Acta Commercii - Independent Research Journal in the Management Sciences
ISSN: (Online) 1684-1999, (Print) 2413-1903

Machine learning and company failure prediction: Evidence from South Africa

Authors:
Nicolene Wesson[1]
Dewald Mienie[1]
Anthea Myatt[1]

Affiliations:
[1]Stellenbosch Business School, Faculty of Economic and Management Sciences, Stellenbosch University, Bellville, South Africa

Corresponding author:
Nicolene Wesson,
nwesson@sun.ac.za

Dates:
Received: 26 Nov. 2024
Accepted: 11 Feb. 2025
Published: 19 Mar. 2025

How to cite this article:
Wesson, N., Mienie, D. & Myatt, A., 2025, 'Machine learning and company failure prediction: Evidence from South Africa', *Acta Commercii* 25(1), a1365. https://doi.org/10.4102/ac.v25i1.1365

Orientation: Machine learning has advanced substantially over the past two decades and exhibits the potential to overcome the limitations of traditional statistical methods for predicting company failure. While extensive research has been conducted globally to predict company failure using machine learning, these techniques are relatively unexplored in an emerging market context.

Research purpose: The accuracy of company failure prediction was assessed when applying an array of fundamental machine learning algorithms in South Africa.

Motivation for the study: Given the significant social and economic impact of company failures, insights are provided into appropriate company failure prediction techniques in an emerging market context.

Research design, approach and method: The study sample consisted of 56 companies (of which 28 were classified as failed) that were listed on the Johannesburg Stock Exchange during 2010–2021. Company failure prediction up to 3 years in advance was measured by applying eight fundamental machine learning techniques and the traditional logit analysis statistical method.

Main findings: Two machine learning algorithms outperformed the traditional method in some years. Furthermore, not all machine learning techniques were suited to predict company failure in all years.

Practical implications: Machine learning is not necessarily more accurate than traditional statistical methods. Applying the appropriate technique in company failure prediction models requires a clear understanding of the available methodologies for the task at hand.

Contribution: This study provides a benchmark for predictive accuracy in the South African context and lays the ground for a more sophisticated ensemble of methods to assess the accuracy of machine learning.

Keywords: company failure; failure prediction; logit analysis; machine learning; South Africa.

Introduction

Company failures have a significant social and economic impact (Huang et al. 2008; Perboli & Arabnezhad 2021). The ability to predict company failure is therefore crucial to all stakeholders, both internal and external to a company. Although each stakeholder has their own role and agenda, all are interested in the best possible state of health of the company (Huang & Yen 2019; Huang et al. 2008; Kim, Cho & Ryu 2020; Naidoo & Du Toit 2007; Perboli & Arabnezhad 2021; Qu et al. 2019).

Historically, company failure prediction was mainly performed by means of statistical methods such as multiple discriminant analysis and logit analysis (Aziz & Humayon 2006; Sewpersadh 2020; Tsai 2014). The seminal works on company failure prediction of Beaver (1966) and Altman (1968) applied univariant and multiple discriminant analysis, respectively. About a decade later, Ohlson (1980) introduced the logit analysis method, which has emerged as the dominant statistical model in company bankruptcy literature (Ding et al. 2023; Jones 2016; Oz & Simga-Mugan 2018; Rodríguez-Masero & López-Manjón 2020).

A more modern alternative to traditional statistical methods is the use of machine learning to predict company failure. Machine learning is a branch of artificial intelligence (AI) and has a long history – starting with the foundational work of Alan Turing, who introduced the Turing test to assess machine intelligence in the 1950s (Alzubi, Nayyar & Kumar 2018). The utility of machine learning algorithms increased enormously with the introduction of deep learning by Geoffery Hinton in 2006 (Alzubi et al. 2018).

Machine learning revolves around the problem of prediction, and its appeal is that it can discover complex patterns in data that were not specified in advance (Mullainathan & Spiess 2017). In this context, predicting company failure is a classification problem where the goal is to predict – based on financial statements – whether a company will fail or not. This problem is therefore well suited for machine learning. The results on the accuracy of machine learning in predicting company failure are, however, mixed (Barboza, Kimura & Altman 2017; Huang et al. 2008; Le & Viviani 2018; Nazareth & Reddy 2023; Perboli & Arabnezhad 2021; Sermpinis, Tsoukas & Zhang 2023).

Furthermore, earlier studies that compared traditional statistical methods with machine learning methods generally applied a limited number of machine learning algorithms and were mostly performed in developed economies (Barboza et al. 2017; Huang et al. 2008; Le & Viviani 2018; Perboli & Arabnezhad 2021; Sermpinis et al. 2023). This study therefore aimed to address the gap in literature by applying an array of fundamental machine learning methods (Aktan 2011) in the context of an emerging market (namely, South Africa).

The ability to predict company failure is of specific relevance in South Africa, given the country's challenging economic climate. The present study builds on the work of Muller, Steyn-Bruwer and Hamman (2009) and Steyn-Bruwer and Hamman (2006). Muller et al. (2009) compared two machine learning algorithms with two traditional statistical methods, while Steyn-Bruwer and Hamman (2006) only applied one machine learning algorithm. Applying an array of fundamental machine learning algorithms in the South African context is yet unexplored.

The purpose of this study was to test the predictive accuracy of an array of fundamental machine learning algorithms in the South African context. Firstly, the Aktan (2011) study (using eight machine learning algorithms) was replicated using South African companies to assess the success of a wide array of machine learning algorithms in predicting company failure. Secondly, the predictive accuracy using the eight machine learning algorithms of Aktan was compared with the predictive accuracy of the logit analysis method, by applying a set of variables that reflect more recent empirical evidence on significant determinants of company failure prediction (Rodríguez-Masero & López-Manjón 2020).

Literature study

Company failure

Researchers' definitions of company failure differ. The seminal study by Beaver (1966) defined failure as a company being unable to meet its financial responsibilities owing to bankruptcy, non-payment of preference dividends or bond default. Other definitions range from judicial bankruptcy to financial distress (Balcaen & Ooghe 2006; Muller et al. 2009). Aktan (2011) considered companies to be distressed if they had negative equity figures or if the Turkish bankruptcy law found them bankrupt. Under this law, a business that loses two-thirds of its capital stock is considered bankrupt even though the business has not gone through the legal process of applying for bankruptcy. Owing to jurisdictions having different rules to classify companies as being in liquidation and/or bankrupt, the definition of company failure in failure prediction studies is often country specific (Balcaen & Ooghe 2006; Kuruppu, Laswad & Oyelere 2003; Muller et al. 2009).

In line with the methodology applied by Muller et al. (2009) and Steyn-Bruwer and Hamman (2006) in the South African environment, the present study defined company failure as a company that cannot continue operating in its current state and therefore encompasses a liquidation, delisting, bankruptcy or restructuring. The terminologies 'failed' and 'non-failed' are applied to classify the companies.

