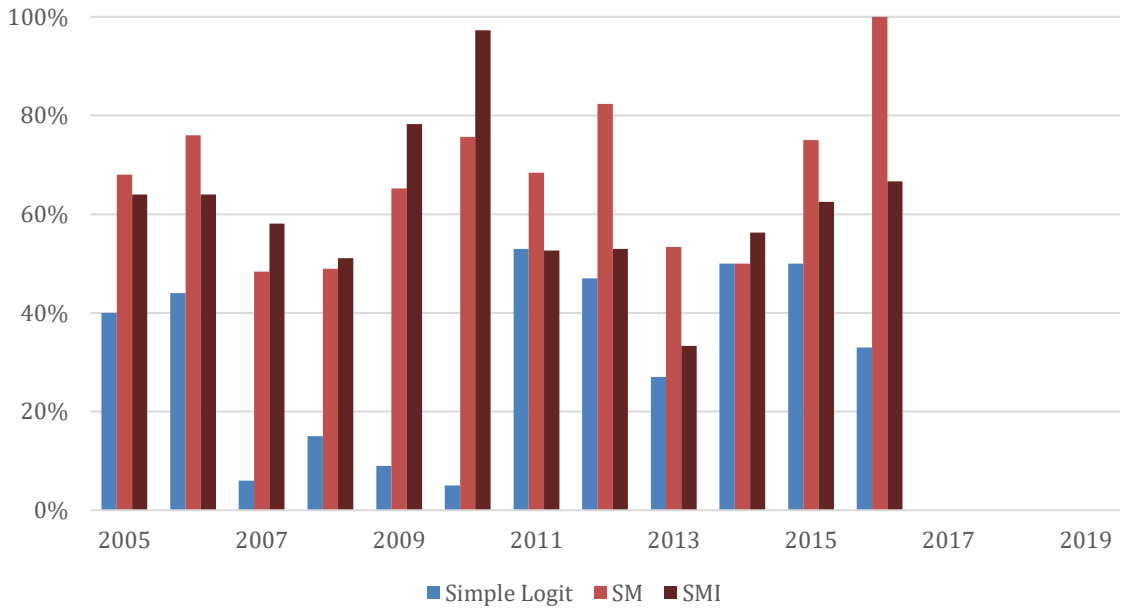


Figure 6.8 | Recall throughout the predicted time period for improved BPNN models vs. Simple Logit
This figure represents the distribution of different BPNN models' Recalls versus the Recall of the Simple Logit throughout the predicted time period.

It is shown in Figure 6.8 that both BPNN improved models outperform the Simple Logit model throughout the whole of the predicting period. The two oversampling models for BPNN are generally good but one outperforms the other in certain periods leading to think that different models might work better, for instance, in economically prosperous periods while others might work in periods where the economy is failing and possibly more bankruptcies exist. Additionally, it is important to evaluate the BPNN improved models in terms of F1-score and in Figure 6.9 those are displayed in comparison with the Simple Logit model. From the graph, it is not easily concluded which model performs better in terms of F1-score as in some periods the Simple Logit outperforms the others while in other periods one of the BPNN improved

models outperforms the rest. Though, on average the Simple Logit model outperforms the deep learning models with an F1-score of 13.53%.

