Investigating Accounting Patterns for Bankruptcy and Filing Outcome Prediction using Machine Learning Models

Derek Snow

Department of Accounting and Finance

University of Auckland

Owen G. Glenn Building

12 Grafton Road, Auckland, 1010

+6422 3922 056

d.snow@auckland.ac.nz

Contents

Tables

Figures

Abstract

I study the use of non-linear models and accounting inputs to predict the occurrence of litigated bankruptcies and their associated filing outcomes. The main purpose of this study is to identify the accounting patterns associated with bankruptcies. The filing outcomes include, among others, how long the bankruptcy process will endure, whether the firm will successfully emerge after the bankruptcy period, whether the bankruptcy is tortious, and whether it will involve an asset sale. The study highlights the importance of previously unidentified accounting variables that are useful in predicting bankruptcies and bankruptcy outcomes. The study categorises predictor variables in accounting dimensions to empirically identify the importance of each dimension to the prediction tasks. The high dimensionality of the gradient boosting machine allows us to identify and explain the nonlinear interactions between a wide range of variables.

I. Introduction and Motivation

This study makes use of a modern gradient boosting machine (GBM), XGBoost, to predict litigated bankruptcies and filing outcomes. A GBM sequentially builds multiple decision tree models from which the final outcome is predicted. To ensure that the best model is used to investigate the variable importance scores, I compare the GBM with four state-of-the-art non-linear models and a Logit model. The overall GBM model predicts bankruptcy with an accuracy of 97% and an ROC AUC[1] of close to 96% compared to the 69% accuracy and 71% ROC (AUC) of a standard Logit Model. The selected models use a wide spectrum of dollar accounting values and ratios as inputs, including price ratios embraced under the 'valuation' category. Consistent with past research, this study reports the ROC (AUC) score, accuracy, cross-entropy, and error rates associated with the performance of the prediction model. Further analysis also includes the use of a confusion matrix.

The models used in this study are different from parametric models and do not rely on significance tests; instead, they rely on a data-centric approach that looks at the predictive ability of a parameter based on the variable selection and ranking in the nodes of the trees. In this chapter, I argue that GBMs provide for an improved analysis of accounting and associated bankruptcy patterns compared to that of linear models. First, these models empirically report on the non-linear relationships of variables. This is an important attribute as financial data is likely to exhibit non-linearities. GBMs do not require one to predefine interactions and polynomial transformations to improve model performance. Furthermore, these models are resistant to multicollinearity issues since they simply ignore weaker correlated variables; they are also resistant to parameter clogging, in that redundant variables are simply ignored, which significantly improves on the stability of the model; the hyperparameters of these models can further be adjusted to lessen model complexity and overfitting (also known as regularisation), all of which contribute to more realistic variable importance measures compared to linear models’ effect and significance measures.

---

[1] ROC AUC (receiver operating characteristics area under curve) plots the true positive relative to the false positive rate with respect to all decision probability thresholds (the threshold is a value from 0%-100% used to classify an observation as 1 as opposed to 0). When Type 1 errors (FP) and Type 2 errors (FN) are minimised across all decision thresholds, this value is maximised. The ROC AUC score therefore provides an integral based performance measure of the quality of the classifier. A value of 50% is expected for random noisy predictions. Generally, values from 80%-100% are considered as great classifiers. It is arguably the best single number machine learning researchers have in measuring the performance of a classifier (Bradley, 1997; Fawcett, 2006; Powers, 2011).

The vast majority of past high dimensional bankruptcy studies limit themselves to theoretically identified variables in prior literature (Jones, 2017; Kim & Upneja, 2014). Due to past literature's lack of identifying and isolating important high dimensional interaction pairs, this study does not limit itself to previously identified variables. This study focuses on a comprehensive range of accounting variables and their transformations. The expectation is that these inputs reflect all the necessary information to closely match the current research benchmarks (Jones, Johnstone, & Wilson, 2017; Volkov, Benoit, & Van den Poel, 2017). The simplicity of accounting measures is beneficial to the theoretical discussions associated with the variables. This study is the first to identify and describe high dimensional interactions. It describes the simultaneous interactions and marginal effects of up to three variables on the response variable.

The use of the XGBoost model (GBM) is largely driven by its success in practical domains; for example, it has been shown to be highly effective in data science competitions (Chen & Guestrin, 2016). The GBM model has the benefit of being interpretable, albeit not as well as logit models; however, the identification of non-linear interactions warrants its use. Multiple studies have shown that ensemble techniques can be used to improve financial distress prediction (Deligianni & Kotsiantis, 2012; Sun & Li, 2012). Many effective machine learning approaches have been used to predict default in recent years including artificial neural networks. In this study, I report on the greater usefulness of ensemble models over deep learning and other sophisticated models. I have tested multiple recurrent neural networks (RNNs), feedforward neural networks (FNNs) and convolutional neural networks (CNNs) architectures. I report the results of the best performing neural network, a deep convolutional neural network (DCNN), previously used for large temporal financial datasets (Chen, Chen, Huang, Huang, & Chen, 2016). I show that this particular neural network model performs worse than the GBM model but outperforms other neural network models.

II. Literature

Bankruptcy prediction research can largely be divided into the identification of 'symptoms' that lead to bankruptcy (Dambolena & Khoury, 1980; Gombola & Ketz, 1983; Scott, 1981) and studies that compare the performance of different bankruptcy prediction models (Altman, 1968; Ohlson, 1980). These two strains of research remain intact in modern bankruptcy prediction research. However, in recent years the traditional methods and processes have been uprooted by the development of advanced machine learning models.

Traditional statistical models have been largely dropped in favour of high dimensional models (Barboza, Kimura, & Altman, 2017; du Jardin, 2017; Jones, 2017; Liang, Lu, Tsai, & Shih, 2016). These advanced models present numerous advantages in flexibility, efficiency, and most importantly, enhanced prediction quality (Jones, 2017). The purpose of this chapter is to identify the accounting-related symptoms of bankruptcy in higher dimensions, and consequently I will give some attention to model performance to confirm the validity of identified predictor variables.

In recent years, the accuracy measure has been largely replaced by the ROC (AUC) score and other metrics (Bauer & Agarwal, 2014). Traditional significance tests of predictor variable performance have also been substituted by higher dimensional classification tree measures such as Gini Importance, Information Gain, and Split Frequency, as well as their relative measure counterparts like Relative Variable Importance (Behr & Weinblat, 2017; Jones et al., 2017; Mselmi, Lahiani, & Hamza, 2017). These are all data-centric approaches that look at the predictive ability of a parameter based on the variables selected and ranked in the nodes of the trees instead of significance tests. The application of these measures has practical and theoretical implications. Kleinberg, Lakkaraju, Leskovec, Ludwig, and Mullainathan (2018), for example showed that machine learning importance measures can be used to understand and improve judges' decision-making in trials.

Recent models perform internal variable selection procedures, mostly removing the need for researchers to prune model inputs before feeding them into an algorithm (Mullainathan & Spiess, 2017). This means that the model can decide from a wide range of variables what it deems to be important without human intervention. With a sufficiently broad set of inputs, researchers can simply copy the model as used in one task and apply it to another, especially when making use of automated freeware to execute the necessary hyperparameter tuning (to optimise the model hyperparameter inputs). This strategy of reusing model architecture and inputs is used in this study to predict filing characteristics such as bankruptcy proceeding durations, survival, filing chapters, asset sales, and tortious claims. The readers mostly interested in the *results* of this study should read the next Contribution and Hypothesis section and, after that, move straight to the first table on page 19. For a further exposition about the research related to the Predictor Variables, Categories, Models, Predictive Power, and Filing Outcomes see the *Literature Addendum* in *Appendix B.*

III. Contribution and Hypothesis

This study contributes to the literature in several dimensions. It is the first study to make use of an XGBoost model to predict bankruptcies and to identify important accounting patterns associated with bankruptcies (Zięba et al., 2016). It is also the first study to implement a DCNN[2] (a biologically inspired variant of multilayer perceptron), which has shown great promise in other domains such as image recognition (Sharif Razavian, Azizpour, Sullivan, & Carlsson, 2014). This study further compares advanced deep learning models with modern decision tree ensemble models. To my knowledge, this is also the first study that uses a stacked model to improve prediction quality. The bankruptcy period spans 37 years (1980 to 2017), which is the longest ranging sample period of all cited literature. Furthermore, few studies acknowledge the theoretical equivalence between dollar-denominated accounting values and accounting ratios in the prediction of bankruptcy using higher dimensional models. In this study, I attempt to show that accounting values can at least be as important as ratios in predicting bankruptcies. It is also the first study to analyse interactions at an interaction depth of three variables.

Furthermore, this is the first attempt to rank the different accounting dimensions according to empirical ranking methods. This study contributes to the literature by using higher dimensional techniques to identify a structural difference in variable and category importance before and after the global financial crises (GFC). Most importantly, it is also the first study that attempts to predict filing outcome responses such as whether the bankruptcy process will endure for longer than a year, whether the firm will successfully emerge after the bankruptcy period, whether the bankruptcy is tortious, and whether or not the bankruptcy will involve asset sales. This is the first step towards successfully using high dimensional models to both improve prediction quality and predictor variable analysis. This study shows, from a categorisation of input variables, that Assets & Liability values, Solvency ratios, and Income values are the most important dimensions unlike past research that emphasises the importance of Profitability, Valuation, and Liquidity values. I put this difference down to the inability of linear models to capture the true reality of high dimensional relationships.

This study is also one of the first to investigate higher dimensional interactions and importance measures. One of the resulting interactions shows that when firms have large R&D programs, they are less likely to become bankrupt, all else equal. Some researchers

[2] Deep Convolutional Neural Network

have historically argued the opposite and said that there is a ‘failure-inducement’ effect in firms’ effort to push for innovation when performance falls (Antonelli, 1989). I find that a handful of variables have a strong association with bankruptcy, many of which have not been noted by past research, such as the level of Stockholder's Equity, Depreciation & Amortization, and the Research & Development to Sales ratio.

This study includes a few additional steps to enhance the robustness of the results. It has the most imbalanced and lowest bankruptcy-to-healthy firm ratio of all decision tree and boosting related bankruptcy studies. It investigates bankruptcy prediction across all industries. The size effects of bankruptcy have been kept to a minimum by establishing minimum constraints on firm size. In addition, more than 10% or 120 of the bankrupt firm-year observations are filed on the premise of tortious claims, 70% of which relates to fraud. Chaudhuri and De (2011) observed that no models have yet been successful in detecting corporate fraud. I similarly find that if the fraud does not go hand in hand with financial distress, it is hard to predict these fraudulent bankruptcies. Therefore, the results reported in this study are much more conservative than those of past research. Following is a summary of key findings:

1. A Gradient Boosting Model (XGBoost) outperforms deep neural networks (DCNN, FFN) in prediction quality as measured by Accuracy and ROC (AUC) scores.
2. In a high dimensional setting, financial ratios have lower aggregate predictive ability over dollar-denominated accounting values as a result of linear constraints imposed on them.[3]
3. Solvency-related accounting-ratios are an important accounting ratio dimension for bankruptcy prediction compared against Profitability, Valuation, Liquidity, and Efficiency ratios.
4. Feature importance changes significantly before and after the GFC.
5. By using that same inputs as the bankruptcy prediction task, the GBM model is able to predict important filing outcomes, such as how long the bankruptcy process will endure, whether the firm will successfully emerge after the bankruptcy period, whether the bankruptcy is tortious, and whether or not it will involve asset sales.

[3] The individual constituents to the ratios are not able to interact with other variables independently leading to a loss of predictive power.

IV. Data

I use a sample of large[4] public firm bankruptcy cases filed under Chapter 11 of the US Bankruptcy Code as obtained from UCLA BRD[5] and a control group of a random sample of large and healthy firms. The vast majority of insolvent companies seek protection under Chapter 11 (Altman, 2002). Although Chapter 11 may be the original filing request, the courts may later decide to do a full asset sale outside of Chapter 11 or ask the company to file under Chapter 7. Those observations are also included in the sample. The sample of firms only comprises publicly listed firms for which financial statements were available. The final sample comprises 33,242 healthy firm years and 1224 bankruptcy firm years from 1977 to 2016, with an average bankruptcy to healthy firm ratio of less than 4%, and a standard yearly deviation of more than 4 percentage points, highlighting the variability of bankruptcies over the sample period.

Large firms have been chosen to limit the noise when identifying the most important accounting value determinants in predicting bankruptcy and filing outcomes at a national level. The purpose of this study is not so much prediction success as it is the identification of important accounting variables and interaction effects. In saying that, good prediction success is necessary to validate the predictive power of variables. Consistent with past literature, firms are considered to be bankrupt if they filed for bankruptcy within one year. The accounting information is obtained from Compustat. In this study, I use simplified and standardised financial information for all firms; it includes accounting information from the Balance Sheet, Income Statement, and Cash Flow Statement. 70% of the bankruptcies occurred after 2000. Half of the bankruptcies emerged after the GFC. The BRD database is unique, in that it includes not just the date of filing, but also the date the case was disposed by the court and information on whether an asset sale transpired, whether the business re-emerged, whether the case has been tried under the law of tort, and lastly, information on chapter of filing. This data in this study therefore allows us to not just predict the occurrence of a bankruptcy but also predict the associated filing outcomes.

To obtain a large enough sample of bankrupt firms without having to deal with small firm bankruptcies, I collect up to three years of data to predict bankruptcies one and two

[4] The BRD database only collects information on large firms. A firm is large if the firm reports assets of more than $250 million as measured in 1979 dollars on the last 10-k filing before the bankruptcy case.

[5] "The BRD contains data on all of the more than one-thousand large public companies that have filed bankruptcy cases since October 1, 1979" http://lopucki.law.ucla.edu/

years in advance. When firms had missing data, I followed longitudinal imputation procedures by comparing multiple methods and selecting and implementing the best method.[6] Consistent with Ohlson (1980) and Jones and Hensher (2004), firms do not get removed from the dataset simply because they are recently or newly listed; as a consequence, a few firms in the sample had only a small amount of data.

Consistent with past research, the bankruptcy event is a binary response. All healthy firms are coded with **0** and failed firms are coded with a **1** in the two preceding years. This is necessary as we want to investigate the performance of the firm before the bankruptcy filing. It is not a requirement for all bankruptcy studies; it is only required for discrete choice models, as it is the only means to incorporate a time dimension. Duration type models such as hazard models are set up to predict time-to-event using survival functions, and for these models it is good practice to only label a firm as bankrupt in the year of the bankruptcy (Beaver, McNichols, & Rhie, 2005; Hillegeist, Keating, Cram, & Lundstedt, 2004). In this study, a firm entering into bankruptcy is labelled as bankrupt for two firm-year observations. This study also identifies the performance of the classifier for data that is coded as bankrupt only in the year before the bankruptcy.

V. Methods

This section provides a short discussion of the bankruptcy prediction methodology used in the study. Readers who are interested in the results of this study should feel free to skip this section and go straight to the first results in *Table 1* on page *19*. The empirical part of the study consists of steps such as the imputation of missing values, the creation of training, validation, and test sets, and finally the development and training of a model using hyperparameter tuning on validation sets. A trained model can be calibrated on a validation set to adjust the model parameters.[7] Each firm-year observation is described according to a set of variables and a response value. The trained algorithm is then used with new inputs against a pure holdout test set to assess its accuracy and ROC (AUC) score, among other things. This paper performs and reports more than ten robustness tests on XGBoost model performance.

Before delving into the prediction models, it is valuable to understand the cross-validation technique used in this study. The bankruptcy data is sorted by time, and the first

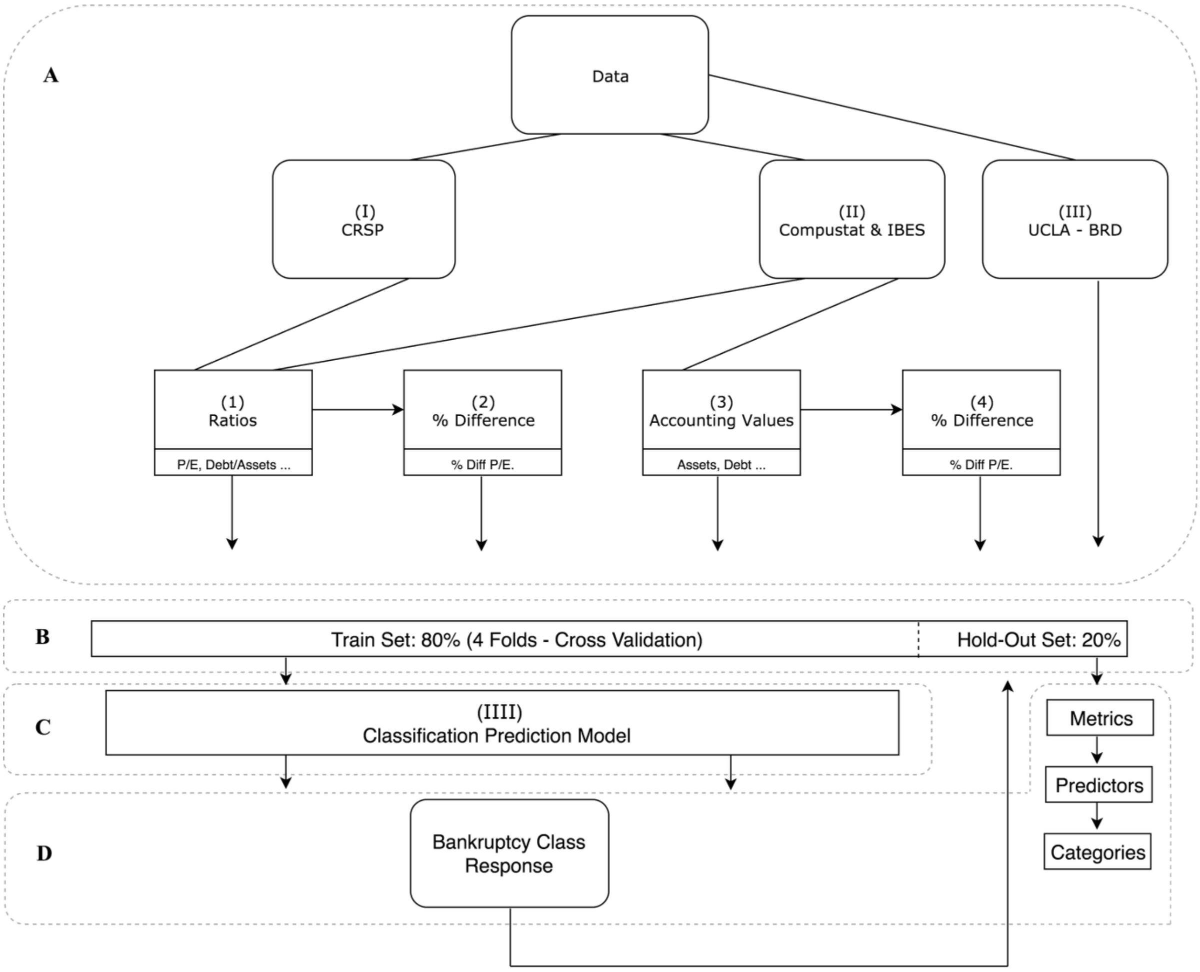
[6] See Appendix XI.D.2 page 182 for a description of the imputation process.
[7] The model parameters include the tree debt, the number of estimators, the learning rate, and many others.

20% of the data is used for training while a random selection of 15% of the remaining data is selected for the validation set. This validation set is used to perform model selection and hyperparameter adjustment after which it gets dropped indefinitely.

Using this approach ensures that the testing data never contains data that is older than the training data, which is a sensible step for preserving prediction integrity. I then use a unique longitudinal blocked form of performance evaluation that suits this form of bankruptcy prediction. This approach creates multiple models in time series to predict the subsequent periods' observation until the last model incorporates all the training data. This chapter uses the same validation approach as the first chapter and like before the final reported metric is the average across all the time splits. For more information on this method, see the cross-validation section XI.D.4 on page 67.

This study goes beyond simple machine learning; it does not only report on how well the prediction fits the test set, but it also sheds light on the most important predictor variables to bankruptcy. The first stage will follow a process similar to standard machine learning and the second part will focus more intensely on what these predictions tell us about bankruptcy and filing outcomes.

Figure 1: Process Tree



(A). I and II are the data sources used in this study (I) CRSP was used to obtain price data, (II) Compustat was used to obtain fundamental accounting information (III) Publicly available information was obtained from the UCLA LoPucki Bankruptcy Database supported with publicly available bankruptcy filings. Items (1) - (4) constitute the final variable groups used in this study. The accounting ratios (1) and values (3) have been transformed to get the respective yearly percentage changes (growth values), (2) and (4). The accounting ratios have been inspired by the WRDS Industry Financial Ratio Manual (2017). (B) identifies the cross-validation split process; details of this process can be found in XI.D.467 as it is more involved than this figure illustrates. (C) identifies the gradient boosting prediction model. (D) is the use of the bankruptcy response variables in the trained model to identify performance metrics on a hold-out set. These metrics are used to identify important variables and variable categories using partial dependence plots.

VI. Prediction

In this classification task I create a classification model (classifier), that assigns an observation to every class based on the learned patterns of a training set. In this study, the outcome to be predicted is 'bankrupt' or 'healthy' firm-years. The training consists of past observations where the classes are known. The model, therefore, learns class associations from the past patterns of explanatory variables commonly called features and maps this input data into a class outcome according to newly learned, weighted, and approximated functions. The XGBoost classification model used in this study is a probabilistic classifier that outputs a probability not unlike a Probit model. Throughout this study, 50% is selected as the decision rule; therefore, the chosen class is the one with the highest probability.

I use several statistics to report the out-of-sample accuracy. The most important metric is the ROC (AUC) score. The ROC (AUC) score is the benchmark statistic in classification research (Bradley, 1997; Fawcett, 2006; Ferri, Flach, & Hernndez-Orallo, 2002). Its use in bankruptcy research has also picked up; in just the last year alone, more than eight studies within neural network and boosted tree model bankruptcy prediction research have made use of this method (*Table A25*, *Table A26*). ROC curves plot the true positive relative to the false positive rate with respect to a threshold probability.[8] An area under the curve greater than 0.8 is considered to be a good classifier (classification model). A visual example of the ROC curve can be found in *Figure 2* on page 20.

The ROC curve is simply the relationship of the true positive rate to the false positive rate with respect to a probability threshold. The diagonal line can be described as the "line of luck" and has an AUC of 0.5. Generally classifiers should perform better than 0.5 to be of ay use at all. An AUC of 1 represents the best possible classification score with no Type I and Type II errors, that is, perfectly predictable. Conventionally ROC (AUC) scores above 0.8 and 0.9 indicate "good" and "great" classifiers, although the interpretation is domain dependent. For a visualisation of the ROC curve, the Type I and Type II errors have to be plotted against all threshold values. The primary ROC curve in this study is reported in *Figure 2.*

[8] The threshold probability is ordinarily set at 50%; this means that if the classification model predicts a 51% chance of a future bankruptcy then the observation would be classified as bankrupt. The 50% probability threshold can be adjusted to best fit the task at hand. For example, increasing the threshold would lead to fewer bankruptcy predictions but better-quality predictions.

Although the ROC (AUC) measure appears in a lot of research areas, it is somewhat limited in that it uses different misclassification cost distributions for different classifiers. An alternative measure has been proposed by Hand (2009) to avoid this limitation. Similar to Jones et al. (2017), I found no significant difference in using the H-score as opposed to the ROC (AUC) score. I therefore only report the more commonly known ROC (AUC) score to avoid confusion. I also present the use of ROC (AUC) measures in conjunction with an inductive technique to identify the importance of groups of predictor variables to explain model success. The ROC curve is further discussed in the Evaluation Section.

This study also reports on accuracy, false positive rate, and cross-entropy (negative average log-likelihood) metrics. The accuracy measure is not well-suited for imbalanced sets and can largely be ignored unless the reader assigns equal importance to the correct prediction of both healthy firms and bankrupt firms in a dataset where less than 4% of the actual observations are bankrupt firm years. The issue with the accuracy measure is that it does not look at class breakdown precision, nor does it provide evidence of true positives or true negatives values. The false positive rate similarly serves a somewhat limited role in this study. It is primarily used as validation metric to ensure that the trained model, from which the variable importance measure is derived, does not mistakenly predict healthy firms as being bankrupt (false positives) as this would undermine the validity of the variable importance measures and resulting variable ranking. The last reported metric is the cross-entropy measure; it serves a purpose similar to the ROC measure but stresses a probability interpretation of model prediction. The cross-entropy measure principally serves as a corroborative measure. For model quality and prediction quality, I urge the reader to focus on the ROC measure. Overall, I maintain that the ROC measure is the most relevant measure in the context of bankruptcy prediction

Further analysis also includes the use of a confusion matrix. This study solves for a binary classification problem that produces a 2×2 matrix. The columns of the matrix represent the predicted values, and the rows represent the actual values for bankruptcy and healthy firm predictions. In the cross-section of the rows and columns, we have the True Positive (TP), False Negative (FN - type II error), False Positive (FP - type I error), and True Negative (TN) values. It is useful for a classification study to produce a classification matrix to aid intuition, especially when the dataset is imbalanced, such as in the case of bankruptcy prediction, where a small minority of the observations are bankruptcies. The confusion matrix is reported in *Table 2* on page 21. Further, I report 13 additional measures for the model as tested in (3), *Table* A30 and *Table A31 on* page 85 and 86.

Accuracy is defined as the percentage of correctly classified instances (observations) by the model. It is the number of correctly predicted bankrupt (true positives) and correctly predicted healthy firm years (true negatives) in proportion to all predicted values. It incorporates all the classes into its measure $(TP + TN)/(TP + TN + FP + FN)$, where *TP*, *FN*, *FP*, and *TN* are the respective true positives, false negatives, false positives and true negatives values for both classes. The measure can otherwise be represented as follows:

$$acc(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=1}^{n_{samples}-1} 1(\hat{y}_i = y_i) \qquad (1)$$

In *Table 1*, I empirically compare the performance of a range of binary classification models. I maintain a consistent framework of common sets of inputs as described in the Data Section. The process also involves identical steps for data handling and parameter estimation. From *Table 23* to *Table 22* on page 62 to 61. I select the best model type which is XGBoost to identify its performance under different conditions; this includes adjustments to the testing procedure, parameters, data assumptions, and sample distributions. These tests allow for more robust results of model performance. However, I also include tests that show how the predictive performance can be enhanced by changing the underlying structure and shape of the data.

Two of the classification models in this table make use of neural networks (NNs). These models are designed to identify latent and highly complex nonlinear relationships in the dataset. These are your quintessential "black-boxes." Researchers specify the architecture of the network and the initial inputs to feed into the first layer of these models; apart from that they play a very limited role. Unlike the XGBoost model, neural networks provide no mathematical formalities of the parameters that define relationships apart from their internal mathematics. Neural networks also have issues with handling data of mixed types such as categorical and continuous. Historically, neural networks have been very cumbersome as they are computationally intensive, but recent advances in processing technology have lightened this burden.

The first NN model used in this study is a Deep Feed Forward Network, which is structurally similar to other researchers' use of Multilayer Perceptron Models (Jones et al., 2017; Mselmi et al., 2017), but it has 5 densely connected hidden layers instead of two. The other NN model is a Deep Convolutional Neural Network (DCNN). This network is a biologically inspired variant of MLP that uses four densely connected hidden layers with the addition of a convolutional layer. The Deep Convolutional Network is traditionally applied to image recognition tasks. Repetitive blocks of neurons are applied across space to learn filters and variables that are associated with the response. The layer structure of the networks has been designed by hand, while the number of neurons has been automatically selected by hyper-parameter search operations. For more information on the construction of these NN variants, see page 64 in the Appendix.

Lastly, as a way to reconcile the performance of these modern models with traditional statistical models in past research, I also include a Logit Model. The XGBoost model outperforms the deep learning models, and the Convolutional Neural Network far outperforms the Feed Forward Network. Given more data, I would expect the gap between the XGBoost model and the DCNN model to progressively decrease. Deep learning models are known to perform especially well with larger datasets. The Logit Model is the worst performing model in this study. Unfortunately, the Logit model does not perform well with too many variables, which often leads to overfitting (Altman, 1968; Ohlson, 1980). Irrelevant variables that enter the global maximum likelihood solution of the Logit Model severely impact the quality of the reduction and model stability, whereas the higher dimensional models are relatively unaffected by noisy variables and outliers. Adding regularisation techniques like L1, L2 and elastic net could also improve to improve the logistic regression prediction performance. A more in-depth theoretical comparison of these models can be found in XI.E on page 70.

Table 1 shows that the XGBoost model, i.e., GBM model performs better than the other models investigated for all metrics considered. It is therefore a valid model to approximate the underlying function and would therefore provide reliable variable importance values. Even without adjusting for the fact that bankrupt firms only account for a very small percentage of the overall firm-years in this sample compared to other studies, this study produced the best accounting-based prediction of all previous studies as measured by both an ROC (AUC) score of 0.958 and a prediction accuracy of 0.976. Furthermore, despite purely making use of accounting data, this study produced the best neural network type model of all previous studies (*Table A25* on page 80) with an AUC of 0.914 and accuracy of 0.95.

