School of Economics and Finance
Faculty of Commerce, Law and Management

A Machine Learning Approach to Corporate Bankruptcy Prediction Using BERT-Based Sentiment Analysis

Master of Commerce (50% Research) in Finance

Lwazi Lungile Mhlambi **[1145170]**

Supervisor: Prof. Yudhvir Seetharam

School of Economics and Finance

March 2023

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Faculty of Commerce, Law and Management

Candidate's Declaration

Name of Candidate:	Lwazi Lungile Mhlambi
Person Number:	1145170
Degree:	Master of Commerce
Supervisor / Co-Supervisor:	Professor Yudhvir Seetharam
School:	School of Economics and Finance
Title of Thesis/Dissertation/ Research Report:	A Machine Learning Approach to Corporate Bankruptcy Prediction Using BERT-Based Sentiment Analysis

Candidate's Declaration:

i. I hereby submit my PhD Thesis/Masters dissertation for examination (circle applicable).

ii. I confirm that my signed declaration in terms of Rule G9.7 is included in each copy of the thesis/ dissertation.

iii. I have checked all copies of my thesis/ dissertation and declare that no pages are missing or poorly reproduced.

iv. I hereby submit a CD/USB containing a PDF version of my submission for examination and nil bound copies.

Candidate's Signature: *L. Mhlambi* **Date:** 31 March 2023

Acknowledgements

"*Mhlambi!*
Hlohlwane,
Nsele eNduna.
Ligwa aliwelwa,
liwelwa zinkonjane.
Ngoba zona zindiza phezulu
Linda kaMkhonto
Zinyane lika Dubul' isbhamu
Elehla ngohologo
Livela eNseleni
Laxoshiswa ngeNkomo
Ngoba lihlabene
Dubul' isbhamu kaNsele
Owazala uNomasonto
uNomasonto wazala uMzuzephi
Wehla uMzuzephi waya kwela maNdebele
Wakhetha ingoduso ebukhosini bakwaMabena
uMbuduma! uNgcagu
Isigwegwe saboBingweni
uMasokasokile
Mbhuduma!
Mabena!
Isihlengehlenge sikaZihle
Ophahlwe uMhluzi neBhalule
Mabena!
uDubul' isibhamu
Onyoni odl' ezinye
Wazidla kwaduma abeSuthu
Bathi uHlohlwane
Ngoba
Siyahlohla
Ngapha siyalwa
Hlohlwane!
Mnyango kawuvalwa
Uvalwa ngamakhanda wamadoda!"

Abstract

The study of bankruptcy prediction has centred on whether firm level information is predictive. Seminal work by Altman (1968) articulates the failure of a business utilising its financial variables that are associated and classified in part to either the liquidity, profitability, solvency, leverage, or activity of a corporation. While this understanding is intuitive, recent studies have broadened the scope of financial ratios used in this prediction as well as incorporated exterior forces affecting the firm, either at an enterprise-wide or an economic-wide level to predict corporate bankruptcy. In the same breath, one cannot ignore the insider knowledge that the leaders and managers of firms would have leading to corporate bankruptcy. Therefore, this provides a curious opportunity in which we can incorporate the sentiment in the analysis provided by the leaders of such firms as an input in predicting the bankruptcy of a given firm. This study applies the Bidirectional Encoder Representations from Transformers (BERT) based sentiment analysis approach to import human sentiment as a variable from corporate disclosure data and apply it to existing corporate bankruptcy models over the period between 1995 to 2022 in South Africa, the United Kingdom and the United States of America.

Table of Contents

Acknowledgements 3
Abstract 4
1. **Introduction** 12
1.1. Background and overview 12
1.2. Motivation for the study 14
1.3. Knowledge gap 15
1.4. Research objective 16
1.5. Problem statement 17
1.6. Feasibility of the study 17
1.7. Hypothesis 18
1.7.1. Primary Hypothesis 18
2. **Literature Review** 20
2.1. Definition of Bankruptcy 20
2.2. Accounting Models Bankruptcy Models 20
2.3. Neural Networks for Bankruptcy Prediction 22
2.4. Text Classification 23
2.5. Summary of the Literature Review 26
3. **Description of data and research approach** 27
3.1. Detailed hypothesis and theoretical framework 27
3.2. Data 27
3.3. Variables 28
3.4. Limitations 31
4. **Results** 32
4.1 Altman Model Analysis 32
References 83

List of tables

Table 1: Mean comparison of South African Sample…….. 32

Table 2: Univariate results of the South African bankruptcy set…….. 33

Table 3: Mean comparison of USA Sample……. 36

Table 4: Univariate results of the American bankruptcy set……. 37

Table 5: Error Analysis of south African data…….. 63

List of figures

Figure 1: Altman Model Predictive Accuracy on the JSE on bankrupt years…… 34

Figure 2: Altman Model Predictive Accuracy on the JSE on non- bankrupt years…… 35

Figure 3: Altman Model Predictive Accuracy on the NYSE on bankrupt years…… 38

Figure 4: Altman Model Predictive Accuracy on the NYSE on non- bankrupt years……... 39

Figure 5: Altman Model Predictive Accuracy on the LSE on bankrupt years……... 40

Figure 6: Altman Model Predictive Accuracy on the LSE on non- bankrupt years…….. 41

Figure 7: Data Completeness Visualiser of the American attributes……. 42

Figure 8: Imputed Plot of our Training data……... 43

Figure 9: Feature correlation heatmap for the American dataset……... 44

Figure 10: PCA results…….. 45

Figure 11: PCA results…….. 45

Figure 12: Box plot for the American dataset……... 46

Figure 13: Accuracy of Models over Bankruptcy Years……... 47

Figure 14: Accuracy of Models over Non- Bankruptcy Years……... 48

Figure 15: Data Completeness Visualiser of the British attributes. …….. 48

Figure 16: Feature Importance of attributes……...……... 49

Figure 17: Imputed Plot of our Training data ……...…… 50

Figure 18: Feature correlation heatmap for the American dataset…… 51

Figure 19: PCA results. ……... ……... 52

Figure 20: Box plot for the Altman British dataset. ……...……... 52

Figure 22: Box and whisker plot of attributes in the British dataset. …… 53

Figure 23: Accuracy of Models over Bankruptcy Years. ……...….. 54

Figure 24: Accuracy of Models over Non- Bankruptcy Years. …… 55

Figure 25: Data Completeness Visualiser of the South African attributes. …….. 56

Figure 26: Data Completeness Visualiser of the South African attributes. …….. 57

Figure 27: Feature correlation heatmap for the American dataset. ……... 58

Figure 28: PCA results. ……...…….. 59

Figure 29: Box and whisker of Altman models. ……...……. 59

Figure 30: Normal distribution of Altman variables in the South Africa. ……. 60

Figure 31: Normal distribution Altman variables pre- bankruptcy in the South Africa. ……61

Figure 32: Feature importance of the South Africa. …….....…….. 61

Figure 33: Accuracy of Models over Bankruptcy Years……...…… 62

Figure 34: Error Analysis South Africa. ……...…… 64

Figure 35: Accuracy of Models over Non- Bankruptcy Years. …… 65

Figure 36: Frequency of Words used in all Countries. …….....…… 66

Figure 37: Frequency of Words in Non Bankrupt Firms in All Countries……. 67

Figure 36: Word cloud in the South African bankrupt dataset……... 68

Figure 37: Word cloud in the South African non- bankrupt dataset……... 69

Figure 38: Word cloud in the British bankrupt dataset……...…… 70

Figure 39: Word cloud in the British non- bankrupt dataset……..... 70

Figure 40: Word cloud in the USA bankrupt dataset……...…….. 71

Figure 41: Word cloud in the USA non- bankrupt dataset……...….. 72

Figure 42: Combined Results of the American Dataset……...…….. 73

Figure 43: Combined Results of the United Kingdom Dataset…….. 75

Figure 44: Combined Results of the South African Dataset……...… 76

List of Variables

ANNs	Artificial Neural Networks
BERT	Bidirectional Encoder Representations from Transformers
BERTNEG	Negative Bidirectional Encoder Representations from Transformers Sentiment
BERTPOS	Positive Bidirectional Encoder Representations from Transformers Sentiment
BioBERT	Biology Bidirectional Encoder Representations from Transformers
BRUPT	Bankrupt
CBoW	Continuous Bag of Words
CNN	Convolutional Neural Networks
COVID-19	Corona Virus Disease 2019
DAPTNEG	Domain Adapted Negative Sentiment
DAPTPOS	Domain Adapted Positive Sentiment
DICTNEG	Dictionary Negative
DICTPOS	Dictionary Positive
EBIT	Earnings Before Interest and Tax
EBITDA	Earnings Before Interest, Tax, Depreciation and Amortization
EMNLP	Empirical Methods in Natural Language Processing
FIN	Financial Variable
FINBERT	Financial Based Bidirectional Encoder Representations from Transformers
GPT-2	Generative Pre-trained Transformer 2
JSE	Johannesburg Stock Exchange
kNN	K-Nearest Neighbour Method

KSB	KSB Annual Report
LASSO	Least Absolute Shrinkage and Selection Operator
LSE	London Stock Exchange
LSTM	Long Short Term Memory
MVE	Market Value of Equity to Total Liability
NCF	Normalised Cost of Failure
NN	Neural Network
NYSE	New York Stock Exchange
PCA	Principal Component Analysis
PELJ	Potchefstroom Electronic Law Journal
RE	Retained Earnings to Total Liabilities
RSA	Republic of South Africa
SALE	Sales Revenue to Total Assets
SME	Small to Medium Size Enterprises
SMOTE	Synthetic Minority Oversampling Technique
SVC	Support Vector Classifier
SVM	Support Vector Machines
Tf-Idf	Term Frequency–Inverse Document Frequency
UK	United Kingdom
UNCTAD	United Nations Conference on Trade and Development
USA	United States of America
W2VNEG	Word to Vector Negative Sentiment
W2VPOS	Word to Vector Positive Sentiment

WC	Working Capital to Total Assets
Word2Vec	Word to Vector
XG	Extreme Gradient
XGB	Extreme Gradient Boosting
Z Score	Standard Score of Standard Deviation

1. Introduction

1.1 Background and overview

The prediction of bankruptcy has been the keen interest of investors, government as well as researchers for many years. Bankruptcy at a firm level has a negative effect on stakeholders and at a wider level, it can exhibit a contagion effect on the economy as a whole, (Mainardes et al., 2020). For the purposes of this paper, bankruptcy refers to the legal procedure by which an organisation is no longer able to honour its financial obligations to its creditors and indicates this by formally filing to a legal body to surrender their estate for the partitioning of its creditors, (Mabe, 2019). This is different from defaulting which is seen as falling behind on payments thus not being able to make payments to creditors for only a certain period of time. This is because having the ability to correctly and adequately detect the early signs of involuntary company failure can provide a valuable opportunity to save that company from such a failure (Bisogno *et al.,* 2018). This, in turn, may help reduce the economic impact that may arise from that company's bankruptcy.

Originally, the study of bankruptcy prediction was developed using models which are based on accounting information of the company. These models date back to the 1930s and were largely dominated by ratio analysis. Studies such as FitzPatrick (1932), Smith and Winakor (1935) and Merwin (1942) were mostly based on univariate analysis and focused on using accounting variables to determine the bankruptcy of the firm. The most prominent study to come from this era was that of FitzPatrick (1932). The author found that in most cases, firms that went bankrupt displayed less favourable ratios than those that did not. This became the fundamental understanding as it supported the logic that "bad" firms that have "bad" characteristics are more likely to fail. Literature has evolved to provide adequate theories and formulae to help us understand the likelihood and stage of corporate bankruptcy.

Beaver (1966) used ratios of the firm and tested their predictive power in classifying firms as being bankrupt or not. In his study, the proportion of net income to debt, sales, and net asset value, as well as cash flows to total debt and assets, showed the highest predictive power of bankruptcy. This is important as this finding also suggested that using multivariate models could lead to an even higher level of predictive power. The first study to explore Beaver's (1966) suggestion was Altman (1968) who, through this inspiration, developed a 5-factor multivariate model to predict the bankruptcy of manufacturing firms. This model would go on to be known as the "Z Score" and

as theorised by Altman (1968), it was found to be predictive. It is important to note, however, that this predictive power would be the strongest (95% predictive) within a year of bankruptcy and the accuracy of this prediction would decline the further along the years the model had to predict. The second problem is that it was built and applied specifically for the manufacturing sector which makes it a victim of its dataset. The data used in developing this model severely constrained the applicability of this model across other industries.

Many developments have emerged in an attempt to reduce the constraints that lead to a lack of cross-industry applicability. Ohlson (1980) built a logit-based model which used less restrictive assumptions than Altman (1968) and was, as a result, more accurate in bankruptcy prediction. It also had an important advantage in including timing as an output of the model. This is because the prediction of bankruptcy tended to be strongest when the financials were released and weaker after pre-financials were released. Later studies, such as Zmijeswki's (1984) probit-based model and Karels and Prakesh's (1987) study, continued with the discriminant multivariate analysis approach. While these newer models were able to improve the predictability, their nature of being discriminate meant that they maintained most of the restrictive assumptions that Altman (1968) had to contend with. Furthermore, the discriminate analysis models had a requirement that the discriminating variables used in this analysis are jointly multivariate normal. This assumption is problematic as it is often violated when financial ratios are used (Ohlson, 1980). Karels and Prakesh (1987) found that this assumption of multivariate normality was essential to a point where any results that came without the inclusion of normality would be erroneous.

In comparison, neural networks are not subject to multivariate normality as they do not adopt the restrictive assumptions of discriminate analysis. Studies by Lacher *et al.,* (1995), Tam and Kiang (1992), and Wilson and Sharda (1994) found that artificial neural networks have a more accurate bankruptcy prediction than traditional statistical methods. Zhang *et al*., (1999) found that the reason for the superior accuracy is because neural networks are the only method that can estimate posterior probabilities when the underlying sample distributions are not known. The application of newly developed bankruptcy models such as those seen by Lacher *et al.,* (1995), Tam and Kiang (1992), and Wilson and Sharda (1994) were done using the actual financial ratios of the firm. Similarly, previous discrete analyses done by the aforementioned authors applied LASSO regressions to financial ratios. This is interesting as the majority of bankruptcy corporate disclosures do not include any financial information which means that the analysis of financial ratios at the point

of bankruptcy is not complete if it is not linked to the written opinions and analysis provided in corporate disclosures.

Li (2011) found that the behavioural patterns observed in textual analysis of public disclosures do not correlate to the information. Similar findings were found by Tetlock et al., (2008) and Mayew et al., (2015) however Li (2011) would go on to accept that the large body of existing research on textual analysis is very much dependent on the methodology used. The argument is that textual disclosures are a new data source and as a result would need a new methodology to correctly analyse these data.

1.2 Motivation for the study

There are multiple motivations for the study. The first motivation has less to do with the theoretical literature of bankruptcy and more to do with the social and economic impact that the bankruptcy of a firm has on our societies. This is because the operations of an organisation are seldomly contained within its bounds, instead, they flow through what is referred to as a supply chain (Park et al., 2020). The economic theorists, some as early as Ricardo (1817), speak to this structure of companies and countries as important as this is how they can specialise and outperform one another. The consequence of a corporation going bankrupt is that it results in a structural break in these value chains which involve many companies, stakeholders, and in some cases entire countries (Park et al., 2020). This economically violent break in supply chain often continues until the market reorganises itself and rebuilds the affected value chains. This happens faster in developed economies due to higher competition and more dynamic institutions that can absorb the effects of a corporate bankruptcy more efficiently, however it tends to become an acute blow in developing economies (Buckley, 2009). The result of this is loss of jobs, decreased economic growth, and the loss of any future growth that the now defunct firm would have contributed. Having the ability to predict such a bankruptcy could have an effect in assisting policymakers, legislators, managers of companies, and involved stakeholders to take steps in reducing, and if possible, circumvent the consequences of corporate bankruptcy.

Studies such as Tinoco and Wilson (2018) have shown that the accounting and market-based models used to predict bankruptcy have fallen short in predicting company failure either at extended time frames or within different industries. The fact that textual analysis gained ground did not do much to assist in predicting corporate bankruptcy as seen by Li (2011). The author stated that much

of what has been textual analysis is dependent on the methodology used. Therefore, this raises a question as to whether it is not possible to use the successful methodological advances in natural language processing as well as sentiment analysis to unlock deeply hidden patterns in human behaviour in line with public corporate disclosures. It is important to note that these new methodologies emerged from a renewed focus on artificial intelligence and natural language processing driven by the current era of digitalisation (Piotrowski, 2012). The reason for this study's devotion with sentiment analysis as a factor in predicting bankruptcy is driven mostly by the vast potential information that sits within public disclosure data which has often gone unused. This is the same realisation that our peers interested in the field of accounting literature have found in that the manager's opinion and tone in Management Discussion and Analysis (MD&A) filings had significant explanatory power in the going concern of a firm (Mayew et al., 2015). From this, we cannot ignore corporate disclosures as a useful data source especially given how limited financial and market-based analysis has been in predicting company failure.

Similarly, the advances in natural language processing and machine learning are only enriched by their applicability in different fields of expertise (Young et al., 2018). The growth in natural language processing and textual analysis has been driven mostly by the fields of computer science and linguistics which were then applied to the field of finance. Therefore, this helps us develop a further motivation to our study in which we aim to explore the success of having a more finance-vernacular designed application of natural language processing and textual analysis.

1.3 Knowledge gap

The common pattern with existing modern machine learning-based bankruptcy models is that they are applied to accounting and market-based models which are numerical in nature and well-structured in format. With the growth of public disclosure data and their digitised accessibility, the integration of these data into existing models has been a challenging interest. A reason for this is that the data used in the case of public disclosures are largely unformatted and unstructured. Another reason is that these data are generally non-numerical. This, therefore, means that extracting and quantifying the data into readable insight is just as much a challenge to a point where a unified output containing numerical and text data is meaningless (Lang and Stice – Lawrence, 2015).

The question may arise as to the value of these data and whether it justifies the effort that is needed in quantifying it. Campbell et al., (2014) found that the information value in firm disclosures is

closely linked to the risk of the firm. This is seen in their finding that the number of risk factors declared in public disclosures by managers is proportional to the risk level of a firm. Similarly, the research paper found that market participants included factors presented in public disclosure into their risk assessment as well as the stock price which further creates the link between public disclosure data and the risk of the firm. The same was seen in Lauran and McDonald (2011) in which the authors used a negative word count and were able to successfully link it to numerical factors such as returns, the volume of trade and earnings, to name just a few. These factors form part of the very accounting and market-based numerical data that historical bankruptcy models are built on.

With all the above considered, the gap in knowledge in bankruptcy prediction sits where existing accounting and market-based models intersect with textual analysis techniques. It is seen that most of the informational value in public disclosure models has the value that can be linked to firm risk and numerical data which characterises the risk of a firm. Consequently, this justifies the need for a unifying model that will be able to utilise structured market and accounting-based data as well as jointly provide insight with unstructured and non-numerical public disclosure data. This framework would be the first to be applied on the JSE as well as the first to compare the framework across different financial markets with different textual approaches.

1.4 Research objective

The purpose of this study is to integrate the observation of existing accounting and market-based data with public disclosure data using the BERT-based model for the prediction of firm bankruptcy. This application will be done on the JSE, NYSE, and LSE. This approach provides the link that is so necessarily needed to assess if we can provide a quantifiable insight that can be linked to financial data. We choose these markets based on a paper by Wako (2021) in which it was found that the United States of America and United Kingdom are the first and second biggest contributors of foreign investment in Africa respectively. Similarly, South Africa is the fourth biggest contributor of foreign direct investment in Africa and simultaneously, the biggest recipient of that foreign direct investment in Africa (UNCTAD, 2022). Thus, the capital investment patterns, economic cycles that underpin this investment and the subsequent business decisions that are made in all these countries are inherently linked. This allows for an analysis of bankruptcy that will be

contextualised by international financial alliances and provide the possibility for cross- country analysis as well.

1.5 Problem statement

Previous studies by Singh and Mishra (2016) show that financial factors alone are inadequate at predicting bankruptcy especially at longer time horizons. It is clear that more creative factors need to be included in the analysis of bankruptcy. Thus, the explanation of bankruptcy could be explored through the inclusion of textual analysis data from public corporate disclosures. It follows that the sentiment within the data could be seen as a higher-order informational data point that could be used in predicting bankruptcy. Current literature uses a Dictionary-based (Loughran and McDonald, 2011) or Word2Vec approach to classify the sentiment of text (Mikolov et al., 2013). These models are relatively rudimentary and limited by scope and contextuality respectively. Newer Transformer-based language models such as BERT have been found to overcome these limitations due to their strong performance in generic applications (Devlin et al., 2018). This study explores the possibility of using the BERT based models on public disclosure data to use sentiment analysis as a factor in improving the prediction of bankruptcies.

1.6 Feasibility of the study

All the models used in this study have been developed and repeatedly tested by seminal papers such as Beaver (1966), Altman (1968) and more recent papers such as Hillegeist et al., (2004) in which the application of bankruptcy prediction tests were applied to accounting data. The same could be said for more market variables in which bankruptcy prediction tests were applied, Shumway (2001). The leap to BERT analysis is bridged by textual analysis and more specifically textual classification which has also been applied repeatedly in various studies such as Loughran and McDonald (2011) where a finance dictionary-based approach was used. We see this text classification again in Mikolov et al., (2013) in which the Word2Vec sentimental classification was used and lastly, we see it more recently used in Kim et al., (2014) where a Contextual Neural Network was used to classify the text. The BERT sentiment analysis model, therefore, is an evolved application of the above-mentioned models. The practical advantage of the BERT sentiment analysis model is that it has been tested on large text corpora and has been found to have reached a state of the art results (Devlin et al., 2018). The data for this study consist of listed stock exchange data from JSE, NYSE, and the LSE. The required market data and financial statement data will be

collected from the Iress database- a financial market database service. The bankruptcy model will be applied and simulated using statistical programming environments such as R in programming languages like python. The BERT machine learning model is pre-trained by Google thus reducing the time of execution of the study which ensures the feasibility of training the data and fine-tuning the model for sentiment analysis. This is done by adding a classification layer on top of the transformer output.

1.7 Hypothesis

1.7.1 Primary Hypothesis

Ho: The bankruptcy of a firm cannot be better predicted by the inclusion of sentiment from corporate public disclosures.

H1: The bankruptcy of a firm can be better predicted by the inclusion of sentiment from corporate public disclosures.

1.7.2 Secondary Hypothesis

Ho: The bankruptcy of a firm cannot be better predicted by the inclusion of sentiment from corporate public disclosures in different markets.

H1: The bankruptcy of a firm can be better predicted by the inclusion of sentiment from corporate public disclosures in different markets.

1.8 Potential benefits of the study

Firstly, the study has the benefit of deepening our understanding of sentiment analysis and its role in understanding the bankruptcy of a firm. It tests whether the textual data that is included in financial public disclosures is of any use in the context of bankruptcy prediction. This has a reach into other related fields in risk and risk management. Computationally, the study will also test the applicability of the BERT sentiment analysis model in the context of finance. This is important as it will provide the answer needed by the field of finance to develop models specific to its interest and whether it should more strongly collaborate with other fields of study or not. The study will test whether the bankruptcy models are applicable in three different countries and provide insights as to the global applicability of bankruptcy prediction using BERT sentiment analysis in major stock exchanges. This is important, especially when testing the robustness of the model and the results it provides. The study should provide an understanding of the different models and their effectiveness in predicting bankruptcy. For one, the question of comparison between pure

traditional methods such as the accounting methods and economic methods and that of mixed models of bankruptcy prediction could be further enlightened by the comparison done in this study. The study will also test the contextual classification ability of the BERT sentiment model and whether it can find insights in a text that is deeper in understanding than what a human being can achieve.

2 Literature Review

A literature review of the study of bankruptcy models dates back to the 1930s and was largely dominated by ratio analysis. These studies were mostly univariate in nature and focused on using accounting variables to determine the bankruptcy of the firm. The most prominent studies to come from this time were FitzPatrick (1932), who found that most of the times, firms that went bankrupt displayed less favourable ratios than those that did not. The two most successful ratios of this time were Net Worth to Debt as well as Net Profits to Net Worth.

2.1 Definition of Bankruptcy

To be able to categorise our study, one must develop a definition of bankruptcy. Fejér-Király (2015) defines bankruptcy as the legal action taken by a firm as a procedure of failing to operate. These are referred to as firms that are in Chapter 7 and 11 court proceedings in the United States of America (Lussier 1995; Baldwin et al., 1997). From this definition, one could link bankruptcy as the legal definition of financial distress (Doumpos & Zopounidis, 1999), and involuntary economic infeasibility (Cannon and Edmondson 2001) which in turn leads to losses to creditors (Lussier 1995; Baldwin et al., 1997).

2.2 Accounting Models Bankruptcy Models

The study of accounting-based models could be divided into three phases as Altman (1968), Ohlson (1980), and Zmijewski (1984) represent the advent of discriminant, logit, and probit-based models respectively. The first is that of Altman (1968) who found the first empirical link between the financial (accounting) characteristics of the firm and bankruptcy. He was able to provide a 95% accuracy of bankruptcy prediction within one year, however, this would drop and become more erratic as the number of years before the bankruptcy was increased. Another fault with this model is that it was only sampled on data constrained to the manufacturing sector thus the cross applicability of this model was rather questionable. Begley (1996) re-estimated the model using data from the 1980s and found the original models to be more predictive than the re-estimated models. Singh and Mishra (2016) found that the overall estimation of Altman's model improves when the coefficients are re-estimated. In the South African context, Muller et al., (2009) found that multiple discriminant analyses together with recursive partitioning tend to lead to higher predictive estimates of bankruptcy on companies listed on the Johannesburg Stock Exchange.

The limitation of discriminate analysis created space for Ohlson's (1980) logit-based model. In this case, instead of five variables, it is found that Ohlson (1980) used nine independent variables that were selected based on their frequency of use in bankruptcy prediction in the literature at the time. From this, Ohlson was able to identify four major factors relating to size, the measurement of financial structure, the performance as well as liquidity as being statistically significant for a sample of 2163 companies. The author was able to predict the bankruptcy of a firm within one-year prior at an accuracy of 95%. Proceeding studies that focused on re-estimating Ohlson's (1980) model found that the re-estimated models are higher in the prediction of bankruptcy than the original as would be seen by Zmijewski (1984), Boritz et al., (2007) and Abdullah (2008). In the South African context, Muller et al., (2009) found that the logit model had one of the best predictive powers of bankruptcy (at 84% accuracy), however it also had the highest normalised cost of failure (NCF) which has high-cost implications to institutional lenders as well financial investors. In the United States of America case, Timmermans (2014) applied the re-estimation of Ohlson to United States of America companies and like Begley (1996), he was able to find that the revised estimations are higher than those of the original models.

