A Brief Review of Methods for the Detection of Accounting Fraud Using Machine Learning Algorithms

Matheus Gonzaga[a], Marcelo A. S. de Almeida[a]

[a]*Data Scientist at Poatek IT Consulting*

Abstract

The complexity of the financial auditing process has been in constant increase over the last few years, so have been the number of cases of financial fraud and the yearly cost of these activities to the public. The techniques of machine learning may present a solution which can enable auditors to make sense of this increased complexity, finding relevant patterns and risk factors in correlations not otherwise possible. This brief review aims to show the many ways in which financial information may be modeled and analyzed with machine learning algorithms, to detect accounting fraud with high degrees of accuracy using heterogeneous data. We start by presenting the motivation for the use of these techniques, and analyze past research works, discussing a wide range of possible use cases in the industry.

1. Motivation

The use of digital analysis in financial auditing is common and widespread. Checking for duplicated payments, missing invoices or check numbers as well as other automatic procedures are all ways to use digital tools to assist in the auditing process.

Since 1997, consideration of fraud is a standard step in financial audit by the Auditing Standards Board. However, this is a difficult task and its highly dependent on the auditors experience Kirkos et al. [5]. Several studies showed that classification models built with machine learning techniques can significantly improve auditors performance. The goal of these models is to provide a probability of fraudulent activities being committed by a given company, so that the auditor may decide on

Email addresses: `matheus.gonzaga@poatek.com` (Matheus Gonzaga), `marcelo.dealmeida@poatek.com` (Marcelo A. S. de Almeida)

which records further investigations should be conducted. In the next section, we discuss some of these studies and their results.

2. Research Works

Green and Choi [3] developed a financial fraud classification model using Artificial Neural Networks (ANN). Using an ensemble of three models built by training an ANN on fraud and nonfraud samples, they found that the model showed significant capabilities when used to find increased risk of misstatements in financial reports, automatically finding risk factors and enabling auditors to direct their attention where it is most needed.

Bell and Carcello [2], using a sample of 77 fraud engagements and 305 nonfraud engagements, developed a logistic regression model capable of estimating the likelihood of a financial report being fraudulent. The data model used was first described by Loebbecke and Willingham [6], and is conditioned on the presence or absence of several fraud-risk factors, such as weak internal control environment, rapid company growth, inadequate or inconsistent relative profitability, management placing undue emphasis on meeting earnings projections, management lying to auditors or overly evasive, the ownership status (public vs. private) of the entity, and an interaction term between a weak control environment and an aggressive management attitude toward financial reporting. Their model was actually more accurate than practicing auditors in assessing risk for the 77 fraud observations.

Ren [7] developed prediction models for the detection of fraudulent companies based on the analysis of different main categories of attributes, including financial indexes, financial risk, and company governance. He used clustering techniques to identify the relevant attributes in each category and then applied logistic classification and a Bayes theorem based machine learning algorithm for the prediction. He found that financial risk and pressure attributes provided the most accuracy for the model, and financial indexes and trading based attributes had low prediction capability, showing that stock market variations, even though generally indicators of a companys financial situation, are not very useful in detecting fraud.

Kirkos et al. [5] used financial ratios, such as Altman's Z score of financial distress Altman [1], debt to equity ratio, conversion rate, and others to identify fraudulent financial statements (FFS). They analyzed a sample of 76 companies, with 38 of those having published indication or proof of participation in the issuing of FFS. They used statistical methods to find the financial ratios that were most informative for the prediction, eliminating unnecessary data. In addition, they applied Artificial

Neural Networks, decision trees, and Bayesian networks, for the identification phase. The algorithm that showed the highest accuracy (90.3%) was the Bayesian network, which, as a result of its implementation properties, also enabled them to see which ratios had the highest influence in the prediction. The results of this study show that there is potential in using published financial statements to build FFS detection models.

Kanapickienė and Grundienė [4] investigated 51 financial ratios from a set of 40 fraudulent financial statements and 125 non-fraudulent financial statements from companies in Lithuania and developed a logistic regression model to predict fraud. The model can be used by external users of financial statements when making decisions for investment and company evaluation.

3. Conclusion

Auditing nowadays has become a demanding task. With the increasing number and complexity of management and other fraud cases, it is hard to detect fraud with classical audit procedures. Machine learning techniques, with accurate classification and prediction capabilities, can facilitate auditors in detecting frauds in financial statements and other financial information. The literature in the area is vast, and there are several studies and cases that lay the groundwork for the application of these techniques in the industry.

Bibliography

[1] Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The journal of finance*, 23(4):589–609.

[2] Bell, T. B. and Carcello, J. V. (2000). A decision aid for assessing the likelihood of fraudulent financial reporting. *Auditing: A Journal of Practice & Theory*, 19(1):169–184.

[3] Green, B. P. and Choi, J. H. (1997). Assessing the risk of management fraud through neural network technology. *Auditing*, 16(1):14.

[4] Kanapickienė, R. and Grundienė, Ž. (2015). The model of fraud detection in financial statements by means of financial ratios. *Procedia-Social and Behavioral Sciences*, 213:321–327.

[5] Kirkos, E., Spathis, C., and Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32(4):995–1003.

[6] Loebbecke, J. and Willingham, J. (1988). Review of sec accounting and auditing enforcement releases. *Unpublished working paper.*

[7] Ren, H. (2006). Investigation research on public company financial report fraud. *Dalian: Dongbei University of Finance.*