Equilibrium. Quarterly Journal of Economics and Economic Policy
Volume 16 Issue 1 March 2021
p-ISSN 1689-765X, e-ISSN 2353-3293
www.economic-policy.pl

ORIGINAL ARTICLE

Citation: Papík, M., & Papíková, L. (2021). Application of selected data mining techniques in unintentional accounting error detection. *Equilibrium. Quarterly Journal of Economics and Economic Policy, 16*(1), 185–201. doi: 10.24136/eq.2021.007

Contact to corresponding author: mario.papik@gmail.com; Department of Economics and Finance, Faculty of Management, Comenius University, Address: Odbojárov 10, P.O.Box 95, 820 05 Bratislava, Slovakia

Received: 08.11.2020; Revised: 18.01.2021; Accepted: 13.02.2021; Published online: 30.03.2021

Mário Papík
Comenius University, Slovakia
orcid.org/0000-0003-2563-7799

Lenka Papíková
Comenius University, Slovakia
orcid.org/0000-0001-6525-6769

Application of selected data mining techniques in unintentional accounting error detection

JEL Classification: *C53; C81; M41; M42*

Keywords: *financial fraud; unintentional accounting errors; financial restatements; decision tree; classification and regression tree; random forest*

Abstract

Research background: Even though unintentional accounting errors leading to financial restatements look like less serious distortion of publicly available information, it has been shown that financial restatements impacts on financial markets are similar to intentional fraudulent activities. Unintentional accounting errors leading to financial restatements then affect value of company shares in the short run which negatively impacts all shareholders.
Purpose of the article: The aim of this manuscript is to predict unintentional accounting errors leading to financial restatements based on information from financial statements of companies. The manuscript analysis if financial statements include sufficient information which would allow detection of unintentional accounting errors.
Methods: Method of classification and regression trees (decision tree) and random forest have been used in this manuscript to fulfill the aim of this manuscript. Data sample has consisted of 400 items from financial statements of 80 selected international companies. The results of developed prediction models have been compared and explained based on their accuracy, sensitivity, specificity, precision and F1 score. Statistical relationship among variables has been tested by correlation analysis. Differences between the group of companies with and without unintentional accounting error have been tested by means of Kruskal-Wallis test. Differences among the models have been tested by Levene and T-tests.

Findings & value added: The results of the analysis have provided evidence that it is possible to detect unintentional accounting errors with high levels of accuracy based on financial ratios (rather than the Beneish variables) and by application of random forest method (rather than classification and regression tree method).

Introduction

Accounting fraud is an intentional attempt to deceive or mislead users of publicly available financial information, particularly investors and creditors, by preparing and disclosing manipulated financial statements (Rezaee, 2005). The 2014 and 2016 Price Waterhouse Coopers (2014; 2016) analysis of economic fraud in the world position accounting fraud to the fifth place of frequency among all economic frauds after theft of company assets, cybercrime, corruption and supplier fraud. Year-on-year, there has been a decline in accounting fraud from 24% in 2011, to 22% in 2014 and 18% in 2016 among all economic frauds.

Accounting frauds have been studied by many authors from different perspectives. Authors such as Beneish *et al.* (1999) have built own simple formulas for prediction of fraudulent financial statements similar to bankruptcy models by Altman (1968). Kotsiantis *et al*. (2006), Ravisankar *et al.* (2011) Liu *et al.* (2015) have used various different data-mining techniques for detection of fraudulent companies based on financial statement data. Other authors such as Humpherys *et al.* (2011) and Throckmorton *et al.* (2015) have tested prediction of fraudulent companies based on linguistic variables in annual reports without using any financial data from these statements. Other authors, such as Ibadin and Ehigie (2019), Paseková *et al.* (2019) or Homola and Paseková (2020), have studied relationship between particular accounting standard and occurrence of accounting fraud or misstatement in companies reporting under this standard.

There is a difference between accounting frauds and unintentional accounting errors which might lead to financial restatements in the future. Unlike accounting frauds, unintentional accounting errors mislead users accidentally in incorrect application of accounting standards and procedures. Li and Zhang (2006) define financial restatement as a correction of a part or whole financial statement that has been previously published by a company. Company publishes financial restatement in a form of corrected balance sheet or income statement as a result of discovery of previously unknown accounting error in its financial reporting.

Even though financial restatements look like less serious mistreatment of publicly available information, some studies have shown that impact of financial restatements on financial markets is like that of intentional fraudulent activities (Bowen *et al.*, 2017). Palmrose *et al.* (2004) have showed that negative impact on market price per share could be in average 9% over in a next two days. The topic of unintentional accounting errors is closely related with aggressive earning management strategies. Sosnowski (2018) has not provided evidence that companies with private equity financing have less aggressive financial reporting. However, Sosnowski (2017) has showed that new stock companies used an aggressive earnings management strategy to increase additional level of financing before process of initial public offering (IPO).

