Chapter 39
Use of Data Reduction Process to Bankruptcy Prediction:
Evidence From an Emerging Market

Morteza Shafiee Sardasht
Islamic Azad University, Mashhad Branch, Iran

Saeed Saheb
Islamic Azad University, Mashhad Branch, Iran

ABSTRACT

Predicting corporate bankruptcy has been an important challenging problem in research topic in accounting and finance. In bankruptcy prediction, researchers often confront a range of observations and variables which are often vast amount of financial ratios. By reducing variables and select relevant data from a given dataset, data reduction process can optimize bankruptcy prediction. This study addresses four well-known data reduction methods including t-test, correlation analysis, principal component analysis (PCA) and factor analysis (FA) and evaluated them in bankruptcy prediction in the Tehran Stock Exchange (TSE). To this end, considering 35 financial ratios, the results of data reduction methods were separately used to train Support Vector Machine (SVM) as the powerful prediction model. Regarding the empirical results, among the aforementioned methods, the t-test lead to the most prediction rate with 97.1% of predictability and PCA by 95.1% provides the next position.

INTRODUCTION

Nowadays, rapid technological developments and changes in the business environment have improved economic circumstances. In this dynamic business environment, increasing the competitiveness of corporations has restricted their access to profit. Thus, the risk of bankruptcy as an important economic phenomenon (Etemadi, Anvary Rostamy, & Dehkordi, 2009) has increasingly soared over the years.

Predicting corporate bankruptcy has been a critical challenging issue (Back, Laitinen, Sere, & van Wezel, 1995; Wu, Tzeng, Goo, & Fang, 2007; Kim & Kang, 2010). According to the Kim and Han (2003) corporate bankruptcy triggers both social and economic costs for management and stockholders.

DOI: 10.4018/978-1-5225-5643-5.ch039

Over years business owners and stakeholders have been looking for a shield to protect themselves against the risks and this has made them sensitive towards using predictive models. Yet, bankruptcy prediction becomes major problem to Academics, researchers and practitioners over the past five decades since Altman (1968) (Shin, Lee, & Kim 2005; Tsai, 2009; Cho, Hong, & Ha, 2010).

According to the Bellovary, Giacomino, and Akers (2007) there are over 150 predictive models available to bankruptcy. They believe that the focus of future research should be on the use of existing bankruptcy prediction models instead of development of new predictive models. Thus, it is noble idea which researchers consider other aspect of bankruptcy problem. Bankruptcy prediction as one of areas of risk management (Ming & Jeong, 2009; Chaudhuri & De, 2011), meets a range of observations and features which are often vast amount of financial ratios. In many cases, not all the measured variables are "important" for understanding the underlying phenomena of interest (Fodor, 2002). Compared to older and smaller data platforms, nowadays, the new ones have led to the information overload challenge in data analysis. The main reason is that most efforts to create a dataset are focused on topics such as storage efficiency and often there is no plan for analyzing this data volume (Kantardzic, 2011). Accordingly, various prediction methods have outlived their efficiency for two reasons. The first is the increasing number of observations and the second - which is more important - is the increasing number of relevant features of an observation.

To deeply analyze a huge amount of information of the corporations is likely to take much time and need many human resources (Tsai, 2009). In fact, several factors such as high volume of financial ratios as well as high time and cost have made bankruptcy predictions inefficient. Data reduction offers researchers a set of analytical tools that make the derivation of meaningful summaries from large datasets possible. As a part of the data preprocessing step in the data mining (DM) projects, data reduction, have disregarded especially in bankruptcy studies.

Over the past years, wide range of bankruptcy works have not carefully concerned about the data reduction process. They usually attempt to use different models to conquer the bankruptcy prediction problem. The purpose of this paper is providing evidence in the effect of data reduction in bankruptcy prediction on the emerging market of Iran. To this end, four well-known data reduction methods including *t*-test, correlation analysis, principal component analysis (PCA) and factor analysis (FA) are addressed and their performance in the bankruptcy prediction evaluated.

The rest of this paper is organized as follows. The next section provides a review of the theoretical literature. Afterward, the application of data reduction methods is examined. This is then followed by the empirical outcomes and discussion. Finally, conclusion is presented in last section.

THEORETICAL LITERATURE

Bankruptcy Prediction

Bankruptcy has been known as one of the major concerns of economic and financial activities, insofar as the number of bankrupt companies is considered as an indicator of the development and robustness of the economy (Zopounidis & Dimitras, 1998). When a company enters a period of financial distress, Shortage of liquidity and pending obligations may arise. Prediction of bankruptcy makes managers and investors able to do preventative measures consist of changing operational policy, financial restructuring