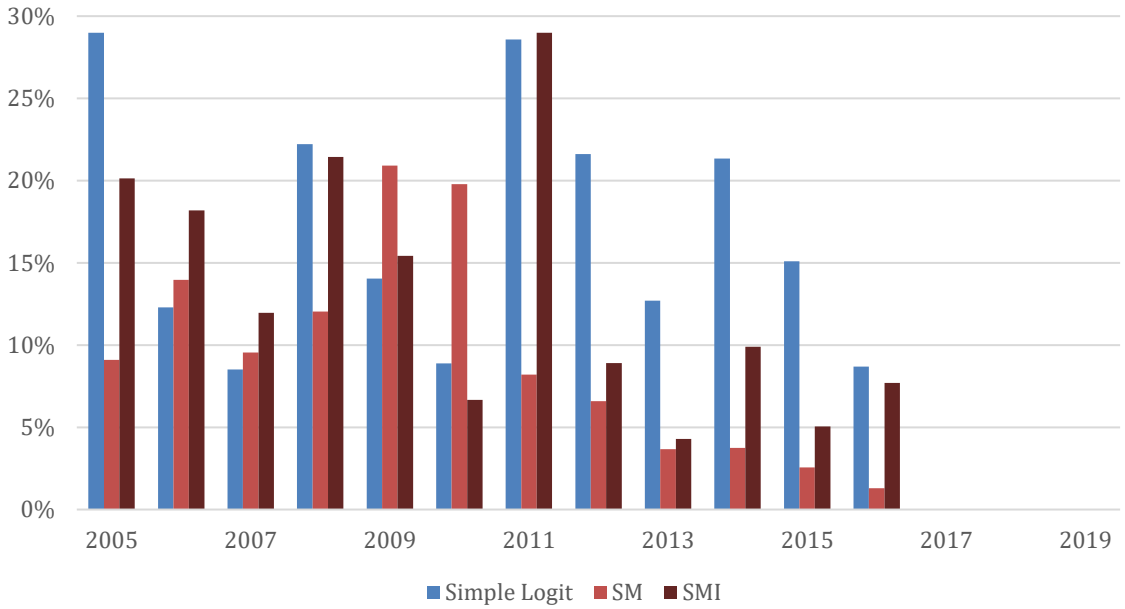


Figure 6.9 | F1-score throughout the predicted time period for improved BPNN models vs. Simple Logit
This figure represents the distribution of different BPNN models' F1-scores versus the F1-score of the Simple Logit throughout the predicted time period.

Similarly important to analyse, the confusion matrices for the improved BPNN models (except BPNN IND), the Simple Logit model and a perfect confusion matrix for the year of 2008 are shown in Figure 6.10. One can observe that both BPNN models outperform the Simple Logit in terms of number of bankruptcies correctly predicted for the year of 2008, though the model with industry effects is significantly more precise and thus would be preferred for this problem. The BPNN improvement models outperform the Logit and SVM TUNI models but are not better than their SM models, in terms of percentage of bankruptcies predicted in 2008.

