

DETECTING FINANCIAL FRAUD USING DATA MINING TECHNIQUES: A LITERATURE REVIEW

Mousa Albashrawi, Manning School of Business, University of Massachusetts Lowell, One University Avenue - Lowell, MA 01854, 978-930-0277, mousa_box@hotmail.com

ABSTRACT

Financial fraud has been a big concern for many organizations across industries; billions of dollars are lost yearly because of this fraud. Data mining techniques play a major role in addressing this continued and growing problem. Reviewing prior research between 2004 and 2015 reveals 40 articles that aim to resolve different types of financial fraud via data mining methods. Concluded points of this review are discussed and communicated for both researchers and practitioners.

INTRODUCTION

Financial fraud has been a big concern for many organizations across industries and in different countries since it brings huge devastations to business. Billions of dollars are lost yearly due to financial fraud; Bank of America, for example, agrees to pay \$16.5 billion for resolving financial fraud case [15]. Also, IRS [16] indicates that Mr. Walker, the founder of Bixby Energy Systems, deceived more than 1,800 investors and committed multi-million dollar fraud. Due to his actions that involve providing false statements of a) his subordinates' salaries and commissions; b) the operational capacity of the firm's core products; and c) an initial public stock offering, he was punished with a sentence of 300 months in prison. Hence, the numbers still indicate this is a growing problem, which needs more attention from professionals and academicians.

Financial fraud detection tools have been brought to the scenic in order to address this problem and to provide reliable solutions to business. Data mining, defined as "a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and subsequently gain knowledge from a large database" [13], is a major contributor for detecting different types of financial fraud through its diverse methods, for example, logistic regression, decision tree, support vector machine (SVM) and naïve Bayes. Some of these techniques outperform the others in specific financial contexts. Glancy and Yadav [14] divide these contexts to three main areas: internal, insurance and credit. However, Jans et al. [24] expand internal fraud and classifies it into two categories: financial statement fraud and transaction fraud. Financial statement fraud, known as management fraud, is defined as "the intentional misstatement of certain financial values to enhance the appearance of profitability and deceive shareholders or creditors" while transaction fraud captures the process of snatching organizational assets. This has motivated us to 1) reveal which context should implement what technique of data mining, 2) unfold what technique can yield high classification accuracy in detecting financial fraud, and 3) provide a new classification framework for financial fraud.

Although detecting financial fraud is considered a high priority for many organizations, the current literature lacks for an up-to-date and in-depth review that help firms make their decisions of selecting the appropriate data mining technique. Ngai et al. [13] provide a well-organized and detailed literature review, ranging from 1997 to 2008, of detecting financial fraud via data mining methods. However, the specified time-period is not able to capture the increasing trend of research in this area, which occurs in 2009, 2011, and 2012. Thus, our primary contribution of this paper is twofold; the first is to provide an up-to-date and comprehensive analysis of this crucial emerging topic as an extension to Ngai et al.'s

[13] review. The second is to provide scholars and practitioners with an excellent source of data mining applications used in financial fraud for their fast access and use. This review is, however, an attempt to leverage our knowledge and to increase our comprehension of data mining applications in financial fraud.

METHOD

A number of keywords was used to identify the pertinent articles, for instance, “detecting financial fraud, financial fraud and data mining, financial fraud detection, and detecting financial fraud via data mining”. Most of the relevant articles were found in MIS related journals, e.g., Decision Support Systems and Expert Systems with Applications but some were found in finance and economic related journals, e.g., the Journal of Risk and Insurance, and Applied Economics. Table 2 lists all journals included in our analysis.

Although the majority of the articles retrieved from Science Direct, the search spanned other online databases (e.g., Emerald, Elsevier, World Scientific, IEEE, and Routledge - Taylor and Francis Group). Our search yielded a sample of 40 relevant articles (39 peer-reviewed journal articles with one conference paper). 11 out of 40 articles were found in Expert Systems with Applications while 6 found in Decision Support Systems. These two journals are considered to have the biggest share (42.5%) of the articles that have addressed this topic (Table 1). However, most of the articles had been conducted in the United States, followed by Taiwan while an equal number of studies conducted in both Spain and China.

Table 1
Distribution of articles by journals (2004–2015)

Journal Title	Frequency	Percentage (%)
<i>Expert Systems with Applications</i>	11	27.50
<i>Decision Support Systems</i>	6	15.0
<i>Knowledge-Based Systems</i>	2	5.0
<i>The Journal of Risk and Insurance</i>	2	5.0
<i>MIS Quarterly</i>	1	2.5
<i>Management Science</i>	1	2.50
<i>Contemporary Accounting Research</i>	1	2.50
<i>Journal of Forecasting</i>	1	2.50
<i>International Journal of Intelligent Systems in Accounting and Finance Management</i>	1	2.50
<i>Journal of Data Science</i>	1	2.50
<i>Managerial Auditing Journal</i>	1	2.50
<i>Computers in Human Behavior</i>	1	2.50
<i>Information Fusion</i>	1	2.50
<i>Journal of Practice & Theory</i>	1	2.50
<i>Computational Intelligence</i>	1	2.50
<i>Journal of Economic Policy Reform</i>	1	2.50
<i>IEEE Transaction on Knowledge and Data Engineering</i>	1	2.50
<i>Journal of Money Laundering Control</i>	1	2.50
<i>International Journal of Pattern Recognition and Artificial Intelligence</i>	1	2.50
<i>Insurance: Mathematics and Economics</i>	1	2.50
<i>International Journal of Information Technology & Decision Making</i>	1	2.50
<i>Applied Economics</i>	1	2.50
<i>European Journal of Operational Research</i>	1	2.50
Total	40	100

