Forecasting Fraudulent Financial Statements with Committee of Cost-Sensitive Decision Tree Classifiers

Elias Zouboulidis[1] and Sotiris Kotsiantis[2]

[1] Hellenic Open University, Greece
eliaszouboulidis@yahoo.com
[2] Department of Mathematics, University of Patras, Greece
sotos@math.upatras.gr

Abstract. This paper uses machine learning techniques in detecting firms that issue fraudulent financial statements (FFS) and deals with the identification of factors associated to FFS. To this end, a number of experiments have been conducted using representative learning algorithms, which were trained using a data set of 164 fraud and non-fraud Greek firms. A random committee of cost-sensitive decision tree classifiers is the best choice according to our experiments.

1 Introduction

Even though it is not a new phenomenon, the number of corporate earnings restatements due to aggressive accounting practices, accounting irregularities, or accounting fraud has raised significantly during the past few years, and it has drawn much awareness from investors, analysts, and regulators. After many high profile accounting frauds and corporate scandals (Enron, WorldCom, Adelphia etc) fraudulent events have been followed by increased governmental intervention and regulation. In 2002, the U.S. congress passed the Sarbanes-Oxley Act to improve the accuracy and reliability of corporate financial reporting and disclosures. Europe also had financial scandals over this same period (with the Parmalat scandal being the most notorious) even if most of which were characteristically different from the US style [8]. In this context, Bollen et. al [5] in order to identify the true causes of Europe's biggest business failures over the past 25 years discovered that high leveraging and management fraud were the only two characteristics common in more than half the cases investigated. However, the authors conclude that even though accounting issues found to play a role in a number of business failures in their study, it is less important compared with large US business failures.

Accounting frauds can be characterized as either fraudulent financial reporting or misappropriation of assets, or both. Fraudulent financial reporting is universally known as cooking the books. Researchers have used a variety of techniques and models to detect accounting fraud in circumstances in which, a priori, is likely to exist. In this study, we perform an in-depth examination of publicly available data from the

I. Maglogiannis, V. Plagianakos, and I. Vlahavas (Eds.): SETN 2012, LNAI 7297, pp. 57–64, 2012.

financial statements of various firms so as to detect FFS by using supervised machine learning methods. The goal of this research is to identify the financial factors to be used by auditors in assessing the likelihood of FFS. The detection of fraudulent financial statements, along with the qualification of financial statements, have also been in the limelight in Greece because of the increase in the number of companies listed on the Athens Stock Exchange (and raising capital through public offerings) and the attempts to decrease taxation on profits. There is an increasing demand for greater transparency, reliability and more information to be incorporated within financial statements.

The following section attempts a brief literature review. Section 3 describes the data set of our study and the feature selection process. Section 4 presents the experimental results for the number of representative compared algorithms and a combining technique that produce better accuracy. Finally, section 5 discusses the conclusions and some future research directions.

2 Literature Review

As Watts and Zimmerman [22] argue the financial statement audit is a monitoring mechanism that helps reduce information asymmetry and protect the interests of the principals, particularly, stockholders and potential stockholders, by providing sound assurance that management's financial statements are free from material misstatements. However, in real life, detecting management fraud is a hard task when using normal audit procedures since there is a lack of knowledge concerning the characteristics of management fraud. Furthermore, given its infrequency, most auditors lack the experience required to detect it. Last but not least, managers purposely try to deceive auditors. Albrecht et al. [1] review the fraud detection aspects of auditing standards and the empirical research conducted on fraud detection. Ansah et al. [2] investigate the relative influence of the size of audit firms, auditor's position tenure and auditor's year of experience in auditing on the likelihood of detecting fraud in the stock and warehouse cycle. They conclude that such factors are statistically significant predictors of the likelihood of detecting fraud, and enlarge the likelihood of fraud detection.

Lin et al [15] developed a Neural Network fraud classification model. Deng [9] used support vector machines (SVMs) to detecting FFS. Hoogs et al. [13] used a genetic algorithm approach to detecting FFS. Bell and Carcello [4] developed and tested a logistic regression to estimate the likelihood of fraudulent financial reporting. Dianmin Yue et al [10] also used Logistic Regression for Detecting Fraudulent Financial Statement of Listed Companies in China.

