

# **Building a Model for the Detection of Fraudulent Accounting Using Textual Information from the MD&A Section of Japanese Companies**

**Keisuke Miyago, Natsuki Sato, Ayuko Komura, and Hirohisa Hirai**

Department of Industrial Engineering and Management

Kanagawa University

Kanagawa, Japan

r201903815ya@jindai.jp, r202170129yo@jindai.jp, komura-a@kanagawa-u.ac.jp,

hirai@kanagawa-u.ac.jp

## **Abstract**

Accounting fraud is defined as "the intentional material misstatement of financial statements or financial disclosures (in notes to the financial statements or SEC filings) or the perpetration of an illegal act that has a material direct effect on the financial statements or financial disclosures". Detecting accounting fraud at an early stage is important, because fraudulent activity reduces public reliability in capital markets and hinders economic development. Therefore, studies have been conducted to establish a model for the detection of accounting fraud using quantitative data, such as accounting and financial information, and qualitative data, which are descriptions in security reports, such as Form 10-K. Previous studies employed quantitative data to develop a fraudulent accounting detection model using logistic regression, with factors affecting the occurrence of accounting fraud as variables. In contrast, one study used Random Forest (RF) to create a list of words with high predictive power in association with accounting fraud (e.g. a list of words related to merger activity or legal issues). These words were used as input variables to build a fraud detection model using a support vector machine (SVM). Previous studies have confirmed that text data are an effective indicator in the construction of fraudulent accounting detection models. However, to the best of our knowledge, no accounting fraud detection model using text analysis exists in the context of Japanese companies. Therefore, the objective of this study was to construct a fraudulent accounting detection model using textual analysis for descriptions in the MD&A section of Japanese companies' annual securities reports. Following the procedure of previous studies, a fraudulent accounting detection model using RF and SVM was constructed for companies listed on the Tokyo Stock Exchange from 2011 to 2017. Firstly, RF was used to determine the rank order of words that can effectively discriminate between fraudulent and non-fraudulent accounting firms. Next, based on the top 200 words, SVM was used to estimate fraudulent and non-fraudulent accounting companies. Accounting fraud and non-fraud firms were also estimated using RF as a baseline model. Our study indicates that the estimation results using SVM had a geometric mean (the geometric mean of the true positive and true negative rates) of 74.29% and a sensitivity (the accuracy of the accounting fraud firms) of 64.58%, showing a higher accuracy than the results obtained when using RF.

## **Keywords**

accounting fraud, textual analysis, management discussion and analysis, machine learning

## **Biographies**

**Keisuke Miyago** is a graduate student of the Department of Industrial Engineering and Management at Kanagawa University, Japan.

**Natsuki Sato** is a graduate student of the Department of Industrial Engineering and Management at Kanagawa University, Japan.

**Ayuko Komura** is an assistant professor in the Department of Industrial Engineering and Management at Kanagawa University, Japan. Her research topics include operating profit stability and management controls. She earned her Ph.D. in business administration at Meiji University. She received a fellowship from the Japan Society for the Promotion of Science between 2017 and 2019.

**Hirohisa Hirai** is a professor in the Department of Industrial Engineering and Management at Kanagawa University, Japan. He earned his Ph.D. in Engineering at Osaka University. His research is an empirical study of management accounting, firm analysis, and applied statistics, particularly firm valuation. He received awards from several academic societies in accounting and management.