Determinants of company failure

Company failure prediction models vary significantly regarding the variables considered and the methodologies applied (Bellovary, Giacomino & Akers 2007). Traditional statistical methods like multiple discriminant and logit analysis models apply pre-defined variables and assign weights based on their significance to calculate a discriminant score for classifying companies as failed or not – with logit analysis also incorporating the probability of company failure (Perboli & Arabnezhad 2021). On the other hand, machine learning models do not apply pre-defined variables: the selected variables are determined by the data (Clement 2020; Shetty, Musa & Brédart 2022). Although models have incorporated more factors over time, increasing factors does not necessarily improve accuracy (Bellovary et al. 2007). No consensus has yet been reached on the optimal explanatory variables for accurately predicting company failure (Kristóf & Virág 2020).

Company failure reflects the inability to meet financial responsibilities and is therefore in essence a 'cash phenomenon' (Steyn-Bruwer & Hamman 2006). The relevance of cash flow as a determinant of company failure has received substantial attention in the accounting and managerial accounting fields (Altman 1968; Beaver 1966; Rodríguez-Masero & López-Manjón 2020; Rujoub, Cook & Hay 1995). Cash flow is crucial to a company's existence, and sufficient cash flow is required to meet the company's financial commitments and day-to-day operation expenses (Rujoub et al. 1995). Furthermore, International Financial Reporting Standards (IFRS) require that all financial statements, except cash flow statements, are prepared by applying the accrual basis of accounting and are therefore influenced by the interpretation of management. Cash flow statements are based on historical events and are therefore free from the judgement of management, making such statements better suited for predicting company failure (Sherman & Young 2016). Notwithstanding the relevance of cash flow in predicting company failure, empirical evidence mainly shows that a combination of both accrual-based and cash flow-based ratios notably enhances the accuracy of company failure prediction models (Bhandari, Showers & Johnson-Snyder 2019; Das 2018; Rujoub et al. 1995; Sharma 2001).

Rodríguez-Masero and López-Manjón's (2020), when applying their logit analysis method, selected their ratios based on the three most frequently used ratios from literature (namely, return on assets, total liabilities to total assets and current liabilities to current assets) and added a cash flow ratio (namely, operating cash flow) as the fourth variable.

Aktan (2011), when applying his machine learning algorithms, initially identified 53 potential variables from literature, and eventually, the 10 most significant financial ratios were selected for inclusion. These ratios included cash to total assets, quick assets to total assets, financial debt to total assets, inventory to net sales, current assets to total assets, total debt to total assets, short-term debt to total assets, return on assets, operating income to total assets and cash flow to total assets.

A case for machine learning

Statistical methods, also generally referred to as traditional methods, rely on statistical models (Perboli & Arabnezhad 2021). As mentioned earlier, statistical models like multiple discriminant analysis and logit analysis employ multiple variables for prediction and assign weights based on their significance. These models are typically linear and provide the link function between a linear combination of predictors and the dependent variable (Gajdošíková & Valášková 2023). The main advantages of statistical methods are their ability to provide a level of certainty (probability) about the result and to convey the contribution of each individual feature to the result (Perboli & Arabnezhad 2021). However, the main criticism of statistical methods is that most models make certain assumptions about the normality of the distributions of the underlying data and that the distribution of financial statement data does not conform to those assumptions (Aktan 2011; Le & Viviani 2018; Ohlson 1980; Sermpinis et al. 2023). For example, even Altman's (1968) popular Z-score model assumes that all variables follow a normal distribution (Shetty et al. 2022), while the logit analysis method does not assume normality or equal covariances between groups (Bisogno, Restaino & Carlo 2018; Jones 2016).

Although statistical models are still widely used in both research and industry, they have become less accurate and require more adaptation to fit different markets and industries (Perboli & Arabnezhad 2021). According to Aktan (2011), the normality assumption of statistical methods has prompted research into financial forecasting and non-parametric models. In addition to not making any assumptions about the distribution of the underlying data, non-parametric models have the advantage that the number of parameters and their structure are determined by the data and are not fixed in advance (Clement 2020; Shetty et al. 2022). Non-parametric models are further able to learn and adapt based on the data and can capture non-linear relationships between variables (Clement 2020; Huang & Yen 2019).

Most of the non-parametric models belong to the domain of data mining in machine learning (Aktan 2011). Extensive research has been conducted on pattern recognition methods in the field of machine learning to overcome the limitations of statistical methods (Perboli & Arabnezhad 2021; Tsai 2014). Owing to its ability to process huge amounts of data while simultaneously accommodating non-linearities in data, machine learning is seen as a powerful branch of AI with wide-ranging applications in both the banking and broader financial industry (Nazareth & Reddy 2023). Opponents of the use of machine learning, however, argue that predictive accuracy is not the only metric of interest, but that there is also a need to understand risk drivers (Li, Crook & Andreeva 2017). Most machine learning algorithms are of the black box type and cannot release this information (Clement 2020). Although machine learning algorithms may display superior performance, most of these models will not be able to inform company management teams how to make improvements (Kim et al. 2020). Kim et al. therefore observed that, while machine learning models cannot be dismissed, an understanding of the appropriate methodology depending on the purpose of the prediction is required.

Machine learning studies on company failure prediction by Aktan (2011) and Tsai (2014) incorporated multiple machine learning algorithms. Aktan successfully used eight machine learning models (namely, Naïve Bayes, Bayesian network, k-nearest neighbour, artificial neural networks, support vector machines and decision trees such as classification and regression tree [CART or CRT], chi-square automatic interaction detector [CHAID] and C4.5) to predict company failure and focused on companies on the Istanbul Stock Exchange. Accuracies of 87.2% – 96.6%, depending on the model used, were reported 1-year preceding failure – with the CRT being the best performer. Tsai (2014) combined multiple machine learning algorithms (also referred to as classifier ensembles) using companies from Australia, Germany and Japan and reported that neural network ensembles resulted in the best predictive accuracies.

Empirical evidence on the failure prediction accuracy of machine learning methods as opposed to traditional statistical methods is, however, mixed. Several studies reported that machine learning is superior to traditional statistical methods (Barboza et al. 2017; Huang et al. 2008; Nazareth & Reddy 2023; Perboli & Arabnezhad 2021), while others have reported that traditional statistical methods outperform machine learning methods (Le & Viviani 2018; Sermpinis et al. 2023). Furthermore, Sermpinis et al. (2021) pointed out that not all machine learning algorithms are suited for failure prediction.