Table 1: XGBoost and Deep Learning Model Performance Comparison

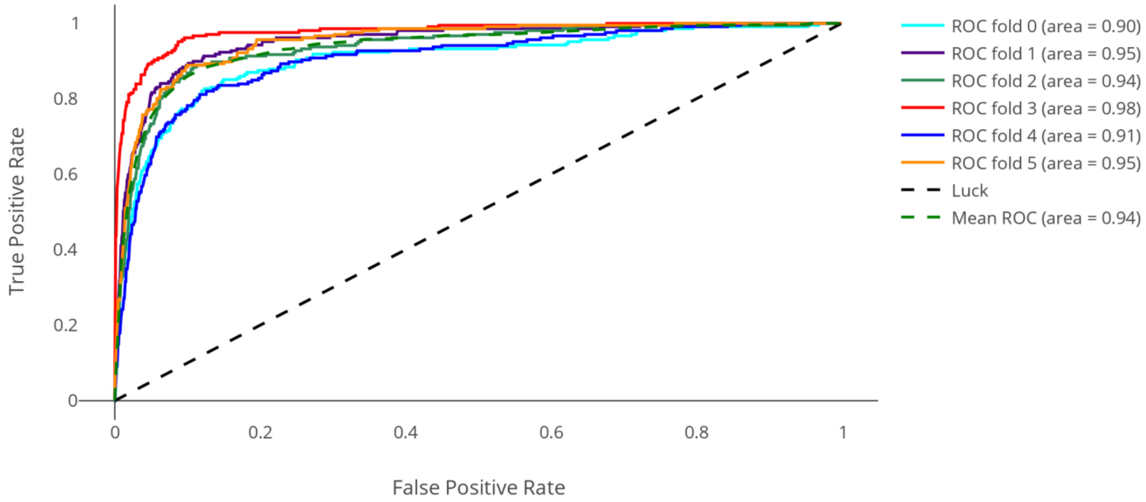
Metrics	XGBoost Model	Deep Feed Forward Network	Deep Convolutional Neural Network	Logit Model
ROC AUC Sore	0.9587	0.8444***	0.9142***	0.7092***
Accuracy Score	0.9755	0.9324	0.9518	0.6856
False Positive Rate	0.0037	0.0666	0.0296	0.2296
Cross-entropy	0.1414	0.5809	0.2996	1.1813

This table illustrates the performance of two deep learning models against the XGBoost Model. The Feed Forward Network is a deep learning network that does not circle back to previous layers. The Convolutional Neural Network is a biologically inspired variant of MLP, popularised by recent image classification studies. The best possible Logit model was established by choosing a selection of the best variables. Further results include the isolation of the 10 best predictor variables (using the Gini Index) in all models; this produced similar results to the above table both in extent and in rank. *p<.1 ** p<.05 *** p<.01. Significance levels are based on a two-tailed Z test to identify the statistically significant difference between all contender models and the best performing model, which is made possible due to the cross-validation process.

The widespread use of the AUC measure in recent studies allows researchers to compare the performance of their model with other studies. The benefit of the metric is that it is somewhat agnostic to different healthy-to-bankrupt firm distributions, at least more so than accuracy measures. In saying that, the problem is that, notwithstanding the recent universal adoption of AUC, it is still hard to compare performance across studies as the sample distribution does have some effect on the performance; other factors include the type of firms, industry, country, sample period, jurisdiction, and the definition of corporate distress.

The average ROC (AUC) of more than ten past decision-tree ensemble studies is around 0.927. The best performing is 0.997 and the worst performing is 0.859. In spite of the conservative sample selection in this chapter, the decision tree ensemble (XGBoost) model used in this study performed better than the average of past reported studies. It is also the best model when compared to other studies that only used accounting values as inputs. The average AUC of eight different neural network studies is 0.850; the best and worst performing past study has an AUC of 0.901 and 0.750 respectively. The DCNN of this model achieved an AUC of 0.9142, making it the best performing neural network of all past research. *Figure 2* below identifies the ROC and the associated AUC of a five-fold cross-validation model. It is the best way to visualise the AUC metric. Both this figure and the aforementioned time-series cross validation table show that there is reasonable amount of variability in the curves for each validation iteration. The dotted green stripes highlight the average AUC and the diagonal line, the line of luck.

Figure 2: The Receiver Operating Characteristic and Area Under the Curve - ROC (AUC)



This figure reports the ROC and the associated area under the curve (AUC). The random ordering or luck line is plotted diagonally through the chart and represents a series of random and noisy predictions. The chart reports five different cross validation folds and the associated performance as well as the mean ROC presented as a dashed green line. The reported mean is much higher than 0.90, which conventionally represents a 'great' classifier.

The results in *Table 2* aggregate multiple predictions to show a contingency table of the different predictions against actual outcomes. *Table 2* presents the precision and improvement score for both healthy and bankrupt firm years. The precision metric is the class-specific accuracy of the predictions made; it is a useful measure to know when there is a class imbalance. The model correctly predicted 258 out of 374 (116+258) predicted bankruptcies. For the purpose of the confusion matrix, the classification threshold is set to minimise the false positives at the expense of a higher false negative rate or recall error. As it happens, a threshold close to 50% is right for this purpose. This decision threshold can, of course, be adjusted to effect change in the table below as the chosen value is wholly up to the researcher. The decision threshold adjustment is possible because the model has a logarithmic loss function that outputs a probability associated with each class, which can simply be adjusted.

Table 2: Healthy and Bankrupt Confusion Matrix

Aggregated Health and Bankrupt Firms Matrix		Predicted		Sample Proportion
		Healthy	Bankrupt	
Actual	Healthy	**29041 - TN**	116 - FP	0.96
	Bankrupt	805 - FN	**258 - TP**	0.03
Precision		0.97	0.69	30220
Improvement		0.01	0.66	-

This bankruptcy prediction task solves for a binary classification problem that produces a 2×2 matrix. The columns of the matrix represent the predicted values, and the rows represent the actual values for bankrupt and healthy firm predictions. In the cross-section of the rows and columns, we have the True Positive (TP), False Negative (FN - type II error), False Positive (FP - type I error)), and True Negative (TN) values. The sample proportion on the far right is equal to all the actual observations of a certain classification divided by all the observations. The *precision* is calculated by dividing the true positives (Bankruptcies) with the sum of itself and the false negatives (Healthy). An example along the second column: $258/(116 + 258) = 69\%$. The improvement is the percentage point improvement the prediction model has over a random choice benchmark.

The good performance in *Table 2* can further be highlighted by drawing up a confusion matrix from random guessing. *Table 3* shows the performance of random guessing based on knowledge of the underlying test distribution. There is a big difference between the distribution of bankruptcy selections in this table compared to the model-predicted table. The performance of the table above is much better than random predictions based on the underlying sample distribution. The random guessing model correctly predicted 37 out of 1063 predicted bankruptcies. This equals a precision of just over 3%, which is much worse than the model's 69%.

Table 3: Random Guessing Confusion Matrix

Aggregated Health and Bankrupt Firms Matrix		Random Guess		Marginal Sum of Actual Values
		Healthy	Bankrupt	
Actual	Healthy	**28131 - TN**	1026 - FP	29157
	Bankrupt	1026 - FN	**37 - TP**	1063
Marginal Sum of Guesses		29157	1063	1063

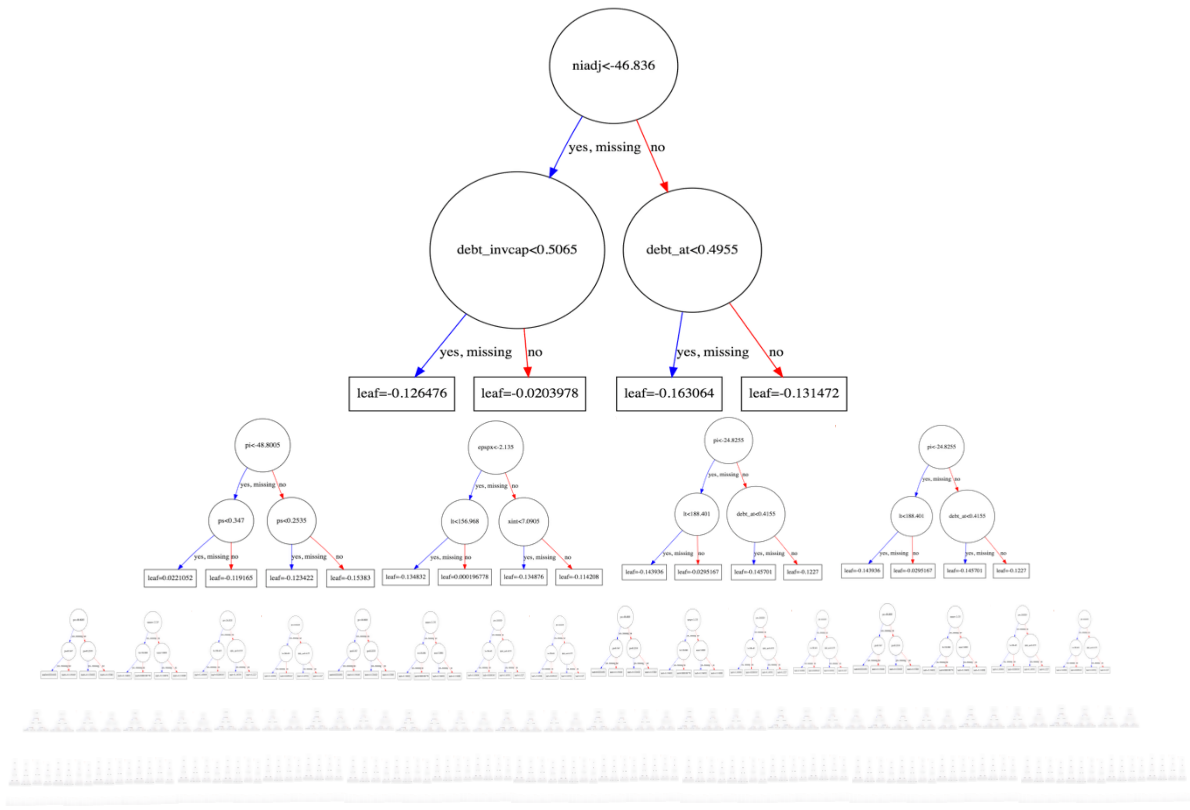
This table is formed by 'randomly choosing the observations' by allocating the observations according to the underlying distribution, as presented by Sample Proportion in *Table 2*.

VII. Variable Importance

In bankruptcy prediction, it is useful to know the relative contribution (RVIs) of all variables on the final prediction outcome. Since the variable importance measures, such as the Gain measure and Gini Index, are relative measures, it is conventional to identify the largest value to be labelled with 100 and then to scale all remaining variables according to the most predictive value (Breiman et al., 1984). The final RVI is the frequency of variable splits weighted by the average squared improvement of the model at each split across all trees (Friedman & Meulman, 2003; Hastie, Tibshirani, & Friedman, 2009). *Figure 3* presents an example of a single decision tree used in this study. At each split, there is a re-estimated contribution that I convert into a percentage for simplicity. In *Figure 3*, I only report each subsequent split to easily fit the full tree at a depth of 12. This figure should give the reader an indication of the internal structure of the 'weak learners' of the XGBoost models. However, in practice, instead of one tree, there are multiple trees. This is called an ensemble. A clear benefit of the XGBoost model is that it allows a wide range of variables to contribute to the overall prediction unlike most conventional models; there is evidently a reasonably even distribution of RVIs across multiple predictor variables (*Table 4*). This is the power of high dimensional input space; multicollinearity does not impair the predictive performance of the model to the extent of conventional linear models.

In *Figure 3*, the first variable that gets split is the net income adjusted for common stock equivalent. The inequality equations represent the split in the branch that leads to differently predicted outcomes. The outcome is reported as the probability of bankruptcy. These numbers can be seen in the second line of nodes. As long as a variable has been used once and contributed to a change in the prediction probability, then it has an RVI above zero. This indicates that the variable adds some predictive power to the overall model. The RVIs reported in *Table 4* as mentioned fall between a value of 0 and 100. The most important variable has a value of 100, and the other variables are scaled to match this level of importance. In this table, it is clear that the relative strength of the variable importance measures differs significantly largely across all variables.

Figure 3: Depth 12 - Decision Tree



This figure illustrates a decision tree with depth 12, the original depth used in this paper. The models in this paper have used as many as 4800 different trees to predict every observation. The above output is an example of one tree. On the right-hand side, the tree shows that when income (niadj) is negative (below 50 million), and debt to assets (debt_at) is small, then the likelihood of bankruptcy is less than if the debt to assets ratio was larger (than 0.4955). On the left-hand side, it shows a similar relationship but with debt to invested capital (debt_invcap). The number of splits and overall contribution of the different predictor variables is used to measure the predictive importance or ability of each feature. Further, the interaction pairs are also calculated by identifying the outcome contributing relationships between different predictor variables. This tree is selected for its symmetrical properties, most trees are asymmetrical and have varying branch lengths.

Table 4 presents the predictive power of the top 50 input variables to the prediction model using Gain as a relative variable importance measure. The table further includes the direction the variable has with the response variable. A positive (+) means that bankruptcy is more likely to occur when the variable increases and a negative (-) means that bankruptcy is less likely to occur when the value increases. Therefore, the minus (-) is a good sign. The strongest variables as identified in *Table 4* are pre-tax income (100) and income before extraordinary expenses (91). The strongest two ratios are the EPS (61) and the Price to Sales ratio (55). It is important to remember that the variables get selected for their contribution to the overall model and that these contributions do not occur in isolation from other variables i.e., the predicted values are primarily a consequence of variable interactions.

Table 4: Predictive Power of Variables

Input Variable	RVI	Post GFC RVI	D	Category
Pretax Income (PI)	100	100	-	Income
Income Before Extraordinary Items (IBC)	84	91	-	Income
EPS(Basic) - Exclude Extra. Items ($&c)	61	27	-	Profitability
Price/Sales	55	57	-	Valuation
Liabilities - Total (LT)	50	19	+	Liability
Interest and Related Expense (XINT)	44	8	+	Expense
Long-Term Debt - Total (DLTT)	44	17	+	Liability
Stockholders Equity - Total (SEQ)	35	22	+	Equity
Total Debt/Total Assets	31	6	+	Solvency
Inventories - Total (INVT)	30	18	+	Asset
Net Income - ADJ for Com Stock Equiv	29	42	-	Income
Depreciation and Amortization (DPC)	26	12	+	Equity
Total Debt/Invested Capital	24	8	+	Solvency
Accounts Payable (AP)	22	14	+	Liability
Research and Development/Sales	21	6	-	Other
Property, Plant & Equip. - Total(Net)	21	17	+	Asset
EPS (Diluted) - Excl. Extra. Items ($&c)	21	25	-	Profitability
Price/Book	21	35	-	Valuation
% Change in Price/Sales	20	37	-	Valuation
Capitalization Ratio	20	32	+	Solvency
Free Cash Flow/Operating Cash Flow	18	11	-	Solvency
Cash and Short-Term Investments (CHE)	18	10	-	Asset
Sales/Stockholders Equity	18	8	-	Efficiency
Cash Balance/Total Liabilities	18	15	-	Solvency
Income Taxes - Total (TXT)	17	15	+	Liability
Capital Expenditures (CAPX)	17	8	+	Asset
% Change in Property, Plant & Equip.	16	7	-	Asset
Sale of Common and Preferred Stock	16	6	-	Asset
Price/Cash flow	16	21	-	Valuation
Total Debt/Capital	16	7	+	Solvency
Cost of Goods Sold (COGS)	16	7	+	Expense
Operating Activities - Net Cash Flow	16	8	-	Cash Flow
Long-term Debt/Invested Capital	15	43	-	Solvency
Operating Income Before Deprec.	15	26	-	Income
After-tax Interest Coverage	14	5	-	Solvency
% Change in Income Before Extrao. Items.	14	2	-	Income
Long-term Debt/Total Liabilities	14	17	+	Solvency

Investing Activities - Net Cash Flow	13	10	-	Asset
Sales/Invested Capital	13	12	-	Efficiency
Long-Term Debt - Reduction (DLTR)	13	35	-	Liability
% Change in Common Equity - Total	13	11	+	Equity
Liabilities - Other (LO)	13	16	+	Liability
Current Assets - Other (ACO)	13	12	-	Asset
Book/Market	13	11	+	Valuation
% Change in Liabilities - Total (LT)	13	1	+	Liability
% Change in Interest and Related Expense	13	8	+	Expense
Assets and Liabilities - Other (Net Change)	13	10	+	Asset
Assets - Total (AT)	13	4	-	Asset

This table contains a list of the variables with the most predictive power as measured by the gain statistic. Gain in this paper is defined as the number of splits the variable undergoes, weighted by the squared improvement of the model that resulted from each spilt. The relative variable importance (RVI) is simply the division of subsequent variable gains by the gain of the most contributing variable scaled by 100. It is calculated from 1979-2016. The Post GFC RVI is the relative gain from 2008-2016. D is the direction of the variable with a bankruptcy outcome. The category is the bucket or dimension in which the variable falls. It is used later to analyse which category has the most predictive power.

The model used in the above table near-randomly discriminates between correlated variable pairs; therefore, the predictive performance of a particular accounting dimension will likely be distributed among a few variables. The process of combining variables in predetermined accounting categories can help to highlight the most important dimensions to predict bankruptcy. The most predictive categories and associated variables are the Assets and Liability category, and more specifically, the Total Liabilities, Long-Term Debt, Accounts Payable, PP&E, Cash, Short-Term investments and Inventory variable inputs. The second most predictive category is Income, more specifically, Pre-tax Income and Income Before Extraordinary Activities input variables. In addition, the third most predictive category is Solvency, more specifically, the Total Debt to Total Assets Ratio, Total Debt to Invested Capital Ratio, and the Capitalisation Ratio and Free Cash Flow to Operating Cash Flow Ratio input variables. The fourth category essential for predicting bankruptcies is Valuation and Profitability, more specifically the Price-to-Book, Price-to-Sales, ROE, and EPS input variables. Other not as essential categories and individual value pairs include Equity - Shareholders Equity, Expense - Interest and Related Expense, Efficiency - Sales to Stakeholders Equity, Other - Research and Development to Sales, Liquidity - the % change in the cash ratio and lastly, Cash Flow - Operating Activities Net Cash Flow. I further show that a very competitive prediction model can be created by using only 50 input variables (*Table 24*, Second Column).

In this study, I include multiple accounting-related variables purely because there is no clear consensus as to what variables and interactions are the most important in lower dimensional models, not to mention higher dimensional models. *Table 6* identifies the characteristics of the most predictive variables to the model. The results in this table support my hypothesis that 70 simple accounting values have improved predictive power over 72 ratios due to the high dimensional interactions that remove the requirement to pre-specify ratios. The issue with ratios in high-dimensional studies is that they self-impose linear restrictions in the relationships between the numerator and denominator. In theory, if you feed the model the raw input to the ratio, it should more easily scour patterns for non-linearity between the inputs, as financial measures are reported to be significantly non-linear (Foster, 1986). As a result of this outcome, I will treat each category as a contributor to prediction success rather than favouring the ratios as the only true variables and regarding the fixed values as 'controls' as done in past studies. The results in Panel B further show that fixed variables are more important than growth variables (% change variables).

Table 5: Variable Type Analysis

Panel A

Type	Importance (%)
Simple	0.57
Ratio	0.43

Panel B

Construction	Importance (%)
Fixed Period	0.87
Growth	0.13

Panel C

Type	Construction	Importance (%)
Fixed Period	Simple	0.50
	Ratio	0.38
Growth	Simple	0.07
	Ratio	0.06

This table computes the importance of the variables grouped by certain characteristics. Panel A groups the variables by whether the value is simple or whether it is a ratio. Panel B looks at whether the value is simple or whether it is a calculation of the change of the variables between time *t-1* and *t*, i.e., a growth measure. Only the most predictive growth variables are included in the model. Panel C is the value type in Panel A grouped with the construction type in Panel B.

Due to the multi-dimensional nature of the model, a good approach to study the most important variables is to group them. In *Table 4*, all measures are grouped according to how they would appear on standardised financial reports. That includes *Assets*, *Liabilities* Expenses, *Income, Equity and Cash Flow*. All other categories are classified according to the following definitions: *Solvency* measures are ratios associated with financial soundness and the ability of the firm to meet its long-term obligations; *Valuation* measures are accounting ratios that identify the firm's attractiveness and whether the stock is under or overpriced; *Profitability* ratios measure the ability of the firm to generate returns; *Efficiency* ratios track the firm's effectiveness in utilizing assets and liabilities; *Liquidity* ratios measure the firm's ability to meet short-term demands; and lastly, *Other* ratios incorporate values such as Research and Development to Sales, and Labour expense to Sales ratios. All predictor variables are exclusively allocated to one of these eleven categories.

Looking back at *Table 4*, it is clear that there is a difference between the variable importance post-GFC (2008) compared to the importance over the full sample period. The best approach would be to identify the categories that show significantly more or significantly less predictive power. After the GFC, the model loading on liability variables is a lot less while valuation measures have increased in importance over equity measures, highlighting the fact that less trust is put on equity as compared to the markets' valuation of the firm. All income measures (PI, IBC, NAIDJ, % IBC) also show increased predictive power after 2008. These differences may be a priori evidence of a structural change in bankruptcy prediction.

In *Table 6*, I empirically investigate the categorical importance post-GFC. I show that Solvency, Income, Valuation, and Profitability measures are more important after the GFC and that Liquidity and Cash-Flow measures are less important after the GFC. These changes make sense, as cash flow and liquidity were not at the heart of the crises. It was largely based on issues of capital structure (Solvency) and unjustified valuations (Valuation). In addition, the prediction algorithm post-GFC gives less attention to the reported valuation (Equity) of firms as it learns that these accounting measures cannot be 'trusted' as much as they could be before the crises. The two measures (3) - (4) in *Table 6* will later be used as three of the nine measures to rank the final categories after the required recategorisation because of strong correlations between some accounting categories.

Table 6: Predictive Power of Categories

Category	(1) R/A	(2) Category Importance (CI) (%)	(3) Relative CI (RCI)	(4) Post GFC RCI
Asset	A	0.18	100	100
Solvency	R	0.17	95	99
Income	A	0.13	70	80
Liability	A	0.11	62	62
Valuation	R	0.11	62	80
Equity	A	0.08	46	32
Profitability	R	0.07	40	44
Expense	A	0.06	31	34
Efficiency	R	0.04	21	21
Other	A	0.02	12	11
Liquidity	R	0.02	12	10
Cash Flow	A	0.01	8	4

This table looks at some predictive metrics of the categorisations used in this study. This table specifically highlights the importance of these categories in predicting bankruptcies before and after the GFC. Here we can compare the difference between the overall Relative Category Importance (RCI) in (3) and the RCI post GFC (4). (1) identifies whether the category constitutes an accounting ratio - R or whether it is a simple accounting value - A. (2) indicates that the category importance (CI) is a summation of the gain measure (PI) of the contained variables. (3) RCI is a relative measure based on the most important category (CI). (4) identifies the respective RCI for the sample period from 2008 onwards.

Decision tree models are mostly robust to multi-collinearity; they will simply choose one of the many correlated variables when deciding upon splitting the tree. This makes sense for a model that is purely interested in boosting prediction performance. Compare this against linear models that rely on all correlated variables for prediction. For decision trees, correlated variables would simply not add to the split process anymore as they do not bring new information that has not already been provided by the first feature. Some boosting models that have several correlated variables as inputs will tend to choose one correlated variable and use it in several trees. The randomised nature of the XGBoost model and associated parameters will, however, randomly select between collinear variables with each subsequent tree and variable selection iteration. This means that across many trees, such as used in this study (4800 trees), the most important correlated variables will on average come out on top. The issue is that the importance still gets distributed across many variables; a prime example is the two top-performing variables, pre-tax income and income before extraordinary expenses, which are both highly correlated (95%+). Therefore, to get a better indication of the overall importance of certain accounting dimensions we can categorise the variable as has been shown in *Table 6.* One of the benefits of this study is that it uses labelled accounting data that makes it easy to understand what variables share similar characteristics. Finally, there can be correlations across categories; the best way to deal with this is to understand that the components of these categories are similar and that they should be combined to form a single category.

One process to identify the success of the categorisation is to PCA transform the variable sets and simultaneously ensure for low correlation between the categories. For the vast majority of the categories, there seems to be no correlation, which is a good sign. However, there is, as should be expected, large collinearity between liabilities and assets. This has to do with the basic accounting equation, where the majority of the interaction is between Liabilities and Assets over time. Further, the Profitability, and Valuation dimensions also seem to be highly correlated. For that reason, it is better to group Assets & Liabilities, and separately, Profitability & Valuation dimensions together. This logic is often left out in the variable discovery task of machine learning bankruptcy studies. Researchers repeatedly forget to look at the correlation between variables. This is an issue because the reported variable importance is misleading if the importance is divided among various categories.

The PCA deviation score measures the standardised correlation deviation between components of a category. In studies where the variables are unknown, a deviation of more than one can be indicative of a category that can be effectively deconstructed into more categories as long as the constituent variables to a category make theoretical sense. In general, a high deviation in correlation is usually a good thing for each category, as uncorrelated variables lead to improved performance. The high reported value for assets and liabilities is likely due to structural differences between current and non-current assets or liabilities. *Table 7* looks at the correlation between categories, after which a decision is made whether or not to combine certain categories for more informed results. The PCA deviation is further used to decide whether a category should be split into smaller subcategories. The results indicate that a few categories should be combined but that none of them has to be broken into smaller constituents.

Table 7: Correlation and Categorisation Analysis

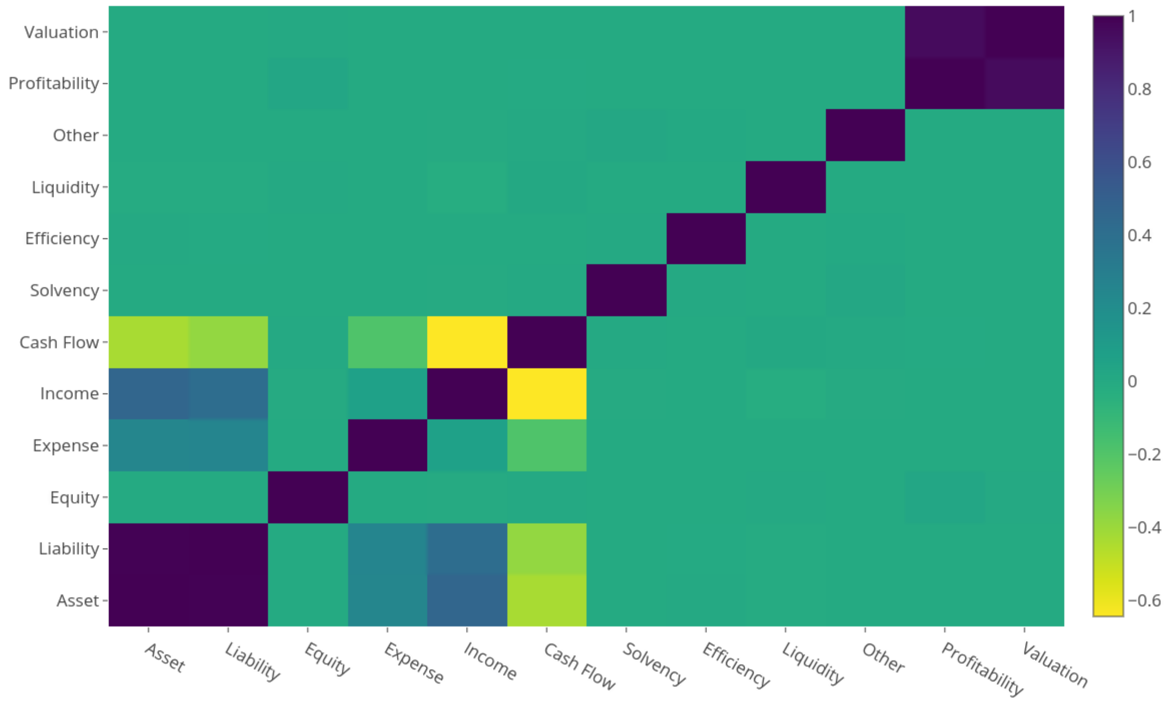
Category	Closely Correlated With	Correlation Score	PCA Category Deviation
Asset	Liability	0.82	1.43
Solvency	Other	0.02	0.49
Income	Cash Flow	-0.56	0.87
Liability	Assets	0.82	1.37
Valuation	Profitability	0.93	0.36
Equity	Profitability	0.03	0.78
Profitability	Valuation	0.93	0.36
Expense	Liability	0.27	0.39
Efficiency	Other	0.01	0.40
Other	Cash Flow	0.03	0.44
Liquidity	Cash Flow	0.01	1.16
Cash Flow	Income	-0.56	0.90

This table identifies the highest correlated category pair and the level of correlation for each category. The process follows a PCA transformation of the first component for each category after which a correlation analysis ensued. The PCA Category deviation is a metric that identifies the variability of the PCA to identify the extent of the diversity inside the category. A measure larger than 1 is indicative of large variability inside a category, bringing the similarity of the variables in the category under question. If all variables in a category can be justified, then a higher PCA Category deviation is usually a good thing, as uncorrelated variables lead to improved performance. A full spectrum of the category correlations is presented in *Figure 4*. Solvency can further be divided into categories such as capital structure, interest coverage, and cash flow ratios. The footnotes of *Table 12* show the ranking order of such a split. The above analysis, however, shows that this is not necessary as Solvency ratios are currently uncorrelated with other categories and such a split may cause new correlation concerns. Furthermore, the variables to the Solvency category are shown to be similar in type as measured by the PCA category deviation score; for that reason, the category should preferably not be split in this study.