The aim of this manuscript is to predict unintentional accounting errors leading to financial restatements based on information from financial statements of companies. This manuscript also determines whether company financial statements include enough information to detect unintentional accounting errors. To fulfill the aim of this manuscript, several prediction models based on classification and regression trees (CART) and random forest have been developed. Results of these models have been subsequently validated by their parameters of accuracy, specificity, sensitivity, precision and F1 score and tested by Levene and T-test. Statistical dependence among variables has been tested by correlation analysis. Differences between the group of companies with and without unintentional accounting error have been tested Kruskal-Wallis test.

The paper is divided into following parts: Section 1 describes literature review, Section 2 contains research methodology, Section 3 presents results of analysis, Section 4 is discussing the results and the last section represents the manuscript concluding remarks.

Literature review

According to the Association of Certified Fraud Examiners (2018), 2 690 fraudulent activities took place in financial year 2018 whilst these fraudulent activates have resulted in losses of more than $7 trillion, with every fifth fraudulent case causing damages above $1 million. Hence the role of corporate internal audit is very important given organizational needs and internal audit structure (Saxunová, 2012). In case there are no internal business controls mechanisms, company should hire external control mechanisms. Regular monitoring by internal of external regulator can prevent occurrence of fraudulent situations (Mariak & Mitková, 2016). On the other hand, activities of external auditors do not conduct for the purpose of detecting fraud and forensic accountants are also limited in these activities (Pavlovič *et al.,* 2019). Without useful tools, which can identify suspicious activities, is mostly random detection of fraudulent financial statements or unintentional errors.

Many prediction models focused on detection of fraudulent financial statements exist based on financial ratios, such as Beneish *et al.* (1999), Kirkos *et al.* (2007), Pai *et al.* (2011), Lin *et al.* (2015), Yao *et al.* (2019) and Wang *et al.* (2020). Beneish *et al.* (1999) has developed M-score by application of eight variables detecting accounting fraud in companies. Kirkos *et al.* (2007) has applied 37 financial variables on a sample of 38 fraudulent and 38 non-fraudulent companies and has achieved accuracy between 74 and 90%, depending on the statistical method applied. This study has therefore implied the importance of data mining techniques in prediction of accounting fraud in companies. Higher accuracy of developed models (between 89 and 93%) has been achieved by Lin *et al.* (2015). The Lin *et al*. (2015) research has used both financial and non-financial variables and has applied following methods: CART, logistic regression and neural network. Similar results have been achieved by Pai *et al*. (2011), who has achieved comparable accuracy (78 to 92%) on a sample of 75 Thai companies. Yao *et al.* (2019) have applied a wide range of methods (Bayesian network, CART, k-nearest points, logistic regression, neural network, or support vector machine) on a sample of Chinese companies. 93 Chinese companies have been studied also in a recent study of Wang *et al.* (2020).

However, there are few studies focusing on detection of financial restatements. Dutta *et al*. (2017) have built pioneering prediction model on 3500 US companies. They used 116 financial variables and different techniques such as Bayesian network, CART, naïve Bayes, neural network or support vector machine. The average accuracy of developed models has been between 60 and 80%. Prediction model able to detect both accounting fraud and financial restatement has been developed by Kim *et al.* (2016). To develop his prediction model, Kim has used Bayesian network, logistic regression and support vector machine methods which achieved high levels of accuracy (82 to 88%).

Papík and Papíková (2020) have built two models on 80 US companies. Linear discriminant analysis model has achieved accuracy 62% and logistic regression model have had over 70%. Research has tried to show that Beneish variables do not provide statistically significant better results than financial features from financial statements.

From the point of view of the technique used, a majority of researches has confirmed that higher accuracy and lower type I and II errors are mostly provided by data-mining techniques, such as Bayesian network, CART, random forest, support vector machine and neural network than classical group selection methods such as discriminant analysis or logistic regression (Kotsiantis *et al*., 2006; Dechow *et al*., 2011; Gepp, 2015; Liu *et al*., 2015).

Moreover, there are differences between data mining techniques themselves. Tang *et al*. (2020) has provided evidence that ensemble learning methods such as random forest or Xboost have higher accuracy than other machine learning techniques.

Research methodology

Following section contains details about collected data sample and application of data-mining techniques in this manuscript.