Ravisankar et al [18] uses data mining techniques such as Multilayer Feed Forward Neural Network, Support Vector Machines, Genetic Programming, Logistic Regression, and Probabilistic Neural Network to identify companies that resort to financial statement fraud. Each of these techniques is tested on a dataset involving 202 Chinese companies and compared with and without feature selection. Probabilistic Neural Network outperformed all the techniques without feature selection.

For Greek data, Spathis et al [20] constructed a model to detect falsified financial statements. He employed the statistical method of logistic regression. Kirkos et al [14] investigate the usefulness of Decision Trees, Neural Networks and Bayesian Belief Networks in the identification of fraudulent financial statements. For both studies [27] and [18] a balanced sample of a total of 76 manufacturing firms was used; 38 firms with FFS were matched with 38 with non-FFS (the sample did not include financial companies).

3 Data Description

Our dataset contains data from 164 Greek listed on the Athens Stock Exchange (ASE) manufacturing firms (no financial companies are included). Auditors checked all the firms in the sample. For 41 of these firms, there was published indication or proof of involvement in issuing FFS. The classification of a financial statement as false was based on the following parameters: inclusion in the auditors' report of serious doubts as to the precision of the accounts, observations by the tax authorities regarding serious taxation intransigencies which drastically altered the company's annual balance sheet and income statement, the application of Greek legislation regarding negative net worth, the inclusion of the company in the Athens Stock Ex-change categories of under observation and negotiation suspended for reasons associated with the falsification of the company's financial data and, the existence of court proceedings pending with respect to FFS or serious taxation contraventions.

The 41 FFS firms were matched with 123 non-FFS firms. All the variables used in the dataset were mined from formal financial statements, such as balance sheets and income statements. This implies that the worth of this study is not restricted by the fact that only Greek company data was used.

The selection of variables to be used as candidates for participation in the input vector was based upon prior research work connected to the topic of FFS. Such work carried out by [7], [15], [20]. Additional variables were also added so as to catch as many as possible predictors not up to that time identified. Table 1 presents a brief description of the financial variables used in the present study.

In an attempt to show how much each attribute influences the induction, we rank the influence of each one according to different statistical measures e.g. Information Gain, Gain Ratio and Relief Score [23]. The attributes that mostly influence the induction are: RLTC/RCR02, AR/TA01, TL/TA02, AR/TA02, WC/TA02, DC/CA02, NFA/TA02, NDAP02 (see ReliefF Score in Table 1). With regard to the remaining variables, it seems that they do not influence the induction.

In general, the identification of the aforementioned variables as crucial factors agrees with the results of previous studies in this field.

Table 1. Research Variables description and Average ReliefF score of each variable

Variables	Variable Description	ReliefF score
RLTC/RCR02	Return on Long -term capital / Return on Capital and Reserves 2002	0.02603371
AR/TA 01	Accounts Receivable/Total Assets 2001	0.02587121
TL/TA02	Total liabilities/Total assets 2002	0.02577709
AR/TA02	Accounts Receivable/Total Assets 2002	0.02257509
WC/TA 02	Working capital/total assets 2002	0.02118785
DC/CA02	Deposits and cash/current assets 2002	0.01364156
NFA/TA	Net Fixed Assets/Total Assets	0.0133596
NDAP02	Number of days accounts payable 2002	0.01085013
LTD/TCR02	Long term debt/total capital and reserves 2002	0.00798901
S/TA02	Sales/total assets 2002	0.00395956
RCF/TA02	Results carried forward/total assets 2002	0.00384807
NDAR02	Number of days accounts receivable 2002	0.00327257
CAR/TA	Change Accounts Receivable/Total Assets	0.00320415
WCL02	Working capital leveraged 2002	0.00254562
ITURN02	Inventory turnover 2002	0.00215535
TA/CR02	Total Assets/Capital and Reserves 2002	0.00208717
EBIT/TA02	Earnings before interest and tax/total assets 2002	0.00206301
CFO02	Cash flows from operations 2002	0.00169573
CFO01	Cash flows from operations 2001	0.0009421
CR02	Current assets to current liabilities 2002	0.00082761
GOCF	Growth of Operational Cash Flow	0.00073566
CAR/NS	Change Accounts Receivable/Net Sales	0.00071853
EBT02/EBIT02	Earnings before tax 2002/Earnings before interest and tax 2002	0.00049986
Z-SCORE02	Altman z-score 2002	0.00047192
CR/TL02	Capital and Reserves/total liabilities 2002	0.00041943

