Expert Systems With Applications 94 (2018) 164–184

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

Systematic review of bankruptcy prediction models: Towards a framework for tool selection

Hafiz A. Alaka[a], Lukumon O. Oyedele[b,*], Hakeem A. Owolabi[c], Vikas Kumar[d], Saheed O. Ajayi[e], Olugbenga O. Akinade[f], Muhammad Bilal[f]

[a] Faculty of Engineering, Environment and Computing, Coventry University, Coventry, United Kingdom
[b] Bristol Enterprise Research and Innovation Centre (BERIC), University of the West of England, Bristol, United Kingdom
[c] Department of International Strategy & Business, The University of Northampton, United Kingdom
[d] Bristol Enterprise Research and Innovation Centre (BERIC), University of the West of England, Bristol, United Kingdom
[e] School of Built Environment and Engineering, Leeds Becket University, Leeds, United Kingdom
[f] Bristol Enterprise Research and Innovation Centre (BERIC), University of the West of England, Bristol, United Kingdom

ARTICLE INFO

Article history:
Received 14 November 2016
Revised 15 October 2017
Accepted 15 October 2017
Available online 26 October 2017

Keywords:
Bankruptcy prediction tools
Financial ratios
Error types
Systematic review
Tool selection framework
Artificial intelligence tools
Statistical tools

ABSTRACT

The bankruptcy prediction research domain continues to evolve with many new different predictive models developed using various tools. Yet many of the tools are used with the wrong data conditions or for the wrong situation. Using the Web of Science, Business Source Complete and Engineering Village databases, a systematic review of 49 journal articles published between 2010 and 2015 was carried out. This review shows how eight popular and promising tools perform based on 13 key criteria within the bankruptcy prediction models research area. These tools include two statistical tools: multiple discriminant analysis and Logistic regression; and six artificial intelligence tools: artificial neural network, support vector machines, rough sets, case based reasoning, decision tree and genetic algorithm. The 13 criteria identified include accuracy, result transparency, fully deterministic output, data size capability, data dispersion, variable selection method required, variable types applicable, and more. Overall, it was found that no single tool is predominantly better than other tools in relation to the 13 identified criteria. A tabular and a diagrammatic framework are provided as guidelines for the selection of tools that best fit different situations. It is concluded that an overall better performance model can only be found by informed integration of tools to form a hybrid model. This paper contributes towards a thorough understanding of the features of the tools used to develop bankruptcy prediction models and their related shortcomings.

1. Introduction

The effect of high rate of business failure can be devastating to firm owner, partners, society and the country's economy at large (Alaka et al., 2015; Edum-Fotwe, Price, & Thorpe, 1996; Hafiz et al., 2015; Xu & Zhang, 2009). The consequent extensive research into developing bankruptcy prediction models (BPM) for firms is undoubtedly justified. The performance of such models is largely dependent on, among other factors, the choice of tool selected to build it. Apart from a few studies (e.g. Altman, 1968; Ohlson, 1980), tool selection in many BPM studies is not based on capabilities of the tool; rather it is either chosen based on popularity (e.g. Abidali & Harris, 1995; Koyuncugil and Ozgulbas, 2012; Langford, Iyagba, & Komba, 1993) or based on professional background (e.g. Altman, Marco, & Varetto, 1994; Beaver, McNichols, & Rhie, 2005; Hillegeist, Keating, Cram, & Lundstedt, 2004; Lin & Mcclean, 2001; Nasir, John, Bennett, Russell, & Patel, 2000). This is because there is no evaluation material which shows and compares the relative performance of major tools in relation to the many important criteria a BPM should satisfy. Such material can provide a guideline and subsequently aid an informed and justified tool selection for BPM developers.

Most prediction tools are either statistical or artificial intelligence (AI) based (Balcaen & Ooghe, 2006; Jo & Han, 1996). The most common statistical tool is the multiple discriminant analysis (MDA) which was first used by Altman (1968) to develop a BPM popularly known as Z model, based on Beaver's (1966) recommendation in his univariate work. MDA, normally used with financial ratios (quantitative variables), subsequently became popular with accounting and finance literature (Taffler, 1982) and many

* Corresponding author.
E-mail addresses: ac7485@coventry.ac.uk (H.A. Alaka), L.Oyedele@uwe.ac.uk (L.O. Oyedele), olugbenga.akinade@uwe.ac.uk (O.O. Akinade).

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

subsequent studies by finance professionals simply adopted MDA without considering the assumptions that are to be satisfied for MDA's model to be valid. This resulted in inappropriate application, causing developed models to be un-generalizable (Joy & Tollefson, 1975; Richardson & Davidson, 1984; Zavgren, 1985). Abidali and Harris (1995), for example, unscholarly employed A-score alongside Z-score (i.e. MDA) in order to involve qualitative managerial variables, alongside quantitative variables, in their analysis when logistic regression (LR) [or logit analysis] can handle both types of variables singularly.