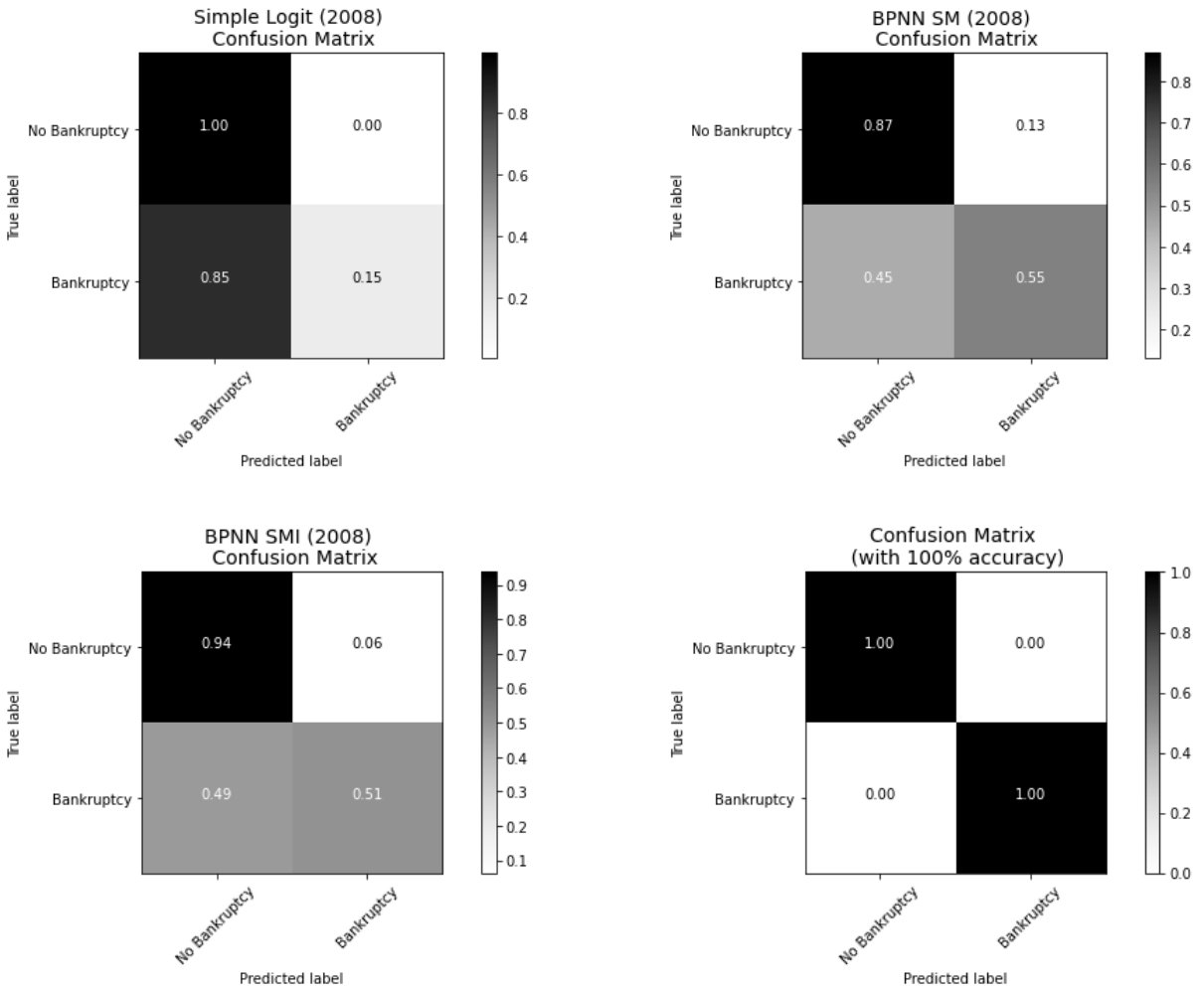


Figure 6.10 | Confusion matrices for improved BPNN models vs. Simple Logit vs. perfect confusion matrix
In this figure the confusion matrices for two different BPNN models in 2008 are shown. The BPNN SM is the model with highest Recall (55%). (Total Non-Bankrupt: 2135, Total Bankrupt: 47)

6.5. Other Results and Variable Predictive Power

As an attempt to improve the aforementioned models an alteration of the threshold was conducted, for models with high Recall and low Precision, the threshold was heightened and vice-versa. Despite the fact that some results were improved, they were not significantly better (only a small percentage increase was shown), thus they will not be displayed in this report.

With regard to the variable importance evaluation, some similarities were encountered throughout time periods and models. For both the Logit and the SVM models in the majority of the rolling windows, the variable with the highest predictive power was the PRICE variable for the year previous to bankruptcy prediction. For the BPNN, there is no clear variable throughout time that diminishes the F1-score to the minimum consistently, though it is clear that at least one of the Campbell et al. (2008) market variables (SIGMA, EXRET, RSIZE or PRICE) was, in the majority of the periods, the one that had the highest predictive power. In sum, in all the three types of model market variables played a significant role in the prediction of corporate bankruptcy. In the Appendix are displayed a table with the variables with the highest predictive power throughout time for the Logit model and, for the machine learning models, grids of the changes in F1-score based on each variable data shuffling.

7. Conclusion

With the purpose of significantly contributing to the already extensive literature of corporate bankruptcy prediction this research had the objective to first and foremost understand if machine learning models and techniques would help improve the results of forecasting models. More specifically it aimed at answering the following three questions: (1) if machine learning models displayed superior results to previously researched models not only in terms of accuracy, which as previously explained is not the best metric for imbalanced datasets but also in terms of actual number of bankruptcies predicted and the precision with which this is conducted; (2) if by using the market variables from Campbell et al. (2008) in machine learning models and techniques one could improve on the basic logit methodology used by the author; (3) finally, if using industry analysis one could improve the results of the models tested.

The first and second question can be answered jointly, as the analysis was conducted fluidly with both questions in mind. The results for the simple machine learning models were unsatisfactory and only displayed improvements when tuning or oversampling techniques were applied. Though one must be aware of the objective and the application of the analysis. If one only cares about correctly predicting as many bankruptcies as possible and does not give importance to how precise these predictions are, the model that displayed highest the Recall value was the Logit based on the Campbell et al. (2008) research with the machine learning technique of oversampling applied. Though if one cares both about Recall and Precision, meaning the number of bankruptcies correctly predicted and how efficient this prediction was, one should take into account the F1-score metric. For this metric, the top three models analysed were the Logit TUNI model, the Simple Logit model, and the SVM model with tuned hyperparameters and industry effects. Though, one of the most balanced models to use in this corporate bankruptcy prediction problem might be the BPNN SMI as it not only displays satisfactory Recall but a high Precision as well. It is one of the models that displays better results for all metrics compared to the initial benchmark, the Altman Z-Score.