From the above two models, we find that in each of the cases, the models were developed on sample data that were non-randomly selected which in theory could result in biased estimations of bankruptcy. Zmijewski (1984) questions the role of oversampling and how it results in "sample-bias" as well as "choice-based sample biases" and whether this may have an effect on overall prediction results when data are not balanced. From the assessment done by Zmijewski (1984), we see that he eliminated the abovementioned biases and used a probit model with three variables to estimate the failure of a firm. Not much research has been done on the application of Zmijewski's (1984) model in South Africa, however, we can use other developing countries as a proxy for how it could be applied and what result it could have in the South African context. This idea has been taken from Oz and Simga-Mugan (2018) who tested a generalisable bankruptcy model on emerging economies. In their assessment, they looked at developing economies, including South Africa, along with Brazil, Hungary, and India to apply the re-estimated Zmijewski (1984) model. The results of this paper do not explicitly provide South African data however they provide emerging market insights which could be applicable in South Africa. The re-estimated Zmijewski (1984) model was found to have high prediction accuracy on both their original applications as well as the re-estimated models.

2.3 Neural Networks for Bankruptcy Prediction

Technological advancements in computation have made newer methods of bankruptcy prediction possible. The fundamental basis for the use of neural networks is taken from the problem that statistical techniques for bankruptcy model creation are done under strict assumptions. Methods such as probit and logistic regressions have been applicable under strict assumptions (Zmijewski, 1984). One such assumption is that of multivariate normality in which it is assumed that factors are normally distributed (Karels and Prakesh, 1987). This assumption of multivariate normality leads to erroneous results in which the residuals of the regression are treated as normally distributed. The probit, logistic as well as any other regression model relies on this strict assumption. It should be noted that this assumption is rather difficult to apply in real world applications as was found by Karels and Prakesh (1987). In these applications, they find that it is a condition that is seldomly met as variables tend to have one form of dependence on each other. Neural networks, on the other hand, are not borne out of these assumptions therefore they do not have this problem. Odom and Sharda (1990) and Tam and Kiang (1992) were able to establish that neural networks could be applied to bankruptcy prediction. Odom and Sharda (1990) did this by creating a neural network to define an organisation in which they monitored for bankruptcy. They did this by using the principal component analysis technique which is a technique for reducing dimensions of a large dataset into summarised information to extract key patterns from that larger dataset. This allowed them discover distress patterns from the neural structure they created. Tam and Kiang (1992) used neural networks to perform discriminant analysis on default data to classify firms as being bankrupt or not. Wilson and Sharda (1994) proposed a method of using neural networks with financial variables for bankruptcy prediction. This therefore created a path for alternative non-linear models that use neural networks in predicting bankruptcy which made such models more mainstream.

Studies exploring artificial neural networks (ANNs) have mixed results. One of the more popular mod is the Gradient Descent Algorithm. An example of this was seen by Lombardo (2022) in which the author used artificial neural networks to predict the bankruptcy using a dataset of 862 public companies. This method is an iterative first-order optimisation algorithm that is used to find the local minima and maximum of a given function. This however has a few limitations. The most prominent of these limitations is that the model can iterate itself into a local optimum which therefore leads to erroneous results and the other problem is that it converges towards an optimum at a relatively slow rate. This increases computing time which could be costly (Taylor et al., 2016). It

has been found as the most popular neural network model for bankruptcy prediction (Aziz and Dar, 2006). Hu and Tseng (2010), through their comparison of models, were able to find evidence to support that artificial neural networks can be stronger at predicting bankruptcy than the classic logit models.

Support Vector Machines (SVM) are a form of supervised learning algorithm that is used for classification and regression. The earliest study to use SVMs for the explicit purpose of predicting bankruptcy is Hang et al., (2004). In their study, they were able to find satisfactory results for the use of SVM models in predicting bankruptcy. In comparison to Altman's (1968) model, the SVM model was found to be more powerful than Altman's model and the best predictor among other machine learning models of firm failure because it had the highest predictive ability (Barboza et al., 2017). The same study was done by Joshi et al., (2018) as well as Gnip and Drotár (2019) and the researchers found the same result.

There is evidence that more modern classification models such as XGBoosts, and Random Forests generalised boosting (Jones, 2015). One such reason is that the more traditional machine learning approaches such as ANN and SVM are known as "black boxes", it means that they handle relationships between data sets in such a complex manner that it is not directly interpretable to humans. As a result, this makes the variables that affect the model's decisions impossible to understand (Jones and Wilson, 2016). When comparing the new age models, XGBoost is better than Random Forests as well as the generalised boosting method (which is a model that combines decisions from multiple decision trees to create one prediction) in bankruptcy prediction. It could predict 99% of the bankruptcies during COVID 19 in the data set of Narvekar and Guha (2021). A similar study by Perboli and Arabnezad (2021) in which the authors focused on short to medium-term bankruptcy prediction used XGBoost as well as random forest methods as part of their prediction of firm failure. They were able to have satisfactory results and because their data set included small to medium size enterprises over the COVID- 19 period, they were able to demonstrate the usefulness of their model for large-scale policy decision making.

2.4 Text Classification

Li (2010) found evidence that text analysis of financial reports holds key information that could be used in conjunction with financial results to better understand the underlying characteristics of

a corporation. Singh and Mishra (2016) and Tinoco and Wilson (2018) also support this notion that financial variables alone are not enough to explain corporate bankruptcy as well as firm performance. Therefore, the primary motivation of understanding text patterns as well as their informational value is due to the majority of annual reports and filings being done in South Africa, the United Kingdom as well as the United States of America having more written text than financial or numerical data.

Text classification models therefore grew from being a niche application of machine learning models and gained prominence over time as a resolution of exploring the informational value of text. The most common use of text classification methods in finance was proposed by Loughran and McDonald (2011) in which they proposed finance-specific corpora that could be used in building a stylised dictionary. This dictionary-based approach relied on creating a reference with keywords that would be used to associate and classify text. The logical limitation of this approach is that it becomes a victim of its own ability to cover the vast keywords, words, and contexts that could be linked to financing. Thus, sentences could be misclassified by this approach. One fix to this problem is using word embedding.

Word embedding is the process by which meaning is assigned to each word in a sentence (Kim and Yoon, 2021). Word embedding has two fundamental approaches. The first is frequency-based models such as Tf-Idf as proposed by Salton and Buckley (1998) with the other being the prediction-based Word2Vec model as proposed by Mikolov *et al.,* (2013). In the case of the Tf-Idf approach, it measures the relevancy of a word in a document with specific reference to a collection of documents. It does this by measuring the frequency as well as the inverse frequency of the document across a collection of documents (Salton and Buckley, 1998). The Word2Vec, on the other hand, places the word in a vector space that is approximate to its semantic relatedness. Naili, Chaibi, and Ghezela (2017) compared each model for topic segmentation and found Word2Vec to be more effective. Cahyani and Patasik (2021) compared Tf-Idf and Word2Vec to represent features in the classification of emotion in text. In this study, they paired the textual models with SVM as well as Multinomial Naïve Bayes methods (which are models that use Bayes' theorem to guess the tag text) to run the classification engine on data from Twitter. It was found from this study that the Tf-Idf classification, when paired with SVM, is superior to word2Vec in classifying emotion from the text.

An improvement of this model is to use convolutional neural networks (CNN) which use pre-trained Word2Vec as arbitrary input of the model. This model works as a sliding filter over the input and finds its result by dynamically comparing neighbouring data points (O'shea and Nash, 2015). Kim *et al.,* (2014) were able to prove that these CNN models paired Word2Vec can be used in classifying sentences using an arbitrary rule. Since CNN are in any case better for image processing and image classification (Yamashita et al., 2018) it is, therefore, no surprise that the major studies linking CNN to bankruptcy are focused on image processing. This is seen in the study by Hosaka (2018) who uses the image processing ability of CNN in classifying firms as being bankrupt or not. In this study, the author presents a set of financial ratios as a greyscale image where each financial ratio is a position on an image set. The author then uses the model's power in sorting image pixels to classify these ratios as being classified as bankrupt or not. This model was found to outperform the more conventional methods such as SVM in its prediction. This is because the author found that CNN models tend to perform better in image recognition applications thus, by converting these financial ratios to greyscale images, the author was able to leverage off of the improved performance of CNN models in image applications.

The cutting edge of text classification is based on Transformer models such as Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer 2 (GPT-2). BERT uses a method of masked language modelling in which it is forced to identify a word by context alone, as a result, it is defined by its surrounding words (Devlin *et al.,* 2018). BERT is therefore considered a bidirectional transformer language (Lin *et al.,* 2019). Stevenson et al., (2021) used the BERT model to predict bankruptcy by textually classifying South American SMEs using textual loan offer statements data. On the other hand, the GPT-2 model works by taking the word vectors as input and calculating the probability of the next word as outputs (Radford *et al.,* 2019). There are many differences between the above models, however, the one major difference between GPT-2 and BERT is that GPT-2 is unidirectional whereas BERT is bidirectional. BERT has been found to be successful in classifying numerical financial results as was found by Wang *et al.,* (2019) as well as being able to provide predictive sentiment of stock market data as was found by Hiew *et al.,* (2019).

2.5 Summary of the Literature Review

Literature has seen the development of bankruptcy models from the simple ratio-based models of Altman (1968) to the development of a model that can rid itself of biases such as Zmijewski (1984). These models were found to be effective in the JSE as much as they were in United States of America listed companies. The problem that came with classical accounting models is that they only made sense provided that the assumption of multivariate normality is maintained for the results of these classical models to hold. As such, neural networks are not built on these assumptions and would provide for more robust results of bankruptcy prediction that are not held back by the assumptions of statistical models. Tam and Kiang (1992) found that neural networks could be used in predicting bankruptcy. Artificial neural networks were found to be better in predicting bankruptcy in comparison to statistical models such as Altman (1968), Ohlson (1980), and Zmijewski (1984). Similarly, the SVM was found to be the most powerful model among all other machine learning models in predicting bankruptcy. Further, it was seen to be more powerful than Altman's model as well. New age machine learning models such as XGBoost, and random forests generalised boosting draw their superiority from higher computing power and larger data training sets. The biggest benefit they have, however, is that they can be controlled internally as opposed to the black box models whose relationships are so complex that they cannot be assessed or understood on a human level. This study will also require a form of unstructured textual analysis which started from the basic dictionary-based model to a more embedded approach in Word2Vec and Tf-Idf. The embedded models offer unique advantages in different scenarios however the Word2Vec model is better in classifying topics whereas the Tf Idf model is better in classifying emotions from word vectors. BERT is a transformer-based model that has been found to take advantage of Word2Vec in creating a bidirectional contextual-based assessment of word vector and has been found to have abilities in classifying the bankruptcy of firms in the South American SME market using loan offer statements as input data.

3 Description of data and research approach

3.1 Detailed hypothesis and theoretical framework

This study aims to predict bankruptcy by including the BERT classified textual analysis to our existing classical accounting models set by Altman (1968) and Maye (2015) to improve the predictive power of bankruptcy. We build the first part of the model by estimating five financial ratios that have been statistically selected by their sensitivity to bankruptcy. The study then uses BERT textual classification along with the SVM, the Hazard as well as the kNN model to be able to classify our textual data to develop sentiment analysis. This BERT classification model is then compared with the prior generation textual classification models of the Dictionary approach, Word2Vec, and Domain Adaption Methods.

3.2 Data

This study has a total of 12176 data points which includes bankrupt and non-bankrupt corporates. The sample in the South African portfolio consists of 1516 companies with 516 being delisted South African companies, 6324 United Kingdom based companies with 317 being delisted and a further 4336 United States of America based companies with 823 being delisted. In the same fashion of Altman (1984), we split the companies in each respective market in half as bankrupt and non-bankrupt firms of each market. The South African group firms were selected from a sample consisting of a mix of delisted and bankrupt firms. The aim of this is to see how successful the model is at separating the two. For the United Kingdom and United States of America based firms, the sample split consists of firms that have ceased operations and those that have a going concern.

We identify bankrupt firms from Iress as well as Bloomberg. In our analysis, we use the year with which a firm files for bankruptcy as the "bankruptcy years" for firms that are in the Johannesburg Securities Exchange (JSE), New York Stock Exchange (NYSE), and London Stock Exchange (LSE). Our sample period is from 1995 to 2022. This period is chosen as it was the dawn of the South African democracy which meant that many of the economic sanctions that would have exacerbated the bankruptcy of many firms in South Africa were relaxed (de Moraes, 2021). The two remaining exchanges also have data over this time period, enabling cross country comparison across all three exchanges. Further, we extract our textual data from the following MD&A sections from annual reports: 10- K, 10- KSB, 10- K405, and 10KSB40 sections of the annual reports for

the United States of America dataset. In the case of South African stocks, we will assess the Non-financial Disclosures of JSE-listed firms in which part of that disclosure has an MD&A section. For stocks listed on the London Stock Exchange, we focus on the Continuous Admission and Disclosure section of their annual financial reports.

3.3 Variables

We follow on by using Altman (1968), Muller et al., (2009) and Mayew (2015) in which we identify five financial variables that we will use for bankruptcy prediction. Therefore, the following variables and the definition therein are used:

- WC denotes working capital to total assets
- RE has retained earnings to total liabilities
- EBITDA is earnings before interest, tax, depreciation, and amortisation to total assets
- MVE is the market value of equity to the total liability
- SALE denotes sales revenue to total assets
- BRUPT denotes the indicator variable which equals one for observations that face bankruptcy within one year (365 days) from the issuance of the annual reports

3.4 Sentiment Analysis Models

3.4.1 Dictionary Based approach

Loughran and McDonald (2011) developed a dictionary with a list of words that were specific to 10-K filings and annual reports. To count the tone of annual disclosures we count the number of positive (DICTPOS) and negative words (DICTNEG) in each disclosure and scale them by number of words in each section of the reports. The dictionary they used is based on a collection of words associated with negative sentiment such as “failure” and “distress” as well as positive words such as “revenue” and “success”. These words are counted within the MD&A disclosure and the overall sentiment is calculated by the frequency in prevalence of either the negative or positive words. Other examples of dictionary-based approaches used in finance is by Kearney and Liu (2014)

found that disclosures, news articles as well as tweets and blogs are applied using field specific dictionaries. One such dictionary is linked to behavioural sciences where the sentiment is geared at understanding whether the observed document is indicative of strong or weak traits; active or passive traits or positive or negative traits.

3.4.2 Word2Vec

The study uses Word2Vec which is a prediction-based word embedding method which trains by predicting the context of a centre word from a continuous bag of words (CBoW). Once trained, each word is compared on a one-to-one level with the word that links to its semantic information. We use Word2Vec word trained on the 10-K filings corpus by Tsai et al., (2016) which was trained for filings between 1996 and 2013. We then use the trained financial sentiment dataset by Malo et al., (2014). The sentences are taken as input and the output is classified as either positive, negative or neutral. We calculate the sentiment of the documents by summing up all the probabilities of all the sentences in the document and normalise them to eventually have a final sentiment score: W2VPOS and W2VNEG.

3.4.3 BERT

We use the original BERT model as provided for by Devlin et al., (2018), however, we fine tune the weight of this transformer-based model FIN- BERT (Araci, 2019) which has been pretrained on financial text. The model being built with a financial corpus is meant to reduce on application times and create more accurate results. The model then takes each sentence as input and then assigns each as a probability of sentiment classified as either positive, negative or neutral. These probabilities are then summed up and normalised to find the final score of each document (BERTPOS and BERTNEG).

3.5 Classification Models

3.5.1 Hazard Model

We use the hazard model to test the accuracy of our prediction. We do this by applying the following logit regression to find the maximum likelihood estimation:

$$\log h_i(t) = \beta X_i(t) \quad (1)$$

where $h_i(t)$ refers to the risk of a firm i at time t. $X_i(t)$ denotes a vector of variables that are known to precede bankruptcies for firm i at time t. Financial variables and MD&A emotions are included in our research. We perform the regression with solely financial variables first (FIN). We then add dictionary-based variables (FIN, DICTPOS, and DICTNEG), Word2Vec-based variables (FIN, W2VPOS, and W2VNEG), BERT-based variables (FIN, BERTPOS, and BERTNEG), domain-adapted BERT-based variables (FIN, DAPTPOS, and DAPTNEG), and domain-adapted BERT-based variables (FIN, DAPTPOS, and DAPTNEG). We calculate the fitted values of log $h_i(t)$ (log-$dh_i(t)$) using the obtained coefficients and classify the observation as bankrupt if log$dh_i(t)$ > 0:5 and non-bankrupt using the obtained coefficients. To reduce the impact of outliers on regression findings, continuous variables are winsorised at a 1 percent level.

3.5.2 *k* Nearest Neighbours and Support Vector Machines

To further improve classification performance, we use the k-nearest neighbour method (kNN) and support vector machine (SVM) algorithms accordingly which are alternative models to linear regression. kNN is a simple nonparametric classification method. First, calculate the Euclidean distance between any two sets of observations. We compute this distance through the following expression:

$$\mathbf{d(X_1, X_2)} = \sqrt{\mathbf{X_1 \cdot X_2}} \quad \mathbf{(2)}$$

where · denotes the inner products between two vectors. In our study, vector $\mathbf{X_i}$ covers variables that occur before bankruptcy. Starting with five financial variables, we move on to sentiment variables derived using several sentiment analysis algorithms. The method then calculates the distances between observations $\mathbf{X_i}$ from the test set and the rest of the training set's observations. It then selects the *k* lowest ranges from observations $\mathbf{X_i}$ and labels them. We used k = 5 in our study.

If the number of bankrupt labels is larger than the number of non-bankrupt labels, the algorithm classifies $\mathbf{X_i}$ as bankrupt. Following that, SVM seeks the hyperplane with the greatest margin of separation that divides the dataset into distinct categories. Assume $\mathbf{X_i}$ is training data, and two classes are labelled with *yi* as +1 or -1, the following minimisation problem with respect to hyper-plane w is solved:

$$\mathbf{min}\,\frac{\mathbf{1}}{\mathbf{2}}||\boldsymbol{\omega}||^{\mathbf{2}} + \mathbf{C}\sum_{\mathbf{i=1}}^{\mathbf{M}} \boldsymbol{\varepsilon_i} \qquad \mathbf{(3)}$$

in which $y_i(\omega X_i + b) \geq 1 - \xi_i$. In this case, $\boldsymbol{\varepsilon_i}$ is a slack variable and C is a regularisation parameter. We set $C = 1e^{-5}$ in our configuration. In addition, we use a linear kernel for SVM classification. Linear kernels reduce computation costs but produce results that are less accurate. As the primary goal of our research is to compare the relative performance of text sentiment classification models, we chose the basic models of kNN and SVM.

3.4 Limitations

- The models are applied and trained mostly in English and thus the nuance that could come from other languages or translation can be minimised.

- The training data set is relatively sizable with more than 300 data points for prediction purposes and more than 1000 data points for training purposes. These models work best for the large training sets, thus the results can easily be influenced by the size of the training data set.

- The computational cost of the study is high thus we rely on outsourced pretrained models..

- The relationships that the models develop between variables become complex and not intuitive. This means that the results the model provide cannot be verified easily due to this phenomenon. This is the black box critique as mentioned by Petch et., (2022)

4 4. Results

4.1 Altman Model Analysis

4.1.1 South Africa Altman Model

The initial univariate study compares the mean to ensure that the chosen sample moves in accordance with BRUPT. Table 1 below looks at the average of each factor in the South African dataset and compares these averages with Bankrupt firms to those that have a going concern. From these results, we see similarities to the result by Altman (1968) in which it is found that the non-bankrupt firms have, on average, higher financial variables to bankrupt firms. This supports the suggestion introduced by Altman (1968) as well as prior papers by FitzPatrick (1932), Smith and Winakor (1935) and Merwin (1942) that firms with better variables perform better than those whose financial variables are comparably poorer. This in itself does not suggest bankruptcy, however, the univariate study does allow for a way to methodologically separate firms in terms of financial performance.

Table 1: Mean comparison of South African Sample

	BRUPT = 1 (1)	**BRUPT = 0 (2)**	**T- Stat (1) - (2)**
WC	-0.7543	-0.3277	-0.4265
RE	-2.7583	-0.9829	-1.7753
EBITDA	-2.1578	-1.1609	-0.9969
MVE	1.9712	24.5830	-22.6118
SALE	0.7783	0.9828	-0.2045

Table 2 looks at the univariate analysis of the factors within the South African financial data set which is important for understanding the initial relationships between the financial factors and bankruptcy. From the set, we note that EBITDA and RE factor have the highest standard deviation

of all of the variables under assessment which is a relationship that looks at the earning potential of a firm as well as the financial performance of the firm. As in the case of Altman (1968) both of these results tend to deviate and are no longer predictable in nature.

Table 2: Univariate results of the South African bankruptcy set

	WC	RE	EBITDA	Equity	MVE	SALE
Count	131.0000	128.0000	130.0000	1.3300e+02	129.0000	131.0000
Mean	-0.8016	-2.9038	-2.2761	3.0606e+05	1.6355	0.7499
Std	4.7262	10.5725	12.2789	1.1452e+06	5.2524	0.9051
Min	-37.2745	-87.9388	-124.0122	-1.0134e+06	-0.8258	0.0000
25%	-0.1302	-1.1851	-0.4975	4.4172e+03	0.0933	0.0939
50%	0.0110	-0.2522	-0.1249	3.0973e+04	0.4495	0.4231
75%	0.1187	0.0101	-0.0314	1.6453e+05	1.0436	1.1652
Max	0.5865	0.1850	0.3620	1.2078e+07	49.9824	3.9427

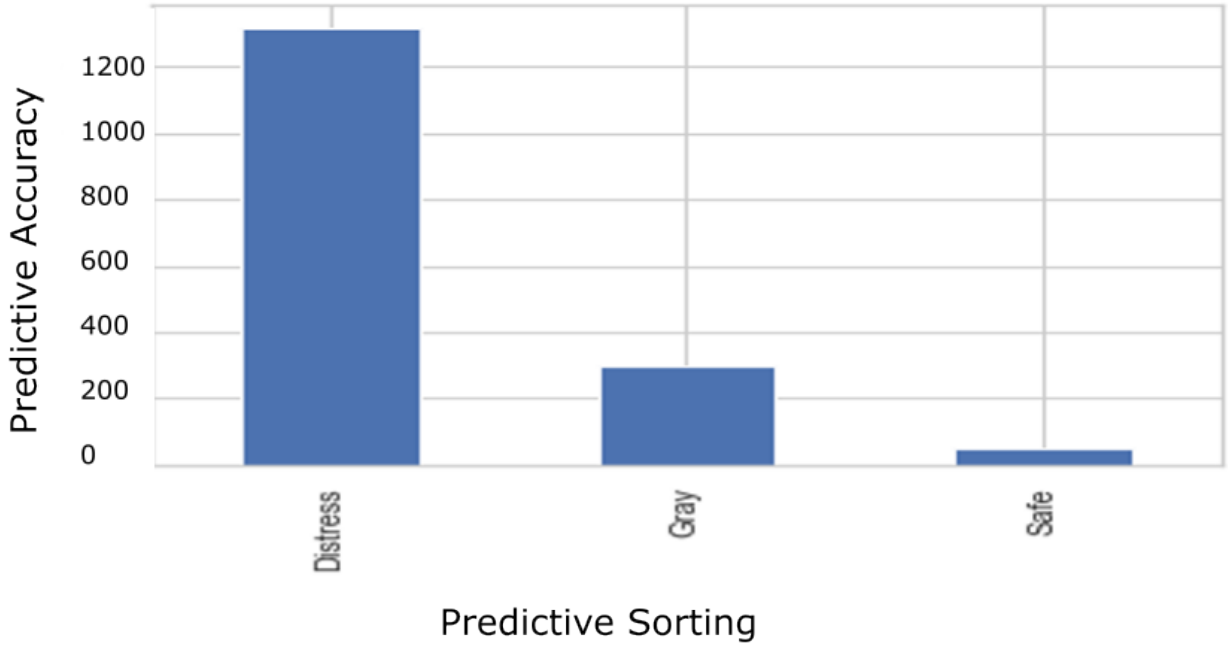
To get an understanding of the actual performance of the original Altman model on the JSE provided data, we look at its predictive accuracy in Figure 1 on a set of only bankrupt companies. We split the model into three categories, namely: Safe, Gray or Distress. These categories are dependent on the score created by the Altman based coefficients which were developed by estimation of firms and industries that survived bankruptcy which is given by the following relationship:

$$\boldsymbol{Z_{Score} = 1.2WC + 1.4RE + 3.3\,EBITDA + 0.6\,MVE + SALE} \qquad \textbf{(4)}$$

The split follows Altman (1968) where it is taken as all scores falling above 2.6 are categorically Safe. Those results that fall between 2.6 and 1.8 are taken as being in Gray and lastly, those results that fall below 1.8 are considered as being in Distress (bankrupt). The results from applying this model on the South African bankrupt-only subset was a predictability of 85% predictive accuracy which is the same result found by the application of the Altman model on the South African context

by Muller (2009). Given the dataset, the model predicted 12% of the dataset as falling in the Gray category and 2% of the data falling in the Safe category.

