

# Data Mining Applications in Accounting and Finance

**D****Wikil Kwak***University of Nebraska at Omaha, USA***Susan Eldridge***University of Nebraska at Omaha, USA***Yong Shi***University of Nebraska at Omaha, USA & Chinese Academy of Sciences, China*

## INTRODUCTION

Data mining has been used in a variety of business applications, such as consumer buying pattern prediction and credit card default prediction, but recent research studies in accounting and finance have applied data mining techniques for classification and prediction of events such as firm bankruptcy and auditor changes. Data mining may be a useful tool for accounting and finance applications because of the huge volume of financial data that impacts business decisions. In this chapter, we summarize three published research studies in which we applied various data mining applications using accounting and other data for classification and prediction decisions, and we identify important issues to consider when applying current data mining tools.

The three research studies we summarize in this chapter are Kwak, Eldridge, Shi, and Kou (2009), Kwak, Eldridge, Shi, and Kou (2011), and Kwak, Shi, and Kou (2012). In these studies, we evaluate the relative usefulness of a variety of data mining approaches primarily by comparing overall accuracy rates for predicting material weaknesses in internal controls (Kwak et al., 2009), auditor changes (Kwak et al., 2011), and bankruptcy (Kwak et al., 2012). We compare the results of multiple criteria linear programming (MCLP) with those of linear discriminant analysis and Decision Tree based See5 (DT) in Kwak et

al. (2009) and with those of classification and regression tree (CART), linear logistic regression (LLR), and Bayesian network in Kwak et al. (2011). The CART, LLR, and Bayesian network applications are three of the 13 data mining models in Witten and Frank's (2005) WEKA Data Mining Workbench, and Kwak et al. (2012) apply all 13 data mining models.

## BACKGROUND

### Accounting Context for Classification or Prediction Decisions

Investors, creditors, and other decision-makers use financial statement and other data about firms to make choices about providing resources to and engaging in business transactions with those firms and to evaluate risks associated with such choices. A firm's deteriorating financial condition is an indicator of increasing risks that may trigger an auditor change, that may be related to material weaknesses in internal controls over financial reporting, and that may result in bankruptcy. Because data mining techniques could be effective tools in predicting these potentially costly events, Kwak et al. (2009), Kwak et al. (2011), and Kwak et al. (2012) explore the relative usefulness of various data mining applications in these settings. In this

subsection of our chapter, we discuss the predictor variables in and supporting literature for these three studies.

Altman (1968), Altman, Haldeman, and Narayanan (1977), and Ohlson (1980) are the most prominent initial bankruptcy prediction studies in academic accounting and finance literature. All three Kwak et al. papers use eleven or more financial statement variables from these bankruptcy prediction studies as indicators of financial distress, noting that bankruptcy is the extreme form of financial distress. The financial statement variables include both financial condition ratios (such as the current ratio and debt-to-assets) and performance measures (such as net income-to-total assets, earnings before interest and taxes-to-total assets, and loss indicator variables). Kwak et al. (2011) and Kwak et al. (2012) both also include a missing dividends indicator variable representing an early stage of financial distress as in Lau (1987).

Kwak et al. (2011) apply the MCLP and three other data mining approaches to evaluate the use of 13 financial statement variables in predicting auditor changes. As indicated above, most of these variables were used in initial bankruptcy prediction studies in which linear discriminant analysis or logistic regression techniques were applied. More recent research studies by Chen, Gupta, and Senteney (2004) and Chen, Yen, and Chang (2009) examine bankruptcy or financial distress prediction and document the statistical significance of an auditor change variable in addition to selected financial statement variables using logistic regression. Thus, Kwak et al. (2011) turn the relationship between financial condition (distress) and auditor changes around to evaluate the predictive ability of financial distress measures for auditor change decisions. Kwak et al.'s (2011) sample includes 790 firms that changed auditors in 2007 or 2008 and 1,132 non-auditor change firms matched on size and industry.