If a researcher feels the need to, Assets and Liabilities can be divided into current and long-term categories to lower the deviation. However, for this study, such a grouping is not necessary as both of those groups are indeed rational constituents of the overall assets and liabilities grouping, and instead of slicing the established categories, the reader can refer back to *Table 4* to gain an understanding of what *type* of assets and liability values are the most important variables. The same can be said for liquidity measure that could be further divided into current ratios and conversion cycle type ratios.

Figure 4 expands on the most correlated categories as reported in *Table 7* to include the correlations of all categories to each other. The figure shows that the categories are largely uncorrelated. There is only a small number of correlated categories that will be dealt with from *Table 12* onwards.

Figure 4: Correlation on the PCA Transformation of Categories



I apply a PCA to the variables in each category and select the first principal component to represent that category. Reported above is the correlations between the first principal components of each category. This table shows that there is minimal correlation between the majority of the categories, but that some categories seem to be strongly correlated such as assets and liabilities and valuation and profitability measures.

Table 8 implements an inductive form of hypothesis testing used by Mullainathan and Spiess (2017); the process works this way: include all possible variables in a machine learning algorithm, and then remove the category or variables you believe to be important, retrain the model, readjust the hyperparameters and look at the model performance without that category or variables. From left to right the models in *Table 9*. exclude Assets and liabilities, Solvency, Income, Equity and Profitability, and Valuation variables while calculating the AUC score at each point with three different measurement methods. I, therefore, start with a complete model and work my way back to identify each group of variables' relative contributions to the ROC.

The approach is useful as it goes beyond the task of predicting and also decomposes the importance of the type of variables in the study, providing added value to the literature. This method solves the issue of multiple correlated variables leading to prediction success. The results of this analysis show that asset & liability related values contribute the most to the model followed by solvency ratios, income values, profitability & valuation ratios, and equity values. This is one of the best tests to show the importance of the different groups of input variables empirically. In contrast to *Table 6*, which is the mere categorical summation of variable importance, this table involves whole new tests. It is also interesting to note that the relative contribution (RVI) falls flats much quicker under this method across categories, possibly highlighting that only a few categories are needed to effectively predict bankruptcies.

Table 10: Reverse Induction Test

ROC (AUC)	(1) Full	(2) A&L	(3) Solvency	(4) Income	(5) Equity	(6) P&V
A - All	0.959	0.942	0.944	0.948	0.955	0.952
B - CV	0.947	0.930	0.935	0.937	0.943	0.941
C - Time CV	0.957	0.940	0.942	0.946	0.954	0.950
D - Average	0.954	0.941	0.944	0.947	0.955	0.951
Relative Contribution (R)	-	100	82	63	22	39

This table compares the various sets of variables against each other using an inductive testing technique to identify the importance of groups of variables that explain the model success. (1) is the performance of a model that contains all the variables. (2) - (6) removes variables that fall within the respective asset & liability, solvency, income, equity, and profitability & valuation categories from the model after which the model is retrained and tested to identify the extent to which each category contributes to the model. The following relative contributions are unreported in the table: expense (R 10), efficiency (R 8), liquidity (R 8), other (R 8), cash flow (R 4). The relative contribution is calculated by using the average of three different performance techniques to ensure the robustness of the results. (A) the first validation method is an 80% train and 20% split result in time-series. (B) is a random 10-fold performance validation split. (C) is a variant of the 10-fold performance split but in time series; it is arguably the most robust method. (D) is the average across all three methods, and the value used to calculate the Relative Contribution which is a normalised value out of 100 of the predictive power lost after dropping the category. All these splits are implemented after the validation and model development steps.

Table 12 reports the ranking of variable categories using nine different methods. This analysis is important to the overall study, as it seeks to identify which groupings of accounting dimensions truly come out on top in predicting future bankruptcies. Similar to Table 11, the new categories that combines assets with liabilities and profitability with valuation ratios are used. The table orders the categories according to the final ranking. All the italicised categories are accounting ratios. This is done because in *Table A27* it can be seen that past literature has focused on ratio measures; the purpose of the table is to compare the results of this study with that of past literature. If we isolate the *ratios*, as has been done in some past papers, then *Solvency* ratios come out on top followed by *Valuation and Profitability*, *Efficiency*, *Other* and *Liquidity* ratios. This result goes against a lot of low-dimensional and even some higher dimensions prediction analysis studies that stress the importance of *liquidity* measures over *Valuation and Profitability ratios* (Kim & Upneja, 2014; Mselmi et al., 2017).

On the full sample, it can be seen that the Asset & Liabilities and Solvency categories are the most important groups to predict bankruptcies. It is interesting to note that cash flow measures are not that important in predicting bankruptcies. Past researchers have also noted that Cash flow measures do not provide incremental prediction power over accrual-based measures (Casey & Bartczak, 1985). It is clear that ratios and fixed values both have high importance in the predictive model.

Table 12: Category Importance Analysis

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Category	CI	RCLI Top 50	Post GFC RCL I	Pot	F	wF	24 CI	PCA 10	RI S	Avg.	Fin.
Assets & Liabilities	1	1	1	1	1	1	1	1	1	1.0	1
Solvency	2	3	2	2	2	2	3	2	2	2.2	2
Income	3	2	3	3	4	5	2	4	3	3.2	3
Valuation & Profitability	4	4	4	4	3	4	5	3	4	3.9	4
Equity	5	5	6	5	5	3	4	6	5	4.9	5
Expense	6	6	5	6	7	6	6	10	7	6.6	6
Efficiency	7	7	7	8	6	7	7	9	6	7.1	7
Other	8	8	8	9	8	8	10	8	8	8.3	8
Liquidity	9	9	9	10	8	9	9	5	10	8.7	9
Cash Flow	10	10	10	7	10	10	8	7	9	9.0	10

This table is an attempt to regroup categories where there is a strong correlation 80% + and to calculate the rank of the categories according to 9 different predictive importance strategies. This table calculates the equal weighted average of nine ranking methods (10). (1) is the normal importance measure (gain measure) calculated for all variables in every category. (2) is the gain measure for newly created categories using only the top 50 variables. (3) is the gain measure after the GFC. (4) is the ranking according to the potency measure, being the category importance weighted by the number of variables in each category. (5) is a measure (FScore) that tallies the amount of variable splits across all trees within each category. (6) measures the FScore weighted by the probability of the splits to take place. (7) is the overall gain measure for a newly created model that only incorporates the 24 best variables. (8) is the importance of the first PCA component for each category. (9) is the relative importance measured by Shapley value contribution. (10) is the equal-weighted rank average for each category. (11) is the final ranking obtained by ranking the raw average. When percentage growth measures were removed from the categories, all categories remained unchanged apart from a swap between *Other* and *Liquidity*. A further split in category where solvency ratios were split between capital structure and coverage and cash flow ratios resulted in the following rank among categories: (1) asset and liabilities (2) income (3) *valuation and profitability*, (4) *capital structure*, (5) equity, (6) *interest coverage*, (7) expense, (8) *efficiency*, (9) *cash flow* ratios, (10) other ratios (11) *liquidity ratios*, (12) cash flow values. The ratio values are italicised.

The vast majority of past high dimensional research has identified Solvency and Solvency related variables as the most important ratios (*Table A27*). The analysis in *Table 12* corroborates its importance. Some studies have, according to this study's analysis, underreported the importance of Valuation and Profitability ratios (Kim & Upneja, 2014; Mselmi et al., 2017). As illustrated by the table, there is very little disagreement in the significance of both Solvency and Valuation & Profitability ratios throughout all methods (1) - (9). The same cannot be said for efficiency and liquidity measures. The final ranking of the Efficiency and Liquidity ratios is quite interesting because, to my knowledge, only two studies have noted that efficiency ratios take importance over liquidity ratios, and both are also high dimensional model studies (Jones et al., 2017; Mselmi et al., 2017). Other than these two studies, very few studies incorporate the efficiency dimension as prediction inputs. Furthermore, only three of the high dimensional studies show the same *ratio* ranking as has been reported in this table in column (11) (Behr & Weinblat, 2017; Jones et al., 2017; Volkov et al., 2017).

VIII. Interaction Analysis

The issue with many machine learning models is that their nonlinearity makes it hard to enforce monotonicity constraints to identify the direction of the relationship between independent variables and the machine-learned response function. In ML, the response can change in a positive or negative direction and at varying rates for changes in an independent variable, making the interpretation of feature importance much harder than simply looking at the coefficients of a linear model. Although the importance or contribution of a feature can be very valuable to understand, the measure does not identify the average direction and size of a variable to a response, nor does it attempt to explain nonlinear movements, which is of interest to researchers and industry experts.

To identify the relationship of variables in a machine learning algorithm, we can make use of a technique called partial dependence. Partial dependence allows us to see into the 'black-box.' Plotting the partial dependence, also known as marginal effect plots, produces information associated with both the direction and the strength of the relationship between explanatory and the outcome variables. Partial dependence plots (PDPs) are the visualisation of fitted functions. PDPs show the effect of variables on the response after accounting for the average effects of all other variables in the model (Friedman, 2001; Friedman & Meulman, 2003). PDPs offer the means of identifying the marginal dependence between the outcome and variables (Hastie et al., 2009). The basic premise of this technique is to obtain a prediction for all unique values of a variable while accounting for the effects of all the other variables to detect nonlinear relationships without the need to pre-specify them.

In this first part of this partial dependence analysis, we will first look at the top variables as highlighted in *Table 4*: Pre-tax Income (PI), Income Before Extraordinary Items (IBC), EPS Excluding Extraordinary Items (EPSPX), Price to Sales (PS), and Total Liabilities (LT). These variables have been individually plotted in Appendix Figure A12, *Figure A13, Figure A14, Figure A15* and *Figure A16.* These figures show their marginal association with the likelihood of bankruptcy. All variables in *Table 4* have followed a similar analysis to identify their respective directions (D), but only the first five are graphed. From the analysis of these five, the following directional relationships has been identified. The likelihood of bankruptcy decreases as PI, IB, EPSPX, and PS increase. The PI, IB, and EPSPX plots display a bimodal distribution around the zero bound, and all of these predictor plots show that extreme negative values are associated with failure.

The relationship for IBC is slightly more complex as values that are only slightly negative are not as concerning and tend to be classified as healthy. Furthermore, the partial dependence plots for PS show a monotonous increase in the likelihood of the firm being healthy as the value increases. A firm with a high PS value is likely to have a good profit margin and likely to be at the top of its industry with a lot of prospective growth. For that reason, it makes sense that the model identifies this positive relationship. As well, if we consider liabilities, it seems that the larger the LT, the more likely the firm is to be classified as bankrupt, all else equal. What is especially clear is that for the very low value of LT, the likelihood of being healthy is very high.

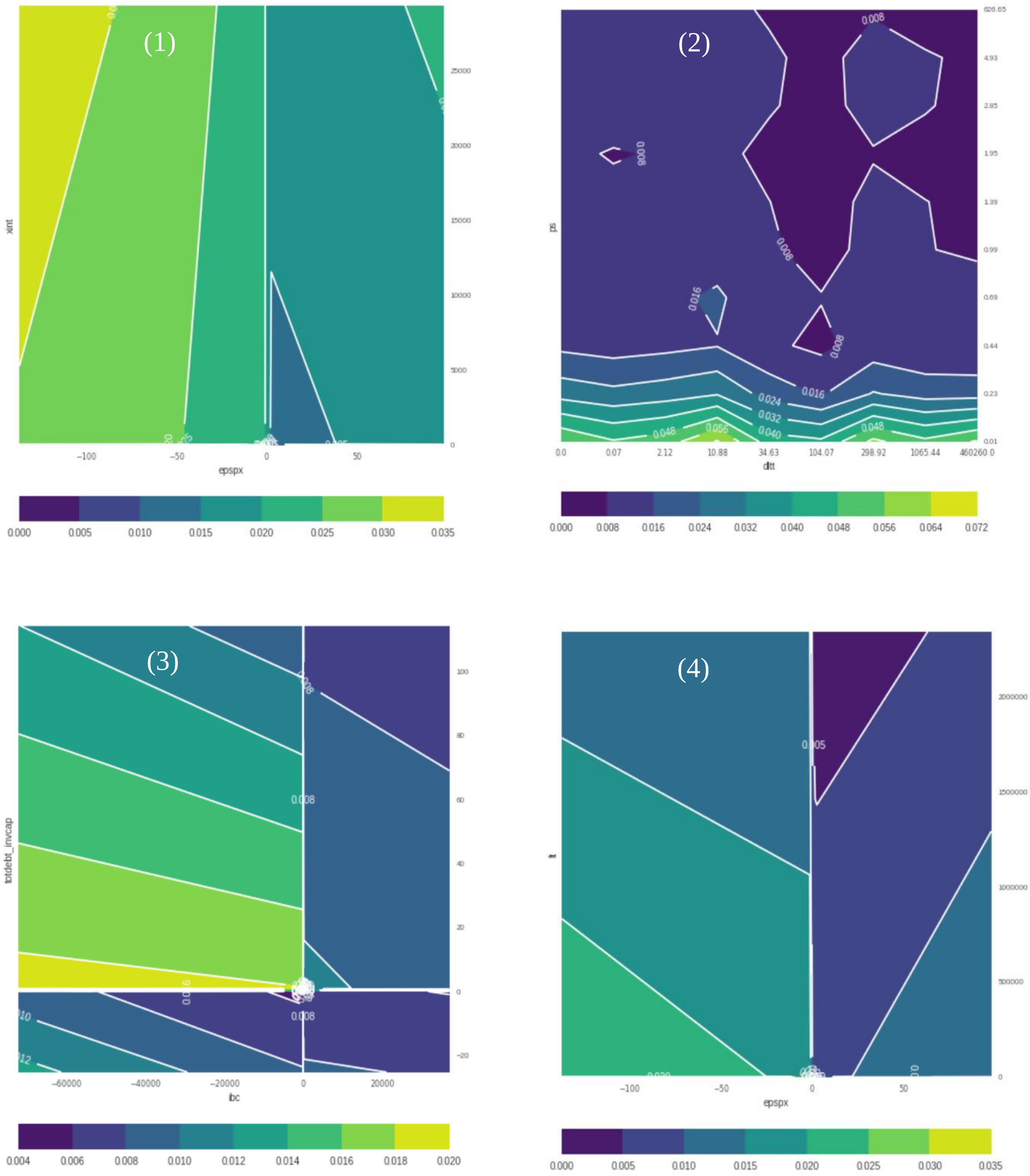
The next set of table interactions are noteworthy, *Table 13* looks at interaction pairs. Since the start of bankruptcy prediction, researchers have been interested in identifying the interactions between variables and how they affect the likelihood of bankruptcy. For example, FitzPatrick (1932) reported that less emphasis should be placed on liquidity ratios for firms that have long-term liabilities. And only in recent years have algorithmic techniques caught up to allow researchers to empirically identify the most important interactions. This is the first study to list the top interactions of a high dimensional model. Also, this is the first study to report interactions as far as depth-3, showing the non-linear relationship between three variables and how that relationship extends to the prediction of bankruptcies. It is also known that interaction effects add substantial explanatory power to bankruptcy models (Jones, 2017). Interactions are not given enough attention in bankruptcy prediction literature. The benefit of the XGBoost model is that it detects important interactions across the full set of explanatory variables. Many of these relationships are relatively self-evident, but some are surprising.

The top left plot in *Figure 5* (1) shows that in situations where EPS is extremely negative and total expenses surpass $5 million, the likelihood for bankruptcy increases significantly all else equal. The relationship between the price-to-sales and long-term debt ratio (2) is also noteworthy as there seems to be a higher dimensional relationship that creates two distinct local maxima. Long-term debt of $10 million and $300 million seem to be two critical points in the prediction model. Values that fall in between $10 million and $300 million combined with a stable PS value are associated with lower rates of bankruptcy. There seems to be a clustered "sweet-spot" that cannot be explained by these variables alone. The plot between Total Debt to Invested Capital (TDIC) and Income Before Extraordinary (IBC).

Items (3) is also interesting as it is quite evident how the variables are playing off each other. Where IBC is larger than zero, it does not matter what TDIC is; for the most part, the likelihood of bankruptcy is not going to increase too much. Value investors have been especially interested in the TDIC ratio. When invested capital is negative, it means that all fixed assets and working capital less interest-free liabilities is negative. This can have both good and bad implications. If a company is growing at its core (e.g., IBC), it means that you do not have to invest any money, and if nothing changes, you can use the extra money from interest-free liabilities to grow your business. Where TDIC is negative you see a smaller likelihood of going out of business, and this shoots up immediately as it becomes positive, just to taper down again as it becomes more positive.

Plot (4) shows that low total liabilities (LT) and low earnings per share (EPS) values are an indication of a future bankruptcy. This is likely due to the failing firm having paid off a lot of debt without being able to pay off the last debt as a result of not achieving good returns (EPS). Firms that can obtain large amounts of LT that have negative EPS are, however, also at higher risk of becoming bankrupt.

Figure 5: Interaction Pair Partial Dependence Plots (Depth Two)



Plot (1) at the top left is the interaction between Interest and Related Expense (xint) and the EPS Excluding Extra. Items (epspx) and resulting response. Top right (2) is the interaction between Price to Sales (ps) and Long-Term Debt (dltt). Bottom left (3) is the interaction between Total Debt to Invested Capital (totdebt_invcap) and Income Before Extraordinary Items (ibc) and the interaction effect on the bankruptcy outcome. Bottom right (4) is the interaction between Total Liabilities (lt) and the EPS Excluding Extra. Items (epspx).

Although there are thousands of relationships, I only present the most predictive interactions in *Table 13*. I further report the direction of the relationship to the response outcome. The results indicate that a small amount of interactions is responsible for the majority of the predictive power of the model.

Table 13: Depth 2 - Interaction Analysis

Term 1	Sign	Term 2	Sign	RII	Gain
ibc	-	totdebt_invcap	+	100	779
epspx	-	lt	+	90	704
epspx	-	xint	+	75	585
dltt	+	ps	-	71	551
debt_assets	+	pi	-	65	509
pi	-	xint	+	54	418
ppent	-	ps	-	50	390
ps	-	ps_prtc	-	43	338
dpc	+	ps	-	36	279

Out of the top 50 variable list, there are millions of ways to conjure up directional relationships. Due to the nature of nonlinear relationships, to conceptually understand the web of relationship, it is best to identify the top interaction pairs. This table represents the most important interaction pairs as measured by the gain statistic at an interaction depth of two. For easier reading, I also report the relative interaction importance (RII). The sign simply indicates the average direction of each variable. The interaction terms are much more informative than single standing variables. Interactions are at the core of what gradient boosting tree models are all about. *Terms 1* are Income Before Extraordinary Items (ibc), EPS Excluding Extra. Items (epspx), Long Term Debt (dltt), Total Debt to Total Assets (debt_at), Pre-tax Income (pi), Property Plant & Equipment (ppent), Price-to-Sales (ps), Depreciation and Amortization (dpc). *Terms 2* are Total Debt-to-Invested Capital (totdebt_invcap), Total Liabilities (lt), Interest and Related Expense (xint), Price-to-Sales (ps), Pre-tax Income (pi), % Change in Price/Sales (ps_prtc). Whether the components to the pairs are in the first or second column is of no consequence; the same value would be reported if the columns swapped, as it is an interaction between values.

In *Table 14*, I further present the interaction results in a more visually appealing way without each variable's directional relationships to bankruptcy. The table also highlights the insignificance of some interactions that are incidental to investigating the top interactions in a cross-tabular fashion.

Table 14: Cross Tab - Top Variable Interactions

	totdebt_invcap	ps	lt	xint	pi
ibc	779	704	63	45	13
pi	66	585	209	338	0
epspx	228	76	551	509	34
dltt	17	418	156	34	14
debt_assets	43	127	239	77	390
ppent	71	279	82	28	61

This table represents the most important interaction pairs as measured by the gain statistic at an interaction depth of two. The table has been constructed to highlight the top ten interactions. For completeness, the surrounding interactions have also been included. Variables vertically follow: Income Before Extraordinary Items (ibc), Pre-tax Income (pi), EPS Excluding Extra. Items (epspx), Long Term Debt (dltt), Total Debt to Total Assets (debt_assets), Property Plant & Equipment (ppent). And horizontally, Total Debt to Invested Capital (totdebt_invcap), Price to Sales (ps), Total Liabilities (lt), Interest and Related Expense (xint).

In *Table 15*, I further highlight the interactions between three variables. Similar to previous tables, for all the interaction signs, + means an increase in the likelihood of becoming bankrupt, - means a decrease in the likelihood of becoming bankrupt. I will describe the most interesting and somewhat unexpected interactions of the table. For the second interaction (2), as income (pi) increases, a higher debt to assets ratio (dept_at) becomes less of a concern. When research and development to sales (rd_sales) is high, the effect of the asset ratio on the outcome is less consequential, whereas the level of income still has a big effect on the outcome. This is very interesting as it shows that firms that have large R&D programs are unlikely to become bankrupt, all else equal. Some researchers have historically argued the opposite and said that there is a 'failure-inducement effect' in a firm's effort to push for innovation efforts when performance falls (Antonelli, 1989). This also makes intuitive sense as management would be less inclined to believe in the future of their company if this value was low. The causal link, like all of these interactions, remains somewhat uncertain. What can be said is that if this relationship with bankruptcy is purely the result of having large amounts of *disposable* income then you would expect other ratios and values in this study to show more importance (advertising/sales, reserves, dividends, purchase of stock), but they do not even feature in the top 50 most important variables. For that reason, I argue that R&D, as a core activity, is essential to the longevity of a company. It might however, also be possible that R&D, is a much stronger signal that a company has additional cash to spend compared to values like advertising, dividends and purchase of stock and that R&D is the first place they would go to cut back on spending when they are in financial distress.

(5) - (6) tracks the relationship between long-term debt (dltt), income (ibc, pi) and price to sales (ps). When a firm's long-term debt is large while simultaneously having negative income and a low price to sales ratios, it is much more likely to be declared bankrupt in the future. These two interactions are collectively much more important than the next biggest interaction. It should be noted that the combined interactions lead to the final classifications and that they should not be used on their own. For example, if a firm has long-term debt and large sales compared to their valuation but are not profitable in the short-run, then they are more likely to be a failing firm, but a simple out-of-sample screening of these types of firms shows that among 'struggling firms,' it picks up companies like the Ford Motor Company and Hewlett Packard (10 Oct 2017). Only time will tell whether these firms are truly in financial distress using such a simple heuristic as the aforementioned interaction, but it is highly unlikely as the fully built out model considers hundreds of thousands of interactions more before making a prediction.

Table 15: Depth 3 - Interaction Analysis

	Term 1	Sign	Term 2	Sign	Term 3	Sign	RII	Gain
(1)	epsfx	-	ibc	-	totdebt_invcap	+	100	456
(2)	debt_at	+	pi	-	rd_sale	-	95	435
(3)	dpc	-	equity_invcap	+	ps	-	92	419
(4)	ibc	-	ps	+	totdebt_invcap	+	88	402
(5)	dltt	+	ibc	-	ps	-	84	383
(6)	dltt	+	pi	-	ps	-	84	382
(7)	ibc	-	ps	-	ps_prtc	-	83	378
(8)	ibc	-	ibc	-	totdebt_invcap	+	79	362
(9)	dltt	+	ps	-	txt	+	76	348
(10)	ibc	-	ppent	+	ps	-	74	336
(11)	at	+	debt_at	+	epspx	-	68	310

Out of the top 50 variable list, there are millions of ways to conjure up directional relationships. Due to the nature of nonlinear relationships, to conceptually understand the web of relationship it is best to identify the top interaction pairs. This table represents the most important interaction pairs as measured by the gain statistic at an interaction depth of three. For easier reading, I also report the relative interaction importance (RII). The sign purely indicates the average direction of each variable. The interaction terms are much more informative than single standing variables. Interactions are at the core of what gradient boosting tree models are all about. Unique *Terms 1* are EPS (Diluted) - Excl. Extra. Items (epsfx), Assets - Total (at). Unique *Terms 2* are Common Equity/Invested Capital (equity_invcap). Unique *Terms 3* are Research and Development/Sales (rd_sale) and Income Taxes - Total (txt).

After surveying past literature, it has been noted that no study has made use of PCA transformations to look at high-level interaction effects between different accounting dimensions. Although this abstracts a lot of the minute interactions away, it still offers an important view of the importance of using different accounting dimensions to predict future bankruptcies. In *Table 16*, I present the most important category interactions. The three most important interactions at depth two are Assets & Liability with the Solvency dimensions (934) and Assets & Liability with the Profitability and Valuation dimension (658). This analysis further emphasises the importance of including fixed accounting values in higher-dimensional bankruptcy models. The analysis shows that the most important interactions occur between a fixed value and ratio dimensions.

Table 16: Cross Tab - Category Interactions

	Asset & Liab	Cash Flow	Efficiency	Equity	Income	Liquidity	Other	Profit & Value	Solvency
Asset & Liab	0	446	31	102	603	573	122	658	934
Cash Flow	446	0	143	267	69	241	188	220	34
Efficiency	31	143	0	279	63	95	157	23	59
Equity	102	267	279	0	335	73	211	420	90
Income	603	69	63	335	0	288	59	578	528
Liquidity	573	241	95	73	288	0	116	283	381
Other	122	188	157	211	59	116	0	465	82
Profit & Value	658	220	23	420	578	283	465	0	88
Solvency	934	34	59	90	528	381	82	88	0
Total	3472	1608	851	1779	2547	2050	1400	2740	2196

This table represents all the interaction pairs between the first PCA component of each category dimension as measured by the gain statistic at an interaction depth of two.

IX. Filing Outcomes

The prediction of all filing outcomes is contingent on a correctly predicted bankruptcy outcome. In this section, I use a GBM model with simple accounting value inputs the year before the filing date to predict bankruptcy outcomes. The summary statistics of filing outcomes can be found in *Table A33*. As mentioned in the *Literature Addendum*, filing outcomes have great economic consequence for creditors and shareholders. Stakeholders not only want to know the likelihood of a litigated bankruptcy occurring, but also all the filing outcomes associated with the predicted bankruptcy. In *Table 17*, I present the performance of five different filing outcome models. The first of these five is the chapter prediction model. It involves a prediction task of whether the bankruptcy will finally be filed under Chapter 7 or Chapter 11. The Chapter prediction model performed the best of all other filing outcomes models. It achieved an AUC of 0.88. The survival prediction model that identifies whether the firm would emerge from bankruptcy performed second best with an AUC of 0.73. The task that attempts to predict whether assets will be sold in a 363 Asset sale or by other means came in third with an AUC of 0.64. The duration task, which involves the prediction of whether the bankruptcy proceedings would endure for longer than one year, came in second to last with an AUC of 0.62. And lastly, the tort task had an AUC score of 0.54, which is only slightly higher than random. All prediction tasks performed better than random guessing.

Table 17: Binary Classification Performance for Predicting Bankruptcy Characteristics

Binary Classification Model	ROC AUC Score	Accuracy Score	Average Precision Score	False Positive Rate	False Negative Rate
Duration	0.62	0.56	0.69	0.66	0.26
Survival	0.73	0.69	0.80	0.61	0.12
Chapter	0.88	0.95	0.70	0.05	0.20
Asset Sale	0.64	0.66	0.39	0.27	0.55
Tort	0.54	0.90	0.17	0.05	0.83

This table reports six important metrics for five alternative classification tests to predict the outcome of predicted bankruptcies. *Duration* classification is the first task to predict the binary outcome. This task involves the prediction of whether or not the disposition will take longer than a year after the initial filing. *Survival* predicts a binary outcome as to whether or not a firm will re-emerge out of bankruptcy and remain in business for longer than 5 years. The *Chapter* task predicts whether the bankruptcy filing will be converted to Chapter 7 or whether it will be a Chapter 11 filing. The *Asset Sale* model predicts whether the debtor will sell all or substantially all the assets during the Chapter 11 proceedings. The *Tort* classification task seeks to predict whether the bankruptcy will occur as a result of tortious actions such as product liability, fraud, pension, environmental, and patent infringement claims. The above metrics have been fully defined in table X.

Table 18 identifies the most important variables to each of the prediction tasks, including the most important categories. To obtain the categories, I apply a PCA to the variables in each category and select the first principal component to represent that category. The model column includes the multiple outcomes for which models were created.

Table 18: List of Each Outcome Model's Most Predictive Variables and Categories

Model	Selection	RI - Rank 1	RI - Rank 2	RI - Rank 3
(1) Duration	*Variable*	Cash and Short-Term Investments (CHE)	Net Profit Margin	Inventories - Total (INVT)
		- 100	+ 86	+ 86
	Category	Asset & Liab. + 100	Profit & Value + 48	Solvency + 15
(2) Survival	*Variable*	Current Liabilities/Total Liabilities	Asset Turnover	Investment and Advances - Equity (IVAEQ)
		- 100	+ 50	+/- 50
	Category	Assets + 100	Solvency + 36	Income + 27
(3) Chapter	*Variable*	Receivables Turnover	Payables Turnover	Current Liabilities - Total (LCT)
		+ 100	- 57	+ 43
	Category	Efficiency + 100	Solvency + 26	Asset + 10
(4) Asset Sale	*Variable*	Current Liabilities/Total Liabilities	Depreciation and Amortization (DPC)	Long-term Debt/Total Liabilities
		+ 100	- 93	- 93
	Category	Asset + 100	Profit & Value + 75	Solvency + 55
(5) Tort	*Variable*	Accruals/Average Assets	Minority Interest - Balance Sheet (MIB)	% Change in Forward P/E to 1-year Growth (PEG) ratio
	RI	- 100	+ 67	+ 67
	Category	Solvency + 100	Profit & Value + 60	Asset & Liab + 60

This table reports the top three *Variables* and the top three *Categories* for five different binary classification tasks including the associated relative importance (RI) among the three values. Number (1) - (5) identifies the different filing outcome-classification models.