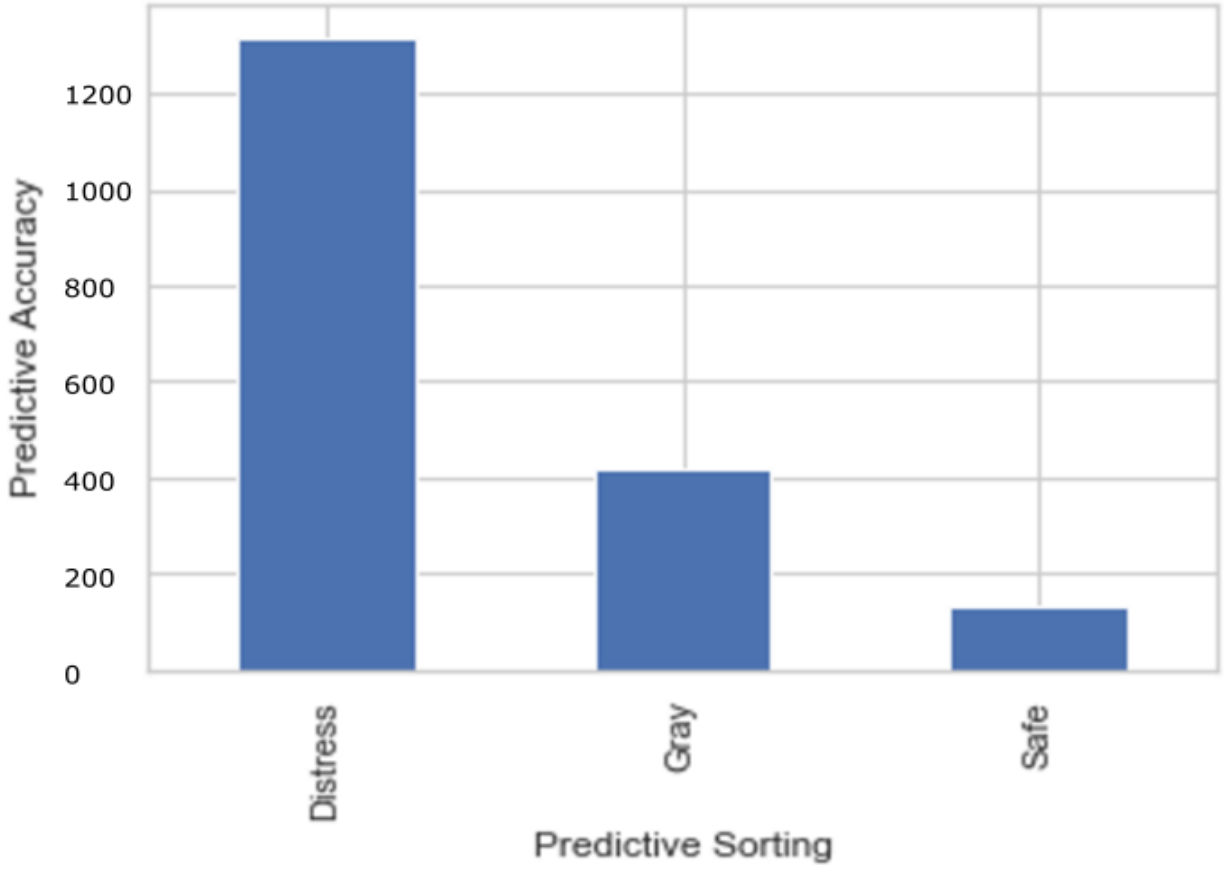
Figure 1: Altman Model Predictive Accuracy on the JSE on bankrupt years



We see that the performance of this model is further in line with Oz and Simga- Mugan (2018) where the Altman model was applied for emerging markets with South Africa being one of their test cases. The reason why the model is not as effective as it would have been with Altman (1968) can be explained best by Oz and Simga- Mugan (2018) who found that the factors that contribute towards bankruptcy tend to lie outside the informational value of financial factors. Instead, they present themselves as various forms of risk, for example, political or regulatory risk which is not expressed numerically. A great example of this can be taken from A.B.E Construction Chemicals which is one of the companies that form part of our observations in the sample. This company had a Z score of 1.53 however much of the reason as to why the company failed was because of changes in the regulatory environment which prohibiting making it hard for the firm to supply their biggest clients at a competitive price. This would not have been an issue the year before because it presented itself as an industry wide shock which was brought forward by the regulatory risk.

We extended the model to provide a comparison on the performance of the base Altman model on years that are non-bankrupt which refers to 2- 3 years prior to the company filing for bankruptcy. We compare the two results and we see the same deterioration in predictive power as seen by Altman (1968), Timmermans (2014) as well as Singh and Mishra (2016). The model was only able to capture 76% of the sample and mislabelling 18% as being Gray and 5% as being safe. This is mostly due to many of the firms having much higher financial performances years leading to their bankruptcy.

Figure 2: Altman Model Predictive Accuracy on the JSE on non- bankrupt years



4.1.2 United States of America Altman Model

The initial univariate study of the United States of America dataset also compares the mean to ensure that the chosen sample moves in accordance with BRUPT as set out by Altman (1968). Table 3 below looks at the mean of each Altman risk factor in the United States of America dataset

and compares these mean values between Bankrupt firms and those that have a going concern. From these results, we see similarities to the result found in the South African dataset as well as by Altman (1968) in which it is found that the non-bankrupt firms have, on average, higher financial variables to bankrupt firms and lower SALE. This supports the results shown by Altman (1968) in which the financial ratios of bankrupt firms exhibit the same characteristics of higher financial working capital, retained incomes assets, total liability ratios and a lower SALE. These results also support Smith and Winakor (1935) and Merwin (1942) that firms with better financials perform better than those whose financial variables are comparably poorer. This does not suggest bankruptcy however the univariate study does allow for a way to methodologically separate firms in terms of financial performance.

Table 3: Mean comparison of United States of America Sample

	BRUPT = 1 **(1)**	**BRUPT = 0** **(2)**	**T- Stat**
WC	-0.4378	0.2317	-0.2061
RE	-2.0032	0.0616	-2.0648
EBITDA	-1.7417	0.0731	-1.8173
MVE	1.2589	6.0448	-4.7859
SALE	1.9698	1.5526	0.4172

The following results in Table are the univariate analysis of our Altman risk factors. From this table, we again see the reoccurrence of the high standard deviation of the RE and MVE risk factors as would be attributed by the South African data as well as Altman. From the finding of Altman (1968) these factors tend to express themselves as being more unpredictable as the financial performance of a company declines in performance, which, when including these factors, could give you a sign of a company in financial distress. This results in the inclusion of these factors that are introduced as a means of capturing the financial distress; these factors, more than any other, are able to do so.

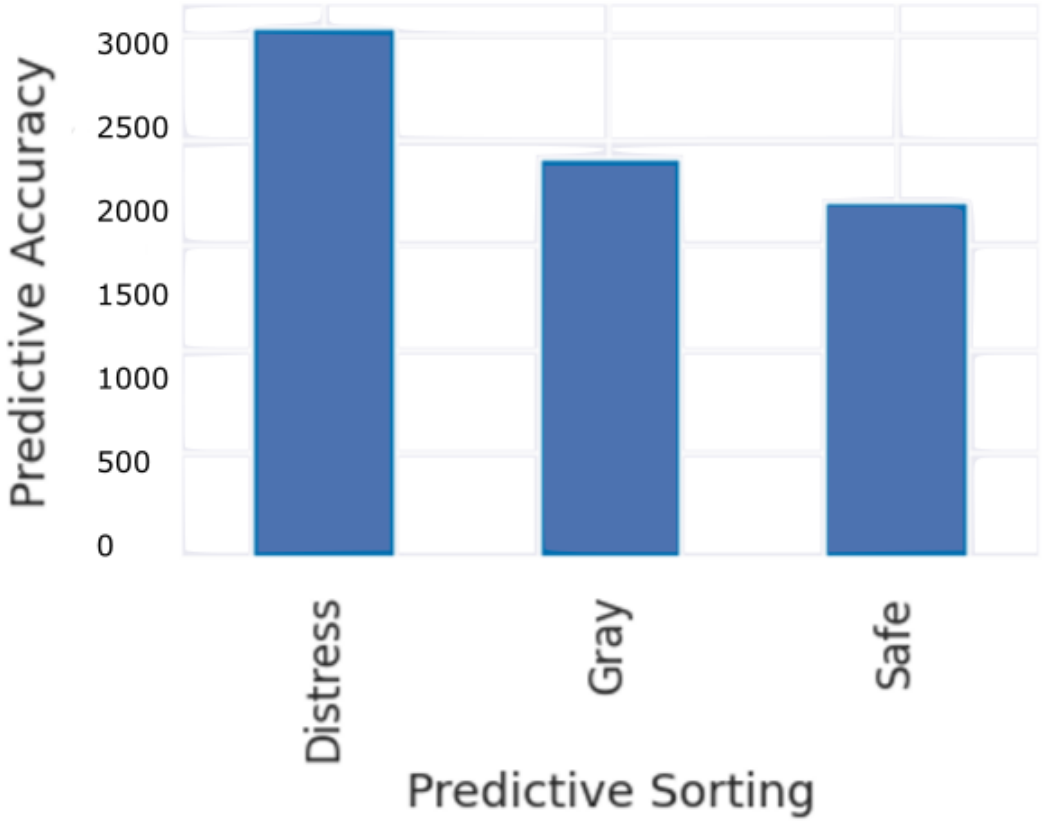
Table 4: Univariate results of the United States of America bankruptcy set

	WC	RE	EBITDA	MVE	SALE
count	4816.0000	4816.0000	4816.0000	4804.0000	4818.0000
mean	0.1889	-0.0700	-0.0425	5.7399	1.5792
std	1.2824	7.7785	6.7079	109.5080	1.3427
min	-72.0670	-463.8900	-463.8900	-3.7351	0.0001
25%	0.0449	0.0000	0.0059	0.4815	1.0156
50%	0.2185	0.0000	0.0566	1.1490	1.1405
75%	0.4201	0.1104	0.1360	2.7813	1.8140
max	28.3360	203.1500	2.3523	6868.5000	37.8070

We test the accuracy of the Altman model by extracting the firms that have a Bankrupt indicator of 1. That is, we create a set of firms that have filed for bankruptcy in the United States of America data set and we apply the Altman formula stated below to obtain our results. We then split this dataset into categories of Safe, Gray or Distress where Gray and Distress represent firms that are collectively performing poorly financially and Safe represents firms that are financially sound. The predictive accuracy of the model is represented in figure 3 below.

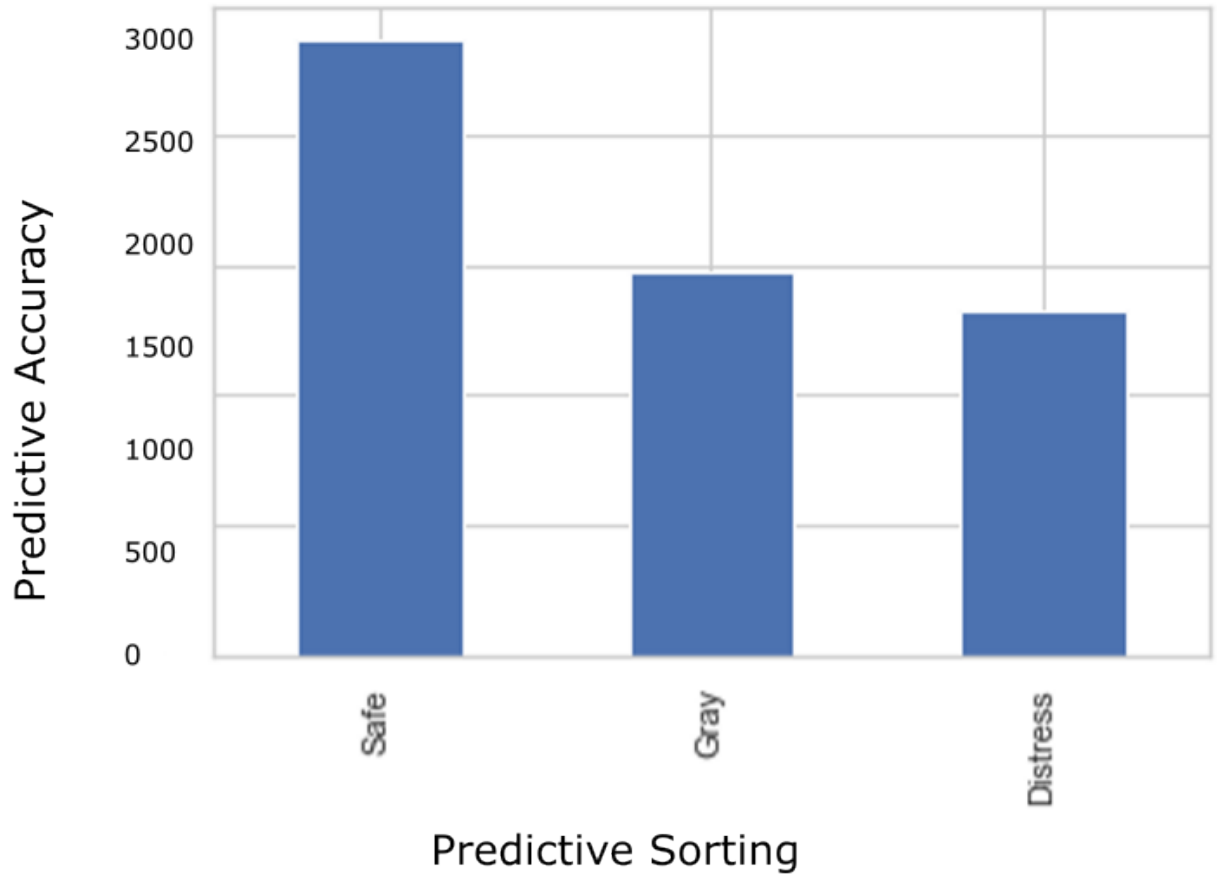
From our estimates, the model is 72% predictive with 223 out 308 companies being correctly classified. The model result has misclassified 28% of the data. These results are favourable to Timmermans (2014) who estimated the Altman model for a more contemporary data set of 2009 to 2014 and found the predictive range to be lower than 80%. An explanation for this was that some of the risk factors that are used by Altman (1968) deteriorated in importance over time due to changes in industry legislation, business practices as well as the formation of entirely new business industries where other factors gain more prominence in bankruptcy prediction.

Figure 3: Altman Model Predictive Accuracy on the NYSE on bankrupt years



As in the case of what was done for the South African data set, we extend our analysis of the model to provide a comparison on the performance of the base Altman model on years that are non-bankrupt which refers to years prior to the company filing for bankruptcy. We compare the two results and we see the same deterioration in predictive power as seen by Altman (1968), Timmermans (2014) as well as Singh and Mishra (2016). The model, in this case, has misclassified most of the data as Safe and only 20% of the data as correctly in distress. This has many implications in as far as placing a question on the Altman measure and its predictive power further along the bankruptcy years. This result is mostly supportive of works by Ohlson (1980) and Zmijewski (1984) who found the model to be predictive based on the sample data used as well as the period that the model is applied. This also supports the case put forward by Begley (1996) who found that the re-estimated model is poor at bankruptcy prediction one year prior to the result.

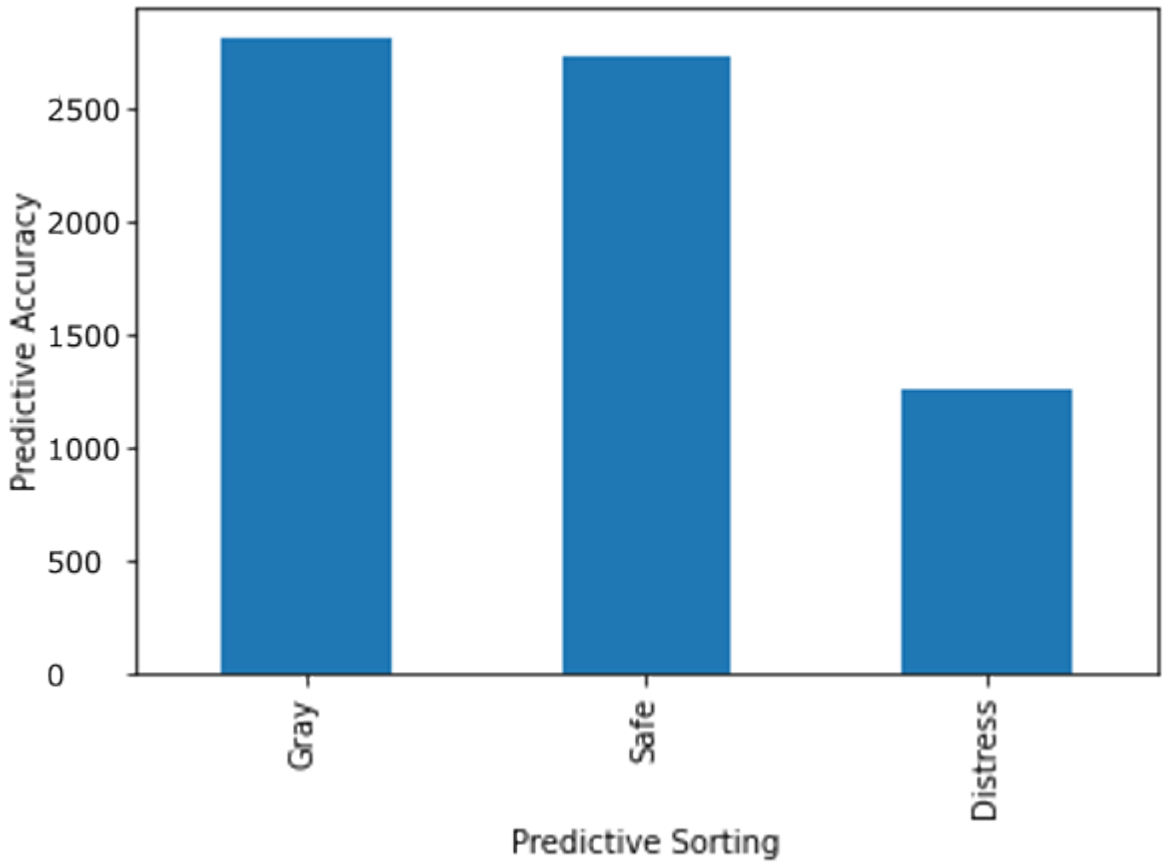
Figure 4: Altman Model Predictive Accuracy on the NYSE on non- bankrupt years



4.1.3 United Kingdom Altman Model

We further apply the Altman (1968) on the United Kingdom dataset and we obtain the results as expressed in Figure 5 and Figure 6. From the results, we see that the model is poorest in the British dataset as it only captures 60% of the data correctly and misclassifies 40% of the data during bankruptcy years. This is consistent with studies by Lacher *et al.,* (1995), Tam and Kiang (1992), and Wilson and Sharda (1994) who found that it is possible that the restrictive assumptions of Altman (1968) are prohibitive in capturing bankruptcy in different markets. Furthermore, we see that the model could possibly have sample bias as it was built based on United States of America data and that this model fails in other datasets. This was the same finding to the findings by Zmijewski (1984) whose question on oversampling results in "sample- bias" could find the model not applicable in different periods or different markets as seen by our results below.

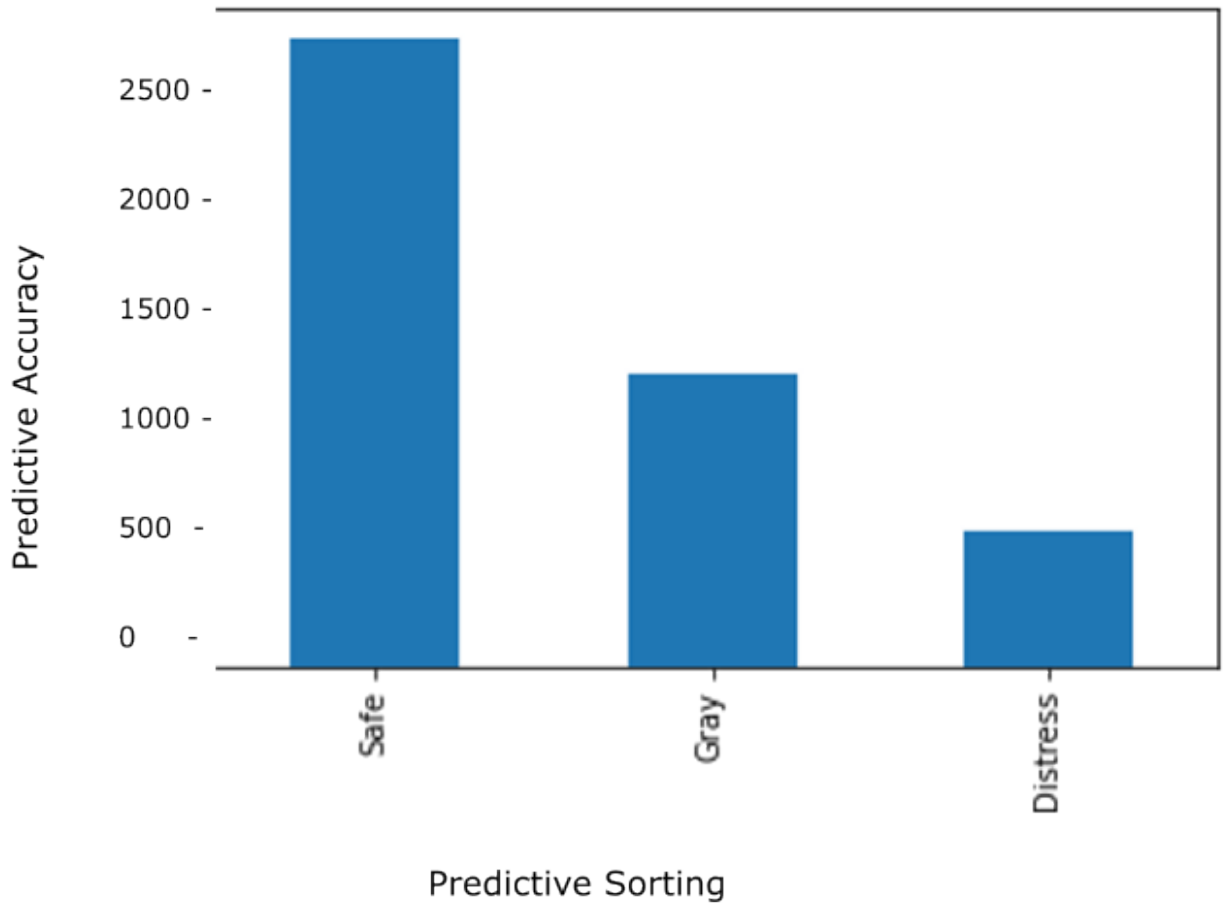
Figure 5: Altman Model Predictive Accuracy on the LSE on bankrupt years



The non-bankrupt data set shows that the model performs even worse in the non- bankrupt years. We see that it misclassified most of the data as Safe. We see the same deterioration in predictive power as seen by Altman (1968), Timmermans (2014) in the South African and United States of America database. As in the finding by Singh and Mishra (2016), the deterioration is probably much more pronounced due to sample bias however the re-estimation of the factors come with less feature importance. The model has misclassified most of the data as Safe at 60% and only 40% of the data as correctly in distress. This places a question on the Altman model and its predictive ability in the critical years outside of bankruptcy. From this result, we see once again the works by Ohlson (1980) and Zmijewski (1984) who found the evidence that the predictive power of the Atman (1968) model was based on the sample data used as well as the period that model was

applied. This, in effect, also echoes works by Begley (1996), the same deterioration in bankruptcy prediction goes for years further before the bankruptcy years.

Figure 6: Altman Model Predictive Accuracy on the LSE on non- bankrupt years



4.2 Machine Learning Model Analysis

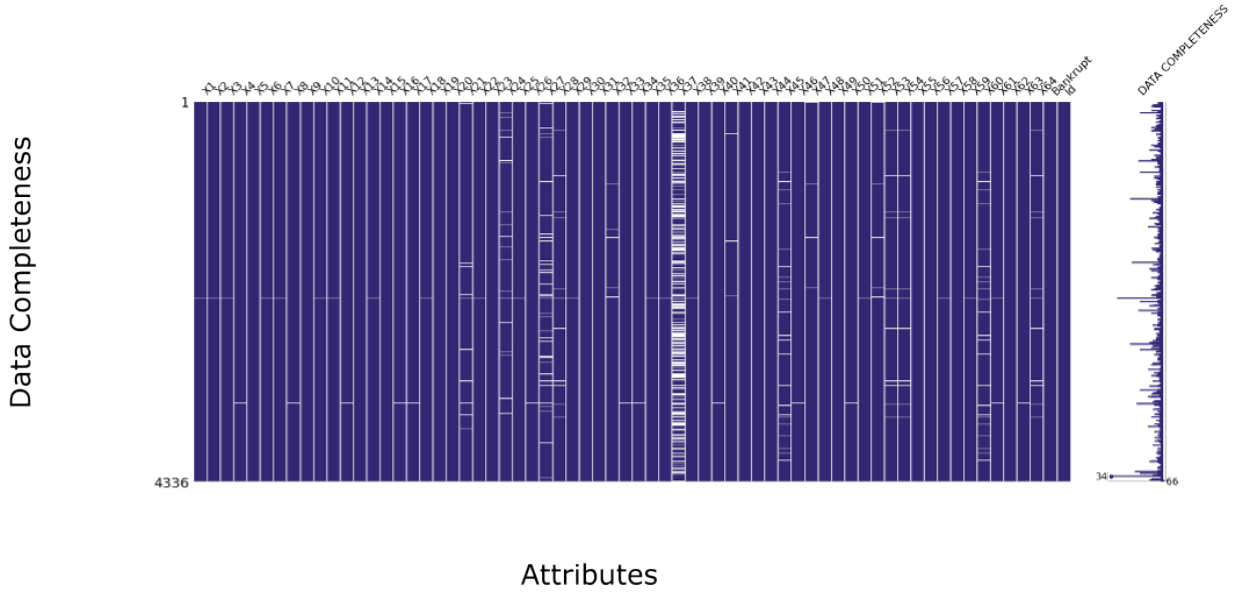
4.2.1 Machine Learning Models on United States of America data

4.2.1.1 Explore The Data

Before we present the results of the machine learning models on the United States of America dataset, we first have to gain an adequate understanding of the data used. The dataset consists of 64 attributes. Five of which were originally defined by Altman, however, the other ratios are included in order to enrich the study and to provide the models with enough data points. We refer to the dataset consisting of synthetic attributes which means that the models are made from mathematically calculated measures which use mathematical operations. Figure depicts the completeness

of our dataset. From the below illustration, the white gaps in each column represent missing data which is labelled as N/A or blank. We pay close attention to attribute X37 which is the current assets less inventories to long-term liabilities ratio. This aforementioned ratio has the most amount of data points missing which could be a problem as our machine learning models would not be able to adequately build meaningful results from these attributes. We consequently have to impute this data in order to replace the values of attribute. This is in line with the approach set out by Narvekar and Guha (2021) in the manner of how they treated attribute selection as well as feature importance.

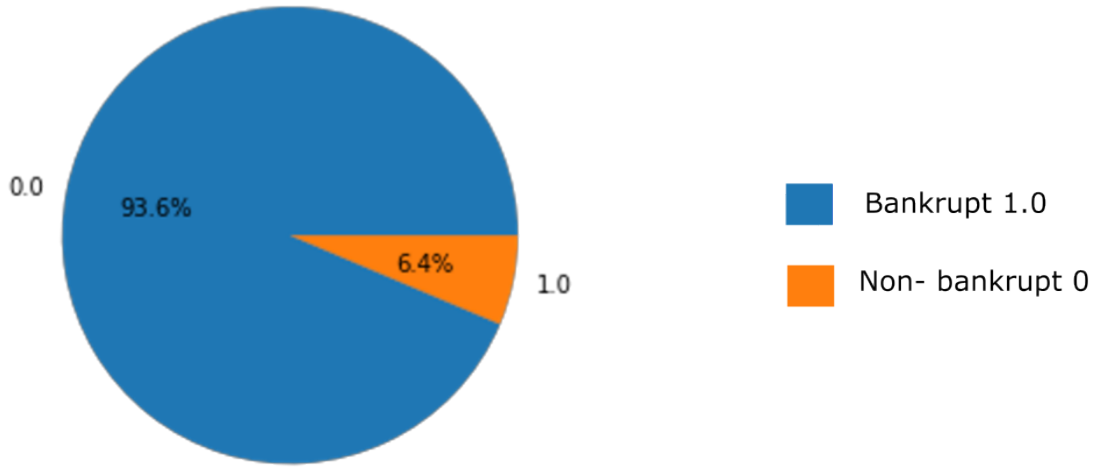
Figure 7: Data Completeness Visualiser of the United States of America attributes.