In regard to the third question, it is clear that adding industry effects improves on the Precision of the models, as all three models had improved results for this metric when industry effects were applied. Though it is not clear whether or not the overall results were improved with this analysis. The results for the number of correctly predicted bankruptcies out of all bankruptcies possible to predict, meaning Recall, are ambiguous, as for some models applying industry effects lead to better results while for others this was not the case. In terms of F1-score, the effect of the industry analysis was positive as each model was improved by using industry

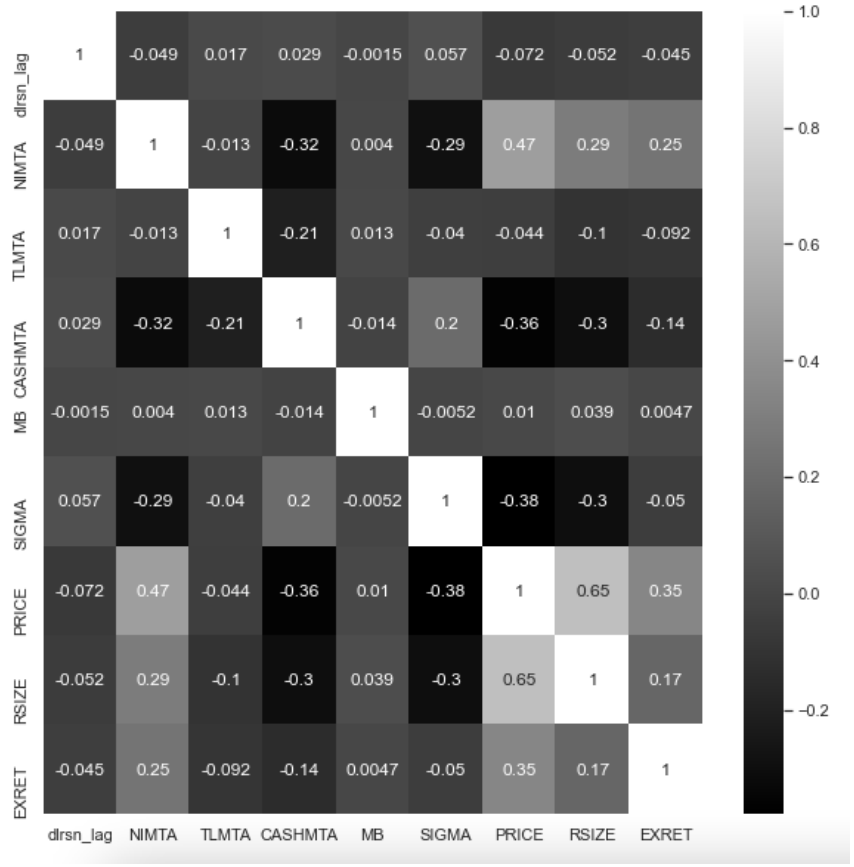
effects, as happened for Precision. Thus, depending once again on the goal of the implementation of the model using industry effects can be beneficial.

In order to understand further the processes of the models applied a variable importance evaluation was also conducted. It was clear from this assessment that the Campbell et al. (2008) market variables (SIGMA, RSIZE, EXRET and PRICE) play a significant role in the prediction of bankruptcy. It was also clear that for the Logit and SVM models the PRICE variable for the year prior to bankruptcy prediction was the most important variable throughout the time period analysed. For the BPNN model, it was only understandable that the market variables played a key role throughout time, but it was not visible that one overpowered the others consistently.

Finally, some limitations and suggestions for further research are important to point out. First, the use of variables limited to the ones from Campbell et al. (2008) is something that could be further researched in comparison with other variables that might lead to better results. Second, the usage of only yearly data might have limited the study and the possible improvement of results using more frequent data should also be used in further research. Thirdly, the usage of only U.S. data means that these results cannot be extrapolated to other geographies, thus the direct use of these techniques with data from other geographical areas is not advised and should be researched first. Fourthly, the analysis was conducted on a rolling window basis and thus an expanding window basis could also be tested as it might carry better results. Finally, further tuning to the machine learning models could be conducted though always bearing in mind the risk of the overfitting of the data.