Data Sample

Firstly, 40 international companies with financial restatements have been identified worldwide in total. Financial information of these companies has been collected for a period of five years prior to publishing financial restatement. Another 40 international companies have been identified as companies without financial restatement as another sample, and these companies have been matched to companies with financial restatements (based on time period, size in terms of value of total assets and industry classification). In overall, 400 financial information has been collected — 200 for companies with financial restatements and 200 for companies without financial restatement. Similar data sample size has been used in studies of Kirkos *et al.* (2007) — 86 companies, Pai *et al.* (2011) — 75 companies, Papík and Papíková (2020) — 80 companies or Wang *et al.* (2020) — 93 companies.

Financial data of selected companies in the data sample has been collected from Security and Exchange Commission (SEC) database and from annual reports of these companies (EDGAR, 2019). Geographically data sample consists of 81% companies from North America, 10% from Western Europe, 8% from Asia and 1% from Africa. In terms of industrial sector classification, most companies are classified as technology providers (43%), followed by retail (16%), finance (11%), energy (6%), mechanical engineering (5%), telecommunications (5%) and health (5%) companies. The remaining companies could be classified as chemical, medicine, agriculture and real estate ones.

The collected financial data has been used for calculation of two types of ratios. The first set of financial ratios (FR) represents standard financial ratios used for detection of accounting fraud or unintentional accounting error by different authors. (Dutta *et al*., 2017; Gepp, 2015; Lin *et al*., 2015) These ratios are listed in Table 1.

The second set of ratios contains variables used in Beneish *et al.* (1999) M-score which are used for detection of accounting fraud based on specific financial ratios. Beneish model uses eight ratios as shown in Table 2. The validity of this model on fraudulent companies has been cross-checked by authors such as Drábková (2015) and MacCarthy (2017). MacCarthy (2017) concludes, based on analysis of Enron case, that Beneish ratios could have detected fraudulent activities in this company around year 2000.

Application of data mining

CART and random forest are commonly used methods to detect accounting fraud in companies. Kotsiantis *et al.* (2006) has achieved the highest level of accuracy (91%) by applying CART. His study has also applied other methods, such as Bayesian network, logistic regression, neural network and support vector machine. A comparably high level of accuracy has been also achieved by studies of Hajek and Henriques (2017), Lin *et al.* (2015), Chen *et al.* (2017), Jan (2018), Yao *et al*. (2019). Random forest method has not been widely used yet, however, few existing studies have achieved comparable accuracy to that of CART (Liu *et al.* (2015) — 88%, Hajek and Henriques (2017) — 88%, Tang *et al.* (2020) — 90%)

The method of CART and random forest have been applied in this manuscript. The process of creating a decision tree for CART starts with a decision node that extends into other decision nodes which visually creates a tree structure. Leaves of the decision tree represent outputs and branches represent attributes, in other words, one decision node is a separate decision which has been based on particular attribute of analysed object. Attributes in this manuscript have been defined as different financial ratios which are differentiated as much as possible so that their correct classification was possible at the end of the decision tree. Random forest method is, furthermore, created by several decision trees (Breiman *et al.,* 1984; Quinlan, 1986). Subsequent outcomes from decision trees are combined into one outcome group model that represents random forest. Random forests, unlike decision tree method, belongs among method group ensemble methods that combines results of several methods, in this case it is combination of decision trees (Breiman, 2001).

Developed model has been cross-validated on 10 folds. Ten models developed in this manuscript have been created on nine training folds which represent 90% of overall sample — 360 financial items (180 from companies with financial restatements and 180 from companies without financial restatement). Model performance has been verified on testing dataset. Testing dataset has been formed by remaining one testing fold which represents 10% of data sample with total 40 financial items (20 from companies with financial restatements and 20 from companies without financial restatement). Final attributes of developed models have been described by chosen parameters as an average parameter of ten testing rounds.

Selected attributes describing model performance of creative models are: Accuracy (TP/(TP + TN)), Sensitivity (TP/(TP+FN)), Specificity (TN/(TN+FP)), Precision (TP/(TP+FP)) and F1 score (2*Sensitivity*Precision/(Sensitivity + Precision)), where True positive (TP) means correctly predicted unintentional error, True negative (TN) stands for correctly rejected unintentional error, False positive (FP) is type I error and False negative (FN) represents type II error.

Results

To identify variables with existing relationships among them, all input variables have been included in correlation analysis. Table 3 shows the results of correlation analysis for all tested variables. Results from correlation analysis show that there is correlation in absolute value greater than 0.5 only among two combinations of variables. The two variables are ROA and DAR with negative correlation coefficient equal to -0.65, and TATA and CAT with positive correlation coefficient equal to 0.67. Even though these correlation coefficients are very high or high, they have not been excluded from dataset. This is because these variables are only random correlations between two variables and not high correlations of one variable with several other variables. For remaining variables, correlation coefficients have varied in range from -0.3 to 0.3 and therefore have been low. Based on these findings, input variables can be considered independent.