The imputation technique we used is the k-Nearest Neighbours technique as it has been done by Lambardo et al., (2022). This method of imputation allows us to find the missing value of training data by computing the mean value of the nearest neighbour. We do the same for our training set. The results are seen in Figure 8 below with 1 representing the total firms in our training set that are bankrupt and 0 representing firms that are not bankrupt. From the data, we note that our data is imbalanced as more than 90% of our training data shows no bankruptcy. This has the consequence of creating inaccuracies when applying our prediction models due to the skewness of the

data. Furthermore, it can create a model that is trained inadequately which can have an impact on the results from our models.

Figure 8: Imputed Plot of our Training data

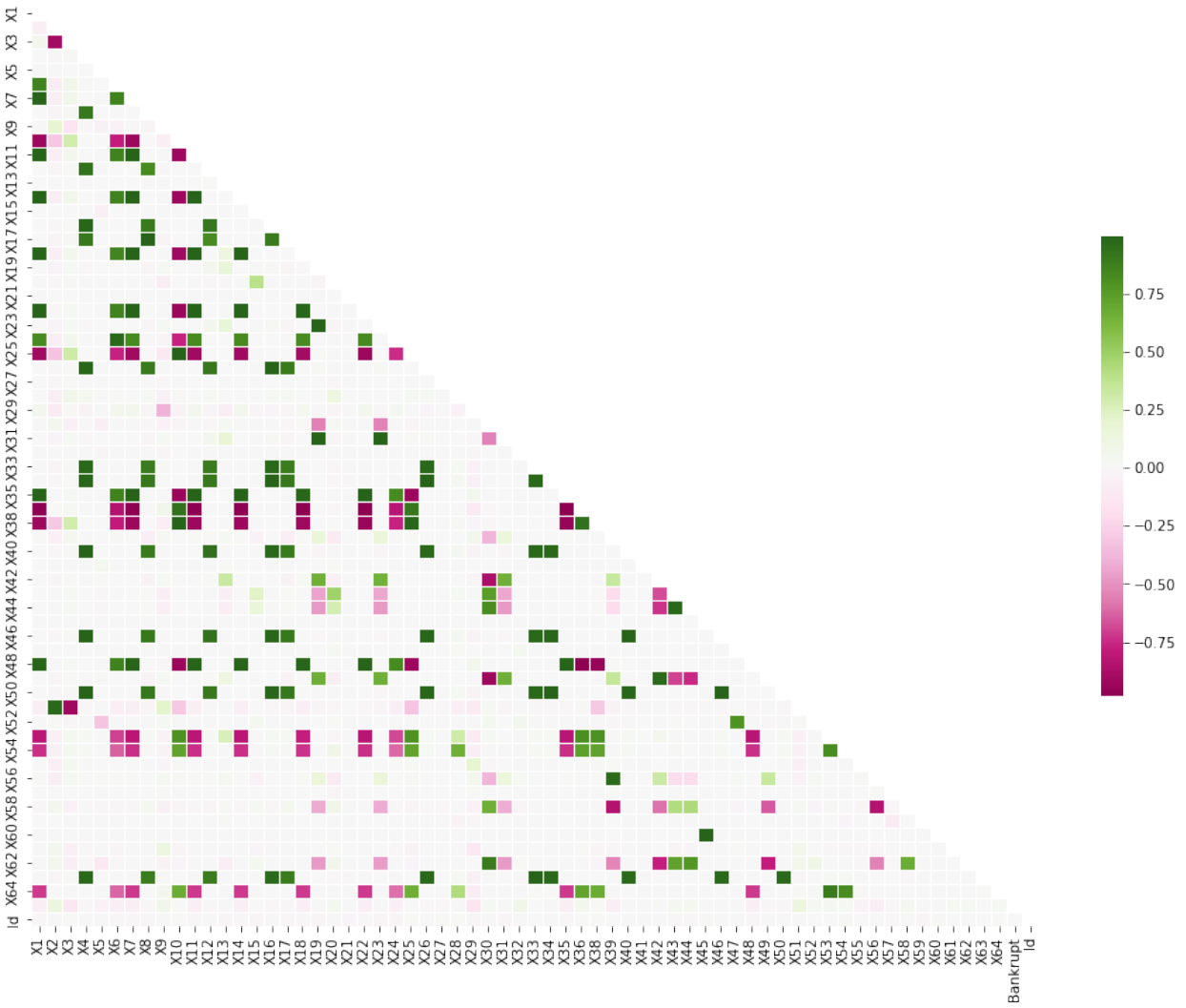


Our study tackles the issue of data imbalance by synthetically generating data to increase the instances of the minority class utilising the Synthetic Minority Oversampling Technique (SMOTE) as seen in Perboli and Arabnezhad (2021), Kim and Yoon (2021) as well as Lambardo *et al.,* (2022). This technique is a set of algorithms that very much like the k- nearest neighbours approach, uses a distance measure based on the feature set to locate the nearest neighbours, *k*, and then changes each attribute one at a time by multiplying each *x* by a random integer (between 0 and 1), creating a new synthetic instance.

We further explore the features of our dataset by looking at the correlations between features. We do this by utilising a heatmap of the features as seen by Figure 9 below. From the map, we see that there are high correlations between our features; however, the explanation of the correlations can be explained in the manner by which some attributes are calculated. One example is that the correlation as seen by attribute 1 and attribute 6 (represented by the net profit to total assets and retained earnings to total assets) which share values in the denominators which accounts for the high strong correlation between the values. Similarly, some features could have above value in

numerator and other features have the same value in the denominator, causing them to be negatively correlated as seen by attribute 9 and attribute 23 represented by sales to total assets and net profit to sales calculations.

Figure 9: Feature correlation heatmap for the United States of America dataset.



We deal with this high correlation of data by reducing the dimension of the data. The approach, as seen by Perboli and Arabnezhad (2021), is to apply Principal Component Analysis (PCA) in which we reduce the dimension of the data from a large one to a smaller one that has most of the information of the original larger dataset. The applied PCA results in Figure 10 shows us that 20% features can explain more than 95% of the variance which is useful for us to reduce the dimension of the data.

Figure 10: PCA results.

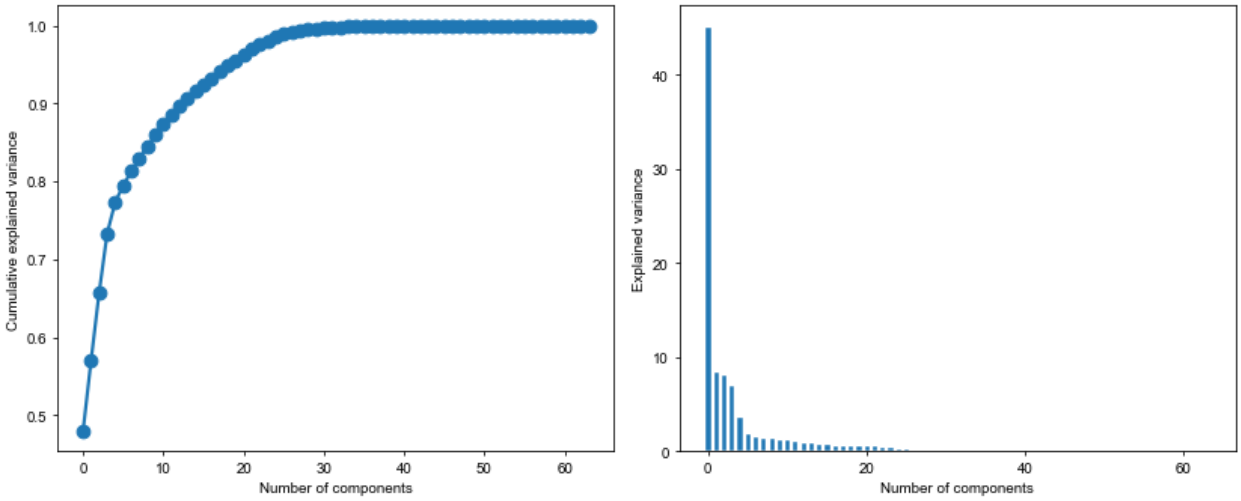
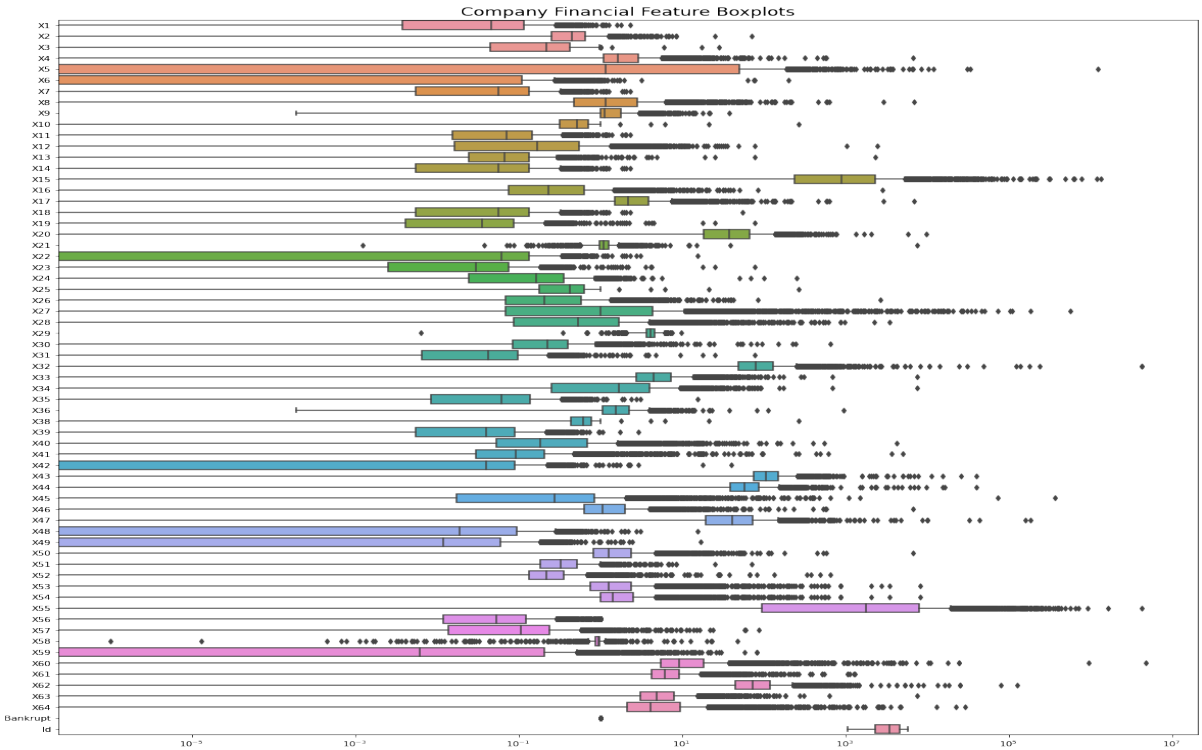


Figure 11 depicts a box and whisker plot of all the attributes and from close assessment, we can denote that some of attributes have extreme outliers that can further have an effect on predictability of our models.

Figure 11: Box plot for the United States of America dataset.



4.2.1.2 Results from the Machine Learning Models

The results from the machine learning models are presented below in Figure 12. From the initial results, we see that the best performing three models in their accuracy of bankruptcy is the random forest, XGB Classifier and k- NN Linear as well as the XGB Classifier. However, the models that tend to outperform other models are Logistic Regression, SVC and Hazard models. This is seen by their relatively higher recall tests and recall precision tests. The recall is calculated as the ratio between the numbers of positive samples correctly classified as positive to the total number of positive samples. This proportion therefore speaks to the quality of the model and when taken with accuracy, the models are a lot more consistent in their correct prediction. For comparison, we look at the same machine learning models on non-bankruptcy years and we see that the models as expected slightly deteriorate with the exception of kNN linear model which in effect performs better with an accuracy of 92% up from 88%. The three best models from the non- bankrupt years are the Logistic Regression, Random Forest Classifier and Hazard model as they have the higher Recall tests of 81%, 72% and 69% respectively, which therefore makes them more consistent in prediction than the other models.

Figure 12: Accuracy of Models over Bankruptcy Years.

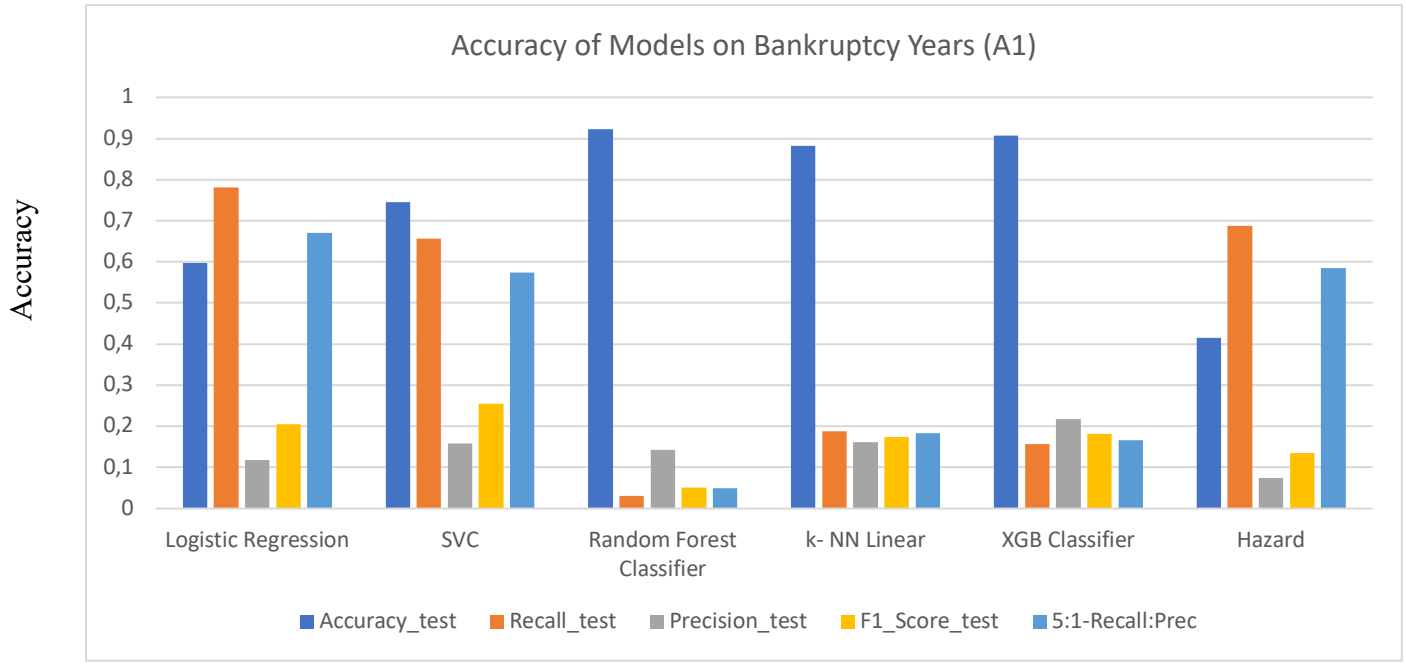
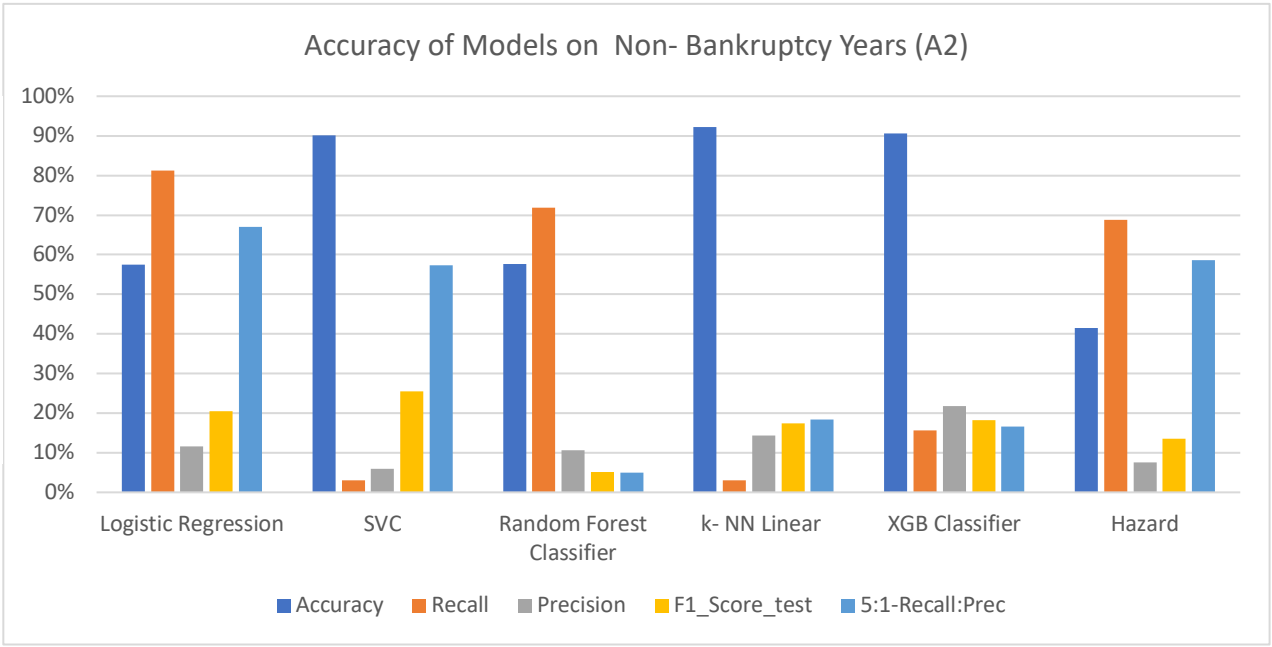


Figure 13: Accuracy of Models over Non- Bankruptcy Years.



4.2.2 Machine Learning Models on United Kingdom data

We utilise the same approach in running our machine learning models for the United Kingdom data. Firstly, we shall explore the data to understand its structure, dimensions and applicability to the model we have selected. This will assist us in ensuring that we get the best results as possible that are free from models. We will then apply performance comparisons of the dataset on the year of bankruptcy as well as the year leading to that bankruptcy. It is through this approach that we will gain an understanding of the performance of each model and how well it is able to predict bankruptcy.

The dataset consists of 64 attributes, five of which were originally defined by Altman, however, we introduce more attributes in order to improve the robustness of the study. Just as in the case of the United States of America dataset, all the attributes are synthetic which means they are derived from arithmetic calculations and operations. We pay our attention to Figure 14 where we see that the data has two attributes that have missing data. Attribute 21 and 37 have the most missing data which could have an effect on the predictability of our overall models. The reason for this is that these relationships might have a lot more importance on the model. Figure 15 looks at the feature importance of the attributes utilising the pretrained classifier and we see that the model has all the

features playing a roughly equal role in describing bankruptcy. We ignore attribute 0 as it is a spurious measure as it represents the identifier attribute which we use for identifying companies. We therefore resolve this issue of missing data by imputing this data in order to replace the missing values of the problematic attributes as done before. This is the same approach set out by Narvekar and Guha (2021) where they utilised the k nearest neighbours approach to calculate the mean of the nearest values and approximate the missing value using this calculation.

Figure 14: Data Completeness Visualiser of the United Kingdom attributes.

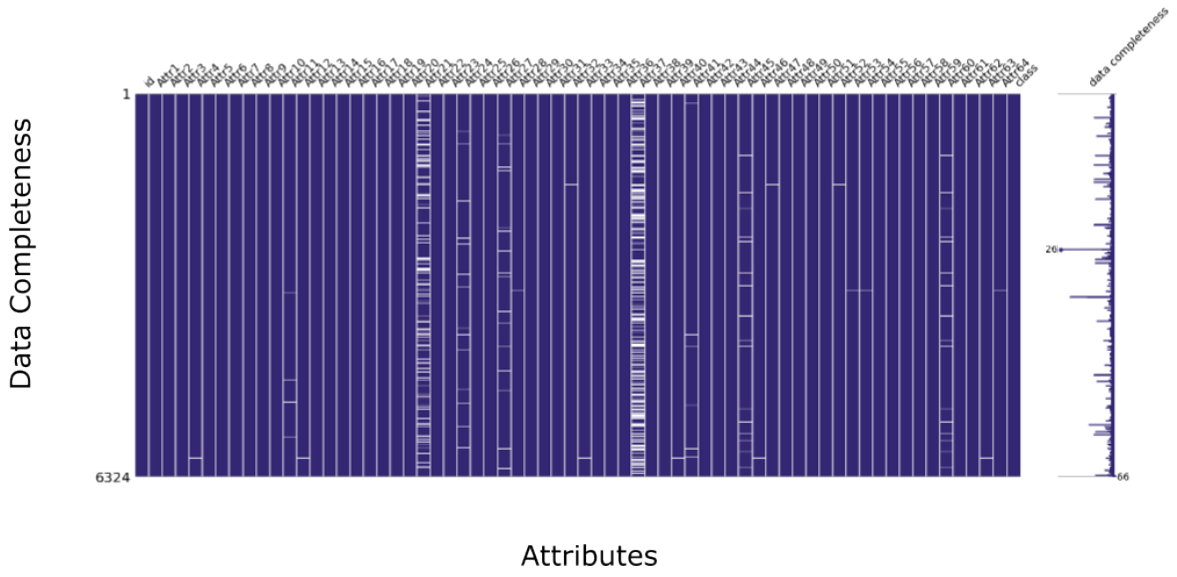
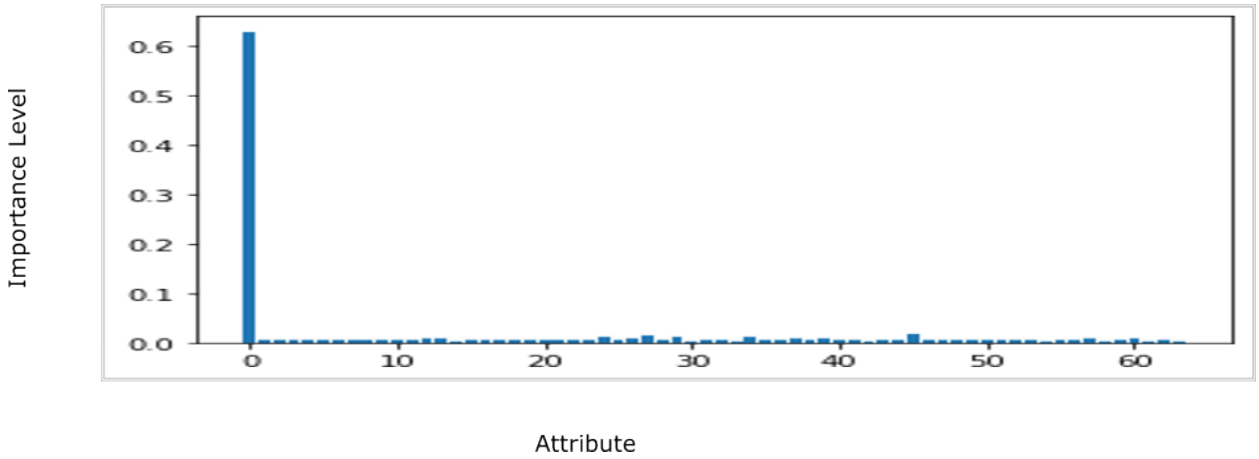
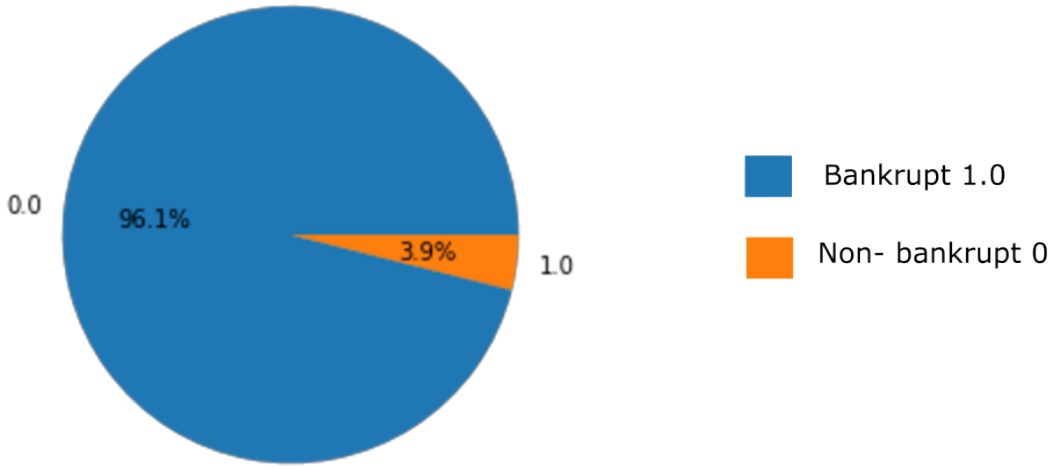


Figure 15: Feature Importance of attributes



Once imputed, we present our training data in figure 16 below. From the representation, we note that our data is imbalanced as more than 95% of our training data is classified as not bankrupt. This has the consequence of creating inaccuracies when applying our prediction models due to the skewness of the data. We have to balance the training data as to reduce the biases towards non bankruptcy due to the sharp skewness of the data.