For each model, the most important variable and accounting category are highlighted. Furthermore, a relative importance measure and the direction to the outcome are also provided. *Table A29* further reports the filing outcome statistics of the most important variables. The results in *Table 18* are interesting for a few different reasons. The first is that these values give the reader some insight into what values the court takes into account or what values are associated with factors the court takes into account before ruling on filing outcomes. The good performance of these models leads me to believe that the inherent characteristics of the firm are important factors that affect the outcome of the bankruptcy. In the next few paragraphs, I will consider the most important variables for each classification model.

(1) Duration: The results of *Table 18* show that firms with increased cash and short-term investments, including increased inventory and a higher net profit margin, will, all else equal, spend more than a year on bankruptcy proceedings. The reason for this is possibly that high net profit firms are more complex to unravel. These firms are also worth saving and spending time on due to the enhanced prospect of creditors reaping the returns in the future if the firm gets sold as a going-concern.

(2) Survival: Firms with low current-to-total liabilities ratios, high investment and advances, as well as a good asset turnover ratio tend to survive the bankruptcy process. If you have proportionally low current liabilities, it means that there are fewer pressing demands in the short-run, which would allow a firm enough time to get back on its feet and re-establish itself as a going-concern. High investment and advances in affiliates, associates, and subsidiaries values indicate a larger interest in the success of the firm. The mere fact that affiliates or subsidiaries exist is enough of an indication that a firm suffering financial pressure can be "rescued" from bankruptcy proceedings by affiliates or subsidiaries. The subsidiaries can be sold to finance the survival of the firm under pressure. A high asset turnover ratio is indicative of a firm generating more revenue per unit of assets. Firms that are efficient, all else equal, will be more successful in emerging from the bankruptcy process; efficiency is likely to be an important factor the court takes into account.

(3) Chapter: If a firm's receivable turnover is low and payables turnover is high while current liabilities are high, then they are more likely to file for Chapter 7 bankruptcy. Firms that pay their average payables more frequently than they collect their average receivables are struggling with short-term liquidity. However, such a pattern or relationship is also indicative of firms giving greater importance to paying back their creditors than to collecting their own debt. Ch. 7 bankruptcy is often referred to as liquidation bankruptcy as these firms are past

the stage of reorganisation. Ch. 7 bankruptcy highlights the process of absolute priority. Trustees are appointed to ensure that the proceeds are paid to specific creditors. It therefore makes sense for these creditors to initiate the Ch. 7 process before the firm remits more money to lower-ranked short-term creditors in the form of payables while ignoring the first priority creditors. Similarly, priority or secured creditors would be concerned with firms that have a disproportionate amount of current liabilities and would initiate action to ensure that they see some of the proceeds coming their way first.

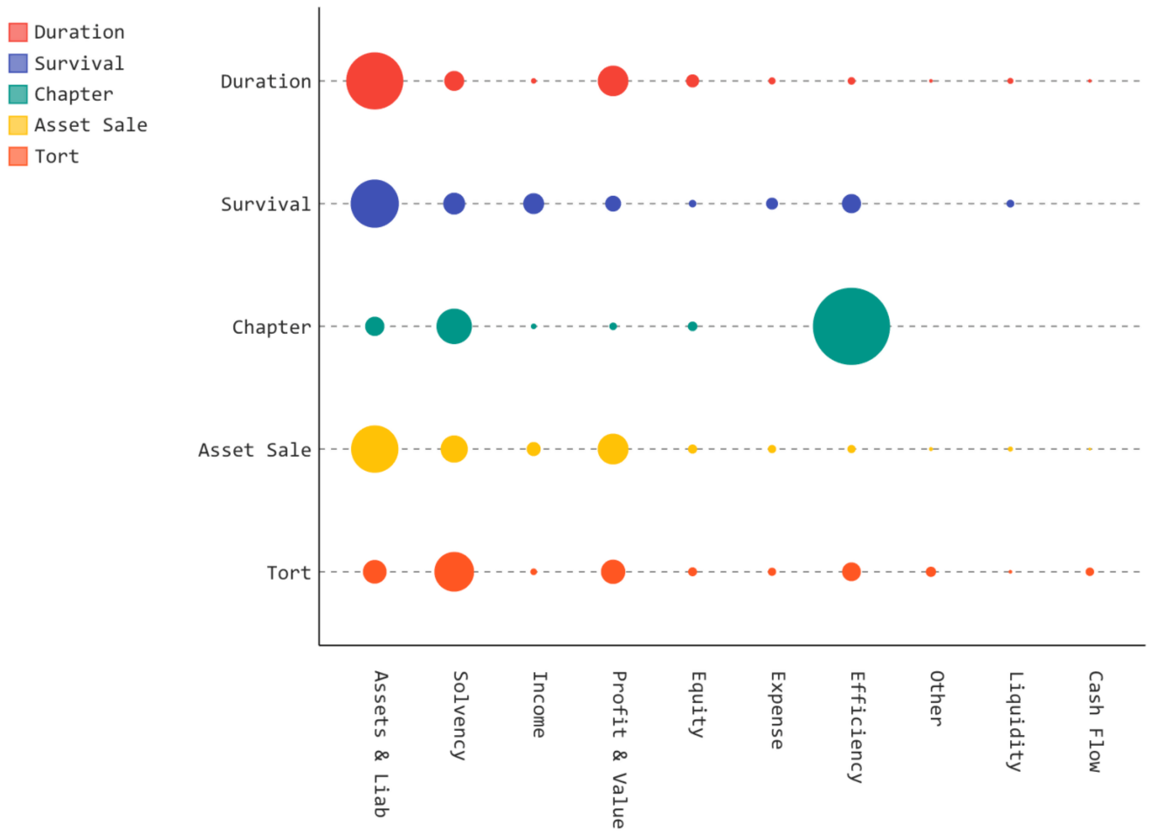
(4) Asset Sale: Firms with high current to total liabilities, and low depreciation and amortisation, and low long-term to total liabilities are likely to sell their assets in the bankruptcy process. The sale of assets is mostly deemed the sale of substantially all the assets of the firm of a Chapter 11 debtor. There are 228 observations of asset sales, and they include the sale of assets under Chapter 7, of which there are 24 observations. These type of asset sales are more commonplace in recent years. Similar to the previous sections, when a firm has a high current liability, the secured creditors put pressure on the company to sell off assets so that they can receive some of the proceeds, as this is the fastest way for them to receive adequate reparations. What is also interesting is that asset sales become more likely when the long-term to total liability ratio is low; this means that when the size of the secured long-term creditors is small, then the small group of creditors have the incentive to sell off the assets to recuperate more money instead of allowing for the reorganisation process to go ahead. This holds especially true when the depreciation and amortisation are low, which is indicative of a lower level of fixed assets making the sale of assets easier than reorganisation.

(5) Tort: Generally speaking, it is not easy to predict tortious events. Chaudhuri and De (2011) note that no models have yet been successful in detecting corporate fraud. More than half of the tortious filings relate to fraud; the other filings relate to environment and product claims. Although tort prediction in this study did not perform well, it showed some interesting associations. When accruals to assets are low and when the total percentage of minority interest is low, then there is a tendency towards a tortious bankruptcy. It could be that management knows that their company is under scrutiny, and for that reason, they underreport their accruals. It could also be the case that potentially tortious companies are selling off subsidiaries to protect them from future claims, increasing their minority interest, but without proper analysis this is mostly speculation.

The last part of this analysis includes a quick categorisation of all the variables to identify which accounting dimensions are the most important for predicting the various filling outcomes. *Figure 6* shows that the Duration of a firm is primarily driven by the Assets

& Liability, and Profitability & Valuation dimensions. Whether a firm would survive or not, is primarily driven by its level of assets. The chapter under filing is driven mostly by Efficiency ratios. And lastly, whether or not an asset sale or a tortious filing would transpire depends on a wide range of categories but primarily by the Assets and Liabilities Dimension.

Figure 6: Bubble Plot and Ranking of each Model's Most Important Categories



This figure reports the relative importance of the five outcome classification models and the associated accounting dimensions. There is a large amount of heterogeneity between the different classification models.

X. Conclusion

This study shows that a Gradient Boosting Tree Model (XGBoost) outperforms some of the latest deep learning networks (DCNN). It also shows that the creation of a meta-model (stacked model using RF, AdaBoost, DCNN, and FNN) outperforms all the individual parts. The study highlights the importance of not only using financial ratios but also including dollar accounting values as inputs to the prediction model. The overall model shows that Assets & Liability values and Solvency ratios are the most important dimensions in predicting bankruptcy. These categories have been ranked using nine different importance metrics. The models developed in this study make use of accounting variables and a conservative sample. Notwithstanding this fact, the DCNN model used in this chapter is the best performing neural network compared to all past studies. Furthermore, the XGBoost model used in this study has comparable performance to the best models of past studies.

I find, similar to more recent literature, that less restrictive, higher-dimensional models with more variables will outperform most linear models in bankruptcy prediction tasks. A significant number of past bankruptcy studies show a small picture of the larger subset of relationships by using restrictive low-dimensional models and a small set of variables. A range of variables has been found to have a strong association with bankruptcy, many of which have not been noted by past research such as the level of stockholder's equity, inventory, depreciation and amortization, and the research and development to sales ratio.

A pitfall of the vast majority of the past higher-dimensional bankruptcy research is that they do not test the correlations between variables before identifying the most important variables, leading to invalid importance measures. A way to deal with the multicollinearity between variables is to categorise variables and to ensure that there is little to no collinearity between the first component PCAs of the distinct categories. According to the categorisations in this study and the associated ranking metrics, the most important dimension out of ten categories in predicting bankruptcy is Assets & Liabilities.

As a result of using a higher dimensional model, this study further reports on the most important interaction pairs that lead to bankruptcy predictions. This study is the first to list the top interaction *pairs* and is also the first study to list the most important interactions between *three* distinct variables. The study found the most important interaction pairs to be between income and the total debt-to-invested capital ratio, and between earning-per-share and the total-liabilities ratio. The most important depth-three interaction is between the earnings-per-share, income, and total debt-to-invested capital ratio.

A significant contribution of this study is the novel focus on litigated bankruptcies, i.e. Chapter 7 and Chapter 11 bankruptcies, which allows for an extended study to, not only predict the future occurrence of bankruptcy, but also to predict binary filing outcomes one year in advance of the filing event. The findings show that many of the filing outcomes can successfully be predicted. It is shown that the level of cash, short-term investments, and inventory strongly affect how long bankruptcy proceedings would endure; and that the current liability to total asset ratio, asset turnover ratio, and investments and advances significantly affect whether a firm would survive the bankruptcy process. Tests before and after the GFC also highlight the dynamic nature of variable importance over the years. The performance of the models changes over time; although the average performance is around 0.957, it ranges from as low as 0.917 to as high as 0.984 ROC (AUC) over different periods. The general trend is that the more data that is included, the better the performance. Other effects of the change in performance over the years include differences in the underlying distribution of bankrupt to healthy firms.

Overall, this study has identified new ways to predict bankruptcies (DCNN); it has shed light on the equivalent importance of accounting values and ratios in bankruptcy prediction when high-dimensional models. The study uncovered variable heterogeneity between time intervals and highlighted the importance of interactions for high dimensional models. It is the first study to predict not only bankruptcy as an outcome but also the associated filing outcomes (duration, survival, asset sale, filing chapter, tort). This study is the first stride towards successfully using high dimensional models to improve both prediction quality and variable analysis.

In the future, it would be interesting to know whether a model can predict the potential refiling of a firm that emerged out of a past bankruptcy. In the same breath, it would be fascinating to see whether the long-term survivability of such firms can accurately be predicted. Other noteworthy prediction tasks could include a classification model that can predict whether a plan for bankruptcy is pre-packaged or whether the plan was simply pre-negotiated, as this can have significant implications for creditors. Future bankruptcy research should also seek to create bankruptcy prediction models using causal specifications. This will be very helpful in developing theory within bankruptcy prediction. It would also be good to replicate the framework in this study to extend the model, to not only predict the future occurrence of bankruptcies, but also predict other risky occurrences such as liquidation events, financial distress, and mergers and acquisitions.

XI. Appendix

A. Summaries

Table A19: Bankruptcy Characteristics Over Defined Periods

(1) Interval	(2) Description	(3) Bankruptcies	(4) Survived (%)	(5) Tortious (%)	(6) Long Legal Process (%)	(7) 363 asset Sale (%)	(8) Total Assets (Billions)
1980s	Oil & Metal	48	0.77	0.13	0.94	0.08	257
1990s	Wholesale	143	0.64	0.08	0.64	0.15	242
Early 2000s	Dotcom Bubble	207	0.62	0.09	0.57	0.24	688
Late 2000s	GFC	115	0.54	0.03	0.44	0.40	1,788
Early 2010s	Oil & Electronic	104	0.55	0.02	0.24	0.28	289

(1) The intervals are selected according to periods of themed bankruptcies. (2) indicates the type of bankruptcies that occurred within the interval. (3) is the sample of bankruptcy filings in each region. (4) is the percentage of firms that emerged out of bankruptcy as a result of an agreed plan to re-emerge out of bankruptcy and remain in business indefinitely. (5) is the proportion of bankruptcy cases that resulted from claims for product liability, fraud, pension, environmental, and patent infringement claims. (6) is the percentage of cases in which the disposition took longer than two years after petition filing. (7) The 363 Sale occurs when the debtor sells off substantially all of its assets during the Chapter 11 proceedings; the court can then distribute the proceeds of sale or convert the case to Chapter 7. (8) is the report of the sum of all firm assets as reported on the most recent 10-k filing prior to the bankruptcy petition.

Table A20: Bankruptcy Characteristics Across Industries

Industry	(1) Bank-ruptcies	(2) Sur-vived (%)	(3) Long Legal Process (%)	(4) Tortious Bank-ruptcy (%)	(5) Average Duration (Days)	(6) 363 Asset Sale (%)	(7) Total Assets (Billions)
Agricultural	3	0.67	0.33	0.00	234	0.67	3
Construction	15	0.87	0.60	0.07	662	0.07	14
Finance	99	0.24	0.54	0.05	617	0.29	2,017
Manufacturing	191	0.72	0.58	0.11	618	0.29	524
Mining	47	0.66	0.30	0.02	299	0.17	75
Retail Trade	78	0.58	0.67	0.03	645	0.29	110
Services	72	0.64	0.42	0.08	411	0.18	122
T, C, E & G	90	0.70	0.57	0.02	570	0.17	375
Wholesale Trade	22	0.64	0.50	0.09	550	0.14	25

This table reports the distribution of various bankruptcy outcomes across industries. It reports the legal descriptive statistics as obtained from the firms' bankruptcy case filings. (1) is the sample of bankruptcy filings in each industry. (2) is the percentage of firms that emerged out of bankruptcy as a result of an agreed plan to re-emerge out of bankruptcy and remain in business indefinitely. (3) is the percentage of cases where the disposition took longer than one year after petition filing. (4) is the proportion of bankruptcy cases that resulted from claims for product liability, fraud, pension, environmental and patent infringement claims. (5) reports the average duration for bankruptcies in the industry. (6) is about the 363 sale that occurs when the debtor sold off substantially all of its assets during the chapter 11 proceedings; the court could then distribute the proceeds of sale or convert the case to Chapter 7. (7) is the sum of all firm assets as reported on the most recent 10-k filing prior to the bankruptcy petition.

B. Literature Addendum

a) Variables and Categories.

Gradient boosting (XGBoost) and many other machine learning algorithms allow for a free complexity parameter (Leathwick, Elith, Francis, Hastie, & Taylor, 2006). XGBoost measures model complexity by the depth of the tree as well as the number of trees and the learning rate. Algorithms gain a lot of strength from being able to choose complexity. These added dimensions of complexity are the hidden strength of the gradient boosting and other machine learning models' flexibility and performance. As a result, researchers can start out with a rich set of inputs and allow the model to decide what inputs to use endogenously. A known issue with added complexity is the proclivity of overfitting. Overfitting can be remedied by testing the model on an 'out-of-sample' dataset as far as the model is concerned (Witten, Frank, Hall, & Pal, 2016). Further explanation of these concepts appears in appendix XI.D.4 on page 67. Apart from allowing researchers to start with a richer collection of potential candidate variables, the enhanced model complexity and endogenous variable selection allow researchers to simply allow the data to identify the most important variables without human intervention.

The next step is to identify a blanket set of variable inputs for the model to analyse. The start of bankruptcy literature can be traced back to a report in the 1930s from the Bureau of Business Research (BRB). The BRB compared the ratios of 29 failing firms with the average ratios across the firms to identify the characteristics associated with failure. This rudimentary report showed the importance of ratios such as working capital to assets, reserves to total assets, net worth to fixed assets, sales to assets and cash to total assets in describing bankruptcies. More than 80 years later, the majority of predictor variable identification literature executes similar tasks. The literature did, however, evolve from simple mean comparisons and univariate statistical techniques between healthy and failed firms to the development of prediction models and statistical significance tests on variables (Altman, 1968; Beaver, 1966; FitzPatrick, 1932). Even in the recent wave of machine learning models, predictor variable identification is still necessary for describing the occurrences of bankruptcies, albeit at a higher dimensional feature space (Behr & Weinblat, 2017; Jones et al., 2017; Mselmi et al., 2017).

After the BRB report, similar exercises soon followed. FitzPatrick (1932) was the first bankruptcy researcher to report that the interactions between variables have important consequences in bankruptcy prediction. He reported that less emphasis should be placed on liquidity ratios for firms that have long-term liabilities. Interactions are notably one of the great strengths of high-dimensional models. Unlike traditional linear models, interactions do not have to be created by hand, nor fathomed by mind like FitzPatrick did (1932). Further pioneering research includes the work of Chudson (1945) who was the first author to do a form of elementary pattern recognition. He showed that specific industry, profitability, and size groups lead to clusters of similar ratios. Future research went on to show that clusters of similar ratios can also be used to predict bankruptcy. These early studies are important theoretical contributors to the recent wave of higher-dimensional studies that use machine learning algorithms to predict bankruptcies.

Although past research has focused on the use of accounting (Zmijewski, 1984; Altman, 1968; Beaver, 1966), market (Duffie, Pedersen, & Singleton, 2003; Ohlson, 1980; Shumway, 2001), corporate governance, industry, analyst, and economic-based variables as inputs (Beaver et al., 2005; Jones & Hensher, 2004; Jones, 2017), this study simply focuses on accounting-based measures for the following reasons. First, additional model variables would consume the variable importance normally associated with account-based variables, due to strong correlations between categories of variables; second, a cornerstone to this study is the identification of symptoms in the US and not prediction accuracy per se; third, the model would also be applied to the prediction of alternative bankruptcy characteristic; and lastly, although past research has shown that models that only focus on accounting information do not provide the best predictive performance (Li, 2012), I believe that this can be attributed to model constraints. More recent high dimensional research shows that a model that incorporates Market Price, Ownership Concentration/Structure Variables, External Rating, Macroeconomic Variables, Executive Compensation, and Accounting Based Variables improves the model quality as measured by the ROC (AUC) score with only 4% as compared to pure Accounting Variables (Jones, 2017). In this study, I argue that all the other values are simply proxies to the pure accounting values. In saying that, the addition of these variables should be attempted in future bankruptcy pattern studies so long as importance measures can be statistically motivated to discriminate between collinear variables to avoid falling into the aforementioned traps.

Beaver et al. (2005) noted that accounting-related variables have performed without issues over a long period. Many empirical studies have led to a great number of accounting-related variables that have been shown to have power in bankruptcy prediction. Aziz, Emanuel, and Lawson (1988) observe that ratios in bankruptcy studies are based on ad hoc pragmatism rather than sound theoretical work. This study consequently includes a blanket of more than 70 accounting ratios. In line with past research, this study includes financial values and ratios such as (1) Solvency ratios (Cash Balance to Total Liabilities, Total Debt to Assets, Capitalisation Ratio), (2) Liquidity ratios (Current Ratios and Cash Ratios), (3) Profitability ratios (Net Profit Margin, Return on Equity), (4) Efficiency Ratios (Asset Turnover, Payables Turnover), and Valuation Ratios (Price/Sales, Price/Book). Consistent with past research, the expectation is that changes in the above ratios will affect the probability of bankruptcy.

An important concept of high dimensional models is that interactions happen in many more ways than just the numerator-denominator interactions such as predetermined in ratios; for that reason, *control* variables in the form of (5) Assets (Total Assets, Current Assets), (6) Liabilities (Total Liabilities, Current Liabilities), (7) Equity (Stockholders Equity, Depreciation and Amortisation), (8) Cash Flow (Operating Activities - Net Cash Flow), (9) Expenses, and (10) Income are also included. These variables were selected based on the standardised financial accounts on Compustat. These variables do not just facilitate controls; they, in fact, have important interaction effects acting as pseudo-ratios without the restrictive linear constraint of the numerator-denominator relationships. The overall interaction effects produce far more interesting and predictive variable combinations than simple numerator-denominator combinations. The expectation is that the fixed values in aggregate will have more predictive power than the ratios. For that reason, apart from just studying the predictive ability of the ratios, attention should also be given to the interaction effects of both the ratios and 'controls.' For all fixed values and ratios, I also include ten of the most promising annual growth values, i.e. percentage changes in the value over the last year.

As important as individual variables are to this study, the vast majority of learning algorithms cannot discriminate between highly correlated variables. As a result, the categorisation of variables becomes an important step in understanding the true relationship between accounting-based values and bankruptcy events. Although a vast majority of the papers on high-dimensional models include category labels, no study attempts to show empirically which categories or dimensions of accounting values and ratios are the most important to predict bankruptcies (Behr & Weinblat, 2017; Jones, 2017; Kim & Upneja, 2014). This paper makes use of nine different techniques to rank-order categories to obtain the final rank of the different dimensions. In this study, the categorisation is achieved by including variables by relatedness of theme rather than construction. For example, the liabilities category would include the dollar change in liabilities, current liabilities, and % change in liabilities values. The MDA approach, which pairs samples of failed and non-failed firms using financial ratios, generally shows that solvency, profitability, and liquidity indicators are the most significant indicators (Almamy, Aston, & Ngwa, 2016). However, the order of importance of these categories is unknown as past studies did not use a standardised set of ratios to measure the health of firms (Altman, 1968). In this study, I attempt to solve this problem by starting with a large base of variables classified into Solvency, Profitability, Liquidity, Efficiency and Valuation Ratio categories and allowing the machine learning model to decide what variables and categories are the most important by analysing the patterns in the data. I also establish categories for the fixed accounting values.

2. Models

Later studies from the 1960s to present contributed to the use of LDA and MDA prediction models instead of mean comparison studies. These historical and subsequent traditional statistical models are not of much consequence to this study apart from attempting to highlight the advantages and disadvantages of this group and new high dimensional machine learning models (see the appendix XI.E on page 70). I further compare the performance of the Logit model to other models in this study. The biggest benefit of the past model comparison studies is that they provide a framework for future bankruptcy studies.

In the early years of bankruptcy research, Merwin (1942) revealed that failing firms showed signs of deterioration as early as five years before failure. A few years later, Altman (1968) showed that by the use of an MDA model that he could predict insolvencies one to two years in advance. In this study, I make use of this knowledge and predict bankruptcies one to two years into the future. Within this definition, it is important to predict the year within which the company failed (Ohlson, 1980).

This study, unlike the majority of studies in this domain, predicts bankruptcies across a broad range of industries. The study also uses an XGBoost model that has many advantages over linear models (Chen & Guestrin, 2016). Recent literature has shown that decision tree ensembles (multiple decision trees) and more specifically boosting ensembles (re-weighing tree importance to assist bad performing trees) almost always come out on top, with the added benefit that they are easier to conceptualise than many black box models such as neural networks (ANN) (Barboza et al., 2017; Jones, Johnstone, & Wilson, 2017b; Olson, Delen, & Meng, 2012; Zięba, Tomczak, & Tomczak, 2016). Olson, Delen and Meng (2012) compared five machine learning models, including ANNs, and found that the decision-tree related models outperformed. Recent evidence by Jones (2017) also highlights this outperformance. In the past, more research has been conducted in the univariate category than in the multivariate category, but that has slowly changed over the years. The structure and internal workings of these models are described in the appendix from page 64 onwards. However, it may be worth the effort to understand some of the concepts.

Due to recent advancements and the re-emergence of artificial neural networks (Barboza et al., 2017; Kim & Kang, 2010; Mselmi et al., 2017; Zhao et al., 2017; Zhou, 2013), I will also provide an alternative model to regular ANNs, in line with popular developments in deep learning[9] research. To do this, I make use of two different models: Feed-forward Neural Networks (Hornik, Stinchcombe, & White, 1989), and Deep Convolutional Neural Networks (DCNN) (Krizhevsky, Sutskever, & Hinton, 2012). I show that DCNNs outperform all past ANNs (refer to literature Table A25). The use of machine learning and deep learning in finance is becoming more common as researchers slowly uncover the nonlinearity of financial data. Callen et al. (1996) showed that machine learning models have long been able to beat time-series models in forecasting. Xiao et al. (2013) demonstrated the power of ensembles in financial market forecasting; they showed that the flexibility of the ensemble approach is key to their ability to capture complex nonlinear-relationships to predict future stock prices. This paper draws inspiration from recent machine learning applications in economics, such as those in papers by Mullainathan et al. (2017) as well as machine learning applications in price behaviour prediction (Bagheri, Peyhani, & Akbari, 2014; Teixeira, De Oliveira, & Adriano Lorena Inacio, 2010) and high-dimensional prediction studies (Jones et al., 2017).

[9] The multi-layered hierarchical representation of data using neural networks.

3. *Predictive Power*

With high-dimensional decision tree models, two potential candidates for measuring impurity to identify variable performance are Entropy and Gini Index. Gini importance makes use of the Gini Index whereas Information Gain makes use of Entropy as error measures. These values are often used interchangeably. A paper by Raileanu and Stoffel (2004) reports that these measures disagree only 2% of the time due to the similar nature in which they are calculated. Both of these measures are used in machine learning research. Gini importance is often preferred because it is less computationally expensive as it does not require a logarithmic calculation like entropy. Both of these measures attempt to measure the decrease in impurity or uncertainty that each variable provides. Both are data-centric approaches that look at the predictive ability of a parameter based on variable selection ranking in the nodes of the trees.

The benefit of the gain measure is that there are some empirically substantiated alternative measures that have been derived from the original measure; this includes measures such as the information gain ratio (Quinlan, 1986) and the expected gain measure (MacKay, 1992). There has also been a recent development in open source packages like Xgbfi that offer alternatives like average gain, which is the gain divided by the number of possible splits taken on a variable or variable interaction. In this study, I therefore make use of the Gain measure. The Gain measure is equal the number of times a variable is selected for splitting weighted by the squared improvement each split adds to the model average over all the trees (Friedman & Meulman, 2003). Friedman (2001) also established a relative measure that is essentially the contribution of each variable scaled to the performance of the best indicator multiplied by 100 (RVI). In this study, I report both the RVI value and the percentage Gain measure. The larger this number, the greater the effect a variable has on the response. The other measure used in this study is Split Frequency, which is the number of times a variable is selected for splitting in all decision trees. Frequency is presented as a simplified measure of gain.

4. *Filing Outcome Prediction*

The Bankruptcy Code ("Code") came into effect in 1978. The act afforded the debtor substantial protection against creditors but to different degrees depending on the filing outcome of the bankruptcy. It is a well-known fact that bankruptcy can be costly; LoPucki and Kalin (2001) show that the direct costs for large public firms can amount to between 1.5-6% of a firm's assets. Given the disparity in cost between bankruptcies, this study argues that it is important to not just predict bankruptcies but also the associated characteristics of the proceedings.

Filing outcomes are important to all creditors and investors. Franks and Torous (1989) note that due to the cost and length of the proceedings in Chapter 11 bankruptcy, the legal and administrative costs may, in many situations, be against the interest of stockholders. Oliver Hart (2000) argues that there should be a greater push towards cash auctions as they are simple and efficient. However, a lot of bankruptcies still occur under Chapter 11 type structured bargaining, which is often costly and time-consuming as a result of the tensions between the parties involved.

The duration of disposition becomes costly for many parties including the executives, stockholders, and creditors. Predicting the duration of a bankruptcy disposition after the filing is therefore of important economic consequence. Results by LoPucki and Doherty (2008) show that case duration is an important determinant of the fees and expenses in large public bankruptcies. Li (1999) asked an important question, "Can this variation in bankruptcy duration be explained by the financial/industrial characteristics of the distressed firms?" In this paper, I seek to answer this question by identifying which accounting-based values have the most predictive power in predicting duration. I similarly ask the same question for all the other filing outcomes.