A descriptive analysis of variables for companies with financial restatement and without financial restatement along with results of individual non-parametric tests after their normalization is shown in Table 4. The results indicate that median of companies without financial restatement except variables DAR, DEPI, DSRI and LVGI has been higher than the median of companies with financial restatement. The higher median of variables DEPI, DSRI and LVGI indicators is in line with original Beneish study. Beneish study assumes that companies manipulating their financial statements achieve higher values of these indicators. On the other hand, remaining Beneish variables do not confirm this outcome.

Results of Kruskal-Wallis test have identified statistical significant differences between companies with financial restatement and without financial restatements for variables DER, FAT, GPTA, NPS and ROA. Companies with financial restatement have achieved lower values of these indicators than companies without financial restatement. Lower values of these indicators could indicate weaker economic performance of these companies, which could subsequently lead them to improvement of their financial results through fraudulent financial reporting.

Table 5 shows values of particular attributes for each of the models and sets of variables. Results indicate that random forest method has achieved better results in accuracy, precision and F1 score than CART method. This manuscript has also provided evidence that variables of only Beneish model have achieved significantly worse results than results of combination of Beneish variables with financial rations, and even worse results compared to those of dataset containing only financial ratios.

Table 6 includes comparison of attributes for individual testing datasets containing Beneish dataset along with remaining two datasets. This comparison has been conducted via Levene test of equality of variances and T-test of equality of mean. Based on results, it is possible to conclude that Beneish variables have achieved worse results than those of dataset with all variables or dataset with only financial ratios. Comparison of all variables to dataset containing only financial ratios has not confirmed any statistically significant differences.

Discussion

The results obtained in this manuscript indicate that it is possible to predict unintentional accounting errors which lead to restating company financial statements. This manuscript, compared to results of other studies, has showed that data mining techniques like CART or random forest are capable of detecting these accounting errors with more that 80% accuracy. This indicates applicability of data mining techniques in detection of errors in financial statements of companies

When compared to results by Papík and Papíková study (2020), this manuscript has obtained improved results in the parameters of accuracy (84% in this study compared to 71% accuracy in the other study), sensitivity (88% compared to 42%) and specificity (80% compared to 79%). These different levels of achieved parameters describing model performance are caused mainly by the methods used in Papík and Papíková's (2020) study. Their study applied discriminant analysis and logistic regression, whilst this manuscript applied data mining techniques. Although based on a smaller data sample, this study has also obtained better results than the CART model designed to predict unintentional accounting errors developed by Dutta *et al*. (2017). For comparison, Dutta's study achieved the accuracy of 78% (compared to 84% accuracy in this study), the sensitivity of 78% (compared to 88% in this study) and the specificity of 69% (compared to 80% in this study). Other significant studies focusing only on detection of financial misstatements (unintentional errors) have not yet been conducted.

This manuscript has also pointed out to differences in detection process of unintentional accounting errors and accounting frauds. Beneish has formed eight variables with high predictive accuracy to detect accounting fraud which has been also verified by other authors. On the other hand, significance of these variables has not been confirmed in the process of unintentional accounting errors detection. Difference between Beneish variables and other financial ratios has been in the range of 15% for parameters of accuracy, sensitivity, specificity, precision and F1 score.

Data sample might be a possible limitation of this manuscript. The data sample consisted of financial data for five consecutive years of 80 companies, which represents 400 input data. The data sample can be still viewed as a relatively small sample. The sample has also consisted of heterogenic companies from different countries which might have led to decreased values of accuracy of developed models.

Conclusions

The aim of this manuscript is to predict unintentional accounting errors leading to financial restatements based on information from financial statements of companies. Unintentional accounting errors, like accounting frauds, have negative impact on market value of a company. Detection of these accounting errors is, therefore, not only a challenge for auditors, but also for various stakeholders who might be negatively impacted by revelation of these errors.

Prediction models developed in this manuscript on a sample of 80 companies have reached accuracy of 72% to 84% and these results are among the highest that have been achieved in this area so far. Despite several existing studies applying ensemble data mining method on detection of accounting errors, this manuscript provides innovative approach and could be considered a novelty in this field — the possibility to detect unintentional accounting errors with high levels of accuracy based on financial ratios (rather than the Beneish variables) and by application of random forest method (rather than CART method).