Kwak et al. (2009) apply the MCLP and Decision Tree based See5 data mining approaches along with linear discriminant analysis to evaluate the use of financial statement and other predictor

variables in predicting material weaknesses in internal controls. Reacting to several major financial frauds, Congress passed the Sarbanes-Oxley Act (SOX) (2002) which included numerous provisions aimed at increasing the quality of financial reporting and investor protection. Section 302 of SOX (2002) requires chief executive and chief financial officers to submit certification reports accompanying their annual filings with the Securities and Exchange Commission (SEC) that include the disclosure of material weaknesses in internal controls over financial reporting. Such material weaknesses could compromise or at least signal red flags regarding the integrity of the resulting financial statements. Although we found no studies that document an association between material weaknesses in internal controls and firm bankruptcy, firms with internal control weaknesses may be more susceptible to firm failure than those without such weaknesses. Therefore, as previously stated, we use a number of financial ratios used in prior bankruptcy prediction studies in our internal control weakness prediction models. We also include as predictor variables proxies for five firm characteristics (business complexity, firm experience, size, profitability, and auditor) based on evidence from Ge and McVay (2005) that four of these five characteristics (all except firm experience) are associated with firms that disclose internal control weaknesses. Because Ettredge, Li, and Scholz (2007) document an association between auditor dismissals and internal control problems, we also include auditor change variables as predictors. Kwak et al.'s (2009) sample includes 1,158 firms that disclosed material weaknesses in their 2003 or 2004 Section 302 filings and 1,158 firms matched on size and industry that did not disclose material weaknesses.

Kwak et al. (2012) apply the WEKA suite of 13 data mining approaches (as in Peng, Kou, Wang, Wang, and Ko, 2009) using 14 financial condition variables (see our initial discussion above), an internal control weakness indicator variable, and annual market returns to predict bankruptcy. Although previous bankruptcy prediction studies

7 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/data-mining-applications-in-accounting-and-finance/107264](http://www.igi-global.com/chapter/data-mining-applications-in-accounting-and-finance/107264)

## Related Content

---

### Online Word-of-Mouth Reviews

Jing Huo, Qinglong Gou, Qi Dai and Zhimin Huang (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 1689-1697).

[www.irma-international.org/chapter/online-word-of-mouth-reviews/107359](http://www.irma-international.org/chapter/online-word-of-mouth-reviews/107359)

### Data-Driven Simulation-Based Analytics

Durai Sundaramoorthi (2014). *Encyclopedia of Business Analytics and Optimization* (pp. 677-689).

[www.irma-international.org/chapter/data-driven-simulation-based-analytics/107271](http://www.irma-international.org/chapter/data-driven-simulation-based-analytics/107271)

### 3PM Revisited: Dissecting the Three Phases Method for Outsourcing Knowledge Discovery

Richard Ooms, Marco R. Spruit and Sietse Overbeek (2019). *International Journal of Business Intelligence Research* (pp. 80-93).

[www.irma-international.org/article/3pm-revisited/219344](http://www.irma-international.org/article/3pm-revisited/219344)

### Sharp Connection: Linking Competitive Intelligence and Intranets

Paul Hendriks and Wendy Jacobs (2004). *Information and Communications Technology for Competitive Intelligence* (pp. 34-56).

[www.irma-international.org/chapter/sharp-connection-linking-competitive-intelligence/22560](http://www.irma-international.org/chapter/sharp-connection-linking-competitive-intelligence/22560)

### Business Analytics in Sport Talent Acquisition: Methods, Experiences, and Open Research Opportunities

Rocio de la Torre, Laura O. Calvet, David Lopez-Lopez, Angel A. Juan and Sara Hatami (2022). *International Journal of Business Analytics* (pp. 1-20).

[www.irma-international.org/article/business-analytics-in-sport-talent-acquisition/290406](http://www.irma-international.org/article/business-analytics-in-sport-talent-acquisition/290406)