Figure 16: Imputed Plot of our Training data

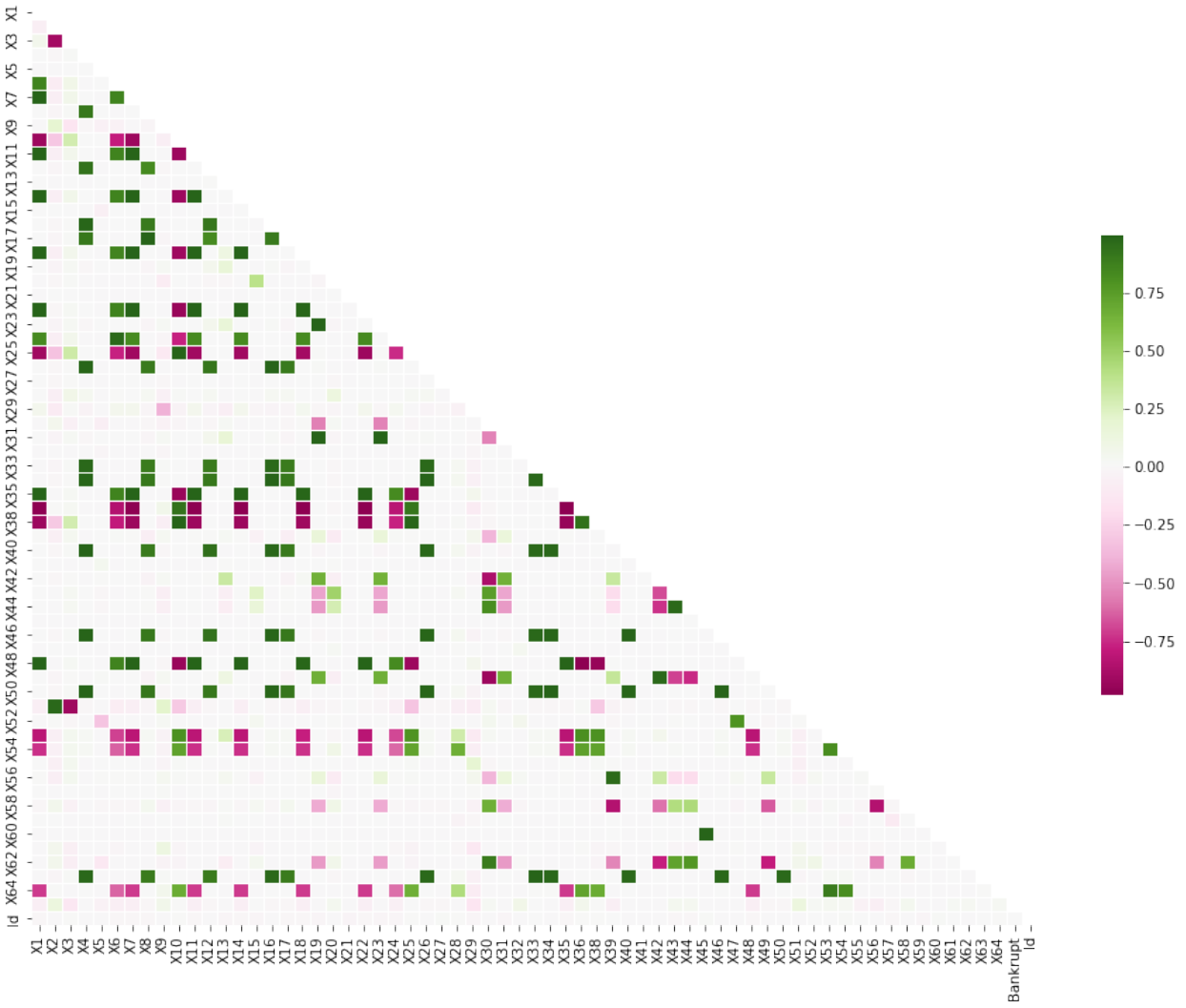


Our study employs the Synthetic Minority Oversampling Technique (SMOTE), which has been used before by Perboli G, Arabnezhad E (2021), Kim and Yoon (2021), as well as Lambardo et al., (2022) to enhance the instances of the minority class. A set of algorithms used in this method, which is very similar to the K-nearest neighbours approach, locate the nearest neighbours, k, using a distance measure based on the feature set, and then change each attribute one at a time by multiplying each x by a random integer (between 0 and 1) to create a new synthetic instance.

We explore the correlations of the dataset attributes in Figure 17 below. From the map, we see that there is high correlation between our features, however, the explanation of the correlations can be explained in the manner by which some attributes are calculated. The deeper pink blocks represent positive correlation and the deeper green blocks represent negative correlation. The main explanation for the cause of the correlation on the the heat map is that some of the attributes in the British dataset are used in calculations of other attributes in which they are a numerator or a denominator of other calculations. Some measures are strongly correlated due to them falling part of other

measures in the financial statements. A great example of this is attribute 53 which is the equity to fixed assets attribute which strongly correlated to attribute 1 which is the net profit to total assets measure. In this case, attribute 53 is essentially a subset of attribute 1 as they effectively measure the same financial indicator, though through a different articulation.

Figure 17: Feature correlation heatmap for the United States of America dataset.



The strong correlations would have an impact on our machine learning models as they would be drawn towards bias by these strongly correlated attributes. We, therefore, deal with the high correlation utilising approach as seen by Perboli G and Arabnezhad E (2021), by reducing the dimensions of our attributes as seen in the United States of America case by applying the Principal Component Analysis (PCA). The applied PCA results in figure 18 show us that 20% of features can explain more than 95% of the variance which is useful in order for us to reduce the dimension

of the data. This allows us to reduce the dimension of the dataset by 80% and dilute some of the strong correlations between attributes.

Figure 18: PCA results.

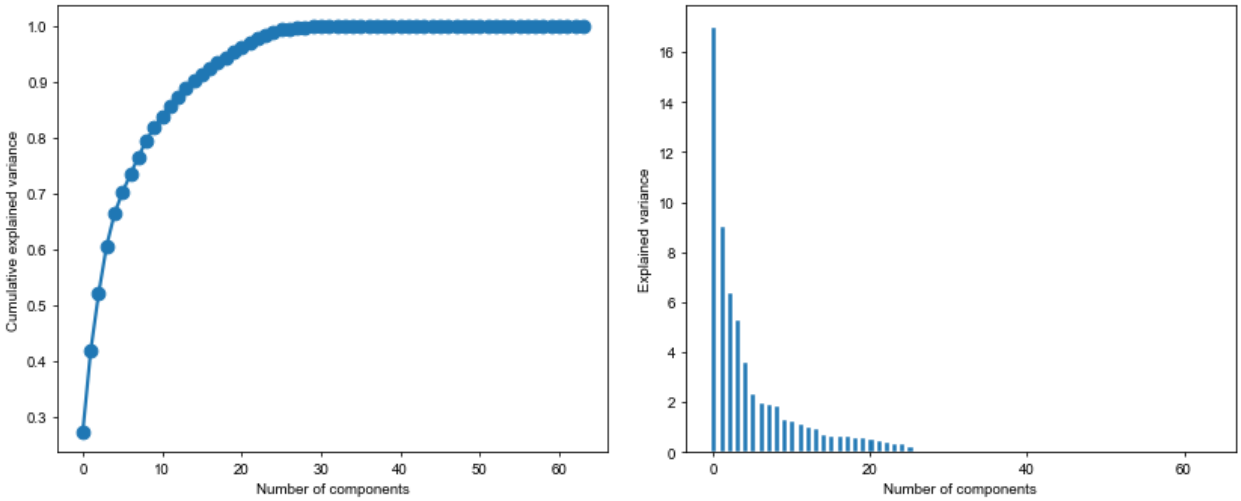


Figure 19 depicts a box and whisker plot of the extreme outliers of the Altman (1968) attributes of our United Kingdom dataset. From this depiction, we clearly see that the attributes have a clustering and extreme outliers for total assets, working capital and total liability and equity measures which are at the core of the Altman's (1968) study. Just like in the case of the correlation between attributes, extreme outliers may also hinder the ability of our model to make accurate predictions of bankruptcy. In order to calculate the distributions for the features of Net Profit, Working Capital, Debt, and Earnings before Interest and Tax, we only selected the Bankrupted companies for the Figure 20 below. Unsurprisingly, while the debt has a long tail on the positive side, profit, working capital, and EBIT all have long tails on the negative side. A company's financial situation would be concerning if it had negative earnings, negative working capital, negative revenue, and unpaid debt. We are not saying that all businesses that meet the aforementioned criteria are more likely to fail as many start-ups fall into this category but still succeed and generate steady income.

Figure 19: Box plot for the Altman United Kingdom dataset.

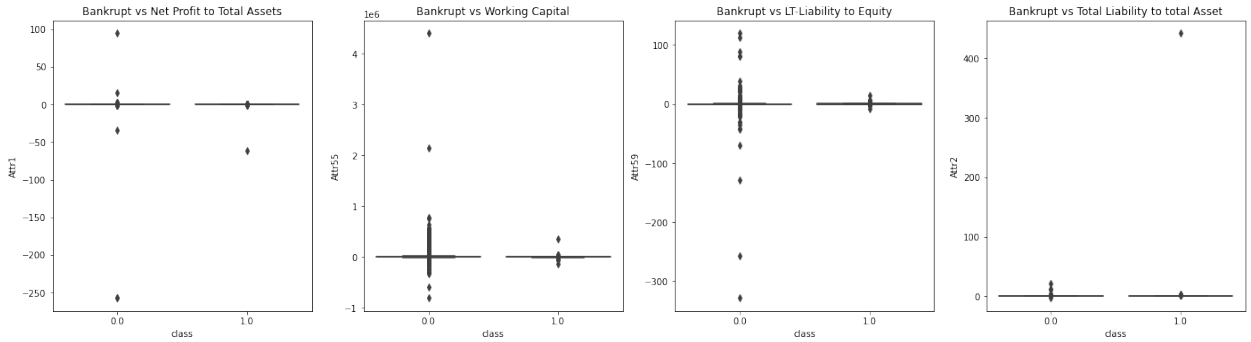


Figure 20: Distribution plots for the Altman British dataset.

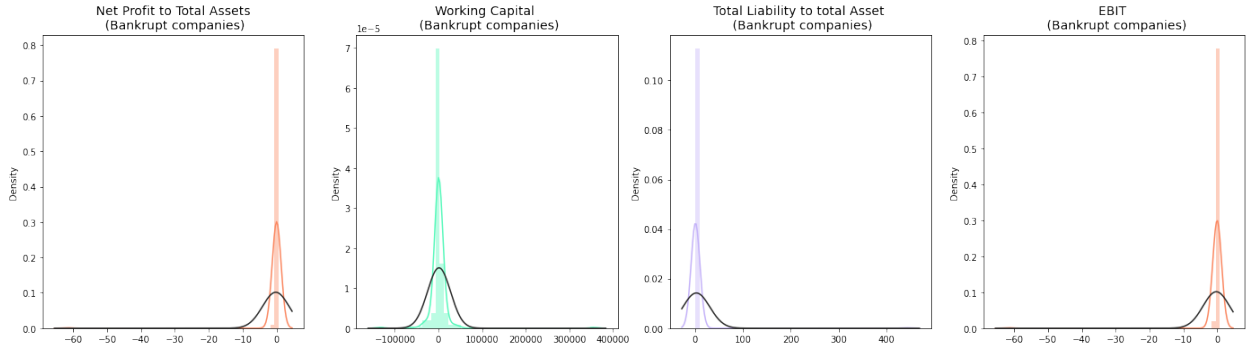
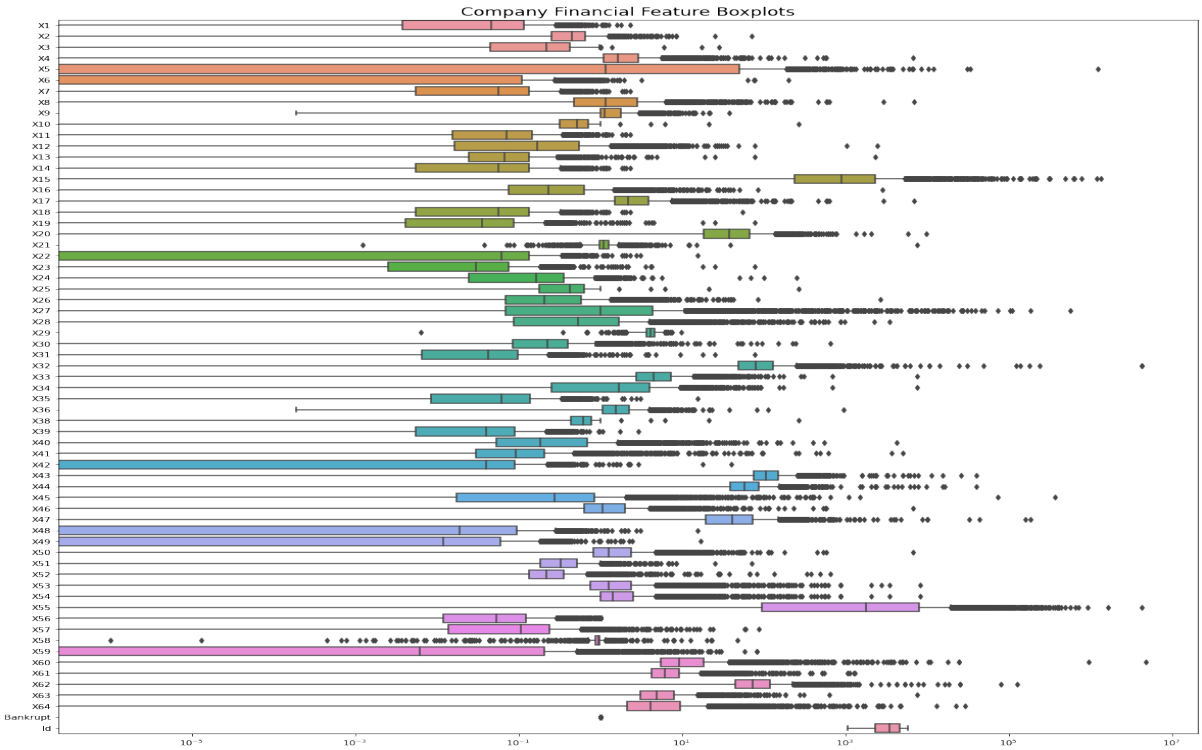


Figure 21 looks at the entire dataset by means of box and whisker and what we can gain from that is that nearly all attributes have extreme outliers with attribute 31, 55 and 61 having the highest outliers. 39% percent of the attributes have attributes falling outside of the 10^3 level distribution whereas many have clustered more compactly nearer their median.

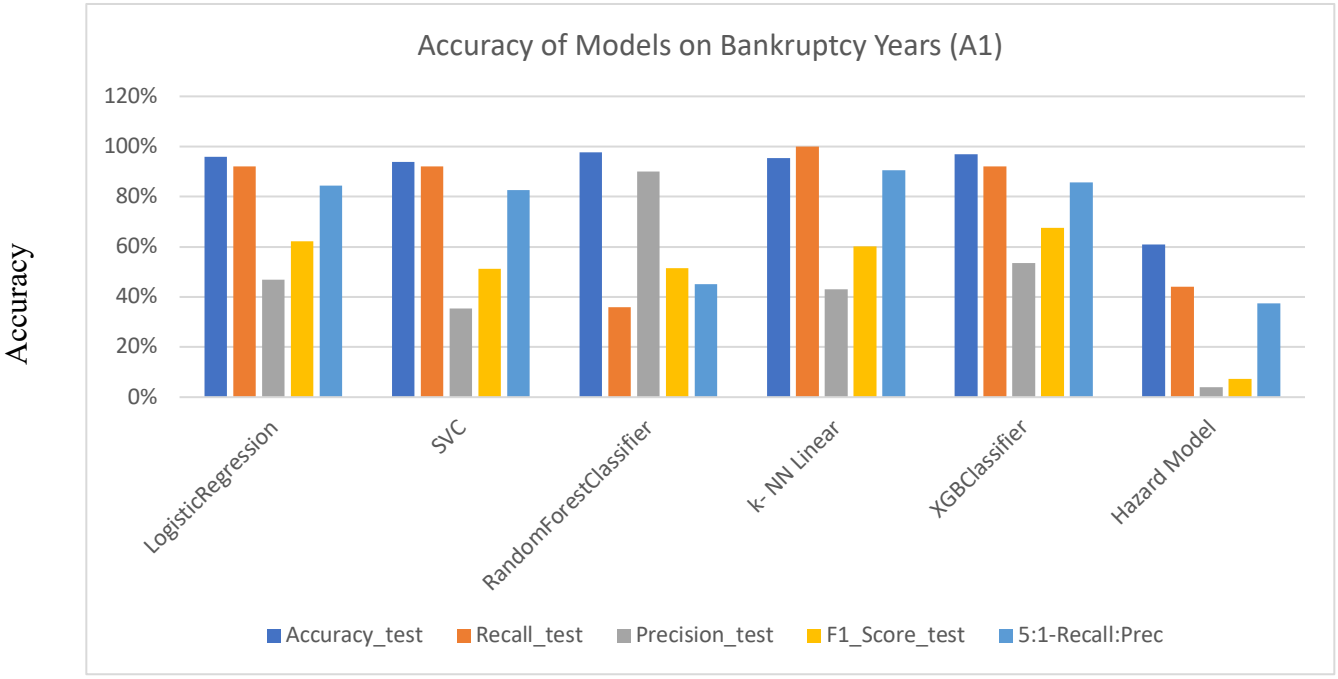
Figure 21: Box and whisker plot of attributes in the British dataset.



4.2.2.1 Results from the Machine Learning Models

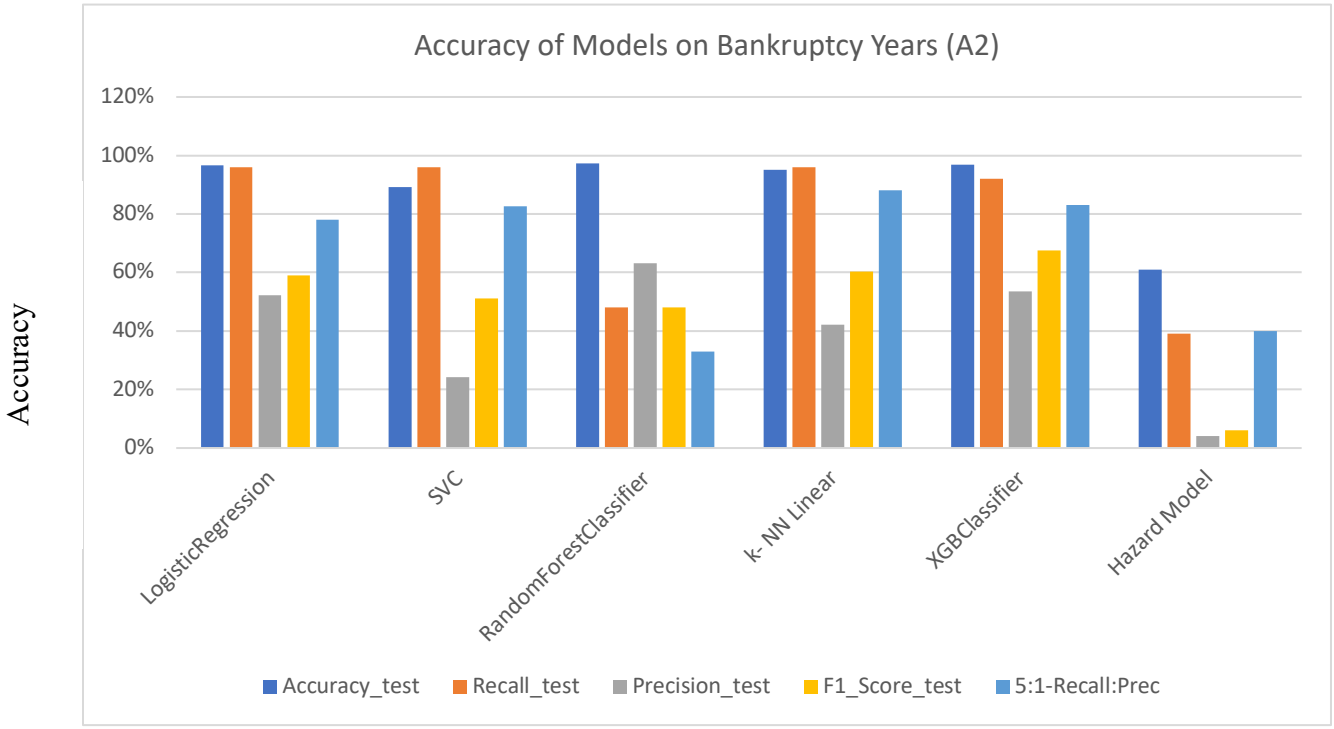
The results from the machine learning models are presented below in Figure 22. From the results, the most accurate measure of bankruptcy is the Random Forest Classifier with 98% correct predictions however it performs the least consistently as seen by its recall value of 36%. The three models with the highest percentage of accuracy and level of consistency are the Logistic Regression, SVC, and the k Nearest Neighbour models with a bankruptcy prediction accuracy of 96%, 94% and 95% respectively as well as having a recall, therefore consistency, of 92%, 92% and 100% respectively.

Figure 22: Accuracy of Models over Bankruptcy Years.



In comparison, we present the results of the previous year as seen in Figure 23. We notice that the machine learning models similarly experienced deterioration in performance as seen in Altman (1968), Narvekar, A., & Guha, D. (2021) and Perboli G and Arabnezhad E (2021) . From this result, we see our three best performing models maintaining their place with accuracies of 97%, 89% and 95% for the Logistic Regression, SVC and k-NN linear models respectively. Similarly, we see their consistency in predicting correctly remaining quite high as well with all three models scoring a consistency level of 96%. This is an increase from the previous year of 2% for the Logistic Regression and SVC models. However, a decrease of 4% for the k-NN linear model had a 100% consistency rate one year ahead. The hazard model performed just as bad as it did in the United Kingdom States dataset with an accuracy of 61%. This makes it the only measure to perform worse than the Altman measures (1968).

Figure 23: Accuracy of Models over Non- Bankruptcy Years.



4.2.3 Machine Learning Models on South African data

As in the previous cases, we utilise the same approach in running our machine learning models on the South African data by exploring the data, eliminating inadequacies in our dataset as well as comparing our various machine learning models to find which models are the best. From figure 24, we see that the dataset has missing attributes in X21 and X37 which we impute in getting nearest average estimations of the data. This will complete the dataset and avoid errors in estimation that may arise from the biases that are created. We do this by applying the K- nearest neighbours as it is the most efficient method to provide us with the close average values that are in line with the values.

Figure 24: Data Completeness Visualiser of the South African attributes.

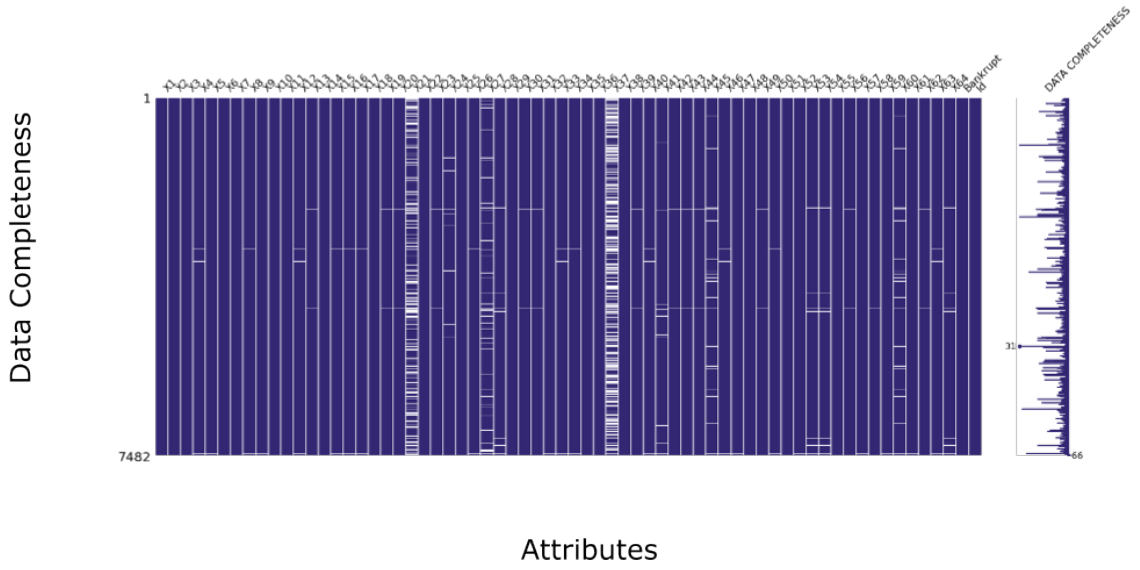
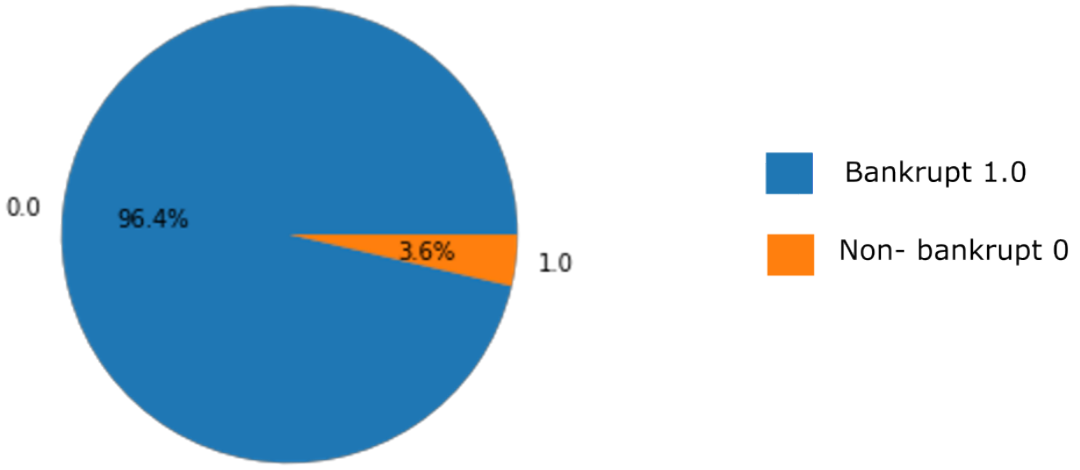


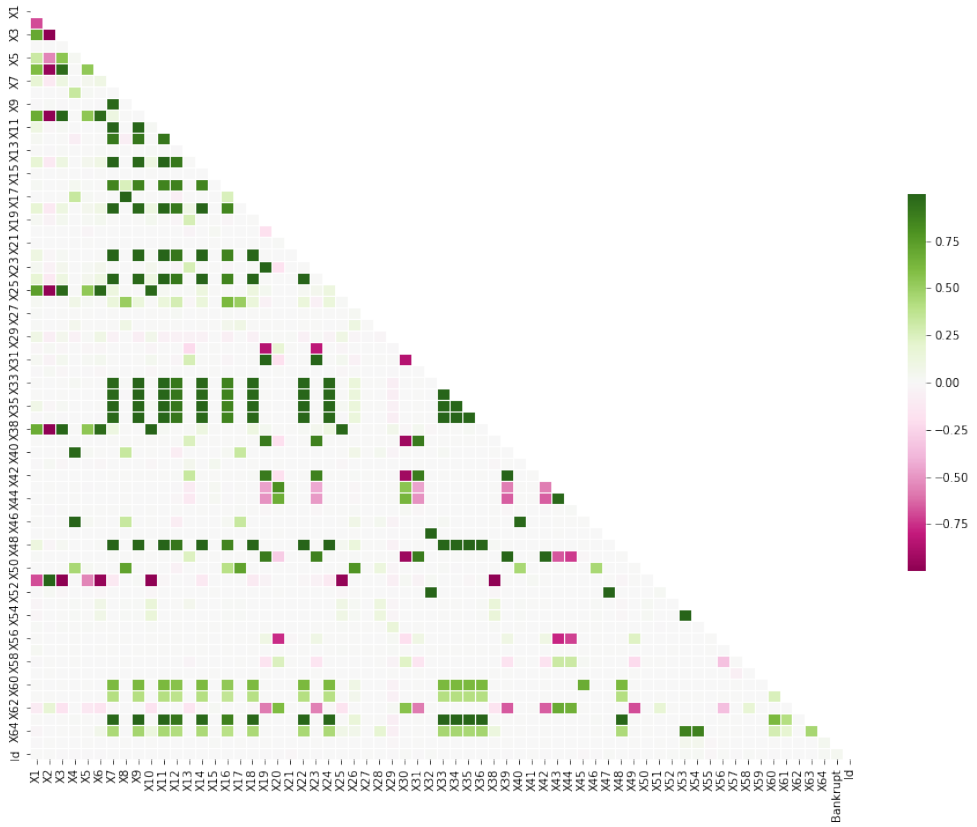
Figure 25 below shows the completion of our training data however, we see that our training data is not as balanced with 96,4% of our data being represented by non-bankrupt data and only 3,6% of the training data being classified as bankrupt. What follows is that any further training based on such data would lean more to a non-bankrupt classification. We apply SMOTE (synthetic minority oversampling technique) to reduce the imbalance by further applying the k- nearest neighbours to oversample our dataset of the minority and to increase the occurrence of the minority data. From this approach, we were able to have a mix of 30% being classified as bankrupt which is 10 times the original amount pre oversampling. This result makes the model have enough data in bankruptcy classification to perform well and provide more prudent results.