Prediction models created by data mining techniques show possible future developments in the field of unintentional accounting errors detection. This manuscript has also provided evidence that models with financial ratios as input features achieve higher accuracy than Beneish variables. Therefore, future studies should also focus on application of various financial ratios. Future studies should also factor in differences arising from geographical, industrial or local accounting specifics. As studies from other fields such as bankruptcy have showed, financial and non-financial parameters can be used to detect accounting errors. Not all information necessary to detect accounting errors might be available in company financial statements. Detection of accounting errors based on quantitative data has not been conducted yet and future studies could focus on quantitative research in this field.

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*(4), 589–609.

Association of Certified Fraud Examiners. (2018). Global study on occupational fraud and abuse. Retrieved from https://s3-us-west-2.amazonaws.com/acfepublic/2018-report-to-the-nations.pdf / (16.02.2020).

Beneish, M. D, Lee, C., Press, E., Whaley, B., Zmijewski, M., & Cisilino, P. (1999). The detection of earnings manipulation. *Financial Analysts Journal, 55*(5), 24–36. doi: 10.2469/faj.v55.n5.2296.

Bowen, R., Dutta, S., & Zhu, P. (2017). Financial constraints and accounting restatements. *University of Ottawa Working Paper*.

Breiman, L., Friedman J., Olshen R., & Stone C. (1984). *Classification and regression trees*. Wadsworth Int. Group.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(5), 5–32. doi: 10.1023/A:1010933404324.

Chen, Y-J., Wu, Ch-H., Chen, Y-M., Li, H-Y., & Chen, H-K. (2017). Enhancement of fraud detection for narratives in annual reports. *International Journal of Accounting Information Systems, 26*. 32–45. doi: 10.1016/j.accinf.2017.06.004.

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R.G. (2011). Predicting material accounting misstatements: Predicting material accounting misstatements. *Contemporary Accounting Research. 28,* 17–82. doi: 10.1111/j.1911-3846.2010.01041.x.

Drábková, Z. (2015). Analysis of possibilities of detecting the manipulation of financial statements in terms of the IFRS and Czech Accounting Standards. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis, 63,* 1859–1866. doi: 10.11118/actaun201563061859.

Dutta, I., Dutta, S., & Raahemi, B. (2017). Detecting financial restatements using data mining techniques. *Expert Systems with Applications 90*, 374–393. doi: 10.1016/j.eswa.2017.08.030.

EDGAR Online (2019). *List of companies*: 10-K.

Gepp, A. (2015). *Financial statement fraud detection using supervised learning methods.* Retrieved from http://epublications.bond.edu.au/cgi/viewcontent.cgi?article=1227&context=theses (16.02.2020).

Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud – a comparative study of machine learning methods. *Knowledge-Based Systems, 128*(5), 139–152. doi: 10.1016/j.knosys.2017.05.001.

Homola, D., & Paseková, M. (2020). Factors influencing true and fair view when preparing financial statements under IFRS: evidence from the Czech. Republic. *Equilibrium. Quarterly Journal of Economics and Economic Policy, 15*(3), 595–611. doi: 10.24136/eq.2020.026.

Humpherys, S. L., Moffitt, K. C., Burns, M. B., Burgoon, J. K., & Felix, W. F. (2011). Identification of fraudulent financial statements using linguistic credibility analysis. *Decision Support Systems, 50*(3), 585–594. doi: 10.1016/j.dss.201 0.08.009.

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications. 32*, 995–1003. doi: 10.1016/j.eswa.2006.02.016.

Kotsiantis, S., Koumanakos, E., Tzelepis, D., & Tampakas, V. (2006). Forecasting fraudulent financial statements using data mining. *International Journal of Computational Intelligence, 3*(2), 104-110.

Ibadin, P. O., & Ehigie, A. H. (2019). Beneish model, corporate governance and financial statements manipulation. *Asian Journal of Accounting and Governance, 12*. 51–64. doi: 10.17576/AJAG-2019-12-05.

Jan, Ch.-I. (2018). An effective financial statements fraud detection model for the sustainable development of financial markets: evidence from Taiwan. *Sustainability*, *10*(2). 513. doi: 10.3390/su10020513.

Kim, Y. J., Baik, B., & Cho, S. (2016). Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning. *Expert Systems with Applications, 62*, 32–43. doi: 10.1016/j.eswa.2016.06.016.

Li, O. Z., & Zhang, Y. (2006). Financial restatement announcements and insider trading. *SSRN Electronic Journal*. doi: 10.2139/ssrn.929539.

Lin, C.-C., Chiu, A.-A., Huang, S. Y., & Yen, D. C. (2015). Detecting the financial statement fraud: the analysis of the differences between data mining techniques and experts' judgments. *Knowledge-Based Systems, 89*, 459–470. doi: 10.1016/j.knosys.2015.08.011.