In a South African context, Steyn-Bruwer and Hamman (2006) reported relatively low predictive accuracies, ranging from 65.7% to 71.2%, when predicting company failure up to 4 years in advance using the CRT decision tree algorithm. Muller et al. (2009) compared two traditional statistical methods (multiple discriminant analysis and logit analysis) with two machine learning models (decision trees and neural networks). They achieved a predictive accuracy of 81.9%, 85.4% and 85.5% for 1, 2 and 3 years before failure, respectively, using neural networks, and predictive accuracies of 72.5%, 77.1% and 78.6% for 1, 2 and 3 years before failure, respectively, in respect of decision trees. Logit analysis outperformed decision trees (with predictive accuracies of 78.9%, 83.0% and 84.7% for 1, 2 and 3 years before failure, respectively), and multiple discriminant analysis was the worst performer.

Overview of machine learning algorithms

This section gives a broad overview of the fundamental machine learning algorithms (also referred to as classifiers) applied in the company failure prediction study by Aktan (2011), as well as in the present study.

The Naïve Bayes classifier calculates the likelihood of a data point belonging to a class by evaluating each feature independently and determining its probability of association with the classification class (Aktan 2011; Sermpinis et al. 2023). In contrast, a Bayesian network represents a probability distribution as a directed acyclic graph (DAG), where each node is a feature, and directed edges indicate dependencies between features without cycles (Aktan 2011).

The k-nearest neighbour algorithm classifies a data point by its proximity to training points in the feature space, assuming that data points of the same class are close together and basing the classification on the classes of the nearest neighbours (Aktan 2011; Sermpinis et al. 2023), while artificial neural networks are inspired by biological neural networks (Aktan 2011; Barboza et al. 2017; Le & Viviani 2018). A neural network consists of an input layer, several hidden layers and an output layer, with neurons in each layer receiving weighted inputs from the previous layer, processing them and passing them forward (Le & Viviani 2018; Shetty et al. 2022). By adjusting these weights, neural networks can learn complex patterns of classification.

Support vector machines are powerful classifiers that project training data points into a hyperspace using a kernel function and then construct a hyperplane by maximising the distance to the nearest data points (Aktan 2011; Barboza et al. 2017; Le & Viviani 2018; Sermpinis et al. 2023; Shetty et al. 2022). This hyperplane serves as the decision boundary for classifying new data points (Shetty et al. 2022).

Lastly, decision trees such as CRT, CHAID and C4.5 use a divide and conquer strategy to classify a data point. Each node in the tree represents a test on a specific feature. The classification process starts at the root and recursively evaluates each decision node until a leaf node is reached that indicates the class of the data point (Aktan 2011).

Research methods and design

Research design

The research methodology process can be described as consisting of different layers, namely, research philosophy, methodological choice, strategy choice, time horizons, and techniques and procedures (Saunders, Lewis & Thornhill 2019). The present study was based in the positivistic paradigm on the ontological foundation that the world is viewed objectively – detached from the values and beliefs of the researcher (Saunders et al. 2019). The study followed a quantitative methodology to perform archival research by analysing secondary data over a research time horizon that was longitudinal in nature. The sample, models and variables applied are elaborated on in the sections that follow.

Sampling

This study used a data set that consisted of 56 companies listed on the Johannesburg Stock Exchange (JSE). Of these, 28 companies represented failed companies that were selected from a population of 252 delisted companies. This was done by searching the Stock Exchange News Service (SENS) for companies that met the following criteria:

- The company delisted during the period 2010–2021.
- The company was undergoing or had undergone business rescue, liquidation or suspension from the JSE for non-compliance.
- The company had at least 4 years of financial information.

The 28 failed companies were matched with 28 non-failed companies based on size and industry classification. All JSE-listed companies that did not meet the criteria of a failed company during the period 2010–2021 were termed 'non-failed' for the purpose of this study.

Models applied

The statistical model used as a baseline in this study is based on the logit analysis model employed by Rodríguez-Masero and López-Manjón (2020). The adapted logit analysis model is (Equation 1):

$$P\left(X\right) = \frac{1}{1 + e^{-\left(b_0 + b_i X_i\right)}} \quad \text{[Eqn 1]}$$

where X_i is an independent variable as defined in Table 1, b_0 is a constant and b_i is a coefficient of each independent variable. The model is evaluated as follows:

- Compute the independent variables, X_i, and multiply each independent variable, X_i, with their corresponding coefficient, b_i;
- Calculate a new variable $y = b_0 + b_i X_i$;
- Calculate the probability of failure for a company as $P(X) = 1/(1+e^{-y})$ and
- Classify the company as failed ($P[X] \leq 0.5$) or non-failed ($P[X] > 0.5$).

The predictive machine learning models used in this study are based on the machine learning models as employed by Aktan (2011), namely the Naïve Bayes, Bayesian network, k-nearest neighbour (k-NN), artificial neural network (ANN) and support vector machine (SVM) models, as well as the CRT, CHAID and C4.5 decision tree models. This Aktan study was replicated because of its inclusion of a broad range of fundamental machine learning algorithms, which may serve as a benchmark for predictive accuracy in the South African context before progressing to more sophisticated ensemble methods.

In line with Aktan (2011), the WEKA software platform was used to train most predictive machine learning models with only the CRT and CHAID decision tree models using the SPSS 15 software by IBM. Given the limited size of the data set, as was also the case with Aktan, the authors preferred to use all data for training and validation. This was done by using the tenfold cross-validation process that further assists in preventing overfitting (Aktan 2011).

Most studies consider three metrics when evaluating a model (Aktan 2011; Tsai 2014), namely predictive accuracy, Type I errors and Type II errors. Predictive accuracy shows how well the model correctly classifies both failed and non-failed companies (Tsai 2014). A Type I error occurs when a company is incorrectly classified as non-failed, while a Type II error occurs when a company is incorrectly classified as failed (Aktan 2011).

Aktan (2011) argued that Type I and Type II errors are not equally important and that a Type I error is more severe. As a result, Aktan introduced a cost matrix that assigns a weight to a Type I error that is 10 times greater than that assigned to a Type II error.

Variables applied

This study used two different sets of variables to evaluate the predictive accuracy of machine learning models in a South African context. The first set contains the 10 variables from the original study by Aktan (2011), which included two cash flow variables: the cash to total asset ratio and the cash flow to total asset ratio. This set was used to evaluate the predictive accuracy of Aktan's models in South Africa and to compare the results with the original study.

For the second variable set, this study applied the four variables of Rodríguez-Masero and López-Manjón (2020) but replaced their current assets to current liabilities ratio with an operating cash flow to current liabilities ratio and added a variable that represents the size of a company (i.e. the log of the square root of total assets). These adjustments to the variables applied by Rodríguez-Masero and López-Manjón are based on empirical evidence of the superiority of the operating cash flow to current liabilities ratio when compared with the current assets to current liabilities ratio (Das 2018) and the size of a company significantly affecting company failure prediction (Ohlson 1980). Furthermore, initial statistical tests in the present study indicated that the predictive performance of the logit analysis model improved in a South African context when applying these adjustments. This second variable set was applied to both Aktan's (2011) machine learning models and the logit analysis method, allowing for a comparison between machine learning and traditional statistical approaches.