Figure 25: Data Completeness Visualiser of the South African attributes.



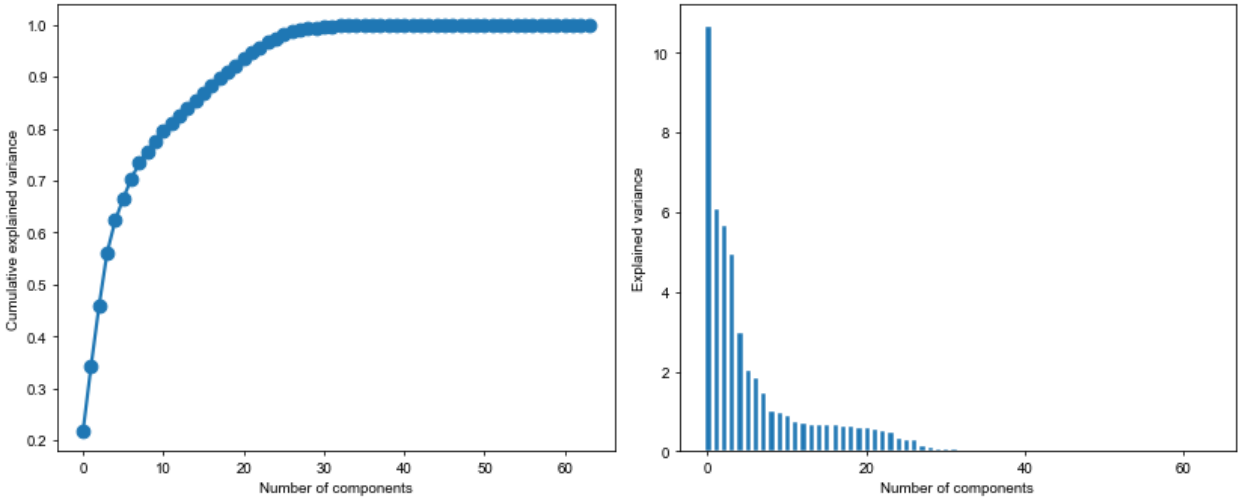
With figure 26, we look at the cross correlation within the South African dataset. It is important to explore this result as previously stated, these cross relationships between datasets have the effect of creating misleading associations, as a result, we may attain misleading results. From the depiction, we multiply positive and negative correlations between datasets as well. This is mostly caused by the mathematical properties in calculating the attributes repeating themselves across different attributes. An example of this is the attribute X25 which has a negative correlation to attribute X2 and a positive correlation to attribute X3. The measure for X25 is the equity less share capital to total assets ratio whereas, attribute X2 and X3 are the total liabilities to total assets ratio and working capital to total assets ratio. It is important to note that in the case of these ratios, shared numerator and denominator values in the measurements result in the correlation as seen in the depiction in Figure 26. The approach we take to deal with this is through dimension reduction by means of the principal component analysis.

Figure 26: Feature correlation heatmap for the United States of America dataset.



The result of the dimension is that most of the data can be explained by 20 attributes with the Altman (1968) being included. This, therefore, allows us to reduce the dimension of the dataset by 80% and dilute some of the strong correlations between attributes.

Figure 27: PCA results.



The extreme outliers of the Altman (1968) attributes in our South African dataset are shown in a box and whisker graphic in Figure 28. This illustration makes it crystal clear that the attributes for the major measurements of the Altman (1968) study - total assets, working capital, total debt, and equity - have a clustering and extreme outliers. Only the bankrupted companies were chosen for Figure 28 below which shows the distributions for the attributes of Net Profit, Working Capital, Debt, and Earnings before Interest and Tax. Profit, working capital and EBIT have lengthy tails on the negative side while debt has a long tail on the positive side.

Figure 28: Box and whisker of Altman models.

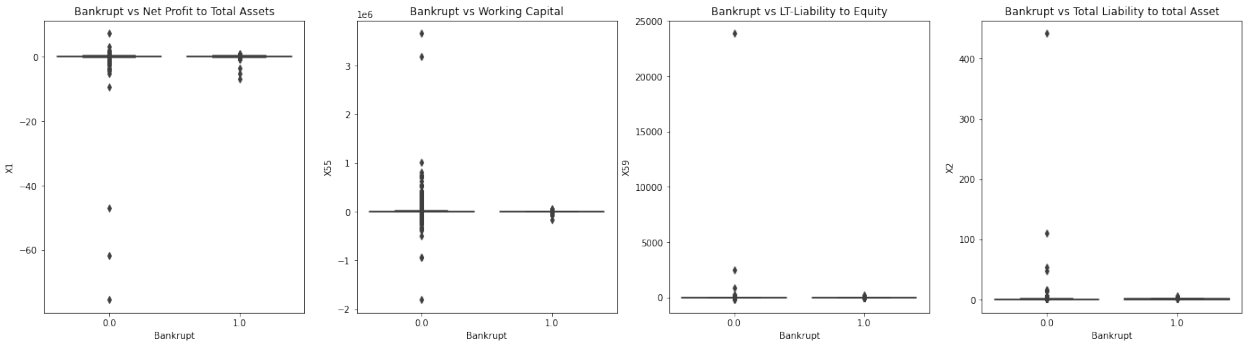
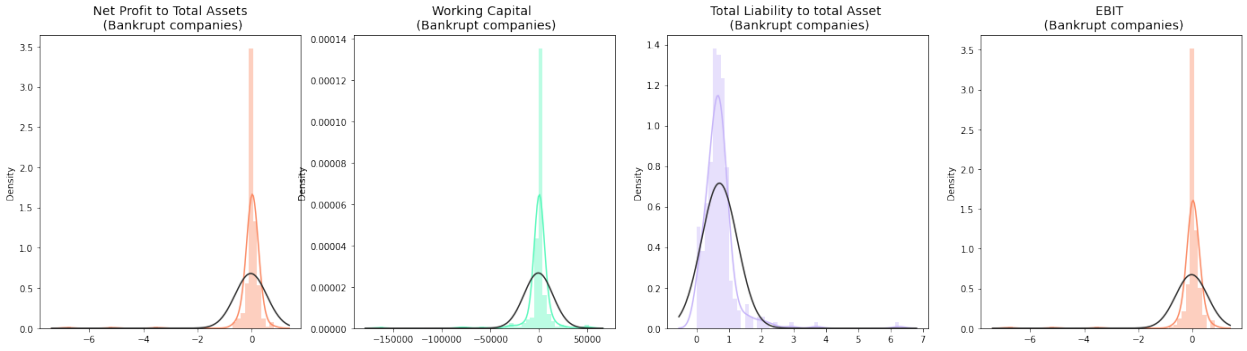


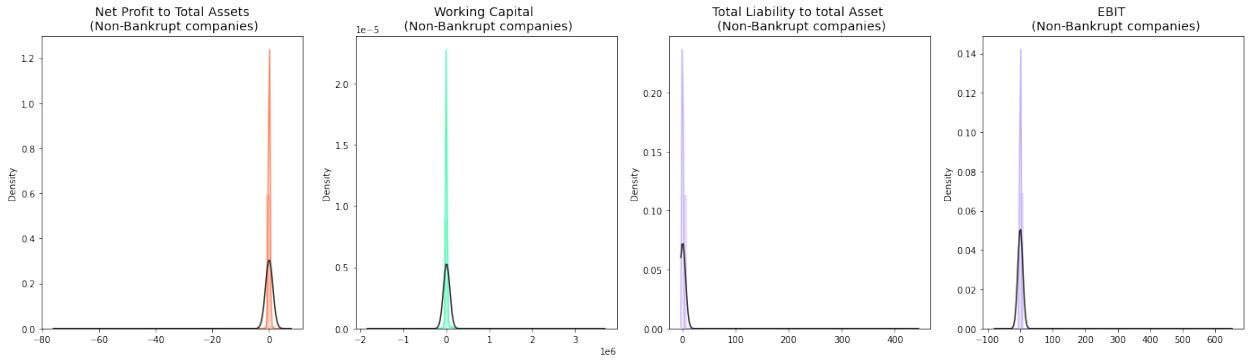
Figure 29 depicts the distribution of the Altman (1968) ratios for bankrupt firms. In the depiction, we see that all of the ratios are normally distributed with relatively long tails to the negative side. The total liability to total asset is skewed to the left with a tail towards the positive side. This is in line with the fact that many of the firms during periods of bankruptcy would experience negative profitability, constrained working capital and decreased earnings, with the share of liability generally being more than assets. This is in line with the fundamental theory by FitzPatrick (1932) wherein firms under bankruptcy have a tendency of having undesirable financial characteristics.

Figure 29: Normal distribution of Altman variables in the South Africa.



We ran the same analysis for the years prior bankruptcy, and we see that the tails are not as pronounced whilst the normal distribution is relatively held. The result of this goes to show how difficult it is to predict a company filing for bankruptcy a year in advance for the South African dataset. The distributions shown below indicate that the financial characteristics are of a firm that has a strong sense of going concern as the profitability, working capital and earnings distribution is not skewed. Similarly, the liability structure of the firm is also not skewed which means that it is not more than assets.

Figure 30: Normal distribution Altman variables pre- bankruptcy in the South Africa.



Along with the distribution of the Altman ratios, we do a feature importance analysis on all the factors we have included in our analysis. From this analysis, we see that many of them have the same importance attribute with 27, 33, 44 and 58 having an above average importance on the dataset. These attributes are represented by the profit on operating activities to financial expenses; operating expenses to short-term liabilities; yearly receivables to sales and total costs to total sales respectively. Many of these features are linked to expenses which could be an indication that the expenses play a relatively big role in leading a firm towards bankruptcy.

Figure 31: Feature importance of South Africa.

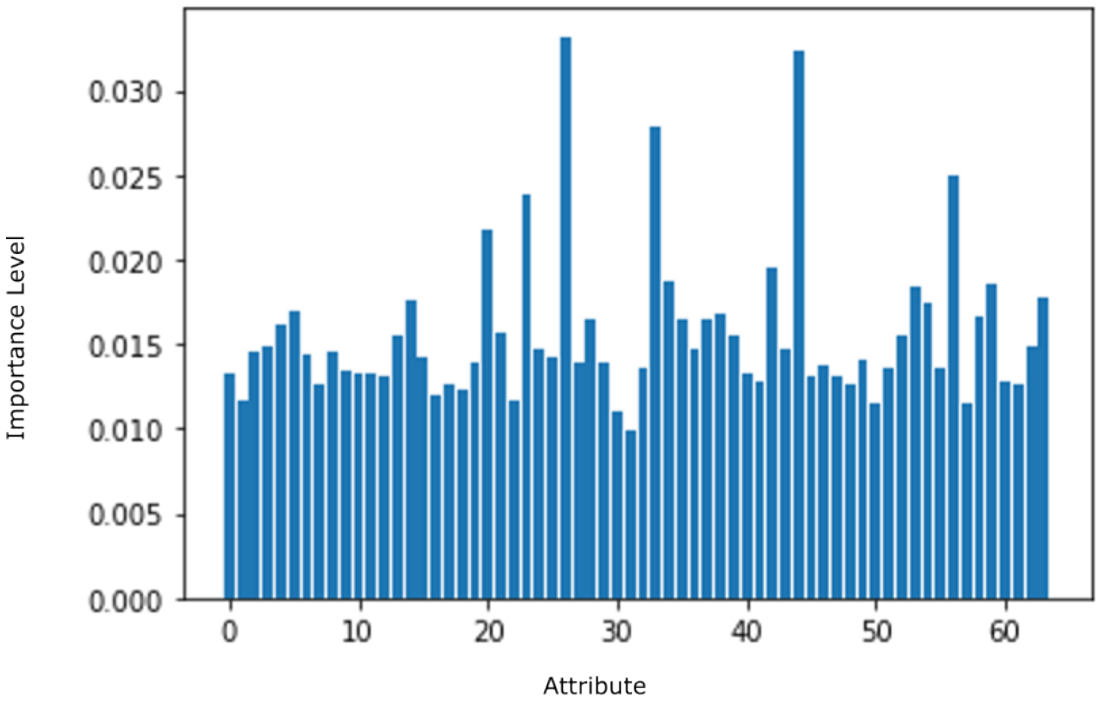
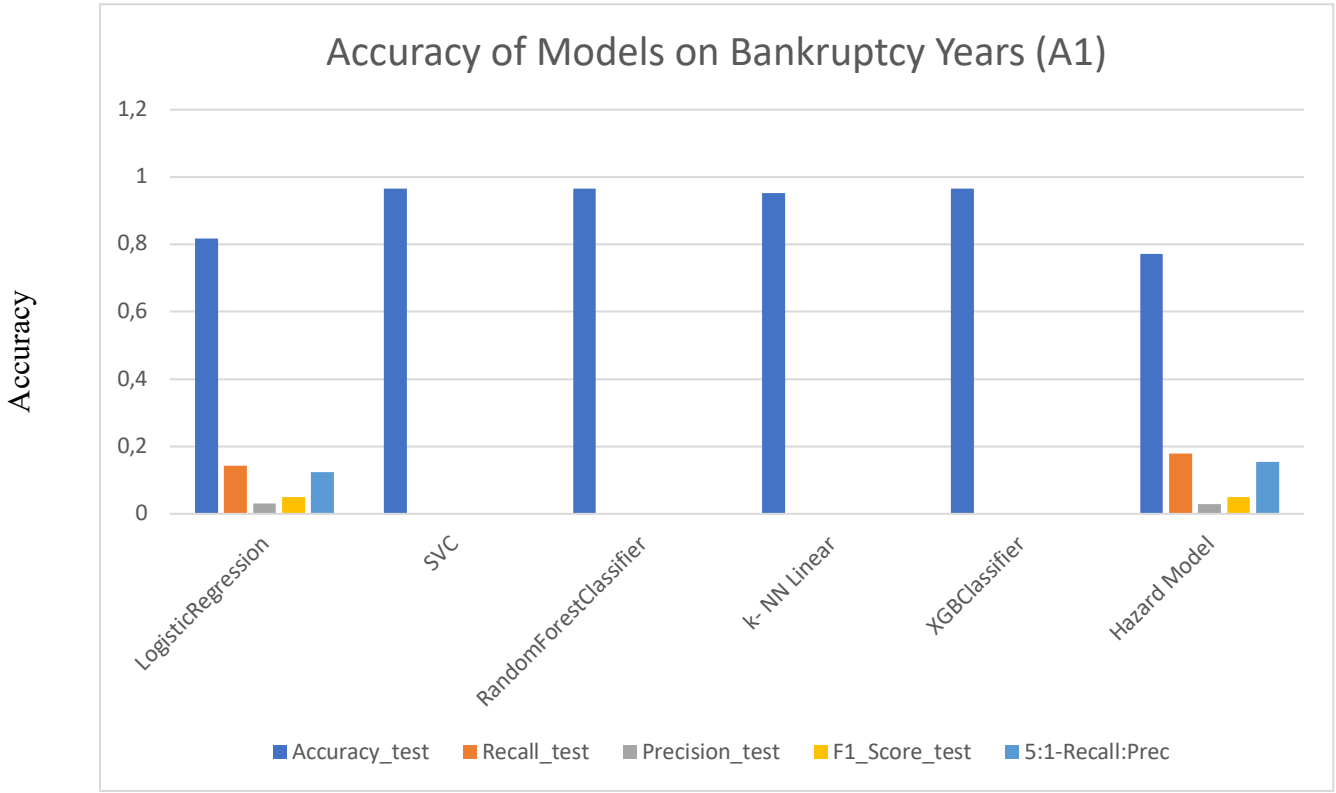


Figure 32 shows the accuracy of the machine learning models on the South African dataset during the bankruptcy years. From the result, we see that the models SVM, Random Forest as well as the XG Boost were among the strongest to perform in as far as having accurate bankrupties. However, they all failed to produce a recall test with all of them having a score of 0%. Therefore, this means that whilst the models were able to accuratley perform the prediction tests, these tests are not reliable for the South African context. We cannot place any reliance on their result. That said, the Logistic Regression as well as the Hazard model were the only two models that had produced a recall test greater than 0%, however,this makes the logistic and hazard models the best performing models of the five.

Figure 32: Accuracy of Models over Bankruptcy Years.



Due to the unreliability of the models, we attempt to understand their overall performance by means of error analysis. Figure 33 breaks down the classification performance of the models. We took the best performing models and compared them to two of the worst performing models in order to observe the differences in errors and to have some understanding as to why the models do not provide reliabile results. From this result, we look at the frequency the model was able to

classify the data accurately as well as how often it did not. The model results can be assessed as falling in one of the four categories explained in Table 6 below.

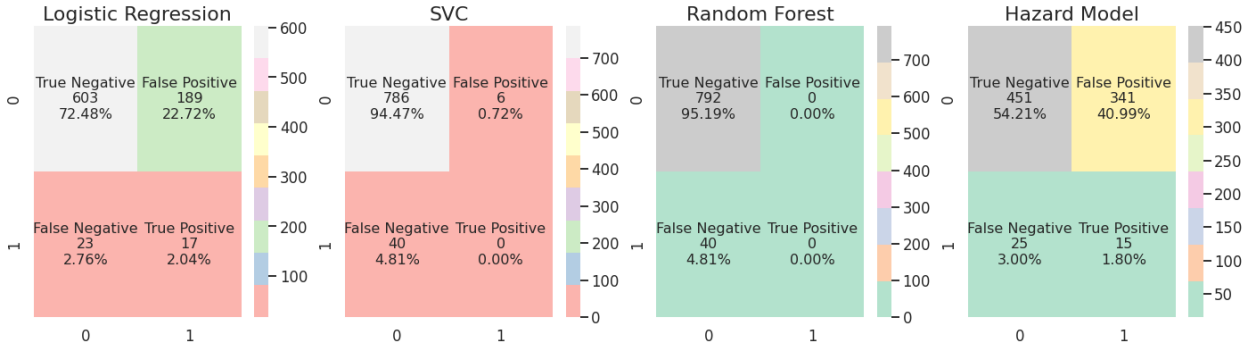
Table 5: Error Analysis of south African data.

True Negative Error	Where the model correctly predicts the non bankrupt classification
False Positive Error	Where the model classifies a company as being bankrupt when it was not bankrupt
False Negative Error	Where the model classifies a company as being non bankrupt when it was bankrupt
True Positive Error	Where the model correctly predicts the bankrupt classification

The logistic and the hazard models show that they perform false positive errors 23% and 41% of the time respectively. This is consistent with their low recall and precision values. Furthermore, this supports their reduced predictive accuracy compared to the SVM and Random Forest models. The logistic and hazard model also had a false negative value of 2,76% and 3% of the time which is relatively lower than the SVC and Random Forest tests. What we obtained from these results is that the models performed accurately and that even though the recall tests were 0%, the error analysis of the more accurate models supports their predictive power.

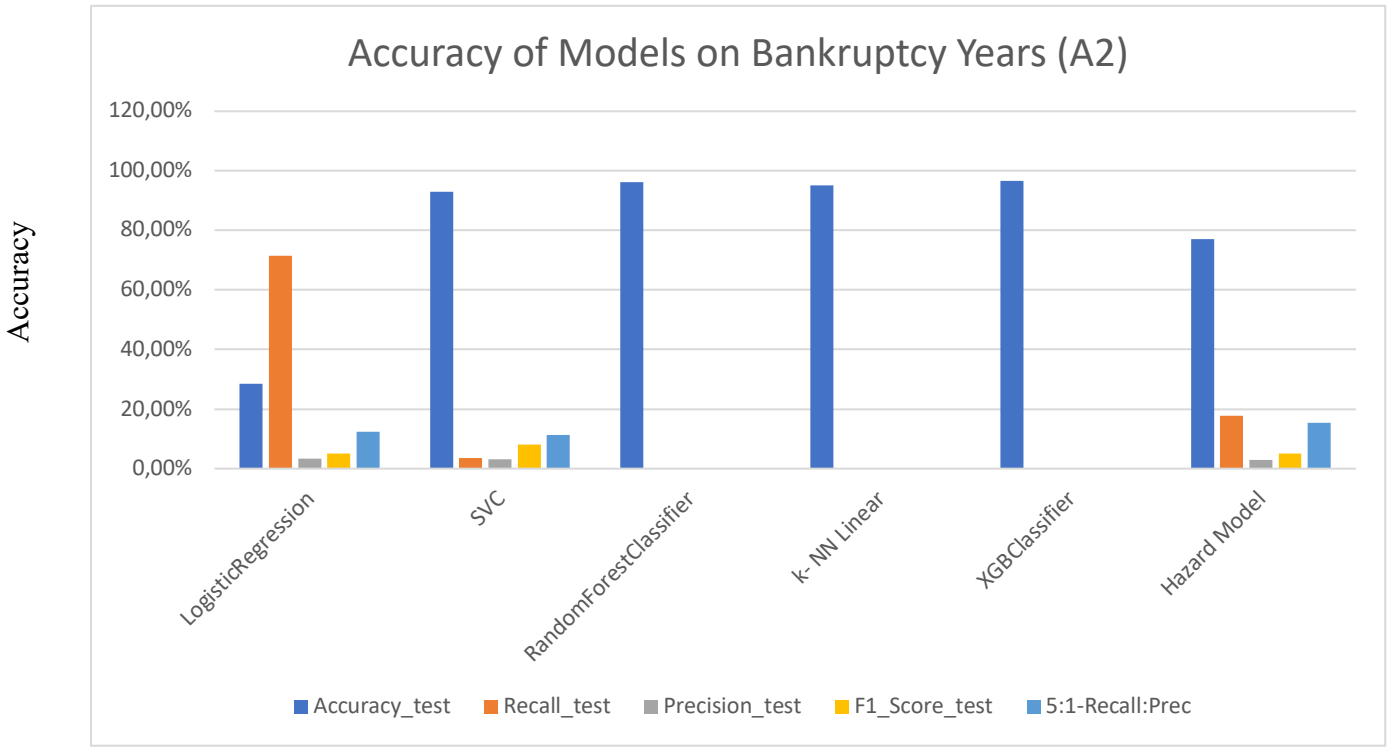
The SVM model was able to correctly allocate no bankrupt data 94% of the time whilst the Random Forest model was able to do so 95%. In the test, we see that models had a true positive error value of 0% because in their sample, there were no bankrupt data to predict. This provides a clear explanation why the models had a 0% recall and precision tests as it was not able to allocate data to perform a test. This shows that there is an issue with how the sample data is structured meaning that the results by all the other models are reliable due to the accuracy holding for the true negative error analysis test.

Figure 33: Error Analysis South Africa.



The same models were run for the non bankrupt period in which we see very consistent results of high accuracy with recall and precision tests on Random Forest, k- NN Linear as well as the XGB Classifier. What is diffierent is that the SVM model had a precision and recall value however it was also just as low. The same argument could be made where the allocation of data in predicting bankrupt data points is non existent in the sample allocation the model was run on, as a result, this had an effect on the recall and precision of the data as there was no data to test on. Ultimately, the model is able to correctly classify non- bankrupt firms correctly as this is the data which the sample allocation provided the models. The SVM, Random Forest, k- NN Linear and XGB models were able to do this true negative error analysis at such a high rate that we can consider the model results as positive regardless of the lack of precision and recall tests to support them.

Figure 34: Accuracy of Models over Non- Bankruptcy Years.