Liu, Ch., Chan, Y., Kazmi, S. H. A., & Fu, H. (2015). Financial fraud detection model: based on random forest. *International Journal of Economics and Finance*, *7*(7), 178–188. doi: 10.5539/ijef.v7n7p178.

Mariak, V., & Mitková, Ľ. (2016). Long-term sustainability of portfolio investments – gender perspective: an overview study. *Oxford Journal of Finance and Risk Perspectives, 5*(1), 219–226.

MacCarthy, J. (2017). Using Altman Z-score and Beneish M-score models to detect financial fraud and corporate failure: a case study of Enron Corporation. *International Journal of Finance and Accounting 6*, 159–166. doi: 10.5923/j.ijfa.20170606.01.

Pai, P.-F., Hsu, M.-F., & Wang, M.-Ch. (2011). A support vector machine-based model for detecting top management fraud. *Knowledge-Based Systems*, *24*(2). 314–321. doi: 10.1016/j.knosys.2010.10.003.

Palmrose, Z. V., Richardson, V., & Scholz, S. (2004). Determinants of market reactions to restatement announcements. *Journal of Accounting and Economics, 37*(1). 59–89.

Papík, M., & Papíková, L. 2020. Detection models for unintentional financial restatements. *Journal of Business Economics and Management*, *21*(1), 64–86. doi: 10.3846/jbem.2019.10179.

Paseková, M., Kramná, E., Svitáková, B., & Dolejšová, M. (2019). Relationship be-tween legislation and accounting errors from the point of view of business representatives in the Czech Republic. *Oeconomia Copernicana*, *10*(1), 193–210. doi: 10.24136/oc.2019.010.

Pavlovič, V., Kneževič, G., Joksimovič, M., & Joksimovič D. (2019). Fraud detection in financial statements applying Benford's law with Monte Carlo simulation. *Acta Oeconomica*, *69*(2), 217–239. doi: 10.1556/032.2019.69.2.4.

Price Waterhouse Coopers (2014). Global Economic Crime Survey 2014. Retrieved from https://www.pwc.at/publikationen/global-economic-crime-survey-2014.pdf (3.01.2020).

Price Waterhouse Coopers (2016). Global Economic Crime Survey 2016. Retrieved from https://www.pwc.com/gx/en/economic-crime-survey/pdf/GlobalEconomicCrimeSurvey2016.pdf (3.01.2020).

Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, *50*, 491–500. doi: 10.1016/j.dss.2010.11.006.

Rezaee, Z. (2005). Causes, consequences and deterrence of financial statement fraud. *Critical Perspectives on Accounting*, *16*(3), 277–298. doi: 10.1016/S1045-2354(03)00072-8.

Saxunová, D. (2012). Investigation of suspected fraud. *International Journal of Business and Management, 1*(2), 347–364.

Sosnowski, T. (2017). Earnings management and the floatation structure: empirical evidence from Polish IPOs. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *12*(4), 693–709. doi: 10.24136/eq.v12i4.36.

Sosnowski, T. (2018). Earnings management in the private equity divestment process on Warsaw Stock Exchange. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *13*(4), 689–705. doi: 10.24136/eq.2018.033.

Yao, J., Pan, Y., Yang, S., Chen, Y., & Li, Y. (2019). Detecting fraudulent financial statements for the sustainable development of the socio-economy in China: A multi-analytic approach. *Sustainability, 11*(6), 1–17. doi: 10.3390/su11061579.

Quinlan, J. R. (1986). Introduction of decision trees. *Machine Learning, 1*, 81–106. doi: 10.1007/BF00116251.

Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: a review. In *Data classification: algorithms and applications*. CRC Press, 37–64. doi: 10.1201/b17320.

Throckmorton, Ch. S., Mayew, W. J., Venkatachalam, M., & Collins, L. M. (2015). Financial fraud detection using vocal, linguistic and financial cues. *Decision Support Systems, 74*. 78–87. doi: 10.1016/j.dss.2015.04.006.

Wang, R., Asghari V., Hsu, Sh.-H., Lee Ch.-J., & Chen, J.-H. (2020). Detecting corporate misconduct through random forest in China's construction industry, *Journal of Cleaner Production*, *268*, doi: 10.1016/j.jclepro.2020.122266.