4 Experimental Results and Proposed Technique

Supervised machine learning is the exploration for algorithms that reason from externally supplied examples to produce general hypotheses, which will make predictions about future examples. For the purpose of this study, a representative algorithm for each learning technique was used. The most commonly used C4.5 algorithm [17] was the representative of the decision trees in our study. Back Propagation (BP) algorithm [23] - was the representative of the ANNs. The 1-NN algorithm was also used as a representative of lazy learners [23]. The Naïve Bayes algorithm [23] was the representative of the Bayesian networks in our study. Finally, the Sequential Minimal Optimization (or SMO) algorithm was the representative of the SVMs [23].

All accuracy estimates were obtained by averaging the results from stratified 10-fold cross-validation in our dataset. It must be mentioned that we used the free available source code for our experiments by the book [23]. The results are presented in Table 2 as far as the total accuracy and the accuracy per class are concerned.

Table 2. Accuracy of simple models in our dataset

	NB	C4.5	1NN	BP	LR	SMO
Total Accuracy	78.7	90.2	80.5	79.3	75.0	75.0
Fraud	56.1	73.2	61.0	58.5	34.1	7.3
Non-Fraud	82.2	95.9	87.0	86.2	88.6	97.4

From Table 2, we can conclude that decision tree outperforms the other models. In an attempt to further improve the accuracy, we try to combine a number of decision tree classifiers.

The most popular ensemble algorithms are bagging [6], boosting [12], decorate [16], rotation forest [19] and random subspace methods [21]. In bagging [6], the training set is randomly sampled k times with replacement, producing k training sets with sizes equal to the original training set. Boosting, induces the ensemble of learners by adaptively changing the distribution of the training set based on the accuracy of the previously created classifiers. There are several boosting variants; AdaBoost [12] is the most well-known. The final classification is obtained from a weighted vote of the base classifiers. On the other hand, in random subspace method [21] the classifier consists of multiple learners constructed by pseudo-randomly selecting subsets of the feature vector, that is, classifiers constructed in randomly chosen subspaces. The main idea of Rotation Forest [19] is to simultaneously encourage diversity by using Principal Components Analysis (PCA) to do feature extraction for each base classifier and accuracy is sought by keeping all principal components and also using the whole data set to train each base learner. Decorate [16] uses a strong learner to construct a diverse committee. This is accomplished by adding different randomly constructed examples to the training set when building new committee members. These artificially constructed examples are given category labels that disagree with the current classification of the committee, thereby directly increasing diversity.

Decision tree classifiers, frequently, employ post-pruning techniques that evaluate the performance of decision trees as they are pruned using a validation set [17]. Any node can be removed and assigned the most frequent class of the training examples that are sorted to the node in question. Thus, if a class is rare, decision tree algorithms often prune the tree down to a single node that classifies all instances as members of the frequent class leading to poor accuracy on the examples of minority class.

A simple method that can be used to imbalanced datasets is to reweigh training examples according to the total cost assigned to each class [3]. The idea is to change the class distributions in the training set towards the most costly class. In our case the instances of the positive (Non Fraud) class are about 3 times more than the instances of the negative class (Fraud). If the number of negative instances are artificially increased by a factor of three, then the learning system, aiming to reduce the number of classification errors, will come up with a classifier that is skewed towards the avoidance of error in the small class, since any such errors are penalized three times more. We implemented an algorithm for building an ensemble of randomizable reweighing base decision tree classifiers. Each of base decision tree classifiers is built using a reweighed different random number seed (but based one the same data). The final prediction is a straight average of the predictions generated by the individual base classifiers. Our approach is schematically represented in Fig. 1. In our case R equals to 15 (5*3). In an ensemble of classifiers about such as a number of sub-classifiers is effective [6], [12], [21]. We could use more sub-classifiers but the accuracy would not be improved enough to worth the additional training time.

