Predicting Business Bankruptcy: A Comprehensive Case Study

Rui Sarmento, FEUP, University of Porto and Gallicyadas, Porto, Portugal
Luís Trigo, FLUP, University of Porto and Gallicyadas, Porto, Portugal
Liliana Fonseca, FEP, University of Porto, Porto, Portugal

ABSTRACT

Managers, investors, financial institutions and government agencies have a major concern on forecasting enterprise bankruptcy. It enables the sustainability assessment of critical suppliers and clients, as well as competitors and the business environment. Throughout the 20th and the 21st century, advances in statistics and computer science fields enabled the development of different trends in financial distress assessment that co-exist today. However, recent Data Mining (DM) techniques are regarded as being the most precise. IT expertise requirements in the constantly evolving DM field may have been a major obstacle to the adoption of these techniques by decision makers. Furthermore, DM software tools that are now widespread offer a broad spectrum of Artificial Intelligence algorithms and the most difficult task may be the decision of selecting the appropriate algorithm. Hence, the adoption of a good workflow method for data processing and analysis is critical for having fast and reliable results. This work presents an overview of the available bankruptcy techniques and provides a comprehensive case study exploring the latest Data Mining techniques.

KEYWORDS
Artificial Intelligence, Artificial Neural Networks, Bankruptcy, Data Mining, Financial Distress Assessment

INTRODUCTION

Accurately predicting the bankruptcy of a company may be impossible, given the market uncertainty and other adverse situations for all companies. However, it is possible to identify those that have a greater chance of not being able to survive the crisis in the near future and, therefore, take preventive measures. Bankruptcy prediction is an important area for decision making of investors, suppliers, clients and even the company itself.

Ratio analysis on bankruptcy prediction dates back to the 1930’s (Bellovary et al., 2007). Beaver (1966) focused on getting statistical evidence from individual ratios with Discriminant Analysis. Altman Z-score published in 1968 made this analysis multivariate and is still very popular today. In the 1980’s decade Artificial Intelligence algorithms started to appear, they grew with the ever increasing massive machine processing resources and now they overrule statistical methods in the literature about financial distress.

It should be noticed that for the decision maker it is important not only the result but also to understand the models implied. That is one of the reasons all these techniques still coexist today and are used depending on the aim and scale of the decision and the cost trade-offs of collecting and processing the data.
The contribution of this work is to present a structured prediction case study for the bankruptcy hypothesis. It traverses a comprehensive Knowledge Discovery in Databases (KDD) workflow to explore different classification algorithms with an illustrative case study.

The complete methodology presented in this document might be useful to a vast variety of decision actors as this is currently an area of Business Intelligence with little comprehensive methods publicly available. Thus, the document is intended to be technically complete and provide an easily repeatable workflow. It begins with a background section where influential historical milestones regarding business bankruptcy prediction are described. A brief introduction about classification algorithms is then made and is followed by a section relating KDD and bankruptcy prediction. Software applications and data selection are the starting point to explore the proposed method.

The initial step of the case study will address the preprocessing stage. The exploratory analysis of the dataset takes a statistical snapshot of all the variables in the model and precedes the application of various techniques to “clean” the data and select the most relevant attributes for the creation of classification models. The data is now ready to be processed by some classification algorithms.

The most difficult decision is usually the selection of the most appropriate algorithm. Thus, several experiments were performed to know which of the classifier algorithms have the best predicting test values. Then, we benchmark the performance of Decision Trees, our best classifier, with the most widely used algorithm today, Artificial Neural Networks.

**RELATED WORK**

In this section, we provide related background to the business bankruptcy prediction task. Additionally, we provide an introduction to classification algorithms data analysts frequently use in prediction tasks. Some of these algorithms will be used in our case study.

Empirical models to predict corporate bankruptcy and bankruptcy theories have been different strands of research. However different paths have a substantial amount of overlap (Scott, 1981).

The literature in the field dates back to the 1930’s with the analysis of single financial ratios for specific purposes and industry (Bellovary et al., 2007). With no advanced statistical methods, analysts only compared failed and non-failed companies and noticed that failed companies had worst performance ratios.

Beaver (1966) introduced a statistical perspective in univariate ratio analysis. From the 30 selected ratios, only six were significant:

- Cash Flow / Total Debt
- Net Income / Total Assets
- (Current + Long-term Liabilities) / Total Assets
- Working Capital / Total Assets
- Current Ratio
- No-credit Interval

The most significant measure was the Cash Flow to Total Debt, and its intuition is that the net income should suffice to pay the total debt. Reduced net income or total debt expansion would result on increasing financial distress.

Altman (1968) went a step further and introduced Multivariate Discriminant Analysis (MDA) to find the linear function of financial ratios that discriminates failed from non-bankrupted companies.
Experience and History: An Experimental Approach to Generational Heterogeneity
[www.igi-global.com/article/experience-and-history/166567?camid=4v1a](www.igi-global.com/article/experience-and-history/166567?camid=4v1a)

Lifestyle Diglossia and Mobile: Ethnography of Multilingual Interaction
[www.igi-global.com/chapter/lifestyle-diglossia-and-mobile/156991?camid=4v1a](www.igi-global.com/chapter/lifestyle-diglossia-and-mobile/156991?camid=4v1a)