Annex

Table 1. Financial ratios and their formulas

Features	Feature description	Formula
CAT	Current Asset Turnover Ratio	Net operating income / Current assets
CR	Current Ratio	Current assets / Current liabilities
DAR	Debt to Asset Ratio	(Total long-term liabilities + Current liabilities) / Total assets
DER	Debt to Equity Ratio	(Total long-term liabilities + Current liabilities) / Total equity
FAT	Fixed Asset Turnover	Net operating income / Fixed assets
GPTA	Gross Profitability Ratio	Gross profit / Total assets
NPS	Net Profit Margin on Sales	Net profit / Sales
ROA	Return on Total Assets	Net profit / Total assets
ROE	Return on Equity	Net profit / Total equity

Source: Dutta *et al.* (2017); Gepp (2015); Lin *et al.* (2015).

Table 2. Beneish variables and their formulas

Features	Feature description	Formula
AQI	Asset Quality Index	((total assets – property, plant and equipment – current assets – securities) / total assets) / ((total assets previous year – property, plant and equipment previous year – current assets previous year– securities previous year) / total assets previous year)
DEPI	Depreciation Index	(depreciation expense / (depreciation expense + property, plant and equipment)) / (depreciation expense previous year / (depreciation expense previous year + property, plant and equipment previous year))
DSRI	Days Sales in Receivables Index	(net receivables / sales) / (net receivables previous year / sales previous year)
GMI	Gross Margin Index	((sales – cost of goods sold) / sales) / ((sales previous year – cost of goods sold previous year) / sales previous year)
LVGI	Leverage Index	((total long-term liabilities + current liabilities) / total assets) / ((total long-term liabilities previous year + current liabilities previous year) / total assets previous year)
SGAI	Sales General and Administrative Expenses Index	(selling, general and administrative expense / sales) / (selling, general and administrative expense previous year / sales previous year)
SQI	Sales Growth Index	sales / sales previous year
TATA	Total Accruals to Total Assets	(income from continuing operations – cash flow from operations) / total assets

Source: Beneish *et al*. (1999).

Table 3. Correlation analysis for all selected features

	TATA	**SQI**	**SGAI**	**ROE**	**ROA**	**NPS**	**LVGI**	**GPTA**	**GMI**	**FAT**	**DSRI**	**DER**	**DEPI**	**DAR**	**CR**	**CAT**
AQI	0.04	0.02	0.01	0.00	-0.15	0.01	-0.02	0.02	0.04	-0.01	0.00	-0.05	0.01	0.12	-0.07	0.01
CAT	**0.67**	0.03	-0.01	0.12	0.03	0.04	-0.02	0.01	0.00	0.08	0.02	0.03	-0.04	0.01	0.09	
CR	0.06	-0.01	-0.06	-0.02	0.03	0.00	-0.02	-0.06	0.00	-0.14	-0.01	-0.15	-0.03	-0.27		
DAR	0.08	-0.01	0.02	0.10	**-0.65**	-0.05	0.07	-0.01	-0.08	0.03	-0.03	0.18	0.01			
DEPI	-0.02	-0.01	0.21	-0.03	-0.01	-0.01	0.01	-0.07	-0.03	0.01	-0.01	0.00				
DER	0.05	-0.02	-0.02	-0.10	0.04	0.02	-0.01	-0.15	-0.09	-0.03	-0.02					
DSRI	0.01	-0.01	0.01	0.00	0.01	0.01	0.00	-0.05	0.01	0.00						
FAT	0.19	0.06	-0.01	0.10	0.14	0.06	-0.06	0.09	0.03							
GMI	0.01	-0.02	-0.04	0.01	0.05	0.03	-0.03	-0.01								
GPTA	-0.03	-0.06	-0.03	0.08	0.13	0.06	-0.06									
LVGI	-0.12	0.00	0.01	-0.12	-0.16	0.00										
NPS	0.09	0.31	0.02	0.15	0.25											
ROA	0.02	0.01	0.00	0.07												
ROE	-0.01	0.01	0.01													
SGAI	-0.03	-0.04														
SQI	0.15															

Source: own calculation in R studio based on SEC data and company annual reports data.