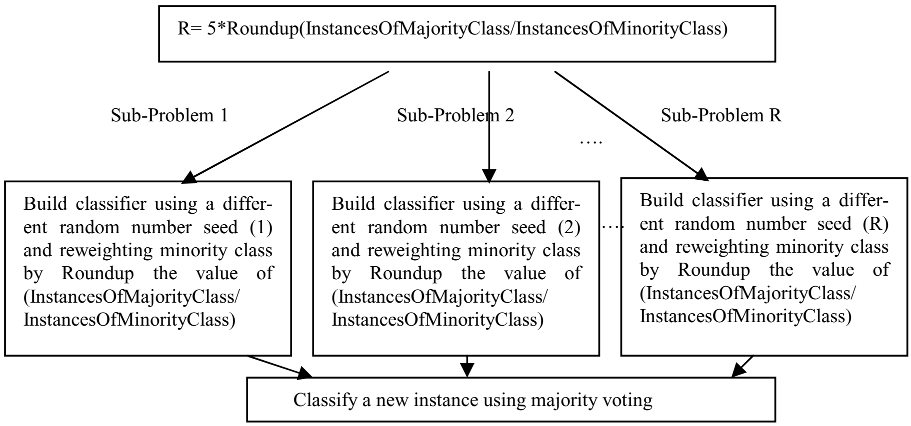

Fig. 1. The presented method

The results of the presented ensemble and that of other well known combining techniques using 25 sub-classifiers are presented in Table 3. It must be mentioned that in the experiment we also include MetaCost [11]. MetaCost is another combining method for making a classifier cost-sensitive. The procedure begins to learn an internal cost-sensitive model by applying a cost-sensitive procedure, which employs a base learning algorithm. Then, MetaCost procedure approximates class probabilities using bagging and then re-labels the training instances with their minimum expected cost classes, and finally relearns a model using the modified training set.

Table 3. Accuracy of ensembles in our dataset

ALGORITHMS	Accuracy	Fraud	Non-Fraud
Adaboost C4.5	92.1	75.6	97.6
Bagging C4.5	92.1	80.5	95.9
Decorate C4.5	86.6	61.0	95.1
Rotation Forest C4.5	85.4	51.2	96.7
Random subspace C4.5	85.4	46.3	98.4
Metacost C4.5	92.1	85.4	94.3
Random Committee of Cost-Sensitive C4.5	94.5	90.2	95.9

The presented algorithm correctly classifies 94.5% of the total sample, 90.2% of the fraud cases and 95.9% of the non-fraud cases. As a conclusion, our approach performs better than other examined ensemble methods.

5 Conclusion

Auditing practices at the present time have to cope with an increasing number of management fraud cases. Supervised machine learning techniques can assist auditors in accomplishing the task of management fraud detection. A relatively small list of financial ratios largely determines the classification results. This knowledge, coupled with machine learning algorithms, can provide models capable of achieving considerable classification accuracies.

Tracking progress is a time-consuming job that can be handled automatically by a learning tool. A screenshot of the implemented tool is presented in Fig 2. While the experts will still have an essential role in monitoring and evaluating progress, the tool can use the data required for reasonable and efficient monitoring.

It must be mentioned that our input vector solely consists of financial ratios. Enriching the input vector with qualitative information, such as previous auditors' qualifications or the composition of the administrative board, could boost the accuracy rate.

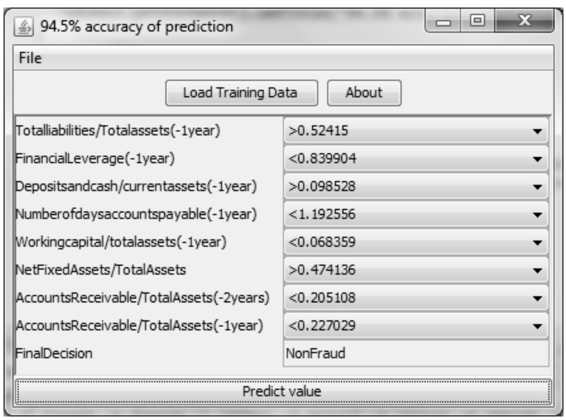

Fig. 2. A screenshot of the implemented decision support tool

References

1. Albrecht, C.C., Albrecht, W.S., Dunn, J.G.: Can auditors detect fraud: a review of the research evidence. Journal of Forensic Accounting 2(1), 1–12 (2001)
2. Ansah, S.O., Moyes, G.D., Oyelere, P.B., Hay, D.: An empirical analysis of the likelihood of detecting fraud in New Zealand. Managerial Auditing Journal 17(4), 192–204 (2002)
3. Barandela, R., Sánchez, J.S., García, V., Rangel, E.: Strategies for learning in class imbalance problems. Pattern Recognition 36(3), 849–851 (2003)
4. Bell, T., Carcello, J.: A decision aid for assessing the likelihood of fraudulent financial reporting. Auditing: A Journal of Practice & Theory 9(1), 169–178 (2000)