The two different sets of independent variables applied data lines from the IRESS Expert database (Table 1).

The variables recorded 1, 2 and 3 years prior to failure were used for companies that failed. Conversely, the variables from 1, 2 and 3 years preceding the most recent data in the data set were employed for companies that remained operational without signs of failure. This approach ensured that the variables corresponded with companies that had remained operational throughout the specified number of subsequent years.

The dependent variable in all the models was a dichotomous variable that could take on the value of either failed (=1) or non-failed (=0) companies.

Ethical considerations

The authors submitted an application for full ethical exemption for two projects to the Stellenbosch Business School Research Ethics Committee, which is a sub-committee of the Social, Behavioural and Education Research Ethics Committee (REC:SBE) of Stellenbosch University. Ethic waivers were granted for the projects with the waiver numbers 25574 and 28764. These waivers have been granted as the projects did not involve direct contact with human subjects. Also, the projects adhered to the ethical standards of the institutional and/or national research committee in all procedures for studies without contact with human subjects.

TABLE 1: Independent variables used in models.

Variable number	Variable description	Aktan (2011) variable	Adjusted Rodríguez-Masero and López-Manjón (2020) variable
X1	Operating income to total assets ratio	✓	✓
X2	Financial debts to total assets ratio	✓	-
X3	Cash to total assets ratio	✓	-
X4	Quick assets to total assets ratio	✓	-
X5	Inventory to net sales ratio	✓	-
X6	Current assets to total assets ratio	✓	-
X7	Total debts to total assets ratio	✓	-
X8	Short-term debts to total assets ratio	✓	-
X9	Net income to total assets ratio	✓	-
X10	Cash flow to total assets ratio	✓	-
X11	Total liabilities to total assets ratio	-	✓
X12	Operating cash flow to total liabilities ratio	-	✓
X13	Operating cash flow to current liabilities ratio	-	✓
X14	Log of the square root of total assets	-	✓

Source: Authors' own compilation from Aktan, S., 2011, 'Application of machine learning algorithms for business failure prediction', *Investment Management and Financial Innovations* 8(2), 52–65; Rodríguez-Masero, N. & López-Manjón, J.D., 2020, 'The usefulness of operating cash flow for predicting business bankruptcy in medium-sized firms', *Revista Brasileira de Gestao de Negocios* 22(4), 917–931. https://doi.org/10.7819/rbgn.v22i4.4079

Results

Three analyses were performed to achieve the aim of the study, namely, to assess the predictive accuracy of machine learning models. Firstly, the Aktan (2011) study was replicated in the South African context by applying eight machine learning algorithms based on the 10 variables of Aktan. Secondly, the eight machine learning algorithms of Aktan, based on five variables adapted from the Rodríguez-Masero and López-Manjón (2020) study, were applied. Thirdly, a logit analysis traditional method, based on the mentioned five variables, was applied. The logit analysis method represented the traditional statistical method for the purpose of comparison with the machine learning methods. In the reported results, the logit analysis method is referred to as the baseline result, and the five variables adapted from Rodríguez-Masero and López-Manjón are referred to as the logit analysis model variables.

The results of the three analyses applied are discussed below in respect of the predictive accuracy for 1, 2 and 3 years preceding company failure.

Predictive accuracy 1 year before company failure

On comparing predictive accuracy when using Aktan's (2011) algorithms and variables on JSE-listed companies 1 year before failure with the results of the original Aktan study, it is immediately evident from Figure 1 that the CRT algorithm had 100% accuracy with no error. This therefore indicates an overfit, which is of no practical use (Aktan 2011). Ignoring the overfitted CRT algorithm result, the CHAID algorithm performed best and even slightly better than in the original Aktan study with an accuracy of 94.64% compared with 92.2%. The Bayesian network algorithm also performed well, although not as well as in the original Aktan study, with an accuracy of 85.71% compared with 91.1%. All the other algorithms performed more poorly in the South African context than in the original Aktan study.

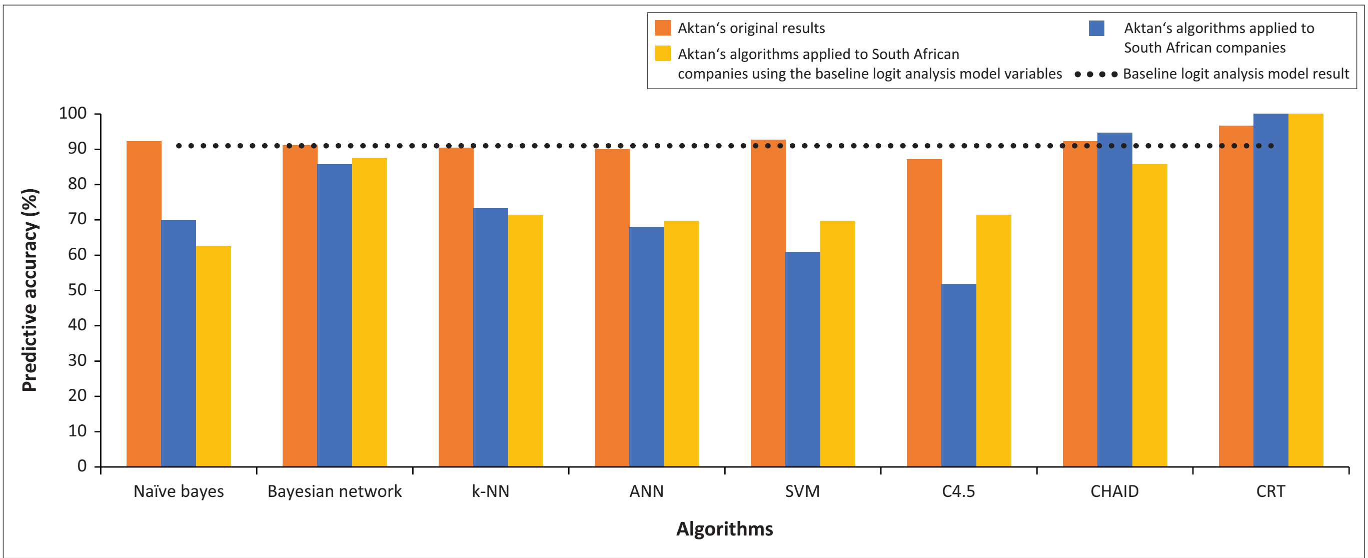
The results of predictive accuracy when using Aktan's (2011) algorithms and the logit analysis model variables 1 year before failure again showed that the CRT algorithm is overfitted (with 100% accuracy with no error). Other than the overfitted CRT algorithm, the rest of the machine learning algorithms performed worse than the baseline logit analysis at 91%, with only Bayesian network and CHAID algorithms coming close at 87.50% and 85.71%, respectively.