8. Appendix

Appendix 1 | Correlation Matrix for the independent variables and the dependent variable ("dlrsn_lag")



Appendix 2 | Variable Importance for Simple Logit model

Year	Variable
2005	PRICE_t+1
2006	PRICE_t+1
2007	PRICE_t+1
2008	PRICE_t+1
2009	RSIZE_t+1
2010	TLMTA_t
2011	PRICE_t+1
2012	PRICE_t+1
2013	PRICE_t+1
2014	PRICE_t+1
2015	RSIZE_t+1
2016	RSIZE_t+1
2017	RSIZE_t+1
2018	TLMTA_t+1
2019	RSIZE_t+1

Appendix 3 | Variable Importance for SVM TUNI model

F1-score	NIMTA_t	NIMTA_t+1	TLMTA_t	TLMTA_t+1	CASHMTA_t	CASHMTA_t+1	MB_t	MB_t+1	SIGMA_t	SIGMA_t+1	RSIZE_t	RSIZE_t+1	EXRET_t	EXRET_t+1	PRICE_t	PRICE_t+1
2005	11,36%	11,67%	10,94%	10,81%	10,98%	11,28%	10,91%	10,87%	11,19%	11,19%	11,63%	6,68%	6,68%	6,20%	6,35%	1,60%
2006	10,22%	8,84%	8,87%	10,40%	10,66%	9,02%	10,53%	10,22%	10,29%	9,82%	9,96%	8,02%	8,02%	6,95%	6,13%	0,57%
2007	14,37%	11,04%	11,11%	10,00%	7,45%	7,50%	5,10%	5,13%	5,16%	5,13%	3,33%	3,39%	3,43%	4,27%	3,80%	3,17%
2008	10,77%	9,49%	8,79%	9,49%	9,49%	10,00%	9,35%	9,40%	9,24%	9,57%	8,18%	6,83%	6,83%	6,22%	6,25%	5,05%
2009	21,15%	22,22%	22,91%	22,56%	20,00%	15,15%	18,63%	17,65%	19,46%	19,90%	13,99%	11,49%	9,72%	8,30%	9,40%	7,31%
2010	20,44%	20,09%	19,82%	20,00%	19,91%	19,53%	19,44%	18,87%	16,19%	14,35%	8,25%	6,00%	2,08%	2,08%	2,06%	1,06%
2011	15,73%	16,00%	16,00%	16,28%	16,67%	16,37%	15,22%	13,83%	13,83%	14,05%	9,96%	8,87%	6,98%	4,29%	4,62%	2,70%
2012	6,54%	6,71%	6,70%	6,70%	6,75%	6,45%	6,25%	6,11%	5,92%	5,94%	5,98%	4,42%	4,39%	4,01%	4,04%	1,57%
2013	5,69%	6,63%	6,59%	4,23%	4,06%	3,94%	6,48%	5,48%	4,46%	4,39%	1,95%	4,71%	3,58%	2,09%	1,54%	0,75%
2014	14,29%	17,24%	16,67%	14,04%	12,70%	12,12%	21,21%	20,34%	13,33%	15,38%	3,96%	1,87%	1,85%	3,54%	1,96%	1,92%
2015	7,27%	7,84%	7,84%	7,41%	4,84%	5,26%	7,08%	7,14%	5,10%	5,23%	2,61%	0,00%	0,00%	0,00%	1,03%	0,00%
2016	15,00%	14,63%	10,00%	13,04%	10,00%	11,43%	10,81%	6,06%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2017	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2018	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2019	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%