Table 4. Descriptive analysis for selected features

Features	W/o financial restatement			Financial restatement			Kruskal-Wall.	
	Mean	Median	Stand. dev.	Mean	Median	Stand. dev.	P-value	Sign.
AQI	1.41	1.01	2.86	1.50	0.99	2.38	0.75	
CAT	-0.34	-0.05	1.75	-0.27	-0.08	0.79	0.65	
CR	1.98	1.45	1.71	1.95	1.42	2.81	0.90	
DAR	0.56	0.51	0.35	0.58	0.52	0.62	0.66	
DEPI	1.48	0.97	5.47	1.04	0.99	0.51	0.26	
DER	**3.74**	**1.00**	**9.44**	**1.88**	**0.93**	**3.88**	**0.01**	**
DSRI	1.04	0.97	0.77	1.07	0.98	0.76	0.73	
FAT	**-0.12**	**-0.11**	**7.22**	**-3.62**	**-0.18**	**22.94**	**0.04**	**
GMI	1.03	1.00	1.91	0.97	0.99	0.90	0.69	
GPTA	**0.32**	**0.26**	**0.27**	**0.26**	**0.23**	**0.20**	**0.01**	**
LVGI	1.09	1.00	0.64	1.96	1.02	10.22	0.23	
NPS	**0.69**	**0.07**	**6.55**	**-1.58**	**0.00**	**13.43**	**0.03**	**
ROA	**0.07**	**0.04**	**0.84**	**-0.28**	**0.00**	**1.54**	**0.00**	***
ROE	0.13	0.09	1.22	-0.11	0.01	2.93	0.29	
SGAI	1.09	1.00	0.63	1.22	0.99	1.95	0.37	
SQI	2.10	1.08	12.81	1.54	1.06	2.41	0.54	
TATA	-0.04	-0.02	0.23	-0.06	-0.03	0.24	0.38	

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Source: own calculation in R studio based on SEC data and company annual reports data.

Table 5. Average model performance results

Features	Method	Accuracy	Sensitivity	Specificity	Precision	F1 score
ALL	DT	72.00%	83.50%	60.50%	67.35%	74.56%
	RF	77.25%	85.00%	69.50%	72.34%	78.16%
Beneish	DT	64.00%	63.50%	64.50%	62.50%	63.00%
	RF	67.75%	72.50%	63.00%	65.91%	69.05%
Financial ratios	DT	80.05%	81.29%	78.82%	78.05%	79.64%
	RF	83.75%	88.00%	79.50%	79.55%	83.56%

Source: own calculation in R studio based on SEC data and company annual reports data.

Table 6. Levene and T-test between model performance measures

	Method	Features	Levene test		T-test			
			F- value	Pr(>F)	T-test	df	p-value	sign
Accuracy	DT	FR vs. Beneish	1.839	0.192	4.539	14.50	0.000	***
		All vs. Beneish	3.267	0.087	-1.445	13.61	0.171	
		All vs. FR	0.165	0.690	2.300	17.79	0.034	
	RF	FR vs. Beneish	1.917	0.183	6.184	14.37	0.000	***
		All vs. Beneish	0.018	0.896	-3.427	17.99	0.003	**
		All vs. FR	2.982	0.101	2.074	14.46	0.056	
Sensitivity	DT	FR vs. Beneish	0.212	0.651	2.308	17.67	0.033	*
		All vs. Beneish	1.710	0.208	-2.193	14.36	0.045	*
		All vs. FR	0.912	0.352	-0.425	15.59	0.677	
	RF	FR vs. Beneish	1.969	0.178	5.247	14.63	0.000	***
		All vs. Beneish	0.792	0.385	-4.448	15.71	0.000	***
		All vs. FR	0.609	0.445	0.836	17.74	0.415	
Specificity	DT	FR vs. Beneish	1.978	0.177	3.477	17.41	0.003	**
		All vs. Beneish	11.882	0.003	0.358	12.62	0.726	
		All vs. FR	7.648	0.013	2.538	14.02	0.024	
	RF	FR vs. Beneish	0.140	0.713	3.003	17.45	0.008	**
		All vs. Beneish	2.138	0.161	-0.833	15.18	0.418	
		All vs. FR	1.053	0.319	1.571	16.74	0.135	
Precision	DT	FR vs. Beneish	2.313	0.146	4.136	15.08	0.000	***
		All vs. Beneish	2.121	0.163	-2.122	15.13	0.051	
		All vs. FR	0.368	0.552	2.497	17.81	0.023	
	RF	FR vs. Beneish	4.427	0.050	5.912	11.94	0.000	***
		All vs. Beneish	1.326	0.265	-3.999	15.40	0.002	**
		All vs. FR	0.868	0.364	1.794	17.17	0.091	
F1 score	DT	FR vs. Beneish	7.238	0.015	4.229	13.86	0.001	***
		All vs. Beneish	11.698	0.003	-0.876	13.06	0.397	
		All vs. FR	0.002	0.964	1.691	17.99	0.108	
	RF	FR vs. Beneish	0.003	0.956	4.921	17.96	0.000	***
		All vs. Beneish	0.841	0.371	-2.611	17.48	0.018	**
		All vs. FR	1.500	0.236	2.142	15.23	0.049	

Meaning: DT – decision tree, RF – random forest, FR – financial ratios, Beneish – Beneish variables, All – financial ratios and Beneish variables
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Source: own calculation in R studio based on SEC data and company annual reports data.