4.2 Textual Analysis

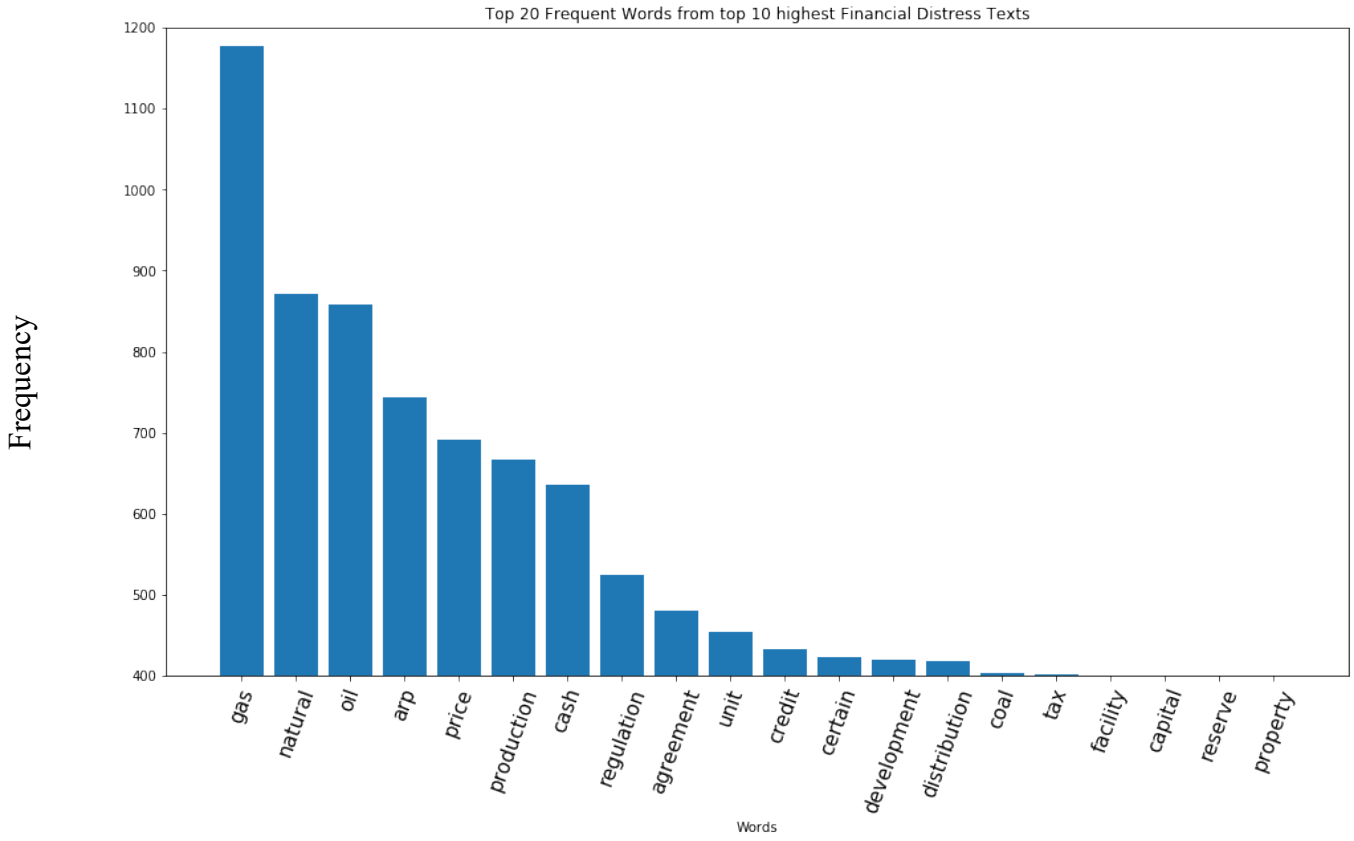
4.2.1 Data Exploration

The following are results emanating from the textual analysis models which were taken from the MD&A section of the United States of America filings, UK report and South African annual reports. The results are broken up into three textual and sentiment analysis models namely; the Dictionary model, Word2Vec, BERT model and the Domain Adaptive model. We will append the machine learning results to the textual analysis models to provide us with an understanding of whether we can more strongly predict corporate bankruptcy. That said, as in the fashion of this research paper, the best way to contextualise the results requires that we first explore the data as it will give us an understanding of how the results come to be.

We mine the text to provide the results in figure 35 below. The effective assessment we do in this case is to understand the frequency of certain words within MD&A disclosures in bankrupt settings in our database. We see the occurrence of gas as the most frequently used wording coming from the United States of America and United Kingdom database which is a strong indicator of its importance in firm bankruptcy. Furthermore, it is a strong indicator of the industries that would go

bankrupt. That is to say that the most likely candidates of firms in the United States of America and United Kingdom that would go bankrupt are those that have the highest linkages to natural gas. That said, the second most occurrence of words in our database is those that are linked to oil which is a feature of importance on all countries observed. That is similar to the notion presented before that the firms and industries that are most linked to oil are more likely to be more exposed to bankruptcy. Lastly, a word feature in the United States of America and United Kingdom though most prevalent in South Africa, is that of regulation. We found that in the case of bankruptcy, regulation - more specifically changes in regulation - have a huge impact in predicting whether a firm is to go bankrupt.

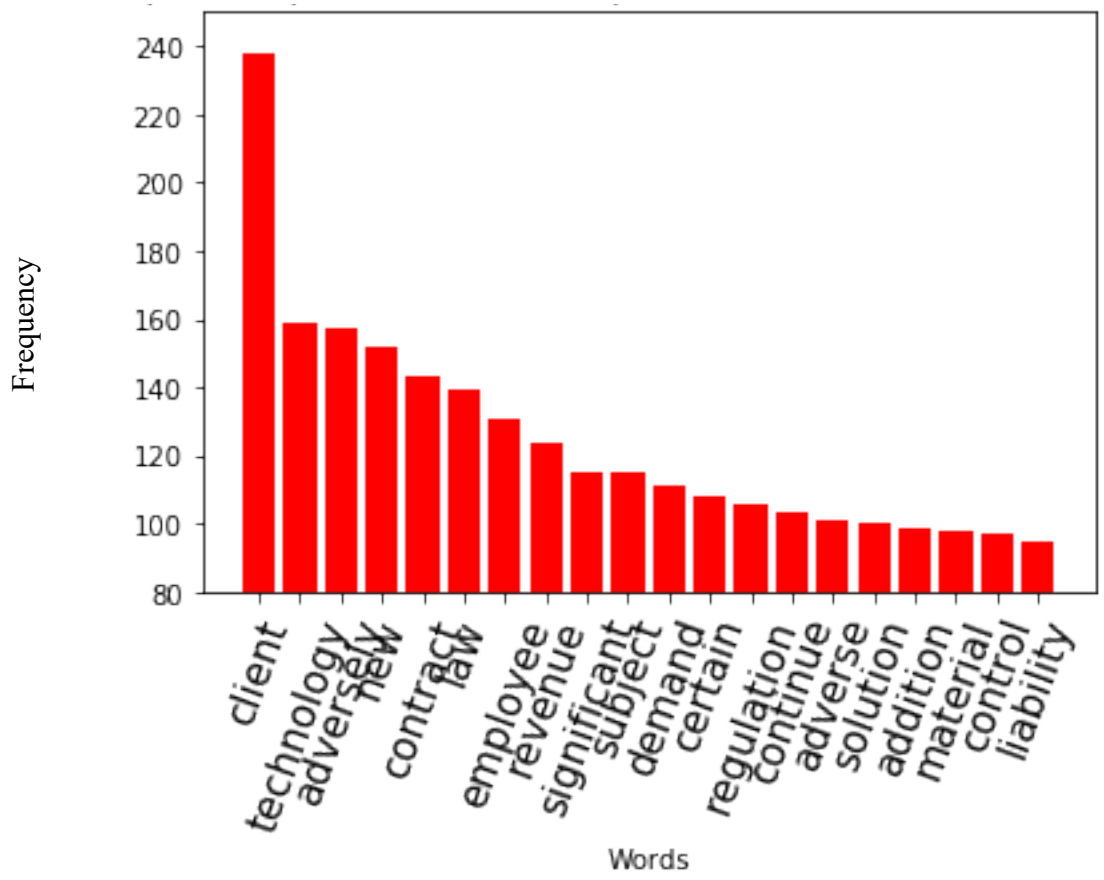
Figure 35: Frequency of Words used in all Countries.



To compare, we mined the most frequently used words in the non-bankrupt dataset and we found that the companies that are client facing and technology focused face the least possibility of bankruptcy. This acts as a counter to industries that tend to be in the oil and gas industry who produce

a raw material which is a necessity to the world at large, therefore, they do not need to focus on clients nor find themselves having to constantly improve in technology.

Figure 36: Frequency of Words in Non Bankrupt Firms in All Countries

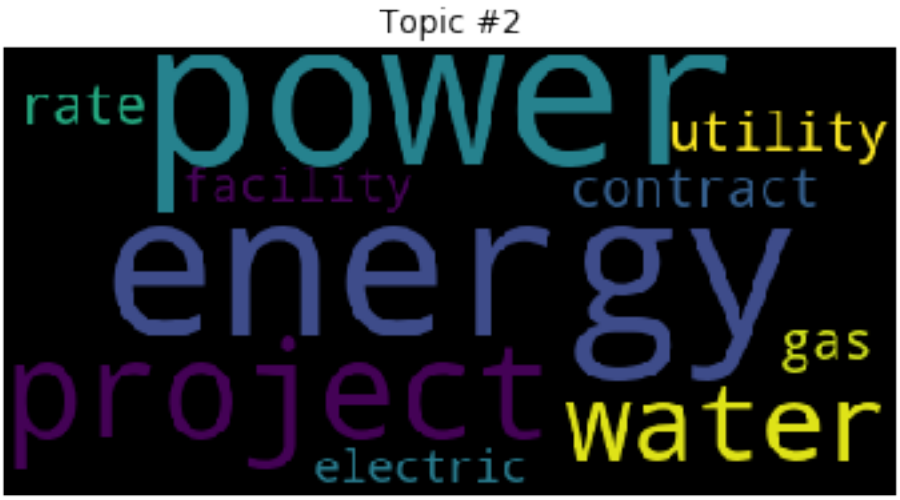


To visualise this data much better, we utilised topic modelling and represent the topics through a representation called Word Cloud in order to see the frequency and importance of certain topics and words, specifically in the different countries under assessment. The larger and more pronounced the word is, the more frequently as well as the more important that topic is in that country. We utilised topic modelling because it is a way of assessing the running themes of bankruptcy much better than counting the frequency of words and interpolating the results from such frequencies.

Figure 37 is a word cloud visualisation of the South African dataset during bankruptcy conditions. From the visualisation, we see very much of the themes being related to power and electricity with

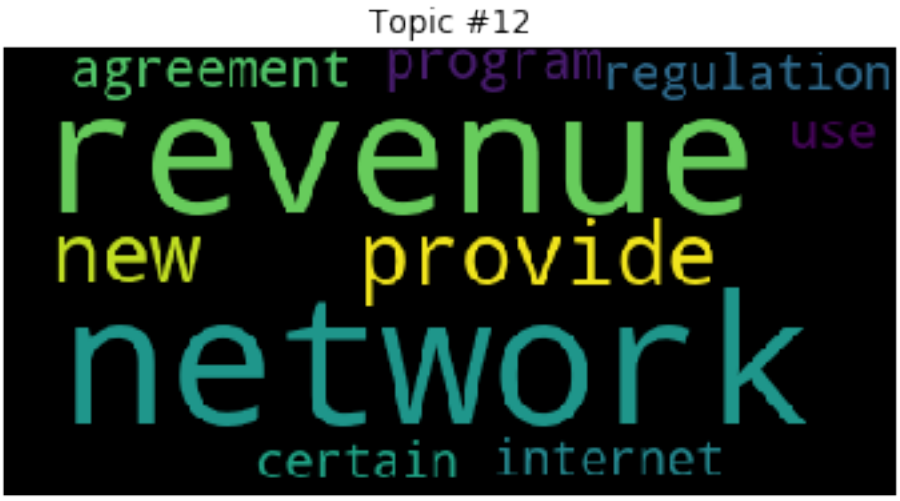
gas, rate and projects being an indicator of the kind of industries that are more frequently exposed to bankruptcy. The price word indicator seemingly got absorbed and is less important in the South African context as would be expected since much of South Africa's industry is less driven by stock market prices but accessibility to power and energy. Therefore, this means that the frequency count done in Figure 34 meant price was pronounced in South Africa only when combining the United States of America and United Kingdom dataset as well.

Figure 37: Word cloud in the South African bankrupt dataset



Under non- bankrupt conditions, we see more positive oriented themes relating to revenue, network and internet. In this depiction, we see that the themes are very much similar to the frequency count done in figure 36 in which the themes related more to technological advancements and client centricity. However, client centricity, very much like price, is merely stated in the combination of the United States of America and United Kingdom dataset and is not much of a central theme in the South African context.

Figure 38: Word cloud in the South African non- bankrupt dataset



In the same fashion, we see that the word cloud topic modelling has a tendency to refer to property, insurance and capital as more central themes under bankruptcy conditions. This suggests a strong dominance of insurance, finance and capital investment firms in the United Kingdom dataset. As a result, property and capital take much more of a greater importance in theme than oil and gas as would be suggested by figure 35 in the frequency count.

Figure 39: Word cloud in the British bankrupt dataset

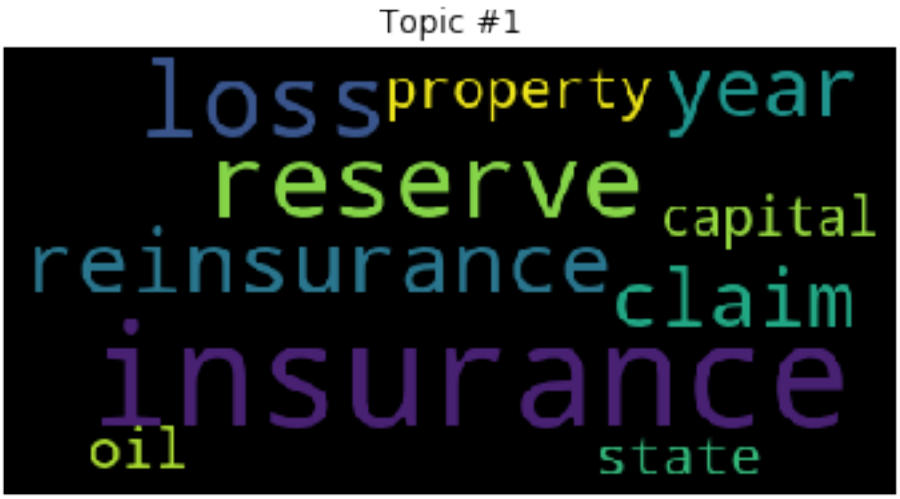


Figure 38 supports the assertion that the British dataset is dominated by insurance, investment and finance firms as a whole thus the central themes running in the United Kingdom dataset are reflective of that. In this case, we can place much assertion to the success of a firm to reflect more of the financial and market health of a corporate. That said, the technology and client centricity themes have been absorbed however it can me denoted that much of the themes surrounding technology and client service are a hardline feature of finance and insurance corporates.

Figure 40: Word cloud in the United Kingdom non- bankrupt dataset

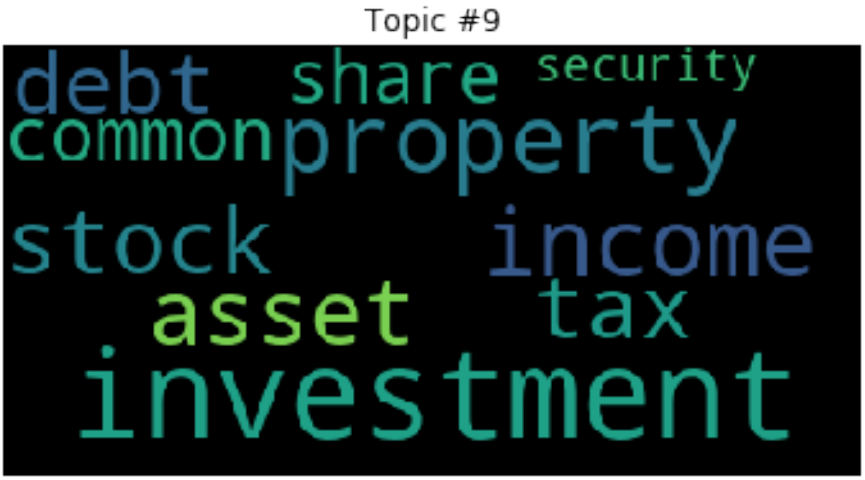


Figure 40, in this case, looks at the central themes of the United States of America dataset under conditions of bankruptcy. From this result, we note that much of the central themes in the US dataset are very much consistent with the frequency count in figure 34. What this showcases is the dominance of the oil, gas and regualtory environment - not only in the United States of America but as a theme in the United Kingdom and South Africa as well. This supports the notion that the American business environment has a global reach therefore effects proportionally in most of the world. We see this in that the central themes that are important to the British and South African corporate space are not as important to the United States of America business environment. However, the reverse is that the themes that fall within the bankruptcy space of the United States of American have a reaching effect on the United Kingdom and South African corporate environment. Nonetheless, we note that oil, price and the regulation are much more central themes to the United States therefore the industries and corporates that deal with these themes in their business operations are much more frequently exposed to bankruptcy.

Figure 41: Word cloud in the United States of America bankrupt dataset

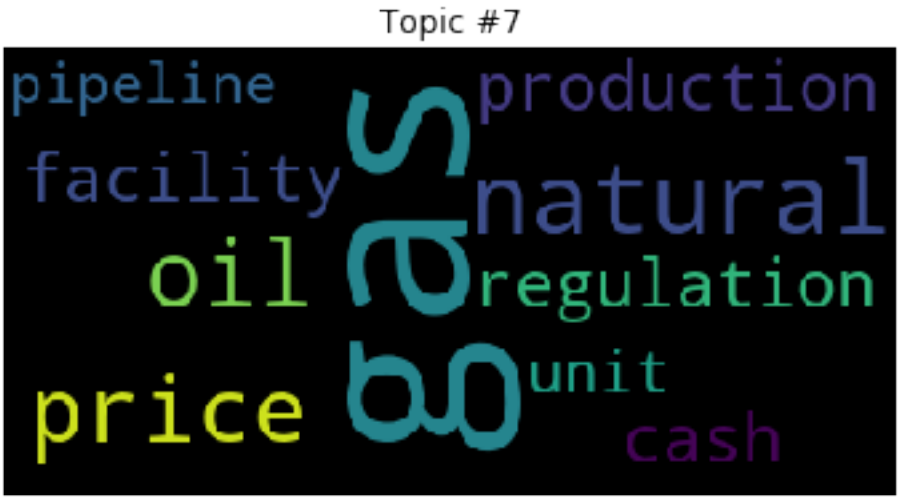
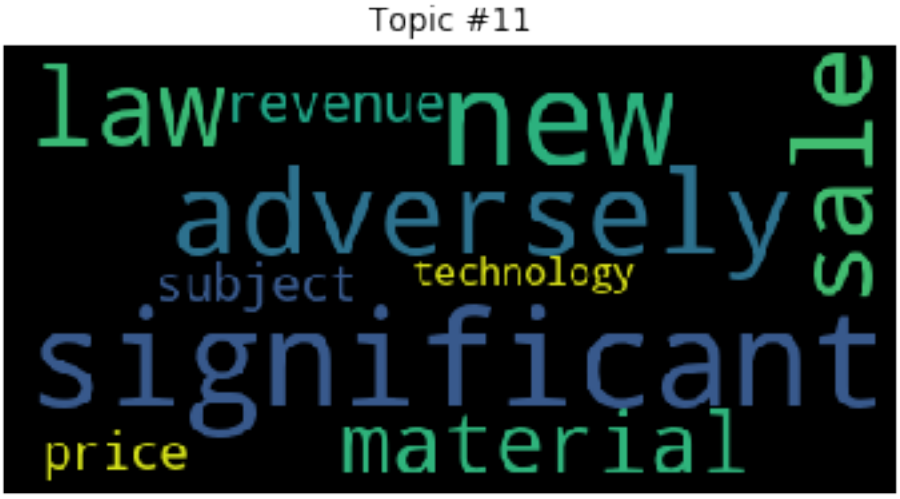


Figure 41 looks at the central themes in the United States of America under non bankrupt conditions and we find that these themes are mostly around law, revenue, sales and technology. From this, we can summise that corporates that have good financial and market performance as well as a strong going concern are less likely to be exposed to bankrtupty. That said, much of the themes highlighted by our topic modelling are also consistent with the non bankupt themes as seen by figure 35's frequency count. Technology and price also appear as central themes with good

alignment with law and regultory environment; this also plays a strong indicator for corporate well- being.

Figure 42: Word cloud in the United States America non- bankrupt dataset



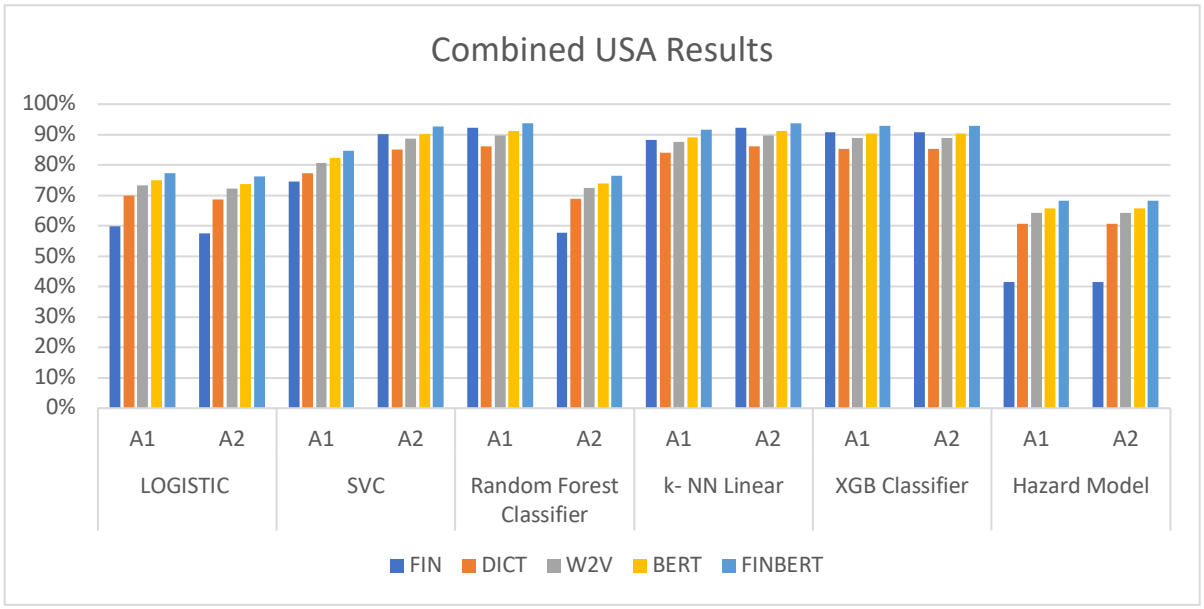
4.3 Combined Bankruptcy results

4.3.1 United States of America

The results, as shown figure 43, are a comparison of the machine learning models appended to the natural processing models for the United States America database. From this depiction, we see that much of the results follow the results found by Kim and Yoon (2021) with the textual analysis improving the results of the machine learning processed financial results. The dictionary approach has an average 7% increase on the entire machine learning models, with the word2vec model providing an additional average of 10% from the original model. The generalised BERT model improves estimation by 12% with the FINBERT model, which is the chosen domain adaptive model, having the biggest improvement in prediction by placing an improvement of 14% on the model's original financial based models. This shows that the more domain adaptive the sentiment analysis is, the stronger the prediction of sentiment is in the United States of America context. This also supports the notion that machine learning models alone are not strong enough to explain bankruptcy, instead, they can be assisted by textual analysis as well.

In the end, the best performing models in the United States of America dataset are the Random Classifier and the k-NN Linear model appended with the domain Adaptive FINBERT model as they received the highest prediction of 94%. This is a 2% increase from the vanilla financial based models. The worst performing model is the Hazard model which was only able to predict 68% of bankruptcies. The model was originally poorer than the rest of the models in predicting bankruptcy with the baseline financial variables - having only predicted 41% of the corporates as being bankrupt. This, however, shows that textual analysis plays a bigger proportional result in predicting bankruptcy where the Hazard model is concerned as it was able to improve the result proportionally higher than any other model, having improved predictions by 27% across all machine learning models.

Figure 43: Combined Results of the United States America Dataset



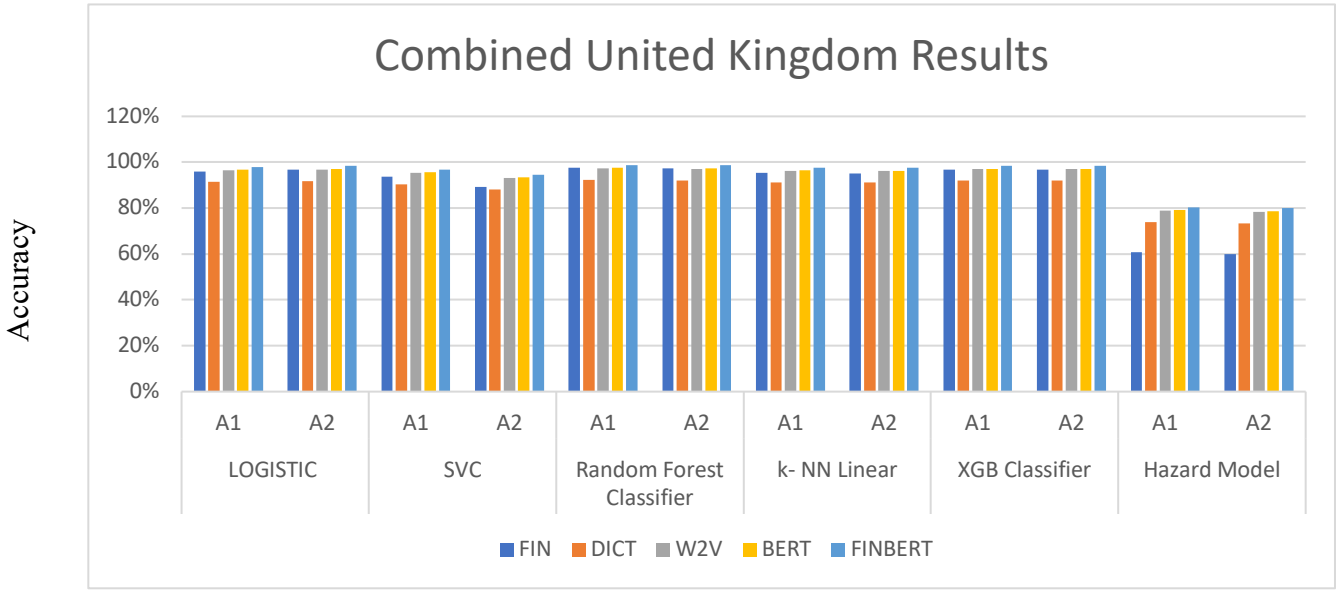
4.3.2 United Kingdom

The same study was done for the United Kingdom as shown in figure 44 which is a comparison of the machine learning models appended to the natural processing models for the United Kingdom database. In general, our results support the notion of sentimental analysis with the textual analysis improving the results of the machine learning processed financial results in the United Kingdom setting. The dictionary approach has an average 2% decrease on the entire machine learning

models. This is due to the insufficiency in predicting sentiment, however, this is also affected by the high prediction rate of the financial results. The word2vec model provides an additional 3% improvement from the financial based models which is in line with the study by Kim and Yoon (2021). The generalised BERT model improves bankruptcy prediction by 4% with the FINBERT model, which is the chosen domain adaptive model, having the biggest improvement in prediction of 5% on the baseline model.

As was the case in the United States of America context, the sentiment analysis models used in the British context improve the prediction of bankruptcy despite the high performance of the financial variables. What this tells us is that textual analysis plays a role in supporting financial variables when assessing firm bankruptcy. Therefore, the best performing model was the Random Classifier appended with the domain adapted FINBERT model which has a prediction accuracy of 99%. Similarly, the poorest performing model was the Hazard model which had a prediction rate of 80%. The reason for this strong performance sits in the financial variables of the United Kingdom however, as Li (2010) found, transparency as well as clear tonality would have the strongest underlying result in textual analysis. With United Kingdom being a dominantly English country as well as the fundamental historical setting of finance internationally, it is seen that textual analysis conducted on English disclosure patterns would have an expected stronger result compared to the United States of America or South Africa who are also English, however, more influenced by multiple cultures. This would therefore constitute different disclosure patterns. This makes United Kingdom one of the best settings to apply English based machine learning sentiment analysis.