Appendix 4 | Variable Importance for BPNN SMI model

F1-score	NIMTA_t	NIMTA_t+1	TLMTA_t	TLMTA_t+1	CASHMTA_t	CASHMTA_t+1	MB_t	MB_t+1	SIGMA_t	SIGMA_t+1	RSIZE_t	RSIZE_t+1	EXRET_t	EXRET_t+1	PRICE_t	PRICE_t+1
2005	9,65%	11,68%	8,49%	9,97%	10,00%	8,07%	10,09%	9,04%	10,81%	10,42%	11,08%	13,28%	9,22%	8,99%	9,44%	11,41%
2006	9,24%	11,72%	10,64%	10,64%	10,63%	10,69%	10,29%	11,07%	9,93%	10,91%	8,54%	7,87%	10,97%	10,60%	9,12%	11,27%
2007	10,19%	6,19%	8,40%	7,17%	7,98%	8,06%	9,00%	10,66%	7,51%	7,91%	8,64%	7,80%	7,95%	11,85%	11,28%	10,04%
2008	16,29%	13,25%	14,29%	12,78%	16,37%	12,69%	15,25%	17,42%	14,66%	16,72%	13,18%	16,34%	11,99%	13,33%	16,72%	15,86%
2009	19,61%	23,62%	19,05%	17,95%	19,01%	19,71%	20,13%	19,54%	21,51%	22,22%	15,77%	13,79%	15,43%	20,00%	19,61%	17,94%
2010	21,05%	23,36%	18,25%	21,26%	20,49%	17,83%	20,79%	18,82%	19,69%	20,30%	17,25%	23,14%	18,59%	18,06%	20,66%	25,23%
2011	8,28%	11,37%	8,48%	10,41%	8,24%	8,86%	9,59%	9,86%	11,11%	9,74%	9,81%	9,82%	10,41%	9,19%	8,75%	7,43%
2012	5,69%	7,89%	6,93%	7,79%	6,34%	5,82%	6,42%	8,82%	7,29%	6,65%	6,96%	6,25%	6,67%	6,12%	7,28%	5,41%
2013	4,31%	3,60%	2,92%	3,04%	4,16%	3,45%	2,86%	3,98%	3,08%	3,86%	3,98%	3,82%	4,84%	2,96%	5,34%	4,21%
2014	3,81%	3,37%	3,94%	3,25%	4,44%	4,76%	8,08%	3,95%	3,91%	4,65%	3,32%	3,16%	4,06%	2,93%	3,15%	5,57%
2015	2,51%	1,83%	1,81%	2,27%	1,86%	2,53%	2,56%	3,17%	1,59%	2,92%	2,37%	2,07%	1,91%	2,49%	1,95%	1,82%
2016	0,81%	1,40%	1,13%	1,02%	0,88%	0,87%	1,46%	1,00%	0,90%	1,32%	1,03%	0,36%	1,55%	1,36%	1,19%	1,36%
2017	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2018	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2019	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%

Appendix 5 | Data Delivery

All data and analyses are available at the following Github Repository: https://github.com/evamferreira/Thesis_Rep

It is divided in two folders: Data and Analysis. In the Data folder one can find the raw data, its treatment and exploration. In the Analysis folder one can find all the analysis conducted separated into four folders (the four types of model): Altman, Logit, SVM and BPNN.

9. References

Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, *32*(8), 1541–1551. https://doi.org/https://doi.org/10.1016/j.jbankfin.2007.07.014

Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, *34*(2), 78–86.

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

Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking and Finance*, *40*(1), 432–442. https://doi.org/10.1016/j.jbankfin.2013.12.013

Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, *4*, 71–111. http://www.jstor.org/stable/2490171

Begley, J., Ming, J., & Watts, S. (1996). Bankruptcy classification errors in the 1980s: an empirical analysis of Altman's and Ohlson's models. *Review of Accounting Studies*, *1*(4), 267–284. https://doi.org/10.1007/BF00570833

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate. *Journal of Political Economy*, *81*(3), 637–654.

Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *Journal of Finance*, *63*(6), 2899–2939. https://doi.org/10.1111/j.1540-6261.2008.01416.x

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, P. W. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1002/eap.2043

Géron, A. (2019). Hands-on Machine LeHands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systemsarning. In *O'Reilly Media, Inc.,*.

Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, *9*, 5–34. https://doi.org/10.1023/B:RAST.0000013627.90884.b7

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems*, *37*(4), 543–558. https://doi.org/10.1016/S0167-9236(03)00086-1

Lee, S., & Choi, W. S. (2013). A multi-industry bankruptcy prediction model using back-

propagation neural network and multivariate discriminant analysis. *Expert Systems with Applications*, *40*(8), 2941–2946. https://doi.org/10.1016/j.eswa.2012.12.009

Merton, R. C. (1974). On the Pricing of Corporate Debt : The Risk Structure of Interest Rates. *The Journal of Finance*, *29*(2), 449–470.

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

Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *IJCNN. International Joint Conference on Neural Networks*, 163–168. https://doi.org/10.1109/ijcnn.1990.137710

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, *18*(1), 109–131. https://www.jstor.org/stable/2490395

Shin, K. S., Lee, T. S., & Kim, H. J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, *28*(1), 127–135. https://doi.org/10.1016/j.eswa.2004.08.009

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

Sloan, R. G. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review*, *71*(3), 289–315.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial Applications of Neural Networks : The Case of Bank Failure Predictions. *Management Science*, *38*(7), 926–947.

Tsai, C. F., & Wu, J. W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, *34*(4), 2639–2649. https://doi.org/10.1016/j.eswa.2007.05.019

Vapnik, V. N. (1998). Statistical Learning Theory. In *New York: Springer*. https://doi.org/10.2307/1271368