The baseline logit analysis performed very well (at 91% predictive accuracy) in the South African context with results mostly on a par with Aktan's (2011) original results. However, when applying the Aktan algorithms in the South African context, the CHAID algorithm performed better than the baseline logit analysis. Furthermore, although certain machine learning results improved when applying the logit analysis model variables when compared with Aktan's algorithms in a South African context, none of these improved results reached the predictive accuracy levels of the baseline logit analysis.

Predictive accuracy 2 years before company failure

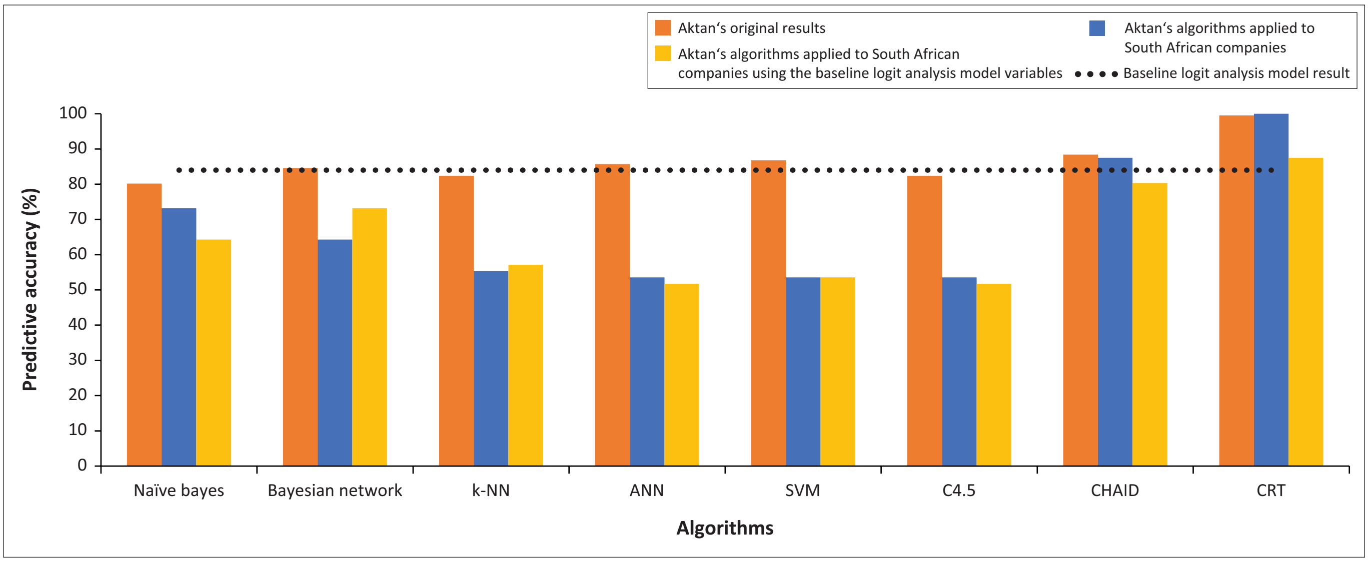
When comparing the predictive accuracy of using Aktan's (2011) algorithms and variables on JSE-listed companies 2 years before failure with the results of the original Aktan study, Figure 2 shows that the CRT algorithm is overfitted (100% accuracy with no error). Ignoring the result of the overfitted CRT algorithm, the CHAID algorithm performed best with an accuracy of 87.5%, which is close to the accuracy of 88.4% of the original Aktan study. In this instance, the Naïve Bayes algorithm performed reasonably well with an accuracy of 73.21% although lower than the 80.2% of the original Aktan study.

The predictive accuracy gained when using Aktan's (2011) algorithms and the logit analysis model variables showed that the 87.5% accuracy of the CRT algorithm outperformed



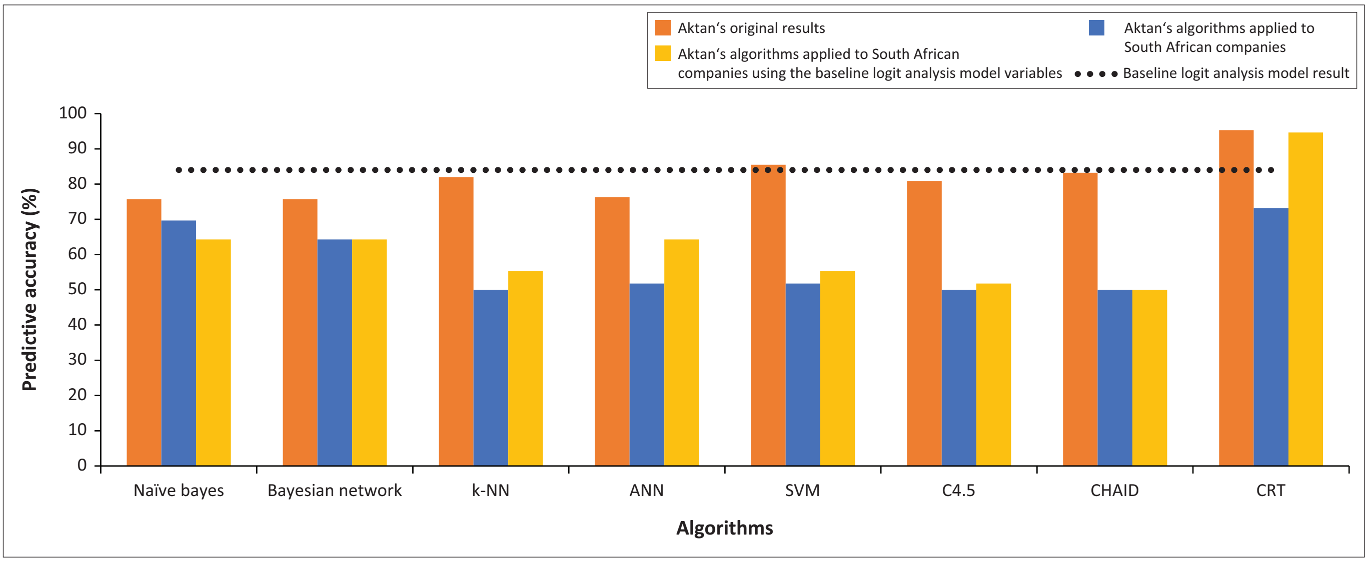
k-NN, k-nearest neighbour; ANN, artificial neural network; SVM, support vector machines; CHAID, chi-square automatic interaction detector, CRT, classification and regression tree.