A 363-asset sale ("asset sale") can enable debtors to expediently and effectively separate a business, which is one of the core goals of the Chapter 11 reorganisation process. An asset sale can be appealing to debtors and creditors alike. In recent years, debtors have increasingly opted to sell their assets rather than restructure under the Chapter 11 process (Baird & Morrison, 2011). A few advantages include the ability of the purchaser to take the assets clear of liens and claims. It is also up to the debtors to pick favourable contracts. It does come with trade-offs such as the negative publicity of selling off assets (Sable, Roeschenthaler, & Blanks, 2006).

The Chapter under which the bankruptcy is filed is also an essential characteristic that shareholders and creditors want to be made aware of as soon as possible. A paper by Bris, Welch and Zhu (2006) shows that Chapter 7 liquidation (cash auctions) can be more expensive in direct costs and as expensive as Chapter 11 bankruptcies in indirect costs. Chapter 7 liquidation does not appear to be more expedient or cheaper than Chapter 11. Moreover, Chapter 11 seems to better preserve assets, allowing creditors to recover more.

The last filing outcome prediction task is whether the firm will file for bankruptcy as a result of tort claims. Torts involve large sums of money that can quickly overwhelm a company (Hardiman, 1985) Tort claimants qualify as creditors under the bankruptcy act and partake in the reorganisation process. Torts in the bankruptcy process are famous for the enormous future liabilities they can entail (Bibler, 1987). It is easy to see how this can undercut commercial creditors and shareholders, and it therefore has significant economic consequence and is worth knowing. This study shows that it is extremely difficult to predict whether a company that is predicted to file bankruptcy will file under tortious Chapter 11 bankruptcy.

A limited number of studies has looked at post-filing resolutions. Barniv, Agarwal, and Leach (2002) showed that knowing the outcome in advance can have immense economic consequences. They noted that significant abnormal returns can be earned if a firm emerges or gets acquired between the filing and disposition resolution, whereas liquidated firms experienced significant negative abnormal returns. The former experienced 155 percent abnormal returns on average, whereas the liquidated firms experienced negative 11 percent returns. Barniv et al. (2002) used a multi-labelled Logit model for predicting firms that will emerge from Chapter 11 bankruptcy. They correctly classified 62% of the firms.

In this study, I show a 70% accuracy using a binary XGBoost classifier to identify firm emergence, which is strictly limited to the use of accounting-related values. I believe that the results obtained in this study can be immensely improved by including additional variables such as the institutional ownership, the filling state, the court and the judge at hand, etc. Barniv et al. (2002) similarly noted that variables such as the resignation of executives, fraud, and other characteristics are important in predicting company survival. More recent research by LoPucki and Doherty (2015) recommended the use of information such as whether firms release press reports regarding bankruptcy and the headquarters' state.

Creditors and investors should not be satisfied with the mere prediction of firm distress and bankruptcy. Stakeholders or prospective stakeholders should have a model whereby they can predict not just the occurrence but also the terms of the bankruptcy in advance of the filing to improve risk management practices. The economic effect of the outcome is largely determined by the characteristics associated with the bankruptcy. As a result, the study extends the prediction to include important bankruptcy outcomes, such as how long the bankruptcy process will endure, whether the firm will successfully emerge after the bankruptcy period, whether the bankruptcy is tortious, and whether it will involve asset sales; all this is done by solely evaluating the accounting variables before the bankruptcy filing.

C. Robustness

1. *Performance-validation*

Table 21 below reports some interesting results for this study. It has to do with the differences in model performance as a result of variations in the train-test splits. In summary, (2) time-split is the longitudinal split in the data to ensure that future information does not leak into the past; (3) is the cross-sectional randomised splits between the train and test sets. In this study, I make use of 10-folds or rounds. Lastly, (4) this time split fold is a unique combination of the TS method and the fold method. It is the most robust and accurate measure of true out-of-sample performance. The full design of this method is presented in the appendix, XI.D.4 on page 67. The time-split fold method, as a result of construction, produces metrics for many sub-periods. It is an effective method to look at the consistency of prediction quality over many years. The deconstructed results of this method can be found in the appendix *Table* A30 and *Table A31*.

Time split fold provides evidence of model performance not just over different time intervals but also for different levels of training data. The results of *Table* A30 on page 85 reflect two important generalisations of machine learning prediction in time series. The first is that with the inclusion of more data, the model tends to perform better (Domingos, 2012). But this does not always hold true for time series data; the reason is that that the learning, i.e., pattern recognition, can occur over different seasonal trends leading to worse future predictions. There are, of course, ways to mitigate this, such as incorporating seasonal indicators as variables in the model. In machine learning, it is desirable that the distribution of the train and test data is the same, but this is not always possible, especially with financial data (Montas, Quevedo, Prieto, & Menndez, 2002). The expectation is that the results will improve if the distribution remains unchanged. Splitting training and test sets by time intervals and checking for parameter stability over time is a very useful exercise in building a robust model and understanding how the model learns and predict. The best reported AUC of 0.984 occurred over the last few years of the 2014-2016 sample. The worst AUC occurred over the 2003-2004 period with an AUC of 0.917 (Table A30). Also note that the performance of the models can be significantly affected by changes in the depth of the tree and adjustments to the underlying sample distribution, see Appendix 2 - Hyper-Parameters.

Table 21: Model Comparison Using Different Performance Validation Procedures

Metrics	(1) All Data	(2) Time-Split (TS)	(3) K-Fold (KF)	(4) Time Split Fold (TSF)	95% Confidence (+/-)
ROC AUC Sore	0.9587	0.9655**	0.9467**	0.9570	0.0142
Accuracy Score	0.9755	0.9837	0.9682	0.9712	0.0163
False Positive Rate (p-value)	0.0037	0.0069	0.0028	0.0039	0.0015
Cross Entropy	0.1414	0.0825	0.1301	0.1052	0.0707

This table compares the performance of the best models that resulted from different out-of-sample performance tests. (1) The original "All Data" model allocates 60% of the observation to the training set, 15% to the development of validation test set and 25% to the test set. The 15% is used to measure and improve the performance of the model. The observations to each of the splits are randomly selected. (2) TS is a simple ordering of the observation in time series and the creation of longitudinal training - 60%, validation - 15% and test set splits -25%; this method ensures that there is no information leakage from the future observations. (3) KF is a randomised cross-sectional method that scrambles the observations and splits them into training and test sets and calculates the average metrics from 10 different iterations or folds. (4) TSF is the most robust method and has also led to the model with the best generalisable performance as evidenced by the battery of metrics - It is a longitudinal blocked form of performance-validation that suits this form of bankruptcy prediction; it uses the strengths of both (2) and (3). All statistical comparisons are made against the model called "All Data." *p<.1 ** p<.05 *** p<.01. Significance levels are based on a two-tailed Z test.

2. *Hyper-Parameters and Other Adjustments*

Table 5 reports a few adjustments to the model and underlying sample, the first being an adjustment to the XGBoost tree-depth parameter. The model development process shows that the model performance is optimised at a tree-depth of 12. It is therefore expected that any deviation from this number, ceteris paribus, would lead to worse model performance; this has indeed been shown to be the case with a significant reduction in the AUC from 0.9587 to 0.9506. The distribution adjustment test shows that the model performs significantly better when there is an equal amount of bankrupt and healthy firm samples, which is indeed the approach the majority of the bankruptcy studies take.

A branch of literature has for some time been consumed by identifying whether market-based or accounting-based measures are better in predicting bankruptcy. Hillegeist et al. (2004) for example show that option pricing models can provide better estimates of corporate bankruptcy than accounting values. Some studies show that by combining the two one can achieve more accurate results (Beaver et al., 2005; Shumway, 2001). The problem with market-based measures is that they may not necessarily be efficient, especially considering the fact that small firms form the majority of bankruptcy samples. The multi-dimensional approach in this study may put this argument to rest by identifying the high dimensional contribution of market-based and accounting-based measures.

In the next column, I remove valuation ratios that have a price component as it can be argued that they are market-based even though they have an accounting component. This did not result in any significant reduction in the AUC. The table shows that the model performs only slightly but non-significantly worse when one excludes price-related variables. A further correlation analysis in *Table 7* shows that valuation and profitability ratios are highly correlated. For that reason, as long as accounting-based profitability ratios are included, valuation ratios do not lead to significantly improved prediction performance. The last form of adjustment is the removal of growth or percentage change variables. This leads to an insignificant increase in the model AUC. The small change seen can be due to the additional noise created by potentially irrelevant growth variables.

Table 22: Model Comparison Adjusting the Type of Inputs and Model Parameters

Metrics	All Data	Six Depth	Distribution Adjustment	Sans Value	Sans Percentage Change
ROC AUC Score	0.9587	0.9506**	0.9661***	0.9560	0.9597
Accuracy Score	0.9755	0.9752	0.9339	0.9823	0.9784
False Positive Rate (p-value)	0.0037	0.0051	0.0201	0.0073	0.0075
Cross Entropy	0.1414	0.1518	0.1741	0.1424	0.1398

This table compares the results of the various model and sample adjustments. The first column includes the original model; the second column reports the performance of a model where the tree depth parameter is changed to six from the original 12. The third column reweights the sample distribution, the fourth removes valuation ratios, and the last column removes growth variables. All statistical comparisons are made against the model called "All Data." *p<.1 ** p<.05 *** p<.01. Significance levels are based on a two-tailed Z test.

3. *Other Decision Tree Ensembles*

Table 23 reports the additional performance of two alternative decision tree ensembles: the AdaBoost model and the Random Forest model. The AdaBoost model performs slightly better than the Random Forest model. The reason for this additional study is to ensure that XGBoost is the best decision tree-based model for the task at hand. The

XGBoost still outperforms these models. By comparing *Table 1* to *Table 23*, it is clear that the decision tree type models perform especially well in bankruptcy prediction. I hypothesize that if larger bankruptcy datasets can be made available, such that the overall sample of bankruptcies reaches the tens of thousands, then the deep learning models will show immense improvement and most likely beat the performance of the decision tree models. For that reason, I included both of these strains of high dimensional models.

The Stacked model in *Table 23* is quite interesting as it is a combination of the AdaBoost, Convolutional Neural Network, Feed Forward Network, and Random Forest Model into one big model that separately weights the importance of each model's predictions using a final Decision Tree model. Therefore, the combination of all models apart from the XGBoost model seems to perform close to the XGBoost model at the expense of being quite inefficient and expensive to run. To my knowledge, this is the first prediction study that has attempted a stacked model in an attempt to improve prediction quality. A further stacked model that includes the XGBoost model as input showed a statistical improvement over the original XGBoost model with an AUC score of 0.9642 and accuracy of 0.9778.

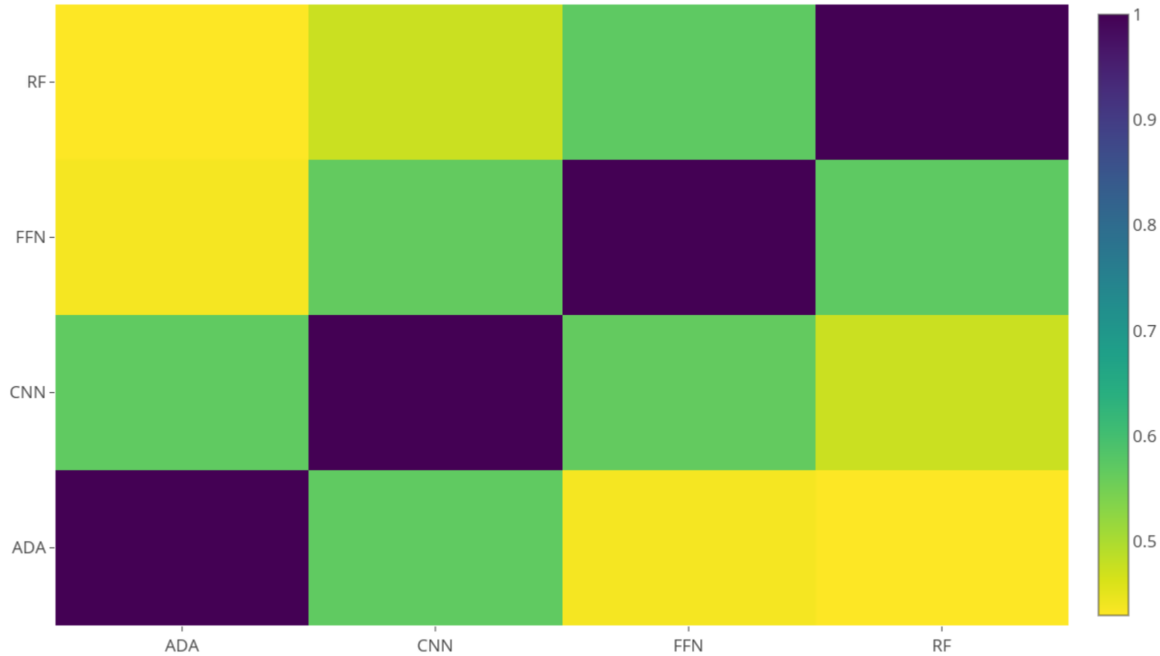
Table 23: XGBoost and Decision Tree Ensemble Model Performance Comparison

Metrics	XGBoost Model	AdaBoost Model	Random Forest Model	Stacked Model
ROC AUC Sore	0.9587	0.9291***	0.9275***	0.9495**
Accuracy Score	0.9755	0.9612	0.9576	0.9681
False Positive Rate	0.0037	0.0185	0.0370	0.0074
Cross Entropy	0.1414	0.1913	0.2409	0.1613

This figure illustrates the performance of the XGBoost model with two other tree ensemble models: AdaBoost and a Random Forest Model. AdaBoost and XGBoost are both ensembles that seek to convert weak learners into a single strong learner. AdaBoost adds weak learners according to performance by changing the sample distribution. In XGBoost, the weak learner trains on the residuals to become a 'strong' learner. Random forests are simply a multitude of decision trees. All three underlying models have decision trees as the base learner. The stacked model is a combination of the AdaBoost, Convolutional Neural Network, Feed Forward Network and Random Forest models into one big model by using the four models' predicted outcomes as inputs to a decision tree model. *p<.1 ** p<.05 *** p<.01. Significance levels are based on a two-tailed Z test.

Stacked models perform especially well when the respective predictions are uncorrelated. *Figure 7* presents a correlation map of these models' predictions. The AdaBoost model (ADA) and Deep Convolutional Neural Network (CNN) is the most correlated model pair. Although similar models in some respects, the AdaBoost and Random Forest models are relatively uncorrelated in their predictions.

Figure 7: Correlation of Predictions Across High-Dimensional Models



This correlation plot shows why the stacked model performed so well as compared to the individual component models. The stacked model is created by combining the AdaBoost, Convolutional Neural Network, Feed Forward Network, and Random Forest models into one big model by using the four models' predicted outcomes as inputs to a decision tree model. Stacking performs well due to its smoothing nature. Stacking is most effective when the based models are less correlated, which is the case for the above models. Stacking is also called meta-ensembles and can be seen as an advanced form of boosting. The results of stacking the models in the correlation map above can be found in *Table 23*.

4. *Time and Variable Variants*

Table 24 onwards only reports the results of the GBM model. The first column repeats the performance of the GBM models as presented in the previous tables. The second column reports the results of a model constructed out of just 50 of the top variables that have been identified in a later section of this study on a validation set using variable selection methods (*Table 4*). It is clear that the model can predict well even with a small number of variables. Further, although significant, the model performance does not change too drastically when we predict bankruptcies one to two years in advance instead of using all observations from both years.

Table 24: Model Comparison Using Different Inputs

Metrics	All Data	50 Variables Model	One Year Before Bankruptcy	Two Years Before Bankruptcy
ROC AUC Sore	0.9587	0.9408***	0.9666**	0.9434***
Accuracy Score	0.9755	0.9700	0.9860	0.9837
False Positive Rate	0.0037	0.0056	0.0010	0.0002
Cross-entropy	0.1414	0.1795	0.1282	0.2206

This table compares the performance of a model that includes only 50 of the most predictive variables as inputs, a model that only includes bankruptcy observations one or two years before the filing. All statistical comparisons are made against the model called "All Data." *p<.1 ** p<.05 *** p<.01. Significance levels are based on a two-tailed Z test.

D. Model Appendix

1. Deep Learning Specifications

The deep feed forward network is a normal sequential model with an input layer followed by four hidden dense layers. The first one has 450 nodes and a ReLu activation function; the second and third have 260 nodes, and the final hidden layer has 240 nodes. I then use a SoftMax activation function that outputs two classes and categorical cross entropy as the loss function while optimising with the Adam algorithm.

The convolutional neural network model uses a 1D convolutional layer with max pooling applied, followed by three hidden dense layers with ReLu activation function. The dense layers have 340, 200, and 200 nodes respectively. The output block gets flattened and a sigmoid activation function is applied. I use stochastic gradient descent as the optimiser and binary cross-entropy as the loss function.

2. Imputation

Missing values are a common issue of data validity in finance prediction tasks. This study empirically compares multiple methods of imputation and selects the best method as revealed by the trained model's performance on the validation set. This study compares the performance of imputing zeros, mean, median, and SVD, KNN, and MICE imputation. This paper finally made use of a KNN - Nearest Neighbour - imputation method, which weights all the samples using mean squared difference on the variable for which a user-specified number of date-preceding rows have observable data. The imputed value is simply selected from the nearest observation with missing values based on the distance between that subject and the target.

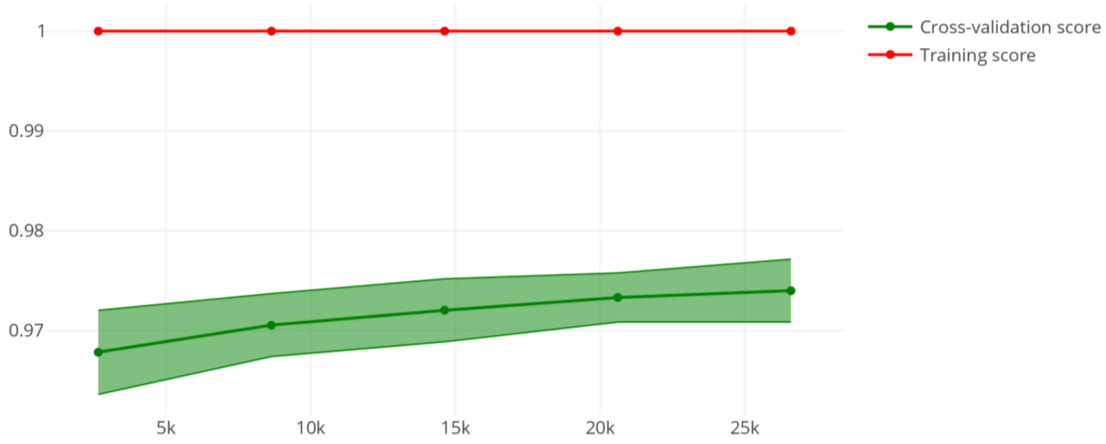
3. Parameter Tuning

Although the parameter tuning procedures can, to a great extent, be automated, it is still worth understanding the underlying implementation. First, it is important to consider that overfitting models to training data reduces their out-of-sample performance. For that reason, regularisation is an important technique to simplify models so that there is a balance between model fit and predictive performance (Friedman, Hastie, & Tibshirani, 2001). For the majority of models, this simplification (regularisation) is achieved by controlling the number of variables with methods such as a stepwise procedure (Whittingham, Stephens, Bradbury, & Freckleton, 2006) or by creating multiple models and comparing them with information measures such as the Akaike's Information Criterion (Anderson & Burnham, 2002).

Alternatively, shrinkage such as using lasso and ridge methods can be used to add terms and down weight contributions. The concept of shrinkage is similar to what is used in GBMs but is incrementally applied to sequential trees. GBM regularisation jointly involves the optimisation of the number of trees, the learning rate, and tree complexity. *Figure A9* tracks adjustments to these parameters. It shows how an adjustment to the model can be more regularised from left to right. The figure also shows how the parameters can be adjusted to increase the recall of bankruptcy prediction at the expense of precision. There is thus a dimensional trade-off between these parameters. The XGBoost implementation of the GBM has many more parameter inputs than that mentioned above. However, the number of trees, the learning rate, and tree complexity are essential parameters in adjusting the model complexity and reducing overfitting. The approach is then to optimise these parameters by testing many parameter combinations to achieve the minimum prediction error on the validation sets.

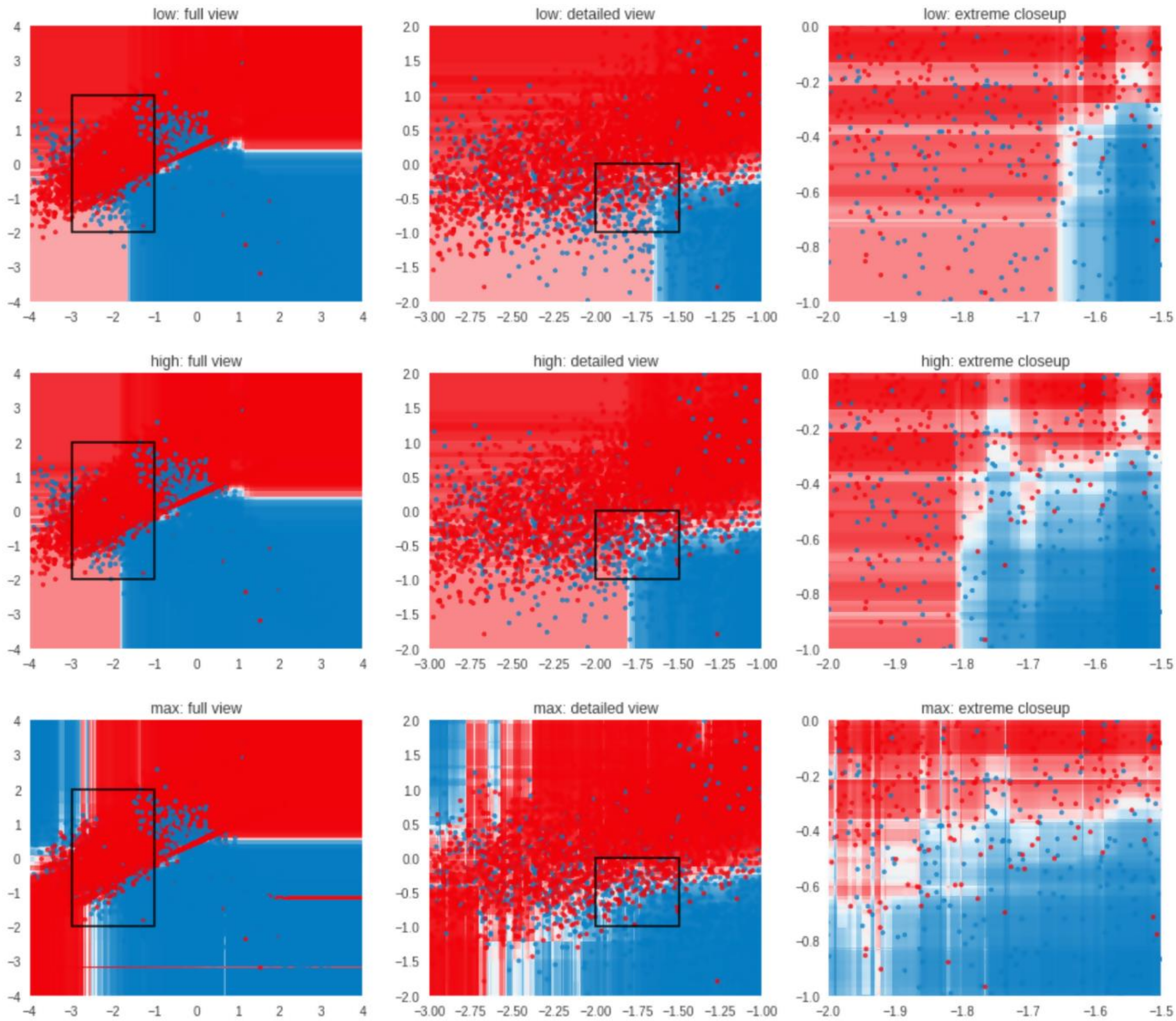
The learning curves in *Figure* A8 have an important function in showing researchers whether more data will lead to better cross-validated accuracy. This figure shows that more data will improve the results of this study. This form of analysis is also interesting as it allows one to gauge whether the models overfit the training set. Although a tree ensemble is more likely to overfit than other models, a training score that immediately moves to 100% accuracy could be indicative of overfitting, and by further adjusting the parameters the researcher can increase the parameter performance. A more promising curve would be the tangential line starting just above 0.99%.

Figure A8: Learning Curve



This figure illustrates the importance of additional observations in improving the out-of-sample cross-validation (multiple splits to obtain a robust average) accuracy score. The general rule in machine learning is that the more data you have, the better the performance of the classification model. This figure also illustrates the importance of testing on a fresh set of data; simply testing on the training set (the dataset used to infer a function using a machine learning algorithm) biases the results. The above measure shows both the training and cross-validated (out-of-sample) score. The above training score is indicative of overfitting. The model can further be improved with regularisation procedures (discussed in the appendix). After the regularisation, the training score is expected to look more like the curved trajectory presented by the adjacent red line.

Figure A9 Parameter Adjustment Decision Boundaries Between 2 PCA Components



From top to bottom, the parameters of the model get adjusted to increase the recalled bankruptcies. The bankruptcy decision boundary starts with the blued-out area and expands as the model adjusts to include more. Left to right shows the boundary at different resolutions. The boundary in this picture involves the total assets and P/E ratio.

4. Validation

A good illustration of the validation technique on time series data can be seen in *Figure A17*. Once all the input data has been gathered, the sample data has to be split into distinct sets to be able to estimate the generalizable prediction success of both classification models. Following research by Tan, Lee, and Pang (2014), all test splits in this study are 'pure hold-out' sets that are not used by the model at any stage apart from testing the final performance. This dataset remains in a ‘lockbox’ until the testing occurs. This concept is quite important as it allows researchers to get feedback from first testing on the validation data without fearing that they are mistakenly ‘datamining’ the test set.

This study makes use of cross-validated metrics to further improve the robustness of the results (Kohavi, 1995). The cross-validation method simply means that multiple test-train sets are used in evaluating model performance. In this study, I use a unique blocked form of cross-validation that is well-suited for longitudinal evaluation (Bergmeir & Bentez, 2012). Using this approach ensures that the testing data never contains data that is older than the training data. This a sensible step for preserving the integrity of the prediction. As more data becomes available, the training set increases, allowing for an improved prediction. The size of the test set stays constant as the final metric is a simple average over the different splits. Although the test set stays constant in size, it shifts forward to test distinct non-overlapping periods. Each of the training splits can then be fed into the machine learning model to predict a range of target values. This value is compared against the test set’s target values to calculate the prediction success metrics. As mentioned, to calculate the final result, I compute an average value across all the splits and calculate the confidence interval.

5. Classifier Design

Machine Learning is defined as the study of inductive algorithms that ‘learn’(Provost & Kohavi, 1998). For this study, it is valuable to have an intuitive grasp of the XGBoost machine learning model. XGBoost is short for Extreme Gradient Boosting, a nonlinear inductive algorithm used to approximate the function between inputs and outputs. The idea behind Gradient Boosting is to “boost” many weak learners or predictive models to create a stronger overall model. A meta-model gets constructed from a large ensemble of weak models. A weak model simply has to predict slightly better than a random guess. To combine the weak learners, one first trains a weak model, *m*, using data samples drawn from some weight distribution. Then one increases the weight of samples that are misclassified by the model *m* and decreases the weight of those classified correctly, after which one trains the next weak learning using samples drawn according to the updated weight distribution. In this way, the algorithm always uses data samples that were hard to learn in previous rounds to train models. This results in an ensemble that is good at learning a large range of seemingly inscrutable patterns in the training data. In this study, decision trees are used as the weak learner. After the weighting process, the sum of all the weak learners is taken to produce the overall prediction.