5. Bollen, L., Mertens, G., Meuwissen, R., VanRaak, J., Scelleman, C.: Classification and Analysis of Major European Business Failures. Maastricht Accounting, Auditing and Information Management Research Center (MARC) of University Maastricht and RSM (2005)
6. Breiman, L.: Bagging Predictors. Machine Learning 24(3), 123–140 (1996)
7. Calderon, T.G., Cheh, J.J.: A roadmap for future neural networks research in auditing and risk assessment. International Journal of Accounting Information Systems 3(4), 203–236 (2002)
8. Coffee, J.: A theory of corporate scandals: Why the USA and Europe differ. Oxford Review of Economic Policy 21(2), 198–211 (2005)
9. Deng, Q.: Application of Support Vector Machine in the Detection of Fraudulent Financial Statements. In: 4th International Conference on Computer Science & Education, pp. 1056–1059 (2009)
10. Yue, D., Wu, X., Shen, N., Chu, C.-H.: Logistic Regression for Detecting Fraudulent Financial Statement of Listed Companies in China. In: International Conference on Artificial Intelligence and Computational Intelligence, pp. 104–108 (2009)
11. Domingos, P.: MetaCost: A General Method for Making Classifiers Cost-Sensitive. In: Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining, pp. 155–164. ACM Press (1999)
12. Freund, Y., Schapire, R.E.: Experiments with a New Boosting Algorithm. In: Proceedings of ICML 1996, pp. 148–156 (1996)
13. Hoogs, B., Kiehl, T., Lacomb, C., Senturk, D.: A Genetic Algorithm Approach to Detecting Temporal Patterns Indicative of Financial Statement Fraud. Intelligent Systems in Accounting, Finance and Management 15, 41–56 (2007)
14. Kirkos, S., Spathis, C., Manolopoulos, Y.: Data Mining Techniques for the Detection of Fraudulent Financial Statements. Expert Systems with Applications 32, 995–1003 (2007)
15. Lin, J.W., Hwang, M., Becker, J.K.: A Fuzzy Neural Network for Assessing the Risk of Fraudulent Financial Reporting. Managerial Auditing Journal 18, 657–665 (2003)
16. Melville, P., Mooney, R.: Constructing Diverse Classifier Ensembles using Artificial Training Examples. In: IJCAI 2003, Mexico, pp. 505–510 (2003)
17. Quinlan, J.R.: C4.5: Programs for machine learning. Morgan Kaufmann, San Francisco (1993)
18. Ravisankar, P., Ravi, V., Raghava, G., Bose, I.: Detection of financial statement fraud and feature selection using data mining techniques. Decision Sup. Systems 50, 491–500 (2011)
19. Rodríguez, J.J., Kuncheva, L.I., Alonso, C.J.: Rotation forest: A new classifier ensemble method. IEEE Trans. Pattern Anal. Machine Intell. 28(10), 1619–1630 (2006)
20. Spathis, C., Doumpos, M., Zopounidis, C.: Detecting falsified financial statements: a comparative study using multicriteria analysis and multivariate statistical techniques. The European Accounting Review 11(3), 509–535 (2002)
21. Ho, T.K.: The Random Subspace Method for Constructing Decision Forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(8), 832–844 (1998)
22. Watts, R.L., Zimmerman, J.L.: Positive Accounting Theory. Prentice-Hall (1986)
23. Witten, I., Frank, E., Hall, M.: Data Mining: Practical Machine Learning Tools and Techniques, 3rd edn. Morgan Kaufmann (2011) ISBN 978-0-12-374856-0 and Techniques (Third Edition), Morgan Kaufmann, (2011), ISBN 978-0-12-374856-0