FIGURE 1: Comparison of the machine learning and logit analysis baseline results for 1 year before company failure.



k-NN, k-nearest neighbour; ANN, artificial neural network; SVM, support vector machines; CHAID, chi-square automatic interaction detector, CRT, classification and regression tree.

FIGURE 2: Comparison of the machine learning and logit analysis baseline results for 2 years before company failure.



k-NN, k-nearest neighbour; ANN, artificial neural network; SVM, support vector machines; CHAID, chi-square automatic interaction detector, CRT, classification and regression tree.

FIGURE 3: Comparison of the machine learning and logit analysis baseline results for 3 years before company failure.

the 84% accuracy of the baseline logit analysis in predicting company failure 2 years in advance. The CHAID algorithm also did well with an accuracy of 80.36%, although this was lower than the accuracy level of the baseline logit analysis model. The accuracy of the other machine learning algorithms was noticeably inferior to that of the baseline logit analysis.

The baseline logit analysis performed relatively well (at 84% predictive accuracy) in the South African context with results mostly on a par with Aktan's (2011) original results. However, when applying the Aktan algorithms in the South African context, the CHAID algorithm performed better than the baseline logit analysis. Furthermore, although most machine learning results deteriorated slightly when applying the logit analysis model variables when compared with Aktan's algorithms in a South African context, the CRT algorithm outperformed the baseline logit analysis.

Predictive accuracy 3 years before company failure

When comparing the predictive accuracy results of Aktan's (2011) algorithms and variables on South African companies 3 years before failure with the results of the original Aktan study (Figure 3), it is clear that the C4.5 and CHAID algorithms had accuracies of exactly 50% (reflecting the proportion of the failed to non-failed companies in the data set) and a Type II error of 100%. This therefore suggests a malfunction of the algorithm (Aktan 2011). In this case, the algorithm indiscriminately classified all companies as failed to mitigate Type I errors, as they carried a higher weight in the cost matrix. Here the CRT algorithm performed best with an accuracy of 73.21% compared with Aktan's result of 95.3%. This was followed by Naïve Bayes and Bayesian networks at 69.64% and 64.29%, respectively. All the other algorithms performed noticeably worse than they did in the original Aktan study.

The predictive accuracy of using Aktan's (2011) algorithms and the logit analysis model variables shows a malfunction of the CHAID algorithm – with an accuracy of exactly 50%, which reflects the ratio of the failed to non-failed companies in the data set (Aktan 2011). Other than the CRT algorithm with an accuracy of 94.64%, the accuracy of the rest of the machine learning algorithms was inferior to the baseline logit analysis at 84%.

The baseline logit analysis performed relatively well (at 84% predictive accuracy) in the South African context with results mostly on a par with Aktan's (2011) original results. Furthermore, the machine learning results improved when applying the logit analysis model variables compared with Aktan's algorithms in a South African context, and the CRT also outperformed the predictive accuracy levels of the baseline logit analysis.

Although Type I and II errors were not reported separately, most machine learning algorithms exhibited lower Type I error rates than Type II error rates, and the baseline logit analysis model had higher Type I error rates compared to Type II error rates. The reported error rates for Type I and II of the machine learning algorithms were affected by the cost matrix defined in Atkan's (2011) methodology. Therefore, these reported error rates were not regarded as a comparable metric for assessing predictive accuracy of the machine learning versus traditional statistical methods in the present study.

Discussion

In summary, the empirical performance of machine learning algorithms compared with traditional statistical methods was not as accurate as anticipated. Comparing the results of Aktan's (2011) algorithms in the South African context with the original Aktan study, most of the machine learning algorithms performed noticeably worse for 1, 2 and 3 years prior to failure. The main exceptions to the outperformance of the original Aktan study are the following:

- In predicting company failure 1 year in advance using Aktan's (2011) original variables, the accuracy of the CHAID decision tree algorithm was greater in the South African context than in Aktan's original study, with an accuracy of 94.64% compared with Aktan's 92.2% (Figure 1).
- In predicting company failure 2 years in advance using the original variables of Aktan (2011), the CHAID decision tree algorithm performed similarly to Aktan's original study with an accuracy of 87.5% compared with Aktan's 88.4% (Figure 2).

When comparing the results of the machine learning algorithms, i.e. the replication of the Aktan (2011) study and the application of the adapted variables from Rodríguez-Masero and López-Manjón (2020) and the traditional logit analysis model, the logit analysis model consistently outperformed the machine learning algorithms across all forecasting horizons except the following:

- In predicting company failure 1 year in advance, the CHAID decision tree algorithm showed a predictive accuracy of 94.64% compared with the logit analysis model's 91% (Figure 1).
- In predicting company failure 2 years in advance, the CHAID decision tree algorithm and the CRT decision tree algorithm (both showing a predictive accuracy of 87.5%) outperformed the logit analysis model's 84.0% (Figure 2).
- In predicting company failure 3 years in advance, the CRT decision tree algorithm showed a predictive accuracy of 94.64% compared with the logit analysis model's 84.0% (Figure 3).

Reasons for the difference in results between the original Atkan (2011) study and the present study may include the different contexts (Türkiye versus South Africa) and the periods covered. Furthermore, Steyn-Bruwer and Hamman (2006) argued, in their study on the performance of decision trees in South Africa, that the lower predictive accuracy of this model compared with global studies can be attributed to their broader definition of company failure. This broader definition of company failure, also used by Muller et al. (2009), Balcaen and Ooghe (2006) and the present study, classified companies as failed when they cannot continue in their current form. This therefore includes liquidation, delisting, bankruptcy and restructuring (Muller et al. 2009), resulting in failed and non-failed companies not being two mutually exclusive groups with a clear-cut off line (Balcaen & Ooghe 2006; Cybinski 2001). In contrast, Aktan used a much stricter definition of company failure wherein companies are classified as failed if they are bankrupt or have lost two-thirds of their capital stock (as dictated by the Turkish bankruptcy law). This more rigorous classification scheme may account for the comparatively inferior performance of Aktan's algorithms when applied in the South African context.

This study has shown that, although machine learning is the modern approach to predicting company failure, it is not necessarily more accurate than traditional statistical methods. In this study, the logit analysis was generally more accurate than the machine learning algorithms in South Africa with accuracies that were on a par with Aktan's (2011) original results. This aligns with the view of Sermpinis et al. (2023) that not all machine learning models are suited for company failure prediction. Indeed, Kim et al. (2020), while not dismissing machine learning, recommended that all models be understood and that the most appropriate methodology be selected for the task at hand.

The black box nature of machine learning algorithms renders the technology opaque. This advanced technology does not provide information to management on the types of improvements required (Clement 2020; Kim et al. 2020; Li et al. 2017) and does not explain the underlying reasons for the suboptimal performance of machine learning algorithms.