To create the overall ensemble model, such as presented by the *Classifier* pseudocode in the Classifier Design section above, we have to define a loss function, *L,* to minimise. This function has to be differentiable as we want to perform a process of steepest descent, which is an iterative process of attempting to reach the global minimum of a loss function by going down the slope until there is no more room to move closer to the minimum. We, therefore, minimise a loss function numerically via the process of steepest descent. For a classification task, we use logistic regression to obtain the probabilistic outputs of the target variable. The focus here is on $f(\boldsymbol{x}_i)$ as this is the compressed form of the predictor of each tree *i*.

$$L(\theta) = \sum_i [\boldsymbol{y}_i \, ln(1 + e^{f(x_i)}) + (1 - \boldsymbol{y}_i) ln(1 + e^{f(x_i)})] \quad (2)$$

$$L(\theta) = \sum_i [\boldsymbol{y}_i \, ln(1 + e^{-\hat{y}\, i}) + (1 - \boldsymbol{y}_i) ln(1 + e^{\hat{y}\, i})] \quad (3)$$

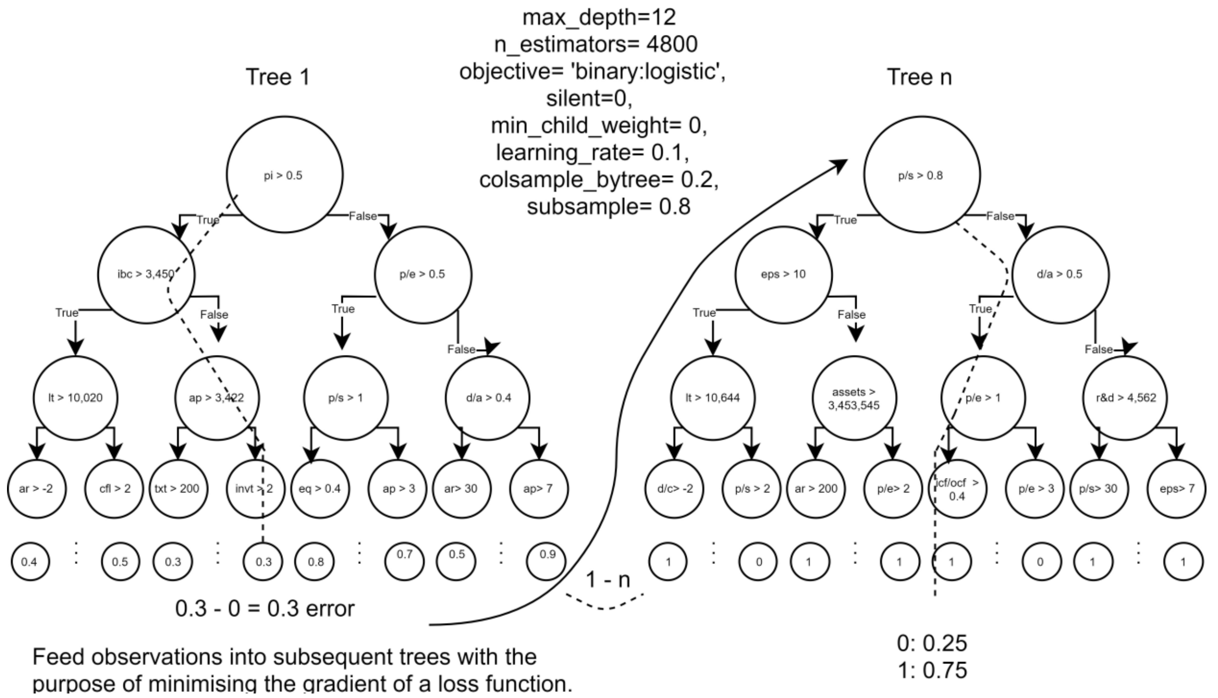
Further, it is necessary to minimise the loss over all the points in the sample, $(\boldsymbol{x}_i, y_i)$:

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} L(\theta) \quad (4)$$

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} L(y_i, f(\boldsymbol{x}_i)) \quad (5)$$

At this point, we are in the position to minimise the predictor function, $f(\boldsymbol{x}_i)$, w.r.t. $\boldsymbol{x}$ since we want a predictor that minimises the total loss of $f(\boldsymbol{x})$. Here we simply apply the iterative process of steepest descent. The minimisation is done in a few phases. The first process starts with adding the first and then successive trees. Adding a tree emulates adding a gradient based correction. Making use of trees ensures that the generation of the gradient expression is successful, as we need the gradient for an unseen test point at each iteration as part of the calculation $f(\boldsymbol{x})$. Finally, this process will return $f(\boldsymbol{x})$ with weighted parameters. The detailed design of the predictor, $f(\boldsymbol{x})$, is outside the purpose of the this section, but for more extensive computational workings, please see the next section.

Figure A10: XGBoost Decision Tree Ensemble



This illustration provides an example of how a group of decision trees is used to predict a target value. In this example, we follow an observation as it makes its way down two decision trees. In this case, a logistic function transforms the output into a probability for two classes. The XGBoost model used in this study is a little more complex than the above illustration, but the above intuition remains at the core of this model.

As soon as the model is fully trained, testing data can be dropped down the model to identify the predicted response variables. In *Figure A10*: XGBoost Decision Tree Ensemble, I created an illustrative example of how an observation runs through the model and how a prediction is made in the classification task. The response variable is classified as either 0 or 1; a healthy firm-year is designated by a 0 and a bankrupt firm-year with a 1. As a result of the logarithmic loss function, the output is a probability associated with each class for every weak learner. The average of the weak learners establishes the final probability. The regression task follows a similar process; the only difference is that there is only one output per observation and the outputs prediction scores get added together to produce the final prediction.

E. Comparing Traditional, Machine Learning and Deep Learning Models

Hastie et al. (2009) note that one of the most important benefits of Gradient Boosting (GBM) machines is that they require very little research intervention. These models are largely unaffected by missing values, outliers, and monotonic transformation of variables. Apart from being able to easily deal with 'dirty' or 'noisy' data, these models are also much more accurate than the traditional alternatives whose performance even after data cleaning and the pre-processing procedure is substandard at best. Further, these models are not impaired by any heteroscedasticity or multicollinearity issues, which is of serious consequence to parametric models (Probit, LDA & related).

The high dimensionality of GBM models allows them to handle many inputs and to remain largely immune to irrelevant inputs. LDAs and MDA models are low dimensional models that make the mistake of unrealistically assuming linear separability and normality of variables (Chandra, Ravi, & Bose, 2009; Neves & Vieira, 2006). In later years the logit model became the more favoured model (Ohlson, 1980; Pervan, Pervan, & Vukoja, 2011). However, the logit model still has many of the same constraints and disadvantages of the MDA model. Logit and MDA models can only handle a small number of variables as a result of overfitting (Altman, 1968; Ohlson, 1980). For these models, irrelevant variables that enter the global maximum likelihood solution can severely impact the quality of the reduction and model stability. Gradient boosting machines simply classify these inputs as redundant; if the model considers inputs to be irrelevant, then they simply get excluded for the final ensemble.

The GBM does not have as many constraints as other models and can allow for thousands of variables and their interaction effects. Some studies have compared machine learning models such as neural networks with logit and MDA models (Altman, Marco, & Varetto, 1994; Jones et al., 2017). It is difficult to study high dimensional relationships with the traditional models. However, in the past studies that attempted comparison studies, the high dimensional models always come out on top. The GBM model has been identified as one of the strongest models used in prediction research (Hastie et al. 2009). Decision tree ensemble models have consistently been shown to outperform conventional and more sophisticated techniques like support vector machines (SVM) and neural networks (NN). (Hastie et al., 2009; Schapire & Freund, 2012). A multitude of literature outside of finance and accounting has identified this outperformance.

Given that GBM significantly improves the prediction quality of test samples, the use of these models in a practical setting is also important to consider. The first evidence of the practicality of these models is the ease with which they can be implemented. Structurally GBMs models have minimal architectural requirements; they can easily be developed and executed by popular statistical packages like R and Python. Apart from having the ability to include numerous variables, they can also rank order them based on their predictive power (Friedman, 2001; Hastie et al., 2009). These models are also easily interpretable by using these relative variable importances (RVI) outputs. The use of these models is widespread across many fields such as satellite image recognition, text and speech recognition, biological sciences, credit risk, and cybersecurity.

Another model used in this study is a Convolutional Neural Network (CNN). The main disadvantage of CNN models is that they require large amounts of data; with only thousands of examples in this study, deep learning is unlikely to beat other advanced models. These models also do not have particularly strong theoretical foundations, which means that determining the hyperparameters or topology of deep learning is a black art with no guiding theory. Furthermore, a big drawback is that what is learned cannot be as easily interpreted, as is the case for decision tree models. In saying that, deep learning, via a process known as feature learning, removes the need for manual feature engineering. Lastly, as evidenced in this study, the architecture can easily be adopted for new problem sets.

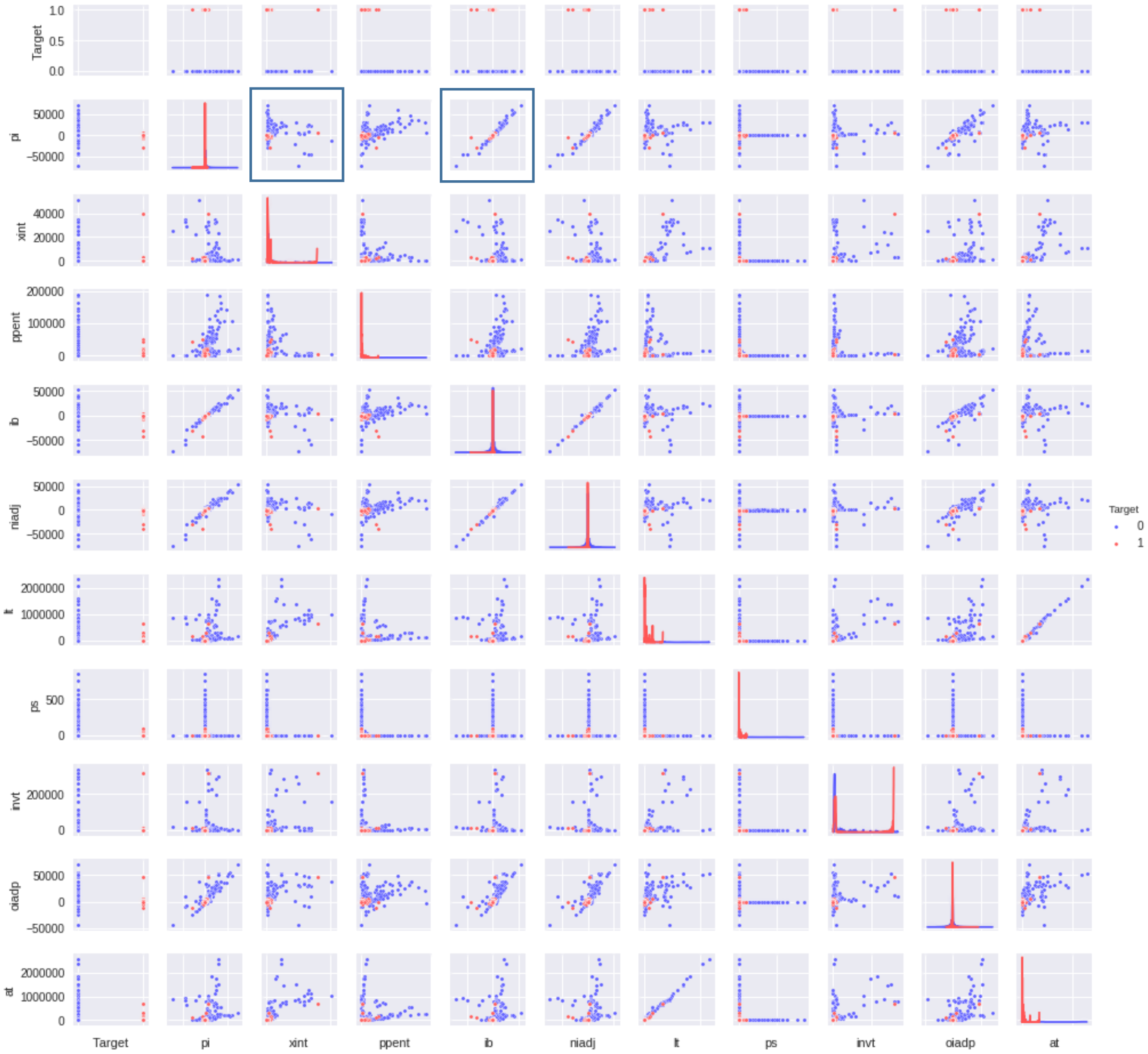
Unlike black-box models, like the CNN model, GBMs are transparent in their methods of inductive learning from data. Further GBMs can be fitted to a small amount of data, whereas the CNN network needs a larger amount of data. Boosting is an interesting feature to the model. It grows the number of trees by sequentially modeling the residuals to include atypical observations that depart from the dominant patterns of the initial trees. In doing this, the algorithm simultaneously reduces the bias and variance of the model. GBMs can also successfully handle different response variables, continuous, discrete, count etc.

Compared with conventional models, there are no p-values to indicate the relative significance of model coefficients; it is also difficult to determine the degrees of freedom in the model. It is questionable as to whether these aspects are a problem or an advantage to the model, as most would be aware there are rigorous debates as to the use of p-value in models (Fidler, Geoff, Mark, & Neil, 2004). Although the GBMs lack simple metrics that can be a disadvantage from a traditional point of view, a large amount of methods of interpretation has developed over the years, with many ongoing developments. These techniques provide equivalent functions to many of the conventional techniques.

F. Extended Analyses

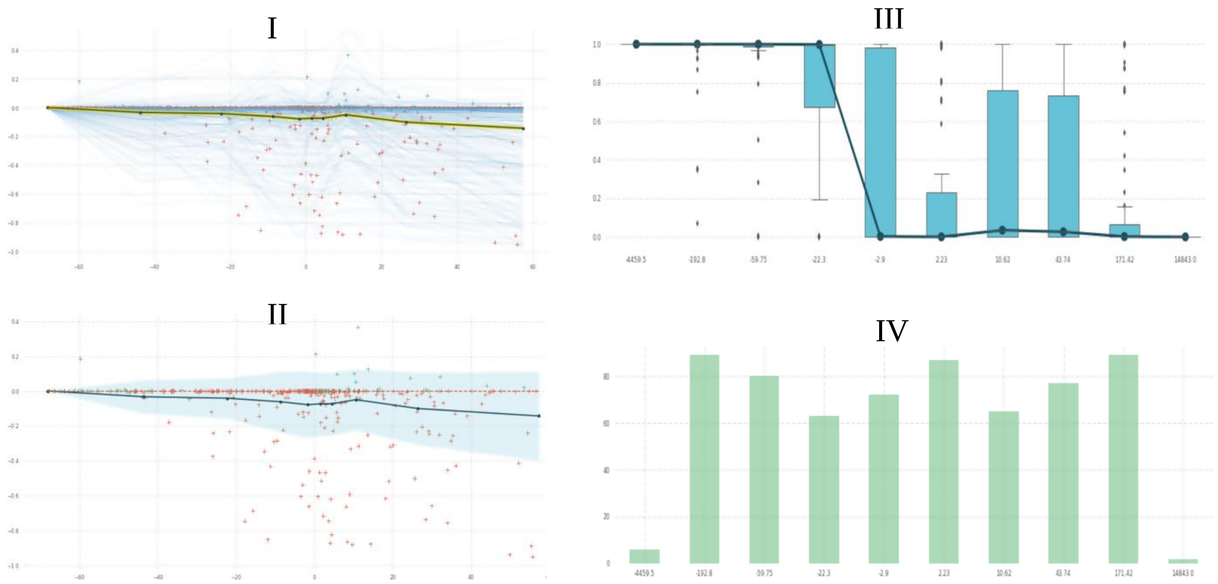
Figure A11: Pair Plots below presents the scatter plots of the most important variables. In this figure, bankruptcies are coded in red and healthy firms in blue. The relationship between pre-tax income (pi) and interest and related expenses (xint) poses as an interesting interaction that would be shown to be one of the most important pairs to predict bankruptcy (See column three, row three in *Table 4*). This pair's predictive power is evident by the dense clustering of bankruptcies around a fixed point, making it easy for a decision tree prediction model to discriminate between bankrupt and non-bankrupt firms. The XGBoost model would seek to sculpt a decision boundary around that point to predict future bankruptcies. This figure illustrates many of these important relationships. The reader should note that this simply shows relationships between pairs in the data, and it does not show the relationship of these variables in a fully trained model. This will be dealt with from *Table 13* onwards. The advantage of the non-linear models used in this study is that they do not just look at the low dimensional relationships of the scatterplots, but they also investigate relationships at extremely high dimensions. In this study, I descriptively report interactions to the depth of three, i.e., up to three variables' non-linear interactions and the predicted response.

Figure A11: Pair Plots



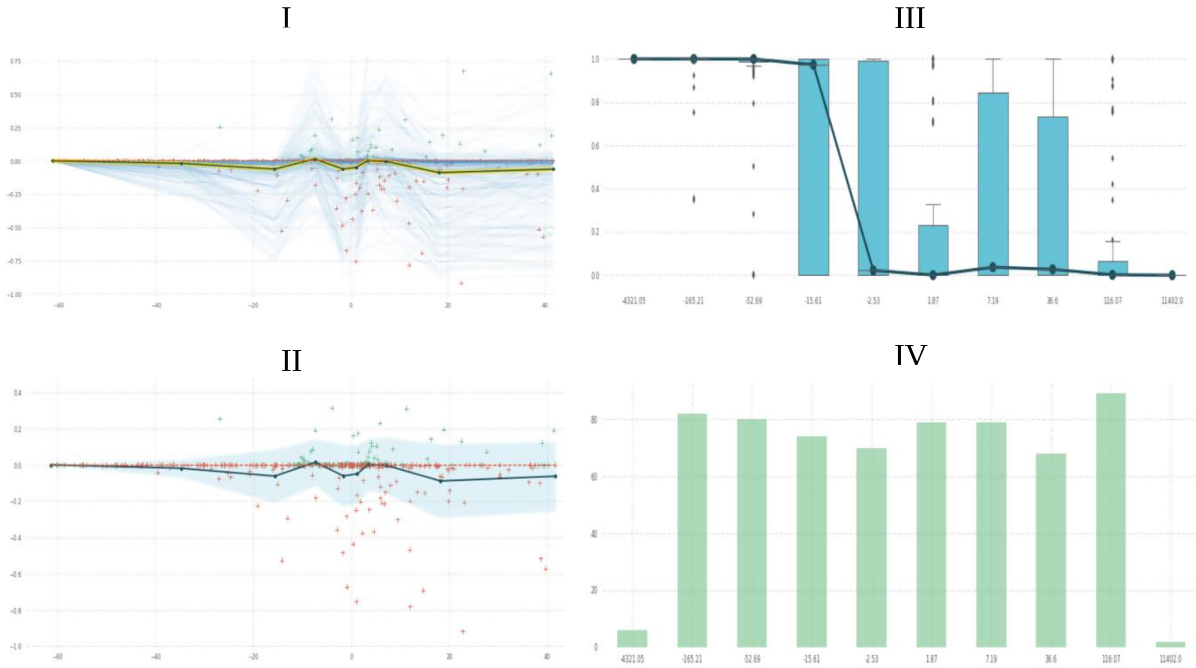
This figure reports the scatterplots of distinct variable relationships that show prominence in their predictive ability. The intersection of the variable with itself plots for the variable's distribution. The red dots are observations labelled as bankruptcies and the blue dots are observations labelled as healthy firms.

Figure A12: Pre-tax Income (PI) Variable Analysis



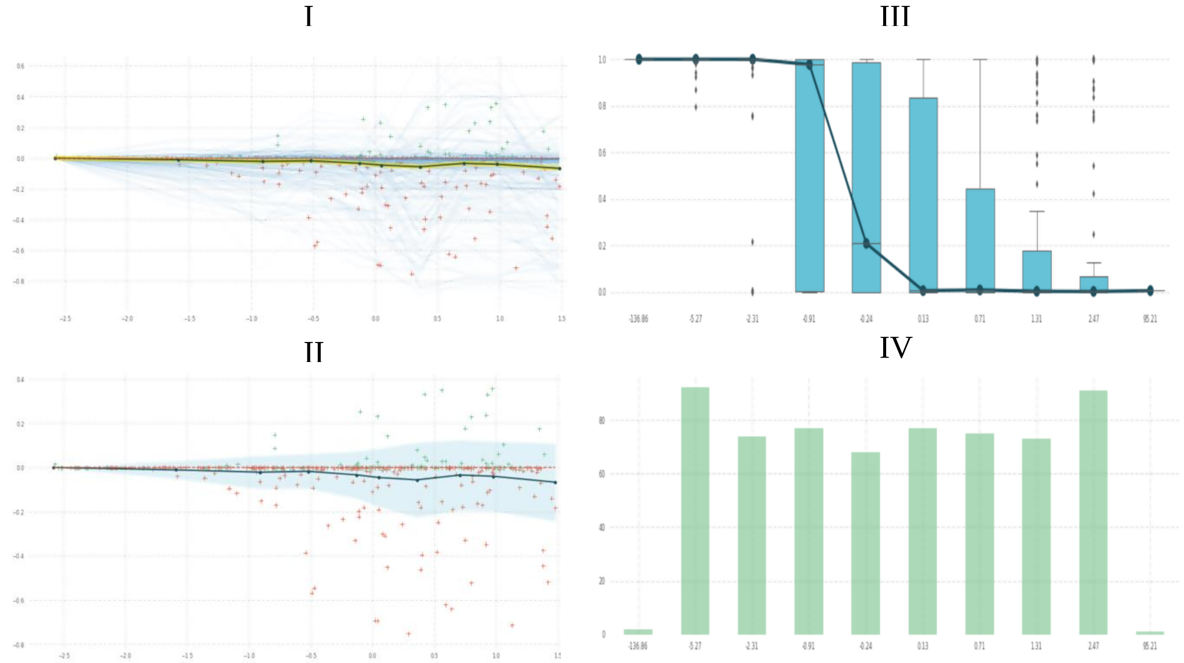
This figure reports the marginal relationship pre-tax income has with bankruptcy prediction. Plot I, II, and III separately show that when the pre-tax-income increases, the likelihood of bankruptcy decreases. Plot I is a simple partial dependence plot that draws lines of all the observational trajectories. Plot II establishes a confidence band and better highlights bankrupt predictions (green dots) and healthy firm predictions (red dots). Plot III is a box plot of an equally balanced bankruptcy and healthy prediction model. Plot IV is a count plot of Plot III. The observations in I and II are Winsorized to improve the look of the plots. Plot III shows that there is a point when the distribution between bankruptcy and healthy firms enlarge at around 10-40 million in income; this is corroborated by a spike in plot I of increased bankruptcy predictions. III shows that negative PI is a potential indicator of future bankruptcy.

Figure A13: Income Before Extraordinary Items (IBC) Variable Analysis



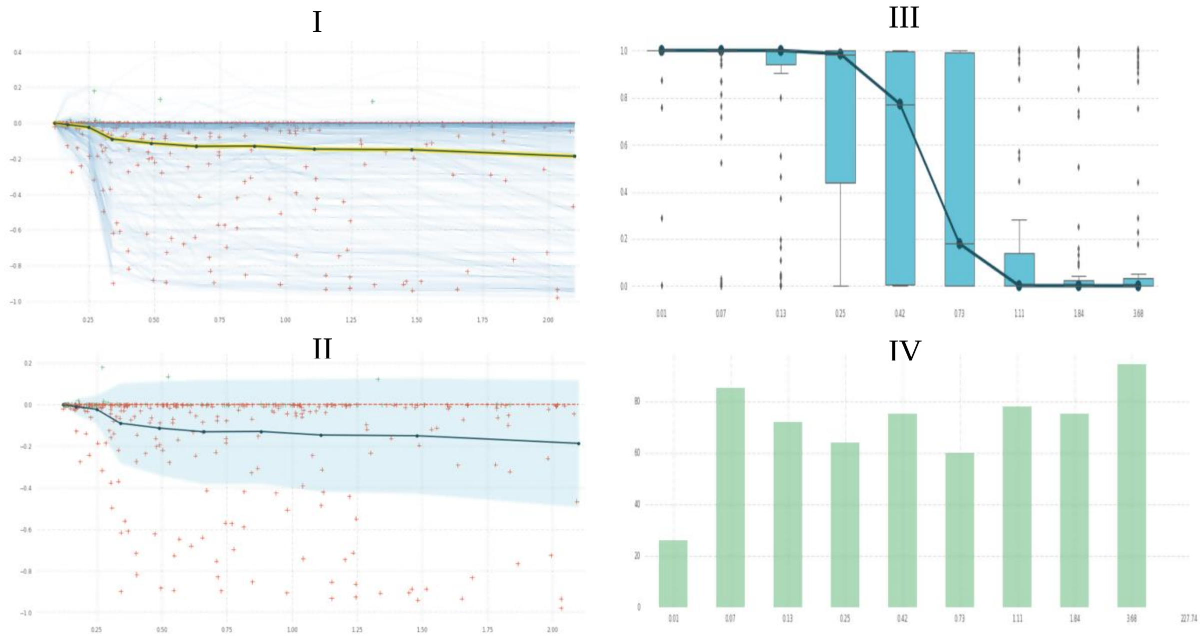
This figure reports the marginal relationship income before extraordinary items (IBC) has with bankruptcy prediction. Plot I, II, and III separately show that when IBC increases the likelihood of bankruptcy decreases. The plot I is a simple partial dependence plot that draws lines of all the observational trajectories. Plot II establishes a confidence band and better highlights bankrupt predictions (green dots) and healthy firm predictions (red dots). Plot III is a box plot of an equally balanced bankruptcy and healthy prediction model. Plot IV is a count plot of Plot III. The observations in I and II are Winsorized to improve the look of the plots. Plot III shows that there is a point when the distribution between bankruptcy and healthy firms enlarges at around 7-35 million in income; this is corroborated by a spike in plot I of increased bankruptcy predictions. III shows that situations of a negative IBC are an indicator of future bankruptcy. This measure seems to be somewhat more volatile than the PI measure. It further highlights in I and II that firms who have small negative IBCs around -10 to -7 are less likely to be bankrupt than those more negative than -10 and more positive than -7 up and till 0.

Figure A14: EPS (Basic) - Exclude Extra. Items (EPSPX) Variable Analysis



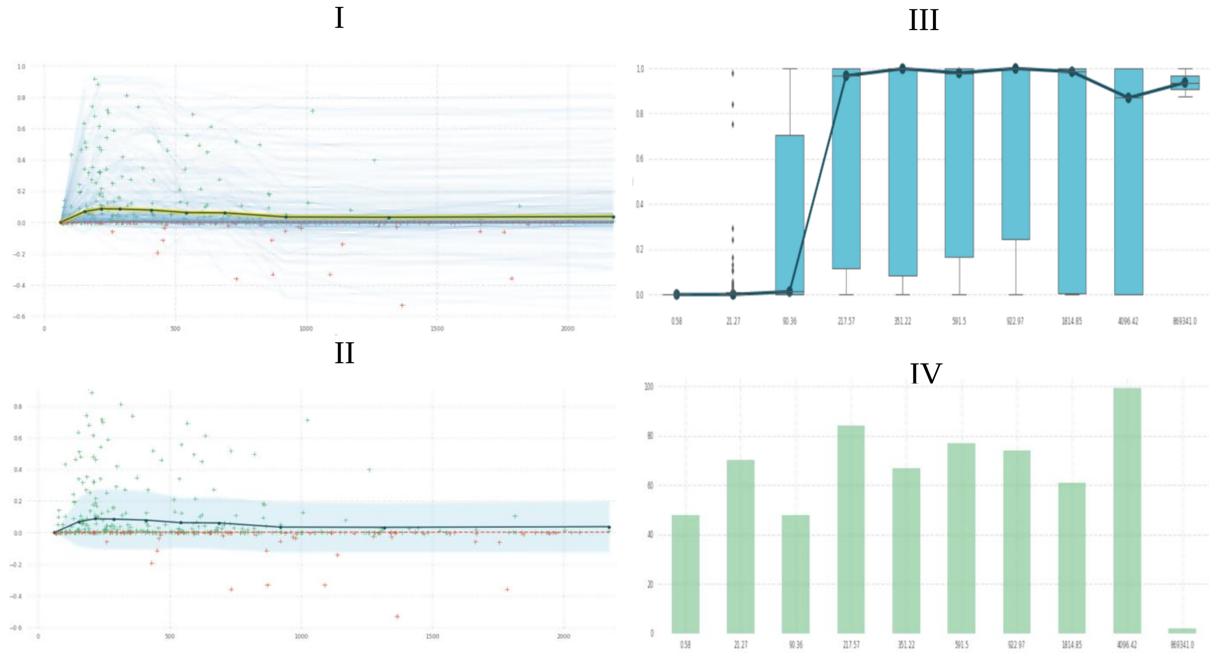
This figure reports the marginal relationship EPS has with bankruptcy prediction. Plot I, II, and III separately show that when EPS increases the likelihood of bankruptcy decreases. The plot I is a simple partial dependence plot that draws lines of all the observational trajectories. Plot II establishes a confidence band and better highlights bankruptcy predictions (green dots) and healthy firm predictions (red dots). Plot III is a box plot of an equally balanced bankruptcy and healthy prediction model. Plot IV is a count plot of Plot III. The observations in I and II are Winsorized to improve the look of the plots. Plot III shows that the distribution gradually centres around a healthy firm prediction as the EPS increases. A negative EPS value seems to be a good indicator of future failure. Although removed from I and II, III shows that situations of increased negative PI are a potential indicator of future bankruptcy.

Figure A15: Price/Sales (PS) Variable Analysis



This figure reports the marginal relationship Price/Sales has with bankruptcy prediction. Plot I, II, and III separately show that when the PS increases the likelihood of bankruptcy decreases. The plot I is a simple partial dependence plot that draws lines of all the observational trajectories. Plot II establishes a confidence band and better highlights bankruptcy predictions (green dots) and healthy firm predictions (red dots). Plot III is a box plot of an equally balanced bankruptcy and healthy prediction model. Plot IV is a count plot of Plot III. The observations in I and II are Winsorized to improve the look of the plots. III shows that situations of increased PI are highly indicative of a healthy firm. A high price to sales firm normally enjoys high-profit margins and are most notably at the top of their respective industries at that point of time, and for that reason, they are unlikely to become bankrupt.

Figure A16: Liabilities - Total (LT) Variable Analysis



This figure reports the marginal relationship liabilities has with bankruptcy prediction. Plot I, II, and III separately show that when LT increases the likelihood of bankruptcy increases. The plot I is a simple partial dependence plot that draws lines of all the observational trajectories. Plot II establishes a confidence band and better highlights bankruptcy predictions (green dots) and healthy firm predictions (red dots). Plot III is a box plot of an equally balanced bankruptcy and healthy prediction model. Plot IV is a count plot of plot III. The observations in I and II are Winsorized to improve the look of the plots. III shows that when a firm has no liabilities, it is extremely unlikely for them to become bankrupt.

