Chapter 11

A Comprehensive Workflow for Enhancing Business Bankruptcy Prediction

Rui Sarmento
LIAAD-INESC TEC, Portugal

Luís Trigo
LIAAD-INESC TEC, Portugal

Liliana Fonseca
University of Porto, Portugal

ABSTRACT

Forecasting enterprise bankruptcy is a critical area for Business Intelligence. It is a major concern for investors and credit institutions on risk analysis. It may also enable the sustainability assessment of critical suppliers and clients, as well as competitors and the business environment. Data Mining may deliver a faster and more precise insight about this issue. Widespread software tools offer a broad spectrum of Artificial Intelligence algorithms and the most difficult task may be the decision of selecting that algorithm. Trying to find an answer for this decision in the relatively large amount of available literature in this area with so many options, advantages, and pitfalls may be as informative as distracting. In this chapter, the authors present an empirical study with a comprehensive Knowledge Discovery and Data Mining (KDD) workflow. The proposed classifier selection automation selects an algorithm that has better prediction performance than the most widely documented in the literature.

INTRODUCTION

Accurately predicting the bankruptcy of a company may be impossible, given the market uncertainty and other adverse situations to which all companies are subjected. 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. That is why integrating the

latest Artificial Intelligence trends in an easy to use decision support tool for this area can further empower a Business Intelligence (BI) system. Hence, Michalewicz, Schmid, Michalewicz and Chiriac (2006) expanded the traditional approach to BI systems introducing Artificial Intelligence in its scope. They presented the Adaptive Business Intelligence concept exploring self-learning adaptive systems that learn with previous decisions and recommend the best actions.

Ratio analysis on bankruptcy prediction dates back to the 1930's (Bellovary, Giacomino & Akers, 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 overrun statistical methods in literature about financial distress.

The contribution of this work is to present a structured prediction method/workflow for the bankruptcy hypothesis. It traverses a comprehensive Knowledge Discovery and Data Mining (KDD) workflow to explore different classification algorithms.

The complete methodology presented in this chapter might be useful to a vast variety of decision actors as this is currently an area of BI with little comprehensive methods publicly available. Thus, the chapter 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 with the explained bridge between KDD and BI and ultimately the connection between KDD and bankruptcy prediction. Software applications and data selection is the starting point to explore the proposed method.

The initial step of the workflow 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 classifiers have the best predicting test values. Then, we benchmark our best classifier J48 performance with the most widely used algorithm today, Artificial Neural Networks, and get better results.

Towards the end of the process, the best of those classifiers should be selected to forecast the value of the target class of the test data set. Finally, a discussion and hypothetical future research paths section is also addressed.

BACKGROUND

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 ratio performances.

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

Related Content

Change Detection in Large Evolving Networks
Josephine M. Namayanjaand Vandana P. Janeja (2019). *International Journal of Data Warehousing and Mining (pp. 62-79).*
www.irma-international.org/article/change-detection-in-large-evolving-networks/225807

Constructing Galois Lattices as a Commonsense Reasoning Process
Xenia Naidenova (2013). *Diagnostic Test Approaches to Machine Learning and Commonsense Reasoning Systems (pp. 34-70).*
www.irma-international.org/chapter/constructing-galois-lattices-commonsense-reasoning/69404

A Fuzzy Portfolio Model With Cardinality Constraints Based on Differential Evolution Algorithms
JianDong He (2024). *International Journal of Data Warehousing and Mining (pp. 1-14).*
www.irma-international.org/article/a-fuzzy-portfolio-model-with-cardinality-constraints-based-on-differential-evolution-algorithms/341268

A Deep Neural Network Model for Cross-Domain Sentiment Analysis
Suman Kumari, Basant Agarwaland Mamta Mittal (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines (pp. 157-173).*
www.irma-international.org/chapter/a-deep-neural-network-model-for-cross-domain-sentiment-analysis/308485

SLOD-BI: An Open Data Infrastructure for Enabling Social Business Intelligence
Rafael Berlanga, Lisette García-Moya, Victoria Nebot, María José Aramburu, Ismael Sanzand Dolores María Llidó (2015). *International Journal of Data Warehousing and Mining (pp. 1-28).*
www.irma-international.org/article/slod-bi/130664