Potential errors and data integrity

This study aimed to replicate the work of Aktan (2011) as reliably as possible, but the authors acknowledge the limitations and potential errors as set out in this section.

The study uses a newer version of the WEKA software platform than was used by Aktan (2011). The software may contain either new bugs or bug fixes that were not present in the version used by Aktan.

Aktan (2011) omitted to specify how the financial information of non-failed companies was used, particularly if the financials of a specific year were chosen or if an average over the period under consideration was used. This study followed the approach of employing the most recent data in the data set for non-failed firms, which may differ from the data used by Aktan.

This study relied on the accuracy of the data obtained from the IRESS Expert database and steps were taken to ensure the integrity of the data. A sample of the data set was manually verified by comparing the values from the IRESS Expert database with the financial statements published on the companies' websites. The whole data set was programmatically analysed for any omissions or formatting issues (e.g. where the data were in a format that could not be interpreted as a number), and these were manually corrected from the financial statements published on the companies' websites. The data of each company were analysed to identify outliers, and these were manually verified and corrected from the financial statements published on the companies' websites.

This study furthermore processed some of the data from the IRESS Expert database and, in so doing, calculation errors may have occurred. Steps were taken to ensure the integrity of the processed data including manually verifying a sample of the calculations. The calculated values were also analysed on a per company basis to identify any outliers and these were manually verified.

Limitations and recommendations

The primary limitation of the study was the number of variables used. While the study replicated the work of Aktan (2011) and used the same set of variables along with a second data set adapted from Rodríguez-Masero and López-Manjón (2020), it should be noted that Aktan selected only 10 out of the 53 potential variables he had initially identified. However, Liang, Tsai and Wu (2015) highlighted that manual variable selection does not always improve machine learning models' accuracies, and these models, especially decision trees, automatically determine the most important variables to use. Le and Viviani (2018) and Clement (2020) highlight this automatic selection of the relevant variables from the data set as the biggest advantage of machine learning. By restricting the variables used in this study to the ones used by Aktan and an adaption from those used by Rodríguez-Masero and López-Manjón – and not using all the information in the available data set – this automatic variable selection advantage has been compromised. It is recommended that this study be replicated with a larger set of variables, incorporating all relevant ratios and values from the available data set. This approach would enable the machine learning algorithms to autonomously select the most important variables, potentially enhancing the overall accuracy of the models.

Another potential limitation of this study was the small data set. However, Aktan (2011) also had a small data set. This study employed tenfold cross-validation to address this limitation and to avoid overfitting as occurred in Aktan's study.

This study chose to replicate the Aktan (2011) study because it had included a broad range of fundamental machine learning algorithms that could serve as a benchmark for predictive accuracy in the South African context. Future studies should build on these results by incorporating more sophisticated ensemble methods.

Conclusion

The use of machine learning models to predict company failure is regarded as a modern and efficient approach surpassing traditional statistical models. Empirical evidence of the predictive accuracy of machine learning models is, however, mixed – while evidence from emerging markets is sparse. This study aimed to test the accuracy of company failure prediction by applying an array of fundamental machine learning algorithms in the emerging market context of South Africa.

To achieve the aim of the study, eight fundamental machine learning algorithms, in line with the methodology of Aktan (2011), were used to assess the accuracy of machine learning in predicting company failure. Furthermore, the accuracy of machine learning compared with traditional statistical methods was tested by applying the Aktan machine learning algorithms and the logit analysis traditional statistical model. In this comparison, a set of variables was applied, which represents more recent empirical evidence on the significant determinants of company failure prediction (Rodríguez-Masero & López-Manjón 2020).

The results showed that not all machine learning algorithms are suited to predicting company failure. Furthermore, it was found that modern machine learning methods were not necessarily more accurate than traditional statistical methods in predicting company failure. The logit analysis model outperformed all machine learning models except two decision tree algorithms (the CHAID 1 and 2 years before company failure, and the CRT 2 and 3 years before company failure). The results further indicate that the algorithms and variables applied in company failure prediction studies may be country specific, and that the definition of company failure (strict or broad) may affect predictive accuracy.

Given the significant social and economic impact of company failures, research into accurately predicting company failure remains an important research avenue and calls for an increased understanding of the models that can be used to perform these predictions.

Acknowledgements

This article is partially based on D.M.'s dissertation entitled 'The cash flow statement as a predictor of company failure: A machine learning approach' towards the degree of Master of Business Administration at the Stellenbosch University in March 2024, with supervisor Prof Nicolene Wesson; and A.M.'s dissertation entitled 'Cash flow as a predictor of company failure' towards the degree of Master of Business Administration at the Stellenbosch University in March 2023, with supervisor Prof Nicolene Wesson.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Authors' contributions

N.W. conceptualised and supervised the project and was responsible for editing and reviewing. A.M. was responsible for obtaining and validating company data, for managing the investigations and performing calculations related to the statistical prediction methods. A.M. was also responsible for reviewing. D.M. took charge of validating company data as well as the investigations and calculations related to the machine learning methods employed. D.M. was also responsible for the original draft.

Funding information

This research received no grants from funding agencies in the public, commercial or not-for-profit sectors.

Data availability

Data supporting this study and findings are available on reasonable request from the corresponding author, N.W.

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency or that of the publisher. The authors are responsible for this article's results, findings and content.

References

Aktan, S., 2011, 'Application of machine learning algorithms for business failure prediction', *Investment Management and Financial Innovations* 8(2), 52–65.

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

Alzubi, J., Nayyar, A. & Kumar, A., 2018, 'Machine learning from theory to algorithms: An overview', *Journal of Physics. Conference series* 1142(1), 012012. https://doi.org/10.1088/1742-6596/1142/1/012012

Aziz, A.M. & Humayon, A.D., 2006, 'Predicting corporate bankruptcy: Where we stand', *Journal of Corporate Governance* 6(1), 18–33. https://doi.org/10.1108/14720700610649436

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

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

Beaver, W.H., 1966, 'Financial ratios as predictors of failure', *Journal of Accounting Research* 4, 71–111. https://doi.org/10.2307/2490171

Bellovary, J.L., Giacomino, D.E. & Akers, M.D., 2007, 'A review of bankruptcy prediction studies: 1930 to present', *Journal of Financial Education* 33(4), 1–42.

Bhandari, S., Showers, V. & Johnson-Snyder, A.J., 2019, 'A comparison: Accrual versus cash flow based financial measures' performance in predicting business failure', *Journal of Accounting and Finance* 19(6), 11–25. https://doi.org/10.33423/jaf.v19i6.2313

Bisogno, M., Restaino, M. & Carlo, A., 2018, 'Forecasting and preventing bankruptcy: A conceptual review', *African Journal of Business Management* 12(9), 231–242. https://doi.org/10.5897/AJBM2018.8503

Clement, C., 2020, 'Machine learning in bankruptcy prediction – A review', *Journal of Public Administration, Finance and Law* 17, 178–197.