Table A25: Neural Network Models Bankruptcy Literature

Reference	**Journal**	**Description**	**Model**	**AUC**
Kim and Kang (2010)	Expert Systems with Applications 37 (2010) 3373–3379	1458 manufacturing firms (2002–2005), half of which went bankrupt (1:1)	Boosted Neural Network	0.750
du Jardin (2017)	Expert Systems with Applications (75) Pages 25-43.	95,910 French Firms (1996 - 2009) 1,920 failing firms (1:0.020)	Feed Forward Neural Network	0.800
Mselmi et al. (2017)	International Review of Financial Analysis (2017) 50: 67-80	212 French firms, half of which is distressed. (1:1)	ANN (MLP)	0.871
Barboza et al. (2017)	Expert Systems with Applications 83 (2017), Pages 405-417	More than 10,000 firm-year observations. 1,796 failed firms (1:0.22)	ANN (MLP)	0.901
Zhou (2013)	Knowledge-Based Systems (2013) 41: 16-25	86,129 US firm year, 918 (1981-2009) bankruptcies (1:0.011)	ANN (MLP)	0.856
Huang et al. (2016)	Kybernetes (45) 2016	270 Taiwanese companies (2004-2014), 90 failed firms (1:0.5)	GRNN model with FOA optimisation	0.903
Jones, Johnstone, and Wilson (2017)	Journal of Business Finance and Accounting (2017) 44: 3–34	30,129 US firm years, 3960 firm-year bankruptcies. (1:0.15)	ANN (MLP)	0.853
This Study		33,242 US large firm years, 1224 firm-year bankruptcies 1977-2016 (1:0.038)	Deep Convolutional Neural Network	0.914

This table reports the results of past neural network research that reported an ROC (AUC) metric to be used for cross-study comparisons.

Table A26: Boosting and Decision Tree Model Literature

Reference	Journal	Description	Model	AUC
Chandra et al. (2009)	Expert Systems with Applications 36 (2009) 4830–4837, C	240 dot-com companies, half of which went bankrupt (1:1)	Boosting	0.900
Olson et al. (2012)	Decision Support Systems 52 (2012) 464–473, A*	1321 US firm years sampled over the period 2005-2011, 100 firms went bankrupt (1:0.082)	Decision Trees	0.947
Kim and Upneja (2014)	Economic Modelling 36 (2014) 354–362., A	142 Restaurant Firms 1988-2010 (1:1)	AdaBoost	0.988
Karas and Reznakova (2017)	Engineering Economics (2017) 28(2): 145-154, B	1540 Construction Firms, 283 went bankrupt (1:0.23)	CART	0.859
Barboza et al. (2017)	Expert Systems with Applications 83 (2017), Pages 405-417, C	More than 10,000 firm-year observations. 1,796 failed firms (1:0.22)	Boosting	0.901
Zieba et al. (2016)	Expert Systems with Applications 58 (2016) 93-101, C	10,174 emerging market firm years, 400 failed 0.041 (2000-2012)	Ensemble XGBoost	0.944
Jones (2017)	Review Accounting Studies (2017) 22:1366–1422, A*	36,209 US firm years, 4460 firm year bankruptcies 1987 to 2013 (1:0.14)	Proprietary Gradient Boosting Machine - TreeNet	0.997
Volkov et al. (2017)	Decision Support Systems 98 (2017) 59–68, A*	19,380 Belgium and Luxembourg Firms, 1,933 bankrupt firms, 2007-2015 (0.11)	Random Forest	0.859
Jones, Johnstone, and Wilson (2017)	Journal of Business Finance and Accounting (2017) 44: 3–34, A	30,129 US firm years, 3960 firm-year bankruptcies. (1:0.15)	Boosting	0.931
This Study		33,242 US large firm years, 1224 firm year bankruptcies 1977-2016 (1:0.038)	Freeware Gradient Boosting Machine - XGBoost	0.957

This table reports the results of past boosting and decision tree ensemble research that reported an ROC (AUC) metric to be used for cross-study comparisons.

Table A27: Literature on Variable and Category Importance for Decision Tree Ensembles

Reference	**Journal**	**Description**	**Selector**	**Category Importance**
Kim and Upneja (2014)	Economic Modelling 36 (2014) 354–362., A	142 Restaurant Firms 1988-2010 (1:1)	Splitter's Level	(1) Solvency, (2) Liquidity, (3) Profitability
Beher and Weinblat (2016)	International Journal of the Economics of Business (2017), 24:2, 181-222, B	Default patterns in European Firms 1,964,374 firm observations from 2010-2011.	Random Forest Variable Importance	(1) Solvency, (2) Profitability, (3) Liquidity
Jones (2017)	Accounting Studies Review (2017) 22:1366–1422, A*	36,209 US firm years, 4460 firm year bankruptcies 1987 to 2013 (1:0.14)	Gain Measure	(1) Governance, (2) Valuation
Volkov et al. (2017)	Decision Support Systems 98 (2017) 59–68, C	19,380 Belgium and Luxembourg Firms, 1,933 bankrupt firms, 2007-2015 (0.11)	Random Forest Variable Importance	(1) Solvency, (2) Profitability, (3) Liquidity
Jones, Johnstone, and Wilson (2017)	Journal of Business Finance and Accounting (2017) 44: 3–34, A	30,129 US firm years, 3960 firm-year bankruptcies. (1:0.15)	Random Forest Variable Importance	(1) Solvency, (2) Profitability, (3) Efficiency, (4) Liquidity
Mselmi et al. (2017)	International Review of Financial Analysis (2017) 50: 67-80, A	212 French firms, half of which is distressed. (1:1)	Stepwise Regression	(1) Solvency, (2) Efficiency, (3) Liquidity, (2) Profitability
This Study		33,242 US large firm years, 1224 firm-year bankruptcies 1977-2016 (1:0.038)	Gain Measure	(1) Solvency, (2) Profitability and Valuation, (3) Efficiency (4) Liquidity

This table reports all past studies that ranked the importance of their variables. I matched the respective variables in each study to a category in the attempt to identify which categories these studies deemed to be more important.

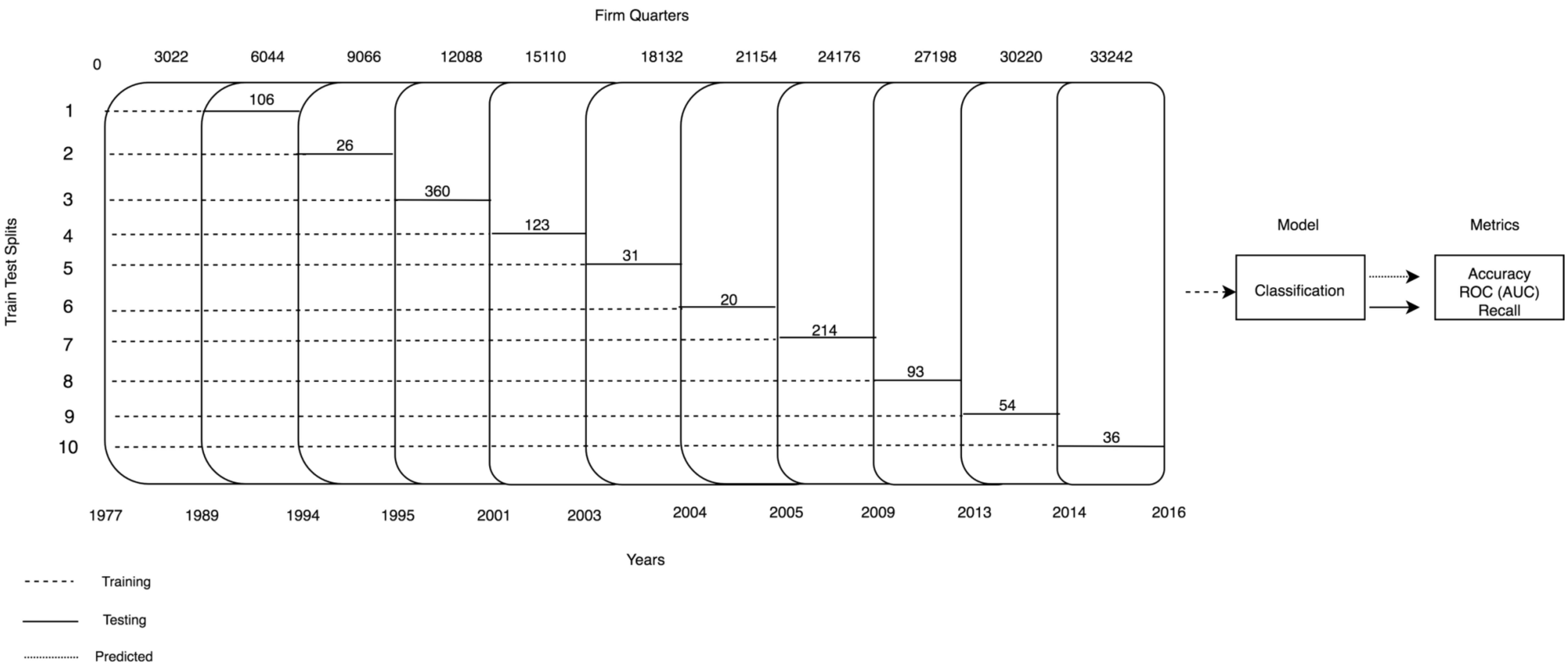
Table A28: Bankruptcy and Healthy Firm Summary Statistics for Important Variables

Table A29: Filing Outcome Summary Statistics

Target	Response	che	npm	lct	at_turn	rect_turn	pay_turn	curr_debt	dpc	accrual	mib
Duration	< 1 Year	163.70	-0.40	510.45	0.89	15.41	12.00	0.42	112.54	0.12	6.36
	> 1 Year	873.82	-1.63	840.72	1.00	17.54	9.31	0.42	120.00	0.06	36.82
Survival	No	1084.19	-0.33	602.56	0.90	15.48	8.24	0.48	71.75	0.07	22.87
	Yes	202.86	-1.53	741.43	0.98	17.23	12.03	0.38	144.90	0.10	22.59
Chapter	11	554.89	-1.08	704.21	0.97	16.95	10.82	0.42	119.17	0.09	23.30
	7	155.81	-0.29	61.10	0.19	1.65	0.75	0.60	17.48	0.05	0.00
Asset Sale	No	262.76	-1.26	718.73	0.93	16.34	11.49	0.40	108.24	0.09	18.88
	Yes	1433.50	-0.43	589.36	1.02	17.22	7.63	0.48	142.71	0.07	34.73
Tort	No	567.78	-1.13	643.65	0.96	17.19	10.59	0.42	113.83	0.09	15.13
	Yes	208.62	-0.11	1323.59	0.84	7.36	10.09	0.41	155.80	0.04	132.19

The following table presents the two most important variables for each prediction task. Instead of reporting the summary statistics for just those two, the summary statistics of the two important variables for all filing outcomes are reported. The full names of the variables in order are: Cash and Short-Term Investments (che), Net Profit Margin (npm), Inventories - Total (invt), Current Liabilities - Total (lct), Asset Turnover (at_turn), Receivables Turnover (rect_turn), Payables Turnover (pay_turn) Current Liabilities/Total Liabilities (curr_debt), Depreciation and Amortization (dpc), Accruals/Average Assets (accrual), Minority Interest - Balance Sheet (mib).

Figure A17: Illustration of Validation in Time Series



Validation is used as an improved means to forecast the accuracy of an inducer by splitting the data into *n* mutually exclusive subsets. To ensure consistent performance measurements on these splits, they should be approximately the same size. In this study, the data splits into four equal-sized sections. And the model is trained and tested on each of these splits. Each time, the model trains on an increasing number of samples ordered by date. This study reports both the overall validated accuracy and breaks the accuracy down to each period and surprise threshold in question. This table does not show a separate process used to do variable selection (the process of removing variables which seem irrelevant for modeling) before the validation. The variable selection is done on a small validation set constituting 15% of the data to ensure that during the development stage there is no "double dipping" into the data. Therefore the model always gets tested on a fresh out-of-sample dataset. Another approach would be to create multiple validation sets and hyperparameter selections for each period.

Table A30: Robustness Table of Validation in Time Series. – Metrics

Train Period	Test Period	Cross-entropy	Brier Score Loss	Accuracy Score	ROC AUC Sore	Average Precision Score	Precision - Bankrupt Firms	Precision - Healthy Firms	False Positive Rate	False Negative Rate	False Discovery Rate
1977 - 1989	1989 - 1994	0.098	0.026	0.967	0.944	0.465	0.563	0.973	0.007	0.745	0.438
1977 - 1994	1994 - 1995	0.023	0.006	0.993	0.983	0.529	0.800	0.994	0.001	0.692	0.200
1977 - 1995	1995 - 2001	0.425	0.089	0.895	0.920	0.672	0.831	0.897	0.004	0.850	0.169
1977 - 2001	2001 - 2003	0.095	0.024	0.971	0.958	0.621	0.688	0.980	0.010	0.480	0.312
1977 - 2003	2003 - 2004	0.042	0.009	0.989	0.917	0.264	0.375	0.991	0.002	0.903	0.625
1977 - 2004	2004 - 2005	0.025	0.005	0.995	0.960	0.465	0.778	0.996	0.001	0.650	0.222
1977 - 2005	2005 - 2009	0.255	0.054	0.937	0.929	0.598	0.780	0.939	0.003	0.850	0.220
1977 - 2009	2009 - 2013	0.085	0.021	0.974	0.953	0.525	0.651	0.978	0.005	0.699	0.349
1977 - 2013	2013 - 2014	0.049	0.011	0.986	0.962	0.566	0.842	0.987	0.001	0.704	0.158
1977 - 2014	2014 - 2016	0.031	0.009	0.988	0.984	0.561	0.500	0.994	0.006	0.472	0.500
1977 - 2014	1989 - 2015	0.105	0.024	0.971	0.957	0.562	0.693	0.975	0.004	0.681	0.307

This table reports an extensive list of model performance metrics over various sample splits. From this table, it is clear that there are large differences between the different periods. The best reported AUC of 0.984 occurred over the last few years of the sample 2014-2016. The worst AUC occurred over the 2003-2004 period. It is worth noting that Congress made amendments to the bankruptcy code in 1994; this could affect the bankruptcy prediction quality from 1995 onwards, at least until new observations are learned (Tabb, 1995). This study shows that the underlying distribution of bankrupt and healthy firms has an impact on the model performance score.

Table A31: Robustness of Validation in Time Series. – Observations

Train Period	Test Period	All Instances	Bankruptcy Sample	Bankrupt Recalled	True Positives	False Positives	Healthy Sample	Healthy Recalled	True Negatives	False Negatives	Bankrupt to Healthy
1977 - 1989	1989 - 1994	3022	106	48	27	21	2916	2974	2895	79	0.04
1977 - 1994	1994 - 1995	3022	26	10	8	2	2996	3012	2994	18	0.01
1977 - 1995	1995 - 2001	3022	360	65	54	11	2662	2957	2651	306	0.14
1977 - 2001	2001 - 2003	3022	123	93	64	29	2899	2929	2870	59	0.04
1977 - 2003	2003 - 2004	3022	31	8	3	5	2991	3014	2986	28	0.01
1977 - 2004	2004 - 2005	3022	20	9	7	2	3002	3013	3000	13	0.01
1977 - 2005	2005 - 2009	3022	214	41	32	9	2808	2981	2799	182	0.08
1977 - 2009	2009 - 2013	3022	93	43	28	15	2929	2979	2914	65	0.03
1977 - 2013	2013 - 2014	3022	54	19	16	3	2968	3003	2965	38	0.02
1977 - 2014	2014 - 2016	3022	36	38	19	19	2986	2984	2967	17	0.01
1977 - 2014	1989 - 2016	30220	1063	374	258	116	29157	29846	29041	805	0.04

This table reports an extensive list of model prediction metrics over various sample splits. From this table, it is clear that there are large differences in the distribution of bankruptcies over the different periods. The least amount of bankruptcies for a validation-period occurred over the last few years of the sample 2014-2016. The most bankruptcies occurred over the period 1995-2001.

Table A32: Table of Financial Ratios and Categorisation

Financial Ratio	Variable Name	Category	Formula
Capitalization Ratio	capital_ratio	Capitalization	Total Long-term Debt as a fraction of the sum of Total Long-term Debt, Common/Ordinary Equity and Preferred Stock
Common Equity/Invested Capital	equity_invcap	Capitalization	Common Equity as a fraction of Invested Capital
Long-term Debt/Invested Capital	debt_invcap	Capitalization	Long-term Debt as a fraction of Invested Capital
Total Debt/Invested Capital	totdebt_invcap	Capitalization	Total Debt (Long-term and Current) as a fraction of Invested Capital
Asset Turnover	at_turn	Efficiency	Sales as a fraction of the average Total Assets based on the most recent two periods
Inventory Turnover	inv_turn	Efficiency	COGS as a fraction of the average Inventories based on the most recent two periods
Payables Turnover	pay_turn	Efficiency	COGS and change in Inventories as a fraction of the average of Accounts Payable based on the most recent two periods
Receivables Turnover	rect_turn	Efficiency	Sales as a fraction of the average of Accounts Receivables based on the most recent two periods
Sales/Stockholders Equity	sale_equity	Efficiency	Sales per dollar of total Stockholders' Equity
Sales/Invested Capital	sale_invcap	Efficiency	Sales per dollar of Invested Capital
Sales/Working Capital	sale_nwc	Efficiency	Sales per dollar of Working Capital, defined as the difference between Current Assets and Current Liabilities
Inventory/Current Assets	invt_act	Financial Soundness	Inventories as a fraction of Current Assets
Receivables/Current Assets	rect_act	Financial Soundness	Accounts Receivables as a fraction of Current Assets
Free Cash Flow/Operating Cash Flow	fcf_ocf	Financial Soundness	Free Cash Flow as a fraction of Operating Cash Flow, where Free Cash Flow is defined as the difference between Operating Cash Flow and Capital Expenditures
Operating CF/Current Liabilities	ocf_lct	Financial Soundness	Operating Cash Flow as a fraction of Current Liabilities
Cash Flow/Total Debt	cash_debt	Financial Soundness	Operating Cash Flow as a fraction of Total Debt
Cash Balance/Total Liabilities	cash_lt	Financial Soundness	Cash Balance as a fraction of Total Liabilities
Cash Flow Margin	cfm	Financial Soundness	Income before Extraordinary Items and Depreciation as a fraction of Sales
Short-Term Debt/Total Debt	short_debt	Financial Soundness	Short-term Debt as a fraction of Total Debt

Financial Ratio	Variable Name	Category	Formula
Profit Before Depreciation/Current Liabilities	profit_lct	Financial Soundness	Operating Income before D&A as a fraction of Current Liabilities
Current Liabilities/Total Liabilities	curr_debt	Financial Soundness	Current Liabilities as a fraction of Total Liabilities
Total Debt/EBITDA	debt_ebitda	Financial Soundness	Gross Debt as a fraction of EBITDA
Long-term Debt/Book Equity	dltt_be	Financial Soundness	Long-term Debt to Book Equity
Interest/Average Long-term Debt	int_debt	Financial Soundness	Interest as a fraction of average Long-term debt based on most recent two periods
Interest/Average Total Debt	int_totdebt	Financial Soundness	Interest as a fraction of average Total Debt based on most recent two periods
Long-term Debt/Total Liabilities	lt_debt	Financial Soundness	Long-term Debt as a fraction of Total Liabilities
Total Liabilities/Total Tangible Assets	lt_ppent	Financial Soundness	Total Liabilities to Total Tangible Assets
Cash Conversion Cycle (Days)	cash_conversion	Liquidity	Inventories per daily COGS plus Account Receivables per daily Sales minus Account Payables per daily COGS
Cash Ratio	cash_ratio	Liquidity	Cash and Short-term Investments as a fraction of Current Liabilities
Current Ratio	curr_ratio	Liquidity	Current Assets as a fraction of Current Liabilities
Quick Ratio (Acid Test)	quick_ratio	Liquidity	Quick Ratio: Current Assets net of Inventories as a fraction of Current Liabilities
Accruals/Average Assets	Accrual	Other	Accruals as a fraction of average Total Assets based on most recent two periods
Research and Development/Sales	RD_SALE	Other	R&D expenses as a fraction of Sales
Avertising Expenses/Sales	adv_sale	Other	Advertising Expenses as a fraction of Sales
Labor Expenses/Sales	staff_sale	Other	Labor Expenses as a fraction of Sales
Effective Tax Rate	efftax	Profitability	Income Tax as a fraction of Pretax Income
Gross Profit/Total Assets	GProf	Profitability	Gross Profitability as a fraction of Total Assets
After-tax Return on Average Common Equity	aftret_eq	Profitability	Net Income as a fraction of average of Common Equity based on most recent two periods
After-tax Return on Total Stockholders' Equity	aftret_equity	Profitability	Net Income as a fraction of average of Total Shareholders' Equity based on most recent two periods

Financial Ratio	Variable Name	Category	Formula
After-tax Return on Invested Capital	aftret_invcapx	Profitability	Net Income plus Interest Expenses as a fraction of Invested Capital
Gross Profit Margin	gpm	Profitability	Gross Profit as a fraction of Sales
Net Profit Margin	npm	Profitability	Net Income as a fraction of Sales
Operating Profit Margin After Depreciation	opmad	Profitability	Operating Income After Depreciation as a fraction of Sales
Operating Profit Margin Before Depreciation	opmbd	Profitability	Operating Income Before Depreciation as a fraction of Sales
Pre-tax Return on Total Earning Assets	pretret_earnat	Profitability	Operating Income After Depreciation as a fraction of average Total Earnings Assets (TEA) based on most recent two periods, where TEA is defined as the sum of Property Plant and Equipment, and Current Assets
Pre-tax return on Net Operating Assets	pretret_noa	Profitability	Operating Income After Depreciation as a fraction of average Net Operating Assets (NOA) based on most recent two periods, where NOA is defined as the sum of Property Plant and Equipment, and Current Assets minus Current Liabilities
Pre-tax Profit Margin	ptpm	Profitability	Pretax Income as a fraction of Sales
Return on Assets	roa	Profitability	Operating Income Before Depreciation as a fraction of average Total Assets based on most recent two periods
Return on Capital Employed	roce	Profitability	Earnings Before Interest and Taxes as a fraction of average Capital Employed based on most recent two periods, where Capital Employed is the sum of Debt in Long-term and Current Liabilities and Common/Ordinary Equity
Return on Equity	roe	Profitability	Net Income as a fraction of average Book Equity based on most recent two periods, where Book Equity is defined as the sum of Total Parent Stockholders' Equity and Deferred Taxes and Investment Tax Credit
Total Debt/Equity	de_ratio	Solvency	Total Liabilities to Shareholders’ Equity (common and preferred)
Total Debt/Total Assets	debt_assets	Solvency	Total Debt as a fraction of Total Assets
Total Debt/Total Assets	debt_at	Solvency	Total Liabilities as a fraction of Total Assets
Total Debt/Capital	debt_capital	Solvency	Total Debt as a fraction of Total Capital, where Total Debt is defined as the sum of Accounts Payable and Total Debt in Current and Long-

Financial Ratio	Variable Name	Category	Formula
			term Liabilities, and Total Capital is defined as the sum of Total Debt and Total Equity (common and preferred)
After-tax Interest Coverage	intcov	Solvency	Multiple of After-tax Income to Interest and Related Expenses
Interest Coverage Ratio	intcov_ratio	Solvency	Multiple of Earnings Before Interest and Taxes to Interest and Related Expenses
Dividend Payout Ratio	dpr	Valuation	Dividends as a fraction of Income Before Extra. Items
Forward P/E to 1-year Growth (PEG) ratio	PEG_1yrforward	Valuation	Price-to-Earnings, excl. Extraordinary Items (diluted) to 1-Year EPS Growth rate
Forward P/E to Long-term Growth (PEG) ratio	PEG_ltgforward	Valuation	Price-to-Earnings, excl. Extraordinary Items (diluted) to Long-term EPS Growth rate
Trailing P/E to Growth (PEG) ratio	PEG_trailing	Valuation	Price-to-Earnings, excl. Extraordinary Items (diluted) to 3-Year past EPS Growth
Book/Market	bm	Valuation	Book Value of Equity as a fraction of Market Value of Equity
Shillers Cyclically Adjusted P/E Ratio	capei	Valuation	Multiple of Market Value of Equity to 5-year moving average of Net Income
Dividend Yield	divyield	Valuation	Indicated Dividend Rate as a fraction of Price
Enterprise Value Multiple	evm	Valuation	Multiple of Enterprise Value to EBITDA
Price/Cash flow	pcf	Valuation	Multiple of Market Value of Equity to Net Cash Flow from Operating Activities
P/E (Diluted, Excl. EI)	pe_exi	Valuation	Price-to-Earnings, excl. Extraordinary Items (diluted)
P/E (Diluted, Incl. EI)	pe_inc	Valuation	Price-to-Earnings, incl. Extraordinary Items (diluted)
Price/Operating Earnings (Basic, Excl. EI)	pe_op_basic	Valuation	Price to Operating EPS, excl. Extraordinary Items (Basic)
Price/Operating Earnings (Diluted, Excl. EI)	pe_op_dil	Valuation	Price to Operating EPS, excl. Extraordinary Items (Diluted)
Price/Sales	ps	Valuation	Multiple of Market Value of Equity to Sales
Price/Book	ptb	Valuation	Multiple of Market Value of Equity to Book Value of Equity

Table A33: Summary Statistics Filing Outcomes

Year	Bankruptcies	Survived	Tortious Bankruptcy	Long Legal Process	Average Duration	363 Asset Sale	Total Assets (Millions)
1980	1	1	0	1	1157	0	514
1981	3	3	0	3	1627	1	4,501
1982	7	5	1	7	920	0	12,212
1983	3	2	0	3	951	0	3,306
1984	5	5	0	5	771	1	8,623
1985	4	4	1	4	830	0	4,515
1986	6	4	2	6	1240	1	19,215
1987	5	5	2	3	597	1	79,864
1988	6	5	0	6	755	0	79,930
1989	8	3	0	7	937	0	44,385
1990	20	15	4	17	845	2	46,403
1991	19	14	2	13	815	1	63,652
1992	16	12	0	10	511	0	44,737
1993	13	10	2	8	523	1	8,683
1994	5	3	0	3	454	0	2,413
1995	9	7	0	7	761	3	13,480
1996	10	4	0	4	356	2	10,216
1997	9	4	0	5	805	3	9,475
1998	16	8	2	10	698	4	13,685
1999	26	15	1	14	519	5	29,450
2000	45	24	2	30	635	12	64,832
2001	56	32	6	33	578	14	221,487
2002	35	23	4	19	546	6	217,744
2003	35	21	5	19	670	13	53,109
2004	14	11	0	5	406	1	21,957
2005	15	11	1	10	612	4	93,852
2006	7	6	0	3	318	0	15,504
2007	10	5	0	4	537	3	70,583
2008	26	13	1	16	726	14	1,245,430
2009	56	33	2	23	398	20	402,871
2010	23	11	0	8	421	9	68,774
2011	10	6	0	4	318	2	57,132
2012	17	7	0	8	344	6	28,226
2013	14	9	0	0	150	5	16,992
2014	10	5	0	4	385	5	53,336
2015	16	6	2	6	348	7	51,247
2016	23	19	0	3	186	4	54,040
2017	14	5	0	0	70	0	28,405

Bibliography*

Abarbanell, J. S., & Lehavy, R. Differences in commercial database reported earnings: Implications for empirical research. *SSRN Electronic Journal,* doi:10.2139/ssrn.228918

Abarbanell, J., & Lehavy, R. (2003). Biased forecasts or biased earnings? the role of reported earnings in explaining apparent bias and over/underreaction in analysts' earnings forecasts. *Journal of Accounting and Economics, 36*(1), 105-146.

Almamy, J., Aston, J., & Ngwa, L. N. (2016). An evaluation of Altman's Z-score using cash flow ratio to predict corporate failure amid the recent financial crisis: Evidence from the UK. *Journal of Corporate Finance, 36*, 278-285.

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

Altman, E. I. (1984). A further empirical investigation of the bankruptcy cost question. *The Journal of Finance, 39*(4), 1067-1089.

Altman, E. I. (2002). *Bankruptcy, credit risk, and high yield junk bonds*. Wiley-Blackwell.

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

Anderson, A., Kleinberg, J., & Mullainathan, S. (2017). Assessing human error against a benchmark of perfection. *ACM Transactions on Knowledge Discovery from Data (TKDD), 11*(4), 45.

Anilowski, C., Feng, M., & Skinner, D. J. (2007). Does earnings guidance affect market returns? the nature and information content of aggregate earnings guidance. *Journal of Accounting and Economics, 44*(1), 36-63.

Antonelli, C. (1989). A failure-inducement model of research and development expenditure: Italian evidence from the early 1980s. *Journal of Economic Behavior & Organization, 12*(2), 159-180.

Aziz, A., Emanuel, D. C., & Lawson, G. H. (1988). Bankruptcy prediction-an investigation of cash flow based models. *Journal of Management Studies, 25*(5), 419-437.

Bagheri, A., Peyhani, H. M., & Akbari, M. (2014). Financial forecasting using ANFIS networks with quantum-behaved particle swarm optimization. *Expert Systems with Applications, 41*(14), 6235-6250.

Bagnoli, M., Beneish, M. D., & Watts, S. G. (1999). Whisper forecasts of quarterly earnings per share. *Journal of Accounting and Economics, 28*(1), 27-50.