AI tools are computer based techniques of which Artificial Neural Network (ANN or NN) is the most common for bankruptcy prediction (Aziz & Dar, 2006; Tseng & Hu, 2010). Simply because it is the most popular architecture, many studies arbitrarily employed the back-propagation algorithm of ANN for bankruptcy prediction (e.g. Boritz, Kennedy, & Albuquerque, 1995; Odom & Sharda, 1990; Tam & Kiang, 1992; Wilson & Sharda, 1994; among others) despite it having a number of relatively undesirable features which include computational intensity, absence of formal theory, "illogical network behaviour in response to different variations of the input values" etc. (Altman et al., 1994; Coats & Fant, 1993, p. 507; Zhang, Hu, Patuwo, & Indro, 1999). Further, Fletcher and Goss (1993) developed an ANN prediction model for a relatively small sample size when ANNs are known to need large samples for optimal performance (Boritz et al., 1995; Ravi Kumar & Ravi, 2007; Shin, Lee, & Kim, 2005).

These improper uses of tools regularly occur because there is no readily available evaluation material or guidelines which can help BPM developers identify which tool best suits what data/purpose/situation. As Chung, Tan, and Holdsworth (2008), p. 20) put it, "*given the variety of techniques now available for insolvency prediction, it is not only necessary to understand the uses and strengths of any prediction model, but to understand their limitations as well*". Hence to ensure a BPM performs well with regards to criteria of preference (e.g. accuracy, type I error, transparency, among others), a model developer has to understand the strength and limitations of the available tools/techniques. This will ensure that the right tool is employed for the right data characteristics, right situation and the right purpose. This study thus aims to develop a comprehensive evaluation framework for selection of BPM tools using a systematic and comprehensive review. The following objectives are needed to achieve this aim:

1. Presentation of an overview of the common tools used for bankruptcy prediction and identification of BPM studies that have used these tools
2. Identifying the key criteria BPMs need to satisfy and how each tool performs in relation to each criterion by analysing the systematic review

The scope of this study is limited to reviewing only popular and promising tools that have been employed for the development of BPMs in past studies since interest in them is high. This is because it is virtually impossible to review all the many tools that can be used for this purpose in this study. In total, two statistical and six AI tools were reviewed. The next section explains the systematic review methodology used in this study with all the inclusion and exclusion criteria. This is followed by a brief description of each of the eight tools. Section four presents the 13 identified key criteria used to assess the tools. Section five discusses the analysis and results of the review in form of tables and charts. Section six presents the proposed tabular and diagrammatic frameworks. This is followed up with a conclusion section.

2. Methodology

This study used a systematic review method to create a guideline for the selection of an appropriate tool for developing a bankruptcy prediction model (BPM). There are so many tools that can be used to develop a BPM that it is virtually impossible to review them all in one study. As a result, the two most popular statistical tools as noted by Balcaen and Ooghe (2006) in their comprehensive review of BPMs were reviewed: multiple discriminant analysis (MDA) and Logistic regression (LR). Also covered in this review are the most popular and promising artificial intelligence (AI) tools as advocated by Aziz and Dar (2006) in their comprehensive review, and Min, Lee, and Han (2006) among others: artificial neural network (ANN), support vector machines (SVM), rough sets (RS), case based reasoning (CBR), decision tree (DT) and genetic algorithm (GA). A process flow of the methodology is presented in Fig. 1.

Systematic review is a well-known method for producing valid and reliable knowledge as it minimizes bias hence its popularity in the all-important medical research world (Schlosser, 2007; Tranfield, Denyer, & Smart, 2003). The inclusion criteria for this study were carefully chosen to allow fair comparison and ensure adequate quality (Khan, Kunz, Kleijnen, & Antes, 2003). To improve validity of this study, only peer reviewed journal articles were considered since they are considered to be of high quality and their contribution considered as very valid (Schlosser, 2007).

Systematic review requires wide literature search (Smith, Devane, Begley, & Clarke, 2011) hence following Appiah, Chizema, and Arthur (2015) approach, which is the most recently published systematic review in the BPM research area, the following databases were considered: Google Scholar; Wiley Interscience; Science Direct; Web of Science UK (WoS); and Business Source Complete (BSC). However, a careful observation revealed Google scholar produced an almost endless result and did not have the required filters to make it very efficient hence it was removed as it was unmanageable. Further observation revealed that (WoS) and BSC contained all the journal articles provided in Wiley and Science Direct; this is probably because the latter two are publishers while the former two are databases with articles from various publishers including the latter two. To increase the width of the search, Engineering Village (EV) database was added to WoS and BSC databases to perform the final search. EV was chosen because articles from the engineering world usually deal with BPM tools comprehensively.

The initial searches in the three databases (WoS, BSC and EV) showed that studies tend to use bankruptcy, insolvency and financial distress as synonyms for failure of firms. A search framework which captured all these words was thus designed with the following defined string ("Forecasting" OR "Prediction" OR "Predicting") AND ("Bankruptcy" OR "Insolvency" OR "Distress" OR "Default" OR "Failure").

To ensure high consistency and repeatability of this study, and consequently reliability and quality (Stenbacka, 2001; Trochim & Donnelly, 2006), only studies that appeared in the three databases were used; this ensured the eradication of database bias (Schlosser, 2007). These databases contain studies from all over the world hence geographic bias was also eliminated. Balcaen and Ooghe (2006) in their comprehensive review of statistical tools in 2006 noted that AI tools, mainly ANN, were gradually becoming adopted in BPM studies. With new tools emerging all the time, a four-year advance from 2006, which would have seen more use of AI tools, is how a start year of 2010 was chosen for this study. The end year is the year this paper was written, 2015.

Generally, the topic of articles that emerge from the search looked okay to determine which ones were fit for this study. However, this was not the case for all articles. Where otherwise, arti-