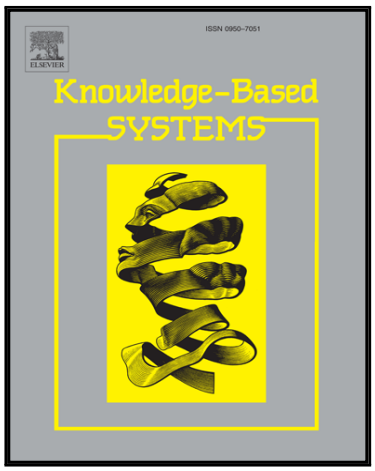

Mining Corporate Annual Reports for Intelligent Detection of Financial Statement Fraud – A Comparative Study of Machine Learning Methods

Petr Hajek , Roberto Henriques

PII: S0950-7051(17)30202-2
DOI: 10.1016/j.knosys.2017.05.001
Reference: KNOSYS 3898

To appear in: *Knowledge-Based Systems*

Received date: 5 December 2016
Revised date: 27 February 2017
Accepted date: 2 May 2017

Please cite this article as: Petr Hajek , Roberto Henriques , Mining Corporate Annual Reports for Intelligent Detection of Financial Statement Fraud – A Comparative Study of Machine Learning Methods, *Knowledge-Based Systems* (2017), doi: 10.1016/j.knosys.2017.05.001

Highlights

- We combine features derived from financial information and managerial comments.
- We employ feature selection and classification using a wide range of machine learning methods.
- Analysts' forecasts of revenues and earnings are necessary to detect fraudulent firms.
- Misclassification cost ratio of 1:2 is based on the loss attributable to financial statement fraud and audit fees.
- Interpretable Naïve Bayes-based models outperform remaining methods in terms of misclassification costs.