Figure 44: Combined Results of the United Kingdom Dataset



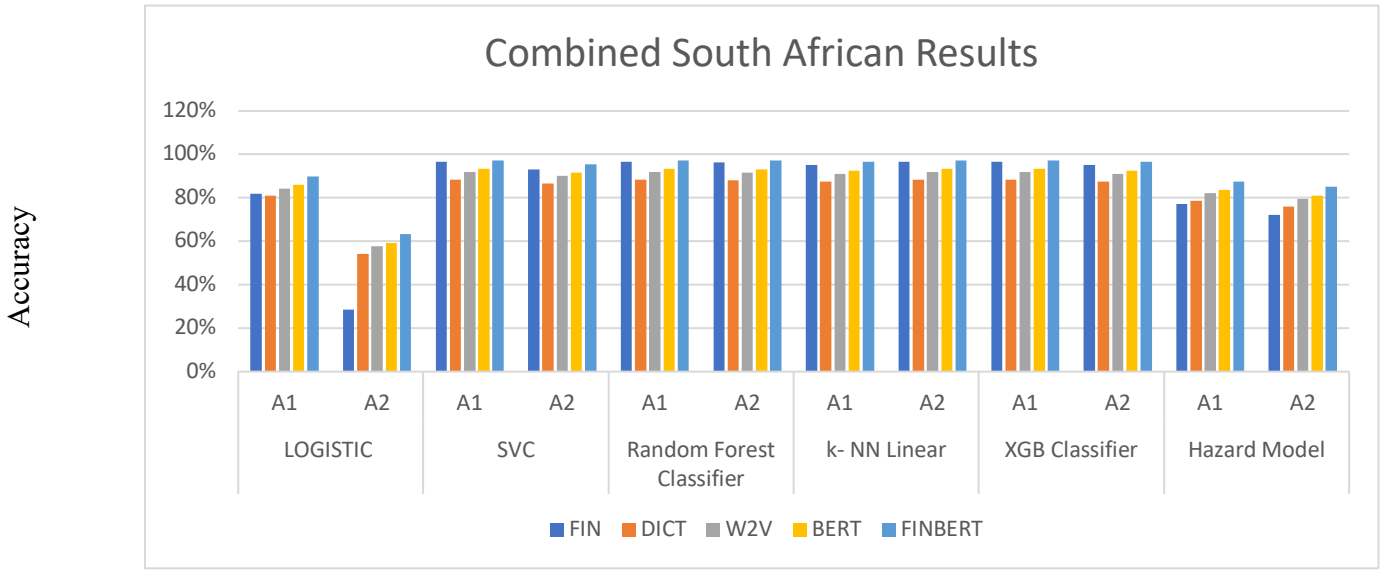
4.3.3 South Africa

The results as depicted in figure 45 are a comparison of the machine learning models appended to the natural processing models for the South African database. The results are supportive of the use of textual analysis in getting a better and more accurate prediction of bankruptcy in the South African context. The dictionary approach has an average of 2% reduction on the entire set of chosen machine learning models. This, therefore, means that the dictionary based sentiment analysis has no fundamental effect in improving corporate bankruptcy prediction in the South African setting. The word2vec model provides an additional 1% from the original models. The generalised BERT model improves estimation by 3% with the FINBERT model. which is the chosen domain adaptive model, having the biggest improvement in prediction by placing an improvement of 5% on the model baseline model. The margins in the South African context are small however, textual analysis still fundamentally improves the predictability of bankruptcy.

The best performing models in the South African context are the Random Classifier, SVM, XGB and k-NN Linear models appended with the FINBERT sentiment analysis model which achieved the same result of 97%. The domain adapted textual analysis seemingly improves results however it does not influence predictability as significantly as one would have hoped in the South African context. A possible reason for this is that the MD&A section in South Africa is neither standardised

nor regulated which allows for a more conservative disclosure with a lot more limited private information sets as was found by Li (2010). The paper by Li (2010) also explains the importance of cultural patterns in the examination of textual analysis which may be a problem as the disclosures assessed in the South African context are in English whilst South Africa in itself is a multilingual society that has African languages, more specifically, isiZulu as a dominant language. This would lead to loss in cultural patterns that would be picked up by our textual and sentiment analysis models, Li (2010). The result of this is that these cultural patterns may not be consistent throughout our dataset and the effect of this is that the models would not be able to develop historical patterns to a point of being able to predict future patterns. This, however, proposes a more curious underlying notion that the models would be better suited in analysing MD&A disclosures that are written in African languages in the South African context. These would have more standardised cultural patterns as well as unlock more private information as studied by Li (2010).

Figure 45: Combined Results of the South African Dataset



4.4 Summary of results

In this section, the predictive accuracy of the Altman (1968) is compared to the machine learning models as well as the natural processing models. The core results of the study show a general improvement of the original Altman model . The original Altman model, as stated before, is a

measurement of bankruptcy using working capital, total assets, equity, retained earnings, sales and earnings before tax and interest as factor inputs to decide if a firm will go bankrupt. As an original model, the Altman factor model was able to capture 85% of corporate bankruptcy during bankruptcy years and 76% during non-bankruptcy years in South Africa. It was also only able to capture 72% of corporate bankruptcy during the bankruptcy years in the United States and only 20% during non-bankruptcy years. Similarly, the original Altman (1968) model was able to capture 60% of the bankrupt firms during bankruptcy years for the United Kingdom and only 40% during non- bankruptcy years.

The machine learning results are unanimously an improvement from the baseline Altman (1968) results. This is seen with the SVM, Random Forest and XGB Classifier being the highest most predictive models with a prediction factor of 96% each for the South African dataset during bankruptcy years. The random forest model, on the other hand, achieved the highest result of 98% for the United Kingdom dataset and 92% for the United States of America dataset during bankruptcy years. This is compared to the 85%, 72% and 60% prediction capability for the Altman (1968) results for South Africa, the United States of America and United Kingdom respectively. The superiority of the machine learning models remains consistent during non-bankruptcy years as we see that the XGB model achieves the highest predictive result for South Africa at 97% predictive value compared to the 72% achieved by the Altman (1968) model. Similarly, we find that the United States of America dataset provided a predictive value of 92% via the k-NN Linear compared to the 20% by the Altman model. Lastly, the strength of the machine learning models is reserved in the United Kingdom dataset with the highest machine learning models being the Logistic models, the Random Forest Classifier and the XGB models. They achieved a height of 97% respectively for the United Kingdom dataset as compared to the 40% achieved during non- bankruptcy years.

We then introduced textual analysis, having been broken into four models namely; Dictionary, Word2Vec, Bert and FINBERT approaches. These models, when appended to the machine learning results ,improve the predictability of bankruptcy substantially and consistently in all three countries. We also see that the transformer models, the BERT and FINBERT models, have a more substantial improvement in bankruptcy prediction than the corpus based models. Lastly, the role of domain adaption is strictly highlighted as we saw a disproportionate improvement in bankruptcy

prediction as a result of sentiment being captured by the FINBERT model more accurately than the generic BERT model.

5 Conclusion

In conclusion, we study whether context specific textual analysis by the BERT model would have an improvement in bankruptcy prediction accuracies in the United States of America, United Kingdom and South African contexts. The fashion by which this was done was to develop the models using financial factors that best describe corporate performance then expose these factors to machine learning models with the ultimate step of appending the financial factors to context-based sentiment analysis models to get a final result. Our study supports the use of sentiment analysis in predicting corporate bankruptcy as we have been able to show that textual analysis, when taken into consideration with financial analysis, has a stronger accuracy in predicting corporate bankruptcy. These results support the notion brought forward by Singh and Mishra (2016) who argued that financial factors alone are not adequate in fully predicting corporate bankruptcy as this is a decision made mostly with other factors other than firm financial performance being taken into consideration.

We find that MD&A sections hold very valuable information about company performance as well as impending director actions in as far as bankruptcy filing is concerned. This supports the findings by Mayew et al., (2015) who found that MD&A sections have a significant effect in understanding corporate performance as well as corporate bankruptcy estimations. We also established that cultural patterns play a substantial thus, the more culturally English the origin of the MD&A disclosure is, the more predictive our sentiment analysis models would be. This is similar to the realisation by Li (2010) who found that cultural patterns have a direct effect in the proficiency of textual analysis. This may be at a country level or in the case of Li (2010), at an even more minute level such as industry or profession. Similarly, we found that the possible case where MD&A disclosures being written in English in countries that may not have been predominantly English, may lead to the necessary patterns required for sentiment analysis being locked away. The suggestion from this applies the sentiment analysis in the language that better exposes the cultural patterns of the country of origin of the MD&A disclosure.

It was theorised in the study that having stronger bankruptcy models would allow for investors, governments as well as researchers to have the opportunity to make adequate decisions regarding the company failure to allow for reduced economic impact that may arise from such a failure. Our results show that sentiment analysis using the BERT and domain adaptive model is consistent with

at least one year prior prediction to the filing of bankruptcy. This is an improvement from the Altman (1968) model and this improvement was consistent in both the generalised BERT model as well as the FINBERT model over bankruptcy and non- bankruptcy years. This would be able to reduce the economic impacts described by Buckley (2009) and Park et al., (2020).

This study was able to close the knowledge gap between text data and financial data from annual reports by combining them into a single interpretable result. This combination remained consistent over the multiple years of study which is in line with findings by Tinoco and Wilson (2018) who found that the financial and market-based models are not predictive over extended time periods.

The study found that the best performing model was the SVM, Random Forest Classifier and XGB model, with the Random Classifier being the best and most consistent model of the three. Whilst comparison of models is a study of the data sciences, we see that this result was consistent with the findings by Guha (2021) who found that the XGB classifier model was a stronger predictor model than the Random Forest Classifier model.

One of the biggest highlights of this study was the importance of large training data sets and the manner in which they were structured which would have an effect on the overall results of the models. Similarly, the models used can only be understood in conjunction with the exploration of the data. One of the biggest challenges faced with this was finding standardised techniques of such data exploration.

Whilst machine learning models have proven to be incredibly powerful for prediction and classification problems, it is seen by this study that the models are incredibly reliant on the quality and size of the data used. The rigorous emphasis on data exploration in this study is all but an important task which only serves to validate the results of the machine learning models. In the case of the Altman original model, the only scrutiny that was applied to the model was its assumptions, not the financial variables or data used to classify the corporate bankruptcy. Furthermore, since the model did not require training, the size of the data was not a necessary feature as the model was based more on a theoretical framework rather than discovered by machine learning iteration. This makes data quality and size one limitation of the study. The second limitation found in this study is the lack of reproducibility of the variable relationships that are developed by the machine learning models. Machine learning models tend to have deep seeded relationships between the factors

and data used which are not interpretable nor known to the programmer. This makes validation counterintuitive, and tests rather than direct observation are required to validate final results.

We have been able to show that bankruptcy prediction can be done to an incredibly high degree with the help of machine learning models as well as sentiment BERT based sentiment analysis. Further studies can best utilise the results of the research in this study to find a more theoretical justification of the models used. This would open up a new branch of finance which is focused on model selection and interpretation. Similarly, the relationship between the Humanities and Finance can further be explored by utilising the same framework to explore sentiment analysis in the field of Finance with the inclusion of other languages, more specifically African languages.

The most impactful limitation of the study is the lack of transparency and interpretability of the complex relationships derived by the machine learning models. This drawback has the effect in which we are trusting the models and their results without a set method to validate those results. The consequence of this is that whilst the results we get from this fit the theory and expectation, there is no true way of fully accepting the outcome of the machine learning models. The second is the prevalence of bias and discrimination in our result. The models are data driven and as a consequence, the results from these models are exposed to the biases prevalent in the data set. This means that certain relationships that eventually come out as a result of machine learning models could simply be the articulation of these biases in the data set. The fundamental problem is that whilst there are methods that can be employed, such as SMOTE (Synthetic Minority Oversampling Technique), which was used in this dissertation, it is incredibly difficult to fully eliminate data bias. Lastly, computational power is an impactful limitation as it affects the efficiency by which results can be found and in some cases outputted. This could mean that certain models which have better results are inaccessible due to computational power and with that their contribution as well.

Each limitation could therefore be seen as a future consideration in which the study of bankruptcy could go. The first, could be to validate some of the models and the results through test split and cross validation models. One could also use much more powerful computational models and test whether their results are first, consistent with those studied in this dissertation and second, provide any further predictive power to bankruptcy. Lastly, the study of sentiment analysis is strongly linked to the study of textual analysis which is language based. The inclusion of African languages

could enrich sentiment analysis as we could be able to ascertain whether their models holds through various cultural backgrounds and language barriers.

References

Abdullah, N. A. H., Halim, A., Ahmad, H., and Rus, R. M. (2008). Predicting corporate failure of Malaysia's listed companies: Comparing multiple discriminant analysis, logistic regression and the hazard model. *International research journal of finance and economics, 15*, 201-217.

Adnan Aziz, M. and Dar, H.A. (2006), "Predicting corporate bankruptcy: where we stand?", *Corporate Governance, Vol. 6 No. 1*, pp. 18-33. https://doi.org/10.1108/14720700610649436

Baldwin, J. R., and Statistics Canada (1997). Failing Concerns [electronic Resource]: Business Bankruptcy in Canada. Statistics Canada*, Micro-Economic Analysis Division*

Begley, J., Ming, J. and Watts, S. (1996) Bankruptcy classification errors in the 1980s: An empirical analysis of Altman's and Ohlson's models. *Rev Acc Stud* ***1,*** *267–284*. https://doi.org/10.1007/BF00570833

Bisogno, M and Restaino, Marialuisa and Carlo, A. (2018). Forecasting and preventing bankruptcy: A conceptual review. *African journal of business management.* 12. 10.5897/AJBM2018.8503.

Boritz, J.E., Kennedy, D.B. and Sun, J.Y. (2007), Predicting Business Failures in Canada. *Accounting Perspectives, 6: 141-165*. https://doi.org/10.1506/G8T2-K05V-1850-52U4

Cahyani, D. E., and Patasik, I. (2021). Performance comparison of TF-IDF and Word2Vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics, 10(5), 2780-2788*. https://doi.org/10.11591/eei.v10i5.3157

Campbell, J.L., Chen, H., Dhaliwal, D.S. (2014). The information content of mandatory risk factor disclosures in corporate filings. *Rev Account Stud 19, 396–455*. https://doi.org/10.1007/s11142-013-9258-3

Colm, K., Sha, L., (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis Volume 33, Pages 171-185, ISSN 1057-5219*. https://doi.org/10.1016/j.irfa.2014.02.006.

Devlin, J., Chang, M. W., Lee, K., and Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *Association for Computational Linguistics.* https://doi.org/10.48550/arXiv.1810.04805.

Drotár, P., Gnip, P., Zoričak, M., and Gazda, V. (2019). Small-and medium-enterprises bankruptcy dataset. *Data in brief, 25*. DOI:https://doi.org/10.1016/j.dib.2019.104360

Emerson, M., Ricardo, M., Nadia, M., The effect of corporate bankruptcy reorganization on consumer behaviour, *European Research on Management and Business Economics*, Volume 26, Issue 2, 2020, Pages 96-102, ISSN 2444-8834, https://doi.org/10.1016/j.iedeen.2020.03.002.

Tseng, F., and Hu, Y. (2010). Comparing four bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks*, Expert Systems with Applications, Volume 37, Issue 3,* Pages 1846-1853, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2009.07.081.

Fejér-Király, G. (2015) Bankruptcy prediction: A survey on evolution, critiques, and solutions. *Acta Universitatis Sapientiae, Economics and Business, 3(1), 93-108*. DOI: 10 1515/auseb-2015-0006

Barboza F., Kimura H., Altman E. (2017) Machine learning models and bankruptcy prediction, *Expert Systems with Applications, Volume 83, Pages 405-417, ISSN 0957-4174*, https://doi.org/10.1016/j.eswa.2017.04.006.

Salton, G. and Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. In *Information Processing and Management*, pages 513-523.

Gnecco, G., Kůrková, V, Sanguineti, M., Can dictionary-based computational models outperform the best linear ones?, *Neural Networks, Volume 24, Issue 8, 2011, Pages 881-887*, ISSN 0893-6080, https://doi.org/10.1016/j.neunet.2011.05.014.

Zhang, G., Hu M., Patuwo, B., Indro, D. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis, *European Journal of Operational Research,Volume 116, Issue 1, Pages 16-32,ISSN 0377-2217*, https://doi.org/10.1016/S0377-2217(98)00051-4

Hiew, G., Huang, X., Mou, H., Li, D., Wu, Q., & Xu, Y. (2019). BERT-based financial sentiment index and LSTM-based stock return predictability. *PeerJ Comput Sci*. arXiv preprint arXiv:1906.09024.

Hillegeist, S.A., Keating, E.K., Cram, D.P. et al. (2004). Assessing the Probability of Bankruptcy. *Review of Accounting Studies* **9,** 5–34. https://doi.org/10.1023/B:RAST.0000013627.90884.b7

Jindal, A., Dua, A., Kaur, K., Singh, M., Kumar, N., and Mishra, S. (2016). Decision tree and SVM-based data analytics for theft detection in smart grid. *IEEE Transactions on Industrial Informatics, 12(3), 1005-1016*

Lee, J., Yoon W., Kim S., Kim D., Kim S., So C, Kang J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, Volume 36, Issue 4, Pages 1234–1240, https://doi.org/10.1093/bioinformatics/btz682

Karels, G.V. and Prakash, A. (1987). Multivariate Normality and Forecasting of Business Bankruptcy, *Journal of Business Finance & Accounting*, 14, 4, 573- 593

Knowles-Barley, S., Kaynig, V., Jones, T. R., Wilson, A., Morgan, J., Lee, D., ... and Pfister, H. (2016). Rhoananet pipeline: Dense automatic neural annotation. https://doi.org/10.48550/arXiv.1611.06973

Kücher, A., Mayr, S., Mitter, C. et al. (2020). Firm age dynamics and causes of corporate bankruptcy: age dependent explanations for business failure. *Rev Manag Sci* **14,** 633–661. https://doi.org/10.1007/s11846-018-0303-2

Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. *Journal of accounting literature, 29(1), 143-165*.

Lombardo, G.; Pellegrino, M.; Adosoglou, G.; Cagnoni, S.; Pardalos, P.M.; Poggi, A. Machine Learning for Bankruptcy Prediction in the American Stock Market: Dataset and Benchmarks. *Future Internet 2022,* 14, 244. https:// doi.org/10.3390/fi14080244

Loughran, T. and McDonald, B. (2011), When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66: 35-65. https://doi.org/10.1111/j.1540-6261.2010.01625.x

Lussier, R. N. (1995). A nonfinancial business success versus failure prediction mo. Journal of Small Business Management, 33(1), 8.

Mabe, Z. (2019). Alternatives to bankruptcy in South Africa that provides for a discharge of debts: lessons from Kenya. Potchefstroom Electronic Law Journal (PELJ), 22(1), 1-34. https://dx.doi.org/10.17159/1727-3781/2019/v22i0a5364

Tinoco, M., Holmes P., Wilson, N. (2018). Polytomous response financial distress models: The role of accounting, market and macroeconomic variables, *International Review of Financial Analysis, Volume 59, Pages 276-289*, ISSN 1057-5219, https://doi.org/10.1016/j.irfa.2018.03.017.

Lang M., Stice-Lawrence L. (2015). Textual analysis and international financial reporting: Large sample evidence, *Journal of Accounting and Economics, Volume 60, Issues 2–3, Pages 110-135, ISSN 0165-4101*, https://doi.org/10.1016/j.jacceco.2015.09.002.

Naili, M., Chaibi, A., Ghezala H. (2017). Comparative study of word embedding methods in topic segmentation, *Procedia Computer Science, Volume 112, Pages 340-349, ISSN 1877-0509*, https://doi.org/10.1016/j.procs.2017.08.009.

Stevenson, M., Mues, C., Bravo B. (2021). The value of text for small business default prediction: A Deep Learning approach, *European Journal of Operational Research, Volume 295, Issue 2,* Pages 758-771, ISSN 0377-2217, https://doi.org/10.1016/j.ejor.2021.03.008.

Merwin, C.L. (1942) Financing Small Corporations in Five Manufacturing Industries, 1926-1936: A Dissertation in Economics. Financing Small Corporations in Five Manufacturing Industries, *National Bureau of Economic Research*, 1926-36.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. https://doi.org/10.48550/arXiv.1301.3781

Muller, G. H., Steyn-Bruwer, B. W., and Hamman, W. D. (2009). Predicting financial distress of companies listed on the JSE-A comparison of techniques. *South African Journal of Business Management, 40(1), 21-32.*

Narvekar, A., and Guha, D. (2021). Bankruptcy prediction using machine learning and an application to the case of the COVID-19 recession. Data Science in Finance and Economics, 1(2), 180-195.

O'Shea, K., and Nash, R. (2015). An introduction to convolutional neural networks. https://doi.org/10.48550/arXiv.1511.08458

Oz, I., and Simga-Mugan, C. (2018). Bankruptcy prediction models' generalizability: Evidence from emerging market economies, *Advances in accounting, Elsevier, vol. 41(C),* pages 114-125. DOI: 10.1016/j.adiac.2018.02.002

Petch, D., and Nelson. (2022) Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology, *Canadian Journal of Cardiology, Volume 38, Issue 2,* Pages 204-213, ISSN 0828-282X, https://doi.org/10.1016/j.cjca.2021.09.004.

Perboli, G., Arabnezhad E. (2021) A Machine Learning-based DSS for mid and long-term company crisis prediction. Expert Syst Appl 174: 114758.

Piotrowski, M. (2012). Natural language processing for historical texts. Synthesis lectures on human language technologies, 5(2), 1-157. https://doi.org/10.2200/S00436ED1V01Y201207HLT017.

Buckley, P. (2009). Internalisation thinking: From the multinational enterprise to the global factory, *International Business Review, Volume 18, Issue 3*, Pages 224-235, ISSN 0969-5931, https://doi.org/10.1016/j.ibusrev.2009.01.006

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1(8), 9.

Lacher, R.C., Coats, P., Sharma, S, Fant L. (1995). A neural network for classifying the financial health of a firm, *European Journal of Operational Research,Volume 85, Issue 1,* Pages 53-65, ISSN 0377-2217, https://doi.org/10.1016/0377-2217(93)E0274-2.

Ricardo, D. (1817). On the Principles of Political Economy and Taxation (John Murray, London). In: Sraffa, P., Ed., The Works and Correspondence of David Ricardo, Vol. 1, *Cambridge University Press, Cambridge,* 1951

Moraes., R. (2021), Transnational Activism and Domestic Politics: Arms Exports and the Anti-Apartheid Struggle in the UK–South Africa Relations (1959–1994), *Foreign Policy Analysis, Volume 17, Issue 4*. orab023, https://doi.org/10.1093/fpa/orab023

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

Smith R. F. & Winakor A. H. (1935). *Changes in the financial structure of unsuccessful industrial corporations*. University of Illinois.

Jones., S, Johnstone D., Wilson R. (2015). An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes, *Journal of Banking & Finance, Volume 56,* Pages 72-85, ISSN 0378-4266,https://doi.org/10.1016/j.jbankfin.2015.02.006.

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks, *Expert Systems with Applications, Volume 117*, Pages 287-299, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2018.09.039.

Tam, K. Y., and Kiang, M. Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. *Management science, 38(7),* 926-947. https://doi.org/10.1287/mnsc.38.7.926

Taylor, G., Burmeister, R., Xu, Z., Singh, B., Patel, A. and amp; Goldstein, T.. (2016). Training Neural Networks Without Gradients: A Scalable ADMM Approach. *Proceedings of The 33rd International Conference on Machine Learning*, in Proceedings of Machine Learning Research 48:2722-2731 Available from https://proceedings.mlr.press/v48/taylor16.html.

Tetlock, P.C., Saar-Tscehansky, M. and Mackassy, S. (2008), More Than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance, 63*: 1437-1467. https://doi.org/10.1111/j.1540-6261.2008.01362.x

Timmermans, M., and Finance, M. (2014). US corporate bankruptcy predicting models. Unpublished Master's thesis, Tilburg University, School of Economics and Management, Tilburg.

Young, T., Hazarika, D., Poria, S., and Cambria E,. (2018). "Recent Trends in Deep Learning Based Natural Language Processing [Review Article]," in IEEE Computational Intelligence Magazine, vol. 13, no. 3, pp. 55-75, doi: 10.1109/MCI.2018.2840738

UNCTAD. (2022). World investment report. Special Economic Zones, United Nations.

Wei, W., et al. "Financial numeral classification model based on BERT." NII *Testbeds and Community for Information Access Research: 14th International Conference,* NTCIR 2019, Tokyo, Japan, June 10–13, 2019, Revised Selected Papers 14. Springer International Publishing, 2019.

Wang, Z., Joshi, S., Savel'ev, S. et al. (2018). Fully memristive neural networks for pattern classification with unsupervised learning. Nat Electron 1, 137–145. https://doi.org/10.1038/s41928-018-0023-2

Mayew, W., Sethuraman, M., and Venkatachalam, M. (2015). MD&A Disclosure and the Firm's Ability to Continue as a Going Concern. *The Accounting Review 1 July 2015*; 90 (4): 1621–1651. doi: https://doi.org/10.2308/accr-50983

Wilson, R.L. and Sharda, R. (1994) Bankruptcy Prediction Using Neural Networks. Decision Support Systems, 11, 545-557. https://doi.org/10.1016/0167-9236(94)90024-8

Yamashita, R., Nishio, M., Do, R.K.G. et al. (2018) Convolutional neural networks: an overview and application in radiology. *Insights Imaging* ***9*****,** 611–629. https://doi.org/10.1007/s13244-018-0639-9

Yoon, K. (2014). Convolutional Neural Networks for Sentence Classification. Proceedings of the 2014 *Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha,* 1746-1751. arXiv: 1408.5882 https://doi.org/10.3115/v1/D14-1181

Young Woong Park, Jennifer Blackhurst, Chinju Paul and Kevin P. Scheibe. (2021). An analysis of the ripple effect for disruptions occurring in circular flows of a supply chain network, *International Journal of Production Research*, DOI: 10.1080/00207543.2021.1934745

Zhi-qiang, J., Hang-guang, F. and Ling-jun, L. (2005). Support Vector Machine for mechanical faults classification. J. Zheijang *Univ.-Sci. A 6, 433–439*. https://doi.org/10.1631/jzus.2005.A0433

Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress *Prediction Models. Journal of Accounting Research*, 22, 59–82. https://doi.org/10.2307/2490859

Zopounidis, C., Doumpos, M. (1999). A Multicriteria Decision Aid Methodology for Sorting Decision Problems: The Case of Financial Distress. Computational Economics 14, 197–218. https://doi.org/10.1023/A:1008713823812