Cybinski, P., 2001, 'Description, explanation, prediction – The evolution of bankruptcy studies?', *Managerial Finance* 27(4), 29–44. https://doi.org/10.1108/03074350110767123

Das, S., 2018, 'Analysis of cash flow ratios: A study on CMC', *Accounting* 4(1), 41–52. https://doi.org/10.5267/j.ac.2017.3.001

Ding, S., Cui, T., Bellotti, A.G., Abedin, M.Z. & Lucey, B., 2023, 'The role of feature importance in predicting corporate financial distress in pre and post COVID periods: Evidence from China', *International Review of Financial Analysis* 90, 102851. https://doi.org/10.1016/j.irfa.2023.102851

Gajdošíková, D. & Valášková, K., 2023, 'Bankruptcy prediction model development and its implications on financial performance in Slovakia', *Economics and Culture* 20(1), 30–42. https://doi.org/10.2478/jec-2023-0003

Huang, S.M., Tsai, C.F., Yen, D.C. & Cheng, Y.L., 2008, 'A hybrid financial analysis model for business failure prediction', *Expert Systems with Applications* 35(3), 1034–1040. https://doi.org/10.1016/j.eswa.2007.08.040

Huang, Y.P. & Yen, M.F., 2019, 'A new perspective of performance comparison among machine learning algorithms for financial distress prediction', *Applied Soft Computing Journal* 83, 105663. https://doi.org/10.1016/j.asoc.2019.105663

Jones, S., 2016, 'A cash flow-based model of corporate bankruptcy in Australia', *Journal of Applied Management Accounting Research* 14(1), 23–37.

Kim, H., Cho, H. & Ryu, D., 2020, 'Corporate default predictions using machine learning: Literature review', *Sustainability (Switzerland)* 12(16), 6325. https://doi.org/10.3390/su12166325

Kristóf, T. & Virág, M., 2020, 'A comprehensive review of corporate bankruptcy prediction in Hungary', *Journal of Risk and Financial Management* 13(2), 1–20. https://doi.org/10.3390/jrfm13020035

Kuruppu, N., Laswad, F. & Oyelere, P., 2003, 'The efficacy of liquidation and bankruptcy prediction models for assessing going concern', *Managerial Auditing Journal* 18(6/7), 577–590. https://doi.org/10.1108/02686900310482713

Le, H.H. & Viviani, J.L., 2018, 'Predicting bank failure: An improvement by implementing a machine-learning approach to classical financial ratios', *Research in International Business and Finance* 44, 16–25. https://doi.org/10.1016/j.ribaf.2017.07.104

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

Liang, D., Tsai, C.F. & Wu, H.T., 2015, 'The effect of feature selection on financial distress prediction', *Knowledge-Based Systems* 73(1), 289–297. https://doi.org/10.1016/j.knosys.2014.10.010

Mullainathan, S. & Spiess, J., 2017, 'Machine learning: An applied econometric approach', *Journal of Economic Perspectives* 31(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Muller, G.H., Steyn-Bruwer, B.W. & Hamman, W.D., 2009, 'Predicting financial distress of companies listed on the JSE – A comparison of techniques', *South African Journal of Business Management* 40(1), 21–32. https://doi.org/10.4102/sajbm.v40i1.532

Naidoo, S.R. & Du Toit, G.S., 2007, 'A predictive model of the states of financial health in South African businesses', *Southern African Business Review* 11(3), 33–55.

Nazareth, N. & Reddy, Y.V., 2023, 'Financial applications of machine learning: A literature review', *Expert Systems with Applications* 219, 119640. https://doi.org/10.1016/j.eswa.2023.119640

Ohlson, J.A., 1980, 'Financial ratios and the probabilistic prediction of bankruptcy', *Journal of Accounting Research* 18(1), 109–131. https://doi.org/10.2307/2490395

Oz, I.O. & Simga-Mugan, C., 2018, 'Bankruptcy prediction models' generalizability: Evidence from emerging market economies', *Advances in Accounting* 41, 114–125. https://doi.org/10.1016/j.adiac.2018.02.002

Perboli, G. & Arabnezhad, E., 2021, 'A machine learning based DSS for mid and long-term company crisis prediction', *Expert Systems with Applications* 174, 114758. https://doi.org/10.1016/j.eswa.2021.114758

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

Rodríguez-Masero, N. & López-Manjón, J.D., 2020, 'The usefulness of operating cash flow for predicting business bankruptcy in medium-sized firms', *Revista Brasileira de Gestao de Negocios* 22(4), 917–931. https://doi.org/10.7819/rbgn.v22i4.4079

Rujoub, M.A., Cook, D.M. & Hay, L.E., 1995, 'Using cash flow ratios to predict business failures', *Journal of Managerial Issues* 7(1), 75–90.

Saunders, M., Lewis, P. & Thornhill, A., 2019, *Research methods for business students*, 8th edn., Pearson, Harlow, UK.

Sermpinis, G., Tsoukas, S. & Zhang, Y., 2023, 'Modelling failure rates with machine-learning models: Evidence from a panel of UK firms', *European Financial Management: The Journal of the European Financial Management Association* 29(3), 734–763. https://doi.org/10.1111/eufm.12369

Sewpersadh, N.S., 2020, 'K-score categorisation of JSE listed sectors under the financial distress continuum theory: A quantitative approach', *Cogent Economics and Finance* 8(1), 1748969. https://doi.org/10.1080/23322039.2020.1748969

Sharma, D.S., 2001, 'The role of cash flow information in predicting corporate failure: The state of the literature', *Managerial Finance* 27(4), 3–28. https://doi.org/10.1108/03074350110767114

Sherman, H.D. & Young, S.D., 2016, 'Where financial reporting still falls short', *Harvard Business Review* 94(7–8), 76–84.

Shetty, S., Musa, M. & Brédart, X., 2022, 'Bankruptcy prediction using machine learning techniques', *Journal of Risk and Financial Management* 15(1), 35. https://doi.org/10.3390/jrfm15010035

Steyn-Bruwer, B.W. & Hamman, W.D., 2006, 'Company failure in South Africa: Classification and prediction by means of recursive partitioning', *South African Journal of Business Management* 37(4), 7–18. https://doi.org/10.4102/sajbm.v37i4.609

Tsai, C.F., 2014, 'Combining cluster analysis with classifier ensembles to predict financial distress', *Information Fusion* 16(1), 46–58. https://doi.org/10.1016/j.inffus.2011.12.001