Abstract

This study compares the performance of six popular statistical and machine learning models in detecting financial statement fraud under different assumptions of misclassification costs and fraud firm to non-fraud firm ratios. The results show, somewhat surprisingly, that logistic regression and support vector machines perform well relative to artificial neural network, bagging, C4.5 and stacking. The results also reveal some diversity in predictors used across the classification algorithms. Out of 42 predictors examined, only 6 are consistently selected and used by different classification algorithms: auditor turnover, total discretionary accruals, Big 4 auditor, accounts receivable, meeting or beating analyst forecasts, and unexpected employee productivity. These findings extend financial statement fraud research and can be used by practitioners and regulators to improve fraud risk models.