Baird, D. G., & Morrison, E. R. (2011). Dodd-frank for bankruptcy lawyers. *Am.Bankr.Inst.L.Rev., 19*, 287.

Baird, J. (2017). Renaissance technologies: Generating alpha without wall street veterans or MBAs. Retrieved from https://digit.hbs.org/submission/renaissance-technologies-generating-alpha-without-wall-street-veterans-or-mbas/

Ball, R., & Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research, 6*(2), 159-178. Retrieved from https://search.proquest.com/docview/1420660825

Barber, B., Lehavy, R., McNichols, M., & Trueman, B. (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *The Journal of Finance, 56*(2), 531-563.

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

Barniv, R., Agarwal, A., & Leach, R. (2002). Predicting bankruptcy resolution. *Journal of Business Finance & Accounting, 29*(3-4), 497-520.

Barron, O. E., Harris, D. G., & Stanford, M. (2005). Evidence that investors trade on private event-period information around earnings announcements. *The Accounting Review, 80*(2), 403-421.

Barth, M. E., & Hutton, A. P. (2004). Analyst earnings forecast revisions and the pricing of accruals. *Review of Accounting Studies, 9*(1), 59-96.

Bartov, E., Givoly, D., & Hayn, C. (2002). The rewards to meeting or beating earnings expectations. *Journal of Accounting and Economics, 33*(2), 173-204.

Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance, 40*, 432-442.

Beaver, W. H. (1968). The information content of annual earnings announcements. *Journal of Accounting Research, 6*, 67-92.

Beaver, W. H., McNichols, M. F., & Rhie, J. (2005). Have financial statements become less informative? evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies, 10*(1), 93-122.

Behn, B. K., Choi, J., & Kang, T. (2008). Audit quality and properties of analyst earnings forecasts. *The Accounting Review, 83*(2), 327-349.

Behr, A., & Weinblat, J. (2017). Default patterns in seven EU countries: A Random Forest approach. *International Journal of the Economics of Business, 24*(2), 181-222.

Berger, P. G., Ham, C. C., & Kaplan, Z. (2016). Do analysts say anything about earnings without revising their earnings forecasts?

Bergmeir, C., & Bentez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences, 191*, 192-213.

Bernard, V. L., & Thomas, J. K. (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics, 13*(4), 305-340.

Bernstein, S., & Sheen, A. (2016). The operational consequences of private equity buyouts: Evidence from the restaurant industry. *The Review of Financial Studies, 29*(9), 2387-2418.

Bhattacharya, N., Sheikh, A., & Thiagarajan, S. R. (2006). Does the market listen to whispers? *The Journal of Investing, 15*(1), 16-24.

Bialik, C. (2017). *Local economic outlook.* Yelp. Retrieved from https://www.yelpblog.com/wp-content/uploads/2017/10/Yelp-Local-Economic-Outlook-Report_October-2017.pdf

Bibler, G. A. (1987). The status of unaccrued tort claims in Chapter 11 bankruptcy proceedings. *Am.Bankr.LJ, 61*, 145.

Booth, A., Gerding, E., & Mcgroarty, F. (2014). Automated trading with performance weighted random forests and seasonality. *Expert Systems with Applications, 41*(8), 3651-3661.

Bouchaud, J., & Cont, R. (1998). A Langevin approach to stock market fluctuations and crashes. *The European Physical Journal B-Condensed Matter and Complex Systems*, *6*(4), 543-550.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7), 1145-1159.

Bradshaw, M. T., Drake, M. S., Myers, J. N., & Myers, L. A. (2012). A re-examination of analysts' superiority over time-series forecasts of annual earnings. *Review of Accounting Studies, 17*(4), 944-968.

Branson, B. C., Lorek, K. S., & Pagach, D. P. (1995). Evidence on the superiority of analysts quarterly earnings forecasts for small capitalization firms. *Decision Sciences, 26*(2), 243-263.

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees* CRC press.

Bris, A., Welch, I., & Zhu, N. (2004). The costs of bankruptcy Chapter 7 cash auctions vs. Chapter 11 bargaining.

Bris, A., Welch, I., & Zhu, N. (2006). The costs of bankruptcy: Chapter 7 liquidation versus chapter 11 reorganization. *The Journal of Finance, 61*(3), 1253-1303.

Brown, L. D. (2001). A temporal analysis of earnings surprises: Profits versus losses. *Journal of Accounting Research, 39*(2), 221-241.

Brown, L. D., Han, J. C., Keon Jr, E. F., & Quinn, W. H. (1996). Predicting analysts' earnings surprise. *The Journal of Investing, 5*(1), 17-23.

Burgstahler, D., & Dichev, I. (1997). Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics, 24*(1), 99-126.

Burgstahler, D., & Eames, M. (2006). Management of earnings and analysts' forecasts to achieve zero and small positive earnings surprises. *Journal of Business Finance & Accounting, 33*(5-6), 633-652.

Callen, J. L., Kwan, C. C., Yip, P. C., & Yuan, Y. (1996). Neural network forecasting of quarterly accounting earnings. *International Journal of Forecasting, 12*(4), 475-482.

Chan, L. K., Karceski, J., & Lakonishok, J. (2007). Analysts' conflicts of interest and biases in earnings forecasts. *Journal of Financial and Quantitative Analysis, 42*(4), 893-913.

Chandra, D. K., Ravi, V., & Bose, I. (2009). Failure prediction of dotcom companies using hybrid intelligent techniques. *Expert Systems with Applications, 36*(3), 4830-4837.

Chaudhuri, A., & De, K. (2011). Fuzzy support vector machine for bankruptcy prediction. *Applied Soft Computing, 11*(2), 2472-2486.

Chen, J., Chen, W., Huang, C., Huang, S., & Chen, A. (2016). Financial time-series data analysis using deep convolutional neural networks. Paper presented at the *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, 87-92.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Paper presented at the *Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 785-794.

Chudson, W. A. (1945). The pattern of corporate financial structure: A cross-section view of manufacturing, mining, trade, and construction, 1937. *NBER Books.*

Claus, J., & Thomas, J. (2001). Equity premia as low as three percent? evidence from analysts' earnings forecasts for domestic and international stock markets. *The Journal of Finance, 56*(5), 1629-1666.

Cornelius Casey, & Norman Bartczak. (1985). Using operating cash flow data to predict financial distress: Some extensions. *Journal of Accounting Research, 23*(1), 384-401. doi:10.2307/2490926

Dambolena, I. G., & Khoury, S. J. (1980). Ratio stability and corporate failure. *The Journal of Finance, 35*(4), 1017-1026.

David R. Anderson, & Kenneth P. Burnham. (2002). Avoiding pitfalls when using information-theoretic methods. *The Journal of Wildlife Management, 66*(3), 912-918. doi:10.2307/3803155

Deligianni, D., & Kotsiantis, S. (2012). Forecasting corporate bankruptcy with an ensemble of classifiers. Paper presented at the *Hellenic Conference on Artificial Intelligence,* 65-72.

Dhar, V., & Chou, D. (2001). A comparison of nonlinear methods for predicting earnings surprises and returns. *IEEE Transactions on Neural Networks, 12*(4), 907-921.

Dietterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning, 40*(2), 139-157.

DiPietro, R. B., Parsa, H. G., & Gregory, A. (2011). Restaurant QSC inspections and financial performance: An empirical investigation. *International Journal of Contemporary Hospitality Management, 23*(7), 982-999.

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78-87.

du Jardin, P. (2017). Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications, 75*, 25-43.

Duffie, D., Pedersen, L. H., & Singleton, K. J. (2003). Modeling sovereign yield spreads: A case study of russian debt. *The Journal of Finance, 58*(1), 119-159.

Easterwood, J. C., & Nutt, S. R. (1999). Inefficiency in analysts' earnings forecasts: Systematic misreaction or systematic optimism? *The Journal of Finance, 54*(5), 1777-1797.

Eliazar, I. (2017). Lindy's law. *Physica A: Statistical Mechanics and its Applications, 486*, 797-805.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology, 77*(4), 802-813.

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics, 116*(1), 1-22.

Fama, E. F., & Blume, M. E. (1966). Filter rules and stock-market trading. *The Journal of Business, 39*(1), 226-241.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, 27(8), 861-874.

Ferri, C., Flach, P., & Hernndez-Orallo, J. (2002). Learning decision trees using the area under the ROC curve. Paper presented at the *Icml, 2*, 139-146.

Fidler, F., Geoff, C., Mark, B., & Neil, T. (2004). Statistical reform in medicine, psychology and ecology. *The Journal of Socio-Economics, 33*(5), 615-630.

FitzPatrick, P. J. (1932). *A comparison of the ratios of successful industrial enterprises with those of failed companies*.

Foster, G. (1973). Stock market reaction to estimates of earnings per share by company officials. *Journal of accounting research, 11*(1), 25-37. Retrieved from http://www.econis.eu/PPNSET?PPN=39273009X

Foster, G. (1986). *Financial statement analysis, 2/e* Pearson Education.

Franks, J. R., & Torous, W. N. (1989). An empirical investigation of US firms in reorganization. *The Journal of Finance, 44*(3), 747-769.

Fried, D., & Givoly, D. (1982). Financial analysts' forecasts of earnings: A better surrogate for market expectations. *Journal of Accounting and Economics, 4*(2), 85-107.

Friedman, J. H., & Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine, 22*(9), 1365-1381.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* Springer series in statistics New York.

Friedrich, R., Peinke, J., & Renner, C. (2000). How to quantify deterministic and random influences on the statistics of the foreign exchange market. *Physical Review Letters, 84*(22), 5224.

George Foster, Chris Olsen, & Terry Shevlin. (1984). Earnings releases, anomalies, and the behavior of security returns. *The Accounting Review, 59*(4), 574-603. Retrieved from https://www.jstor.org/stable/247321

Givoly, D., & Lakonishok, J. (1984). The quality of analysts' forecasts of earnings. *Financial Analysts Journal, 40*(5), 40-47.

Glaeser, E., Kim, H., & Luca, M. (2017). Nowcasting the Local Economy: Using Yelp Data to Measure Economic Activity (No. 24010). *National Bureau of Economic Research, Inc.*

Gu, S., Kelly, B. T., & Xiu, D.Empirical asset pricing via machine learning. *SSRN Electronic Journal,* doi:10.2139/ssrn.3159577

Graham, B., & Dodd, D. L. (1934). *Security analysis: Principles and technique* McGraw-Hill.

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The economic implications of corporate financial reporting. *Journal of Accounting and Economics, 40*(1), 3-73.

Grossman, S. J., & Zhou, Z. (1993). Optimal investment strategies for controlling drawdowns. *Mathematical Finance, 3*(3), 241-276.

Han, Y., & Zhou, G.Taming momentum crashes: A simple stop-loss strategy. *SSRN Electronic Journal,* doi:10.2139/ssrn.2407199

Hand, D. J. (2009). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning, 77*(1), 103-123.

Hardiman, A. (1985). Toxic torts and chapter 11 reorganization: The problem of future claims. *Vand.L.Rev., 38*, 1369-2013.

Hart, O. (2000). Different approaches to bankruptcy (No. w7921). *National Bureau of Economic Research*.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Overview of supervised learning. *The elements of statistical learning* (pp. 9-41). New York: Springer.

Hellstrom, T., & Holmstrom, K. (1998). Predicting the stock market. *Unpublished Thesis, Malardalen University, Department of Mathematics and Physics, Vasteras, Sweden,*

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

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks, 2*(5), 359-366.

Hu, Y., Feng, B., Zhang, X., Ngai, E., & Liu, M. (2015). Stock trading rule discovery with an evolutionary trend following model. *Expert Systems with Applications*, *42*(1), 212-222.

Huang, T., Leu, Y., & Pan, W. (2016). Constructing ZSCORE-based financial crisis warning models using fruit fly optimization algorithm and general regression neural network. *Kybernetes, 45*(4), 650-665.

Hutson, J. K. (1983). TRIX-triple exponential smoothing oscillator. *Technical Analysis of Stocks and Commodities,* , 105-108.

Jerome H. Friedman. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189-1232. doi:10.2307/2699986

Johnson, W. B., & Zhao, R. (2012). Contrarian share price reactions to earnings surprises. *Journal of Accounting, Auditing & Finance, 27*(2), 236-266.

Jones, C. P., & Litzenberger, R. H. (1970). Quarterly earnings reports and intermediate stock price trends. *The Journal of Finance, 25*(1), 143-148.

Jones, S. (2017). Corporate bankruptcy prediction: A high dimensional analysis. *Review of Accounting Studies, 22*(3), 1366-1422.

Jones, S., & Hensher, D. A. (2004). Predicting firm financial distress: A mixed logit model. *The Accounting Review, 79*(4), 1011-1038.

Jones, S., Johnstone, D., & Wilson, R. (2017). Predicting corporate bankruptcy: An evaluation of alternative statistical frameworks. *Journal of Business Finance & Accounting, 44*(1-2), 3-34.

Joret, G., Micek, P., Milans, K. G., Trotter, W. T., Walczak, B., & Wang, R. (2016). Tree-width and dimension. *Combinatorica, 36*(4), 431-450.

Jung, J., Concannon, C., Shroff, R., Goel, S., & Goldstein, D. G. (2017). *Simple rules for complex decisions*

Kaminski, K. M., & Lo, A. W. (2014). *When do stop-loss rules stop losses?* doi://doi-org.ezproxy.auckland.ac.nz/10.1016/j.finmar.2013.07.001

Kang, J. S., Kuznetsova, P., Luca, M., & Choi, Y. (2013). Where not to eat? improving public policy by predicting hygiene inspections using online reviews. Paper presented at the *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing,* 1443-1448.

Karas, M., & Reznakova, M. (2017). Predicting the bankruptcy of construction companies: A CART-based model. *Engineering Economics, 28*(2), 145-154. doi:10.5755/j01.ee.28.2.16353

Kasparov, G. (2010). The chess master and the computer. *The New York Review of Books, 57*(2), 16-19.

Kasznik, R., & McNichols, M. F. (2002). Does meeting earnings expectations matter? evidence from analyst forecast revisions and share prices. *Journal of Accounting Research, 40*(3), 727-759.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., . . . Liu, T. (2017). LightGBM: A highly efficient gradient boosting decision tree. Paper presented at the *Advances in Neural Information Processing Systems,* 3149-3157.

Kim, K. (2003). Financial time series forecasting using support vector machines. *Neurocomputing, 55*(1), 307-319.

Kim, M., & Kang, D. (2010). Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications, 37*(4), 3373-3379.

Kim, S. Y., & Upneja, A. (2014). Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. *Economic Modelling, 36,* 354-362.

Kinney, W., Burgstahler, D., & Martin, R. (2002). Earnings surprise “materiality” as measured by stock returns. *Journal of Accounting Research, 40*(5), 1297-1329.

Kirt C. Butler, & Larry H. P. Lang. (1991). The forecast accuracy of individual analysts: Evidence of systematic optimism and pessimism. *Journal of Accounting Research, 29*(1), 150-156.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics, 133*(1), 237-293.

Kogan, S., Levin, D., Routledge, B. R., Sagi, J. S., & Smith, N. A. (2009). Predicting risk from financial reports with regression. Paper presented at the *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics,* 272-280.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Paper presented at the *Ijcai,, 14*(2) 1137-1145.

Kreeger, J. C., Parsa, H. G., Smith, S. J., & Kubickova, M. (2018). Calendar effect and the role of seasonality in consumer comment behavior: A longitudinal study in the restaurant industry. *Journal of Foodservice Business Research, 21*(3), 342-357.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. Paper presented at the *Advances in Neural Information Processing Systems,* 1097-1105.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* Springer.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *J Stat Softw, 36*(11), 1-13.

Laney, D. B. (2017). *Infonomics: How to monetize, manage, and measure information as an asset for competitive advantage.* Routledge.

Langevin, P. (1908). Sur la thorie du mouvement brownien. *CR Acad.Sci.Paris, 146*(530-533), 530.

Latane, H. A., & Jones, C. P. (1977). Standardized unexpected earnings—A progress report. *The Journal of Finance, 32*(5), 1457-1465.

Lawrence D. Brown, Gordon D. Richardson, & Steven J. Schwager. (1987). An information interpretation of financial analyst superiority in forecasting earnings. *Journal of Accounting Research, 25*(1), 49-67. doi:10.2307/2491258

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: An analysis using boosted regression trees. *Marine Ecology Progress Series, 321*, 267-281.

Lhabitant, F. (2011). *Handbook of hedge funds* John Wiley & Sons.

Li, J. (2012). Prediction of corporate bankruptcy from 2008 through 2011. *Journal of Accounting and Finance, 12*(1), 31-41.

Li, K. (1999). Bayesian analysis of duration models: An application to chapter 11 bankruptcy. *Economics Letters, 63*(3), 305-312.

Lian, J., Zhang, F., Xie, X., & Sun, G. (2017). Restaurant survival analysis with heterogeneous information. Paper presented at the *Proceedings of the 26th International Conference on World Wide Web Companion,* 993-1002.

Liang, D., Lu, C., Tsai, C., & Shih, G. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research, 252*(2), 561-572.

Liu, H., & Motoda, H. (2012). *Feature selection for knowledge discovery and data mining* Springer Science & Business Media.

LoPucki, L. M., & Doherty, J. W. (2008). Professional overcharging in large bankruptcy reorganization cases. *Journal of Empirical Legal Studies, 5*(4), 983-1017.

LoPucki, L. M., & Doherty, J. W. (2015). Bankruptcy survival. *UCLA L.Rev., 62*, 969.

LoPucki, L. M., & Kalin, S. D. (2001). Failure of public company bankruptcies in Delaware and New York: Empirical evidence of a race to the bottom. *Vand.L.Rev., 54*, 231.

Luca, M. (2011). *Reviews, reputation, and revenue: The case of yelp.com.* Retrieved from http://econpapers.repec.org/paper/hbswpaper/12-016.htm

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science, 62*(12), 3412-3427.

Lundberg, S. M., & Lee, S. (2017). A unified approach to interpreting model predictions. Paper presented at the *Advances in Neural Information Processing Systems*, 4768-4777.

Luo, T., & Stark, P. B. (2015). Nine out of 10 restaurants fail? check, please. *Significance, 12*(2), 25-29.

Lys, T., & Sohn, S. (1990). The association between revisions of financial analysts' earnings forecasts and security-price changes. *Journal of Accounting and Economics*, *13*(4), 341-363.

Lys, T., & Soo, L. G. (1995). Analysts' forecast precision as a response to competition. *Journal of Accounting, Auditing & Finance, 10*(4), 751-765.

MacGregor, N., & Lo, D. (2015). The effect of chain size and customer type on company survival: Evidence from multi-unit restaurants. *Academy of Management Proceedings, 2015*(1), 17991.

MacKay, D. J. (1992). The evidence framework applied to classification networks. *Neural Computation, 4*(5), 720-736.

Matras, K. (2016). Zacks investment research - video: Zacks earnings ESP (expected surprise prediction). Retrieved from https://search.proquest.com/docview/1838754510

Mejia, J., Mankad, S., & Gopal, A. (2015). More than just words: Latent semantic analysis, online reviews and restaurant closure. *Academy of Management Proceedings, 2015*(1), 13912. doi:10.5465/AMBPP.2015.13912abstract

Merwin, C. L. 1. (1942). *Financing small corporations in five manufacturing industries, 1926-36*. United States: Retrieved from http://catalog.hathitrust.org/Record/001128380

Meyer, P. A. (1967). Price discrimination, regional loan rates, and the structure of the banking industry. *The Journal of Finance, 22*(1), 37-48.

Mian, S. (2013). *Zacks earnings ESP*. Zacks. Retrieved from https://staticzacks.net/pdf/EESPreport1V3.pdf

Michael J. Gombola, & J. Edward Ketz. (1983). A note on cash flow and classification patterns of financial ratios. *The Accounting Review, 58*(1), 105-114. Retrieved from https://www.jstor.org/stable/246645

Montas, E., Quevedo, J. R., Prieto, M. M., & Menndez, C. O. (2002). Forecasting time series combining machine learning and Box-Jenkins time series. Paper presented at the *Ibero-American Conference on Artificial Intelligence,* 491-499.

Mselmi, N., Lahiani, A., & Hamza, T. (2017). Financial distress prediction: The case of French small and medium-sized firms. *International Review of Financial Analysis, 50*, 67-80.

Mukherjee, S., Golland, P., & Panchenko, D. (2003). Permutation tests for classification, 11-21

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31*(2), 87-106.

National Restaurant Association. (2017). *Restaurant industry outlook*. Retrieved from http://www.restaurant.org/Downloads/PDFs/News-Research/2017_Restaurant_outlook_summary-FINAL.pdf

Neves, J. C., & Vieira, A. (2006). Improving bankruptcy prediction with hidden layer learning vector quantization. *European Accounting Review, 15*(2), 253-271.

O'Brien, P. C. (1998). Discussion of international variation in accounting measurement rules and analysts' earnings forecast errors. *Journal of Business Finance & Accounting, 25*(9-10), 1249-1254.

Ohlson, J. A. (2009). Financial ratios and the probabilistic prediction of bankruptcy. *Financial Accounting and Investment Management, 18(1)*, 363-385. Retrieved from http://www.econis.eu/PPNSET?PPN=603823629

Olson, D. L., Delen, D., & Meng, Y. (2012). Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems, 52*(2), 464-473.

Parsa, H. G., Self, J. T., Njite, D., & King, T. (2005). Why restaurants fail. *Cornell Hotel and Restaurant Administration Quarterly, 46*(3), 304-322.

Parsa, H. G., Self, J., Sydnor-Busso, S., & Yoon, H. J. (2011). Why restaurants fail? part II- the impact of affiliation, location, and size on restaurant failures: Results from a survival analysis. *Journal of Foodservice Business Research, 14*(4), 360-379.

Parsa, H. G., van der Rest, Jean-Pierre I, Smith, S. R., Parsa, R. A., & Bujisic, M. (2015). Why restaurants fail? part IV: The relationship between restaurant failures and demographic factors. *Cornell Hospitality Quarterly, 56*(1), 80-90.

Patel, C. (1980). *Technical trading systems for commodities and stocks*. Trading Systems Research.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications, 42*(1), 259-268.

Payne, J. L., & Thomas, W. B. (2003). The implications of using stock-split adjusted I/B/E/S data in empirical research. *The Accounting Review, 78*(4), 1049-1067.

Pervan, I., Pervan, M., & Vukoja, B. (2011). Prediction of company bankruptcy using statistical techniques–Case of croatia. *Croatian Operational Research Review, 2*(1), 158-167.

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies, 2,* 37-63.

Provost, F., & Kohavi, R. (1998). Glossary of terms. *Journal of Machine Learning, 30*(2-3), 271-274.

Puga, D. (1999). The rise and fall of regional inequalities. *European Economic Review, 43*(2), 303-334.

Purdy, D., Chen, L., & Sumers, T. R. (2017). *Cascaded Boosted Predictive Models.*

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81-106.

Raileanu, L. E., & Stoffel, K. (2004). Theoretical comparison between the Gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence, 41*(1), 77-93.

Ramnath, S., Rock, S., & Shane, P. (2008). The financial analyst forecasting literature: A taxonomy with suggestions for further research. *International Journal of Forecasting, 24*(1), 34-75.

Ramsey, J. B., & Zhang, Z. (1997). The analysis of foreign exchange data using waveform dictionaries. *Journal of Empirical Finance, 4*(4), 341-372.

Robert N. Freeman, & Senyo Y. Tse. (1992). A nonlinear model of security price responses to unexpected earnings. *Journal of Accounting Research, 30*(2), 185-209. doi:10.2307/2491123

Rose, P. S., & Giroux, G. A. (1984). Predicting corporate bankruptcy: An analytical and empirical evaluation. *Review of Financial Economics, 19*(2), 1.

Sable, R. G., Roeschenthaler, M. J., & Blanks, D. F. (2006). When the 363 sale is the best route. *J.Bankr.L.& Prac., 15,* 2.

Schaaf, A. H. (1966). Regional differences in mortgage financing costs. *The Journal of Finance, 21*(1), 85-94.

Schapire, R. E., & Freund, Y. (2012). *Boosting: Foundations and algorithms* MIT press.

Scott, J. (1981). The probability of bankruptcy: A comparison of empirical predictions and theoretical models. *Journal of Banking & Finance, 5*(3), 317-344.

Sharif Razavian, A., Azizpour, H., Sullivan, J., & Carlsson, S. (2014). CNN variables off-the-shelf: An astounding baseline for recognition. Paper presented at the *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 806-813.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business, 74*(1), 101-124.

Siegert, S., Friedrich, R., & Peinke, J. (1998). Analysis of data sets of stochastic systems. *Physics Letters A, 243*(5-6), 275-280.

Soffer, L. C., Thiagarajan, S. R., & Walther, B. R. (2000). Earnings preannouncement strategies. *Review of Accounting Studies, 5*(1), 5-26.

Somnath Das, Carolyn B. Levine, & K. Sivaramakrishnan. (1998). Earnings predictability and bias in analysts' earnings forecasts. *The Accounting Review, 73*(2), 277-294. Retrieved from https://www.jstor.org/stable/248469

Stauth, J. (2013). Trading (and predicting) earnings surprises. *Thomson Reuters*. Retrieved from https://blog.quantopian.com/wp-content/uploads/2013/05/Trading_Earnings_surprises-jess.pdf

Stehman, S. V. (1997). Selecting and interpreting measures of thematic classification accuracy. *Remote Sensing of Environment, 62*(1), 77-89.

Stickel, S. E. (1995). The anatomy of the performance of buy and sell recommendations. *Financial Analysts Journal, 51*(5), 25-39.

Sun, J., & Li, H. (2012). Financial distress prediction using support vector machines: Ensemble vs. individual. *Applied Soft Computing, 12*(8), 2254-2265.

Tan, C., Lee, L., & Pang, B. (2014). The effect of wording on message propagation: Topic- and author-controlled natural experiments on twitter. *arXiv Preprint arXiv:1405.1438,*

Taylor, D. C., & Aday, J. B. (2016). Consumer generated restaurant ratings: A preliminary look at OpenTable.com. *Journal of New Business Ideas and Trends, 14*(1), 14-23.

Teixeira, L. A., & De Oliveira, Adriano Lorena Inacio. (2010). A method for automatic stock trading combining technical analysis and nearest neighbor classification. *Expert Systems with Applications, 37*(10), 6885-6890.

Trueman, B. (1990). Theories of earnings-announcement timing. *Journal of Accounting and Economics, 13*(3), 285-301.

Volkov, A., Benoit, D. F., & Van den Poel, D. (2017). Incorporating sequential information in bankruptcy prediction with predictors based on markov for discrimination. *Decision Support Systems, 98*, 59-68.

W. S. Hopwood, J. C. Mckeown, & P. Newbold. (1982). The additional information content of quarterly earnings reports: Intertemporal disaggregation. *Journal of Accounting Research, 20*(2), 343-349. doi:10.2307/2490744

Waymire, G. (1986). Additional evidence on the accuracy of analyst forecasts before and after voluntary management earnings forecasts. *Accounting Review, 61*, 129-142.

Weiss, L. A., Bhandari, J. S., & Robins, R. (2000). An analysis of state-wide variation in bankruptcy rates in the united states. *Bankr.Dev.J., 17*, 407.

Whittingham, M. J., Stephens, P. A., Bradbury, R. B., & Freckleton, R. P. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology, 75*(5), 1182-1189.

William A. Collins, & William S. Hopwood. (1980). A multivariate analysis of annual earnings forecasts generated from quarterly forecasts of financial analysts and univariate time-series models. *Journal of Accounting Research, 18*(2), 390-406. doi:10.2307/2490585

William H. Beaver. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4,* 71-111. doi:10.2307/2490171

William Kross, Byung Ro, & Douglas Schroeder. (1990). Earnings expectations: The analysts' information advantage. *The Accounting Review, 65*(2), 461-476. Retrieved from https://www.jstor.org/stable/247634

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data mining: Practical machine learning tools and techniques, *Morgan Kaufmann*.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation, 1*(1), 67-82.

Womack, K. L. (1996). Do brokerage analysts' recommendations have investment value? *The Journal of Finance, 51*(1), 137-167.

Xiao, Y., Xiao, J., Lu, F., & Wang, S. (2013). Ensemble ANNs-PSO-GA approach for day-ahead stock E-exchange prices forecasting. *International Journal of Computational Intelligence Systems, 6*(1), 96-114.

Youn, H., & Gu, Z. (2010). Predict US restaurant firm failures: The artificial neural network model versus logistic regression model. *Tourism and Hospitality Research, 10*(3), 171-187.

Zhang, M., & Luo, L.Can user generated content predict restaurant survival: Deep learning of yelp photos and reviews. *SSRN Electronic Journal,* doi:10.2139/ssrn.3108288

Zhao, D., Huang, C., Wei, Y., Yu, F., Wang, M., & Chen, H. (2017). An effective computational model for bankruptcy prediction using kernel extreme learning machine approach. *Computational Economics, 49*(2), 325-341.

Zhou, L. (2013). Performance of corporate bankruptcy prediction models on imbalanced dataset: The effect of sampling methods. *Knowledge-Based Systems, 41,* 16-25.

Zięba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic variables generation in application to bankruptcy prediction. *Expert Systems with Applications, 58*, 93-101.

Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, *22*, 59-82.

*Includes all the citations for three chapters to a doctoral thesis.