RESULTS

Classification Framework Based on Application

According to the reviewed forty articles in this area, it is possible to classify financial fraud into four major categories, namely, financial statement fraud, bank fraud, insurance fraud, and other related financial fraud (Table 2).

Table 2

Classification of fraud types examined by data mining methods in one decade

Fraud Type (application)	Amount	Description
Financial statement fraud	13	It is the act of intentional or irresponsible conducts. This produces materially misleading financial statements [4].
Bank fraud	11	“Whoever knowingly executes, or attempts to execute, a scheme or artifice— (1) to defraud a financial institution; or (2) to obtain any of the moneys, funds, credits, assets, securities, or other property owned by, or under the custody or control of, a financial institution, by means of false or fraudulent pretenses, representations, or promises” [6]. Bank fraud is sub-categorized here into credit card fraud, money laundering, and fraudulent bank account.
Insurance fraud	9	This term is broadly labeled as insurance abuse, especially in practice [38]. Insurance fraud includes here automotive insurance fraud, health insurance, and crop insurance fraud.
Other related financial fraud	7	Other financial fraud category includes general financial fraud, fraudulent financial reporting, financial fraud by top management, tax fraud and transaction fraud.
Total	40	

Table 3 further provides in-depth analysis by observing the frequency of the sub-categories of financial fraud types in the 10-year period.

Table 3

Further break-down for fraud types with corresponding data mining techniques

Techniques	Fraud Types					
	Financial statement fraud	Bank fraud			Insurance fraud	Other related financial fraud
		Credit card fraud	Money laundering	Fraudulent Bank account		
Logistic regression	2	2			4	
Neural networks	2	1			1	2
Decision trees	1	2				2
Discriminant analysis	1	2				1
Bayesian networks	1	1		1	1	1
SVM	2	2				3
Nearest neighbor	1	1				
UTADIS	1					
Association rules		1		1		
Rule-based filtering		1				
Dempster–Shafer adder		1				
Naïve Bayes		1			1	2
Three-phase cutting plane algorithm	1					

A multiple-criteria index		1	
Text mining			1
Process mining			1
Random forests		1	1
Response surface method	1		
Genetic programming	1		
GMDH	1		
Bagging and Stacking	1		
Stochastic gradient boosting			1
Rule ensemble			1
MetaFraud framework			1
GHSOM	1		
Network analysis		1	
CART	1		
Self-organizing map		1	
Probit model			2

LIMITATIONS AND CONCLUSION

This review suffers from some limitations. First, it does not consider all sub-categories of financial fraud, i.e., advanced-fee fraud that targets a very large number of people who looks for “work-from-home” jobs. This fraud deceives people to pay fee in advance so that they get the offer but once the fee is collected, they do not realize the expected benefits. Second, a decade review may not be sufficient to address this growing problem as it started when the business started. Third, the forty articles explored may not reveal the entire story of data mining usage in the domain of financial fraud; several online databases need to be included in the sample for more powerful presentation and analysis.

It is, however, crucial to have a wide-ranging review on detecting financial fraud to increase the understanding and to expand the knowledge of this area among researchers and professionals. This review sheds light on different important and valuable aspects of financial fraud detection:

- It provides a fast and easy-to-use source either for scholars or practitioners who are interested in this topic.
- It shows that logistic regression, decision tree, SVM, NN and Bayesian networks have been widely used (usage percentage > 50%) to detect financial fraud even though they are not always associated with the best results of classification.
- This review classifies financial fraud into four major types - financial statement fraud, bank fraud, insurance fraud, and other related financial fraud. This broad classification can enable us to further categorize any new type of financial fraud. And it is obvious that financial statement fraud is being the most type of financial fraud to be examined. This reflects its importance, which in turn, should make practitioners more cautious when they scan or process their financial statements.
- There has been a huge increase of research conducted to address this topic in the years of 2009, 2011 and 2012. These three years account approximately for 50% of the publications in the 10-year period. More notably, the amount of research increased by 67% in 2011 compared to the previous year.
- It is possible to conclude that the countries (United States, Taiwan, Spain, and China) indicated to have the biggest portion (70%) of publications about this topic, are being more exposed to it.

The highlighted aspects through this review can provide organizations with useful information regarding financial fraud and data mining techniques. Organizations may be able to select the suitable technique once considering its particular usage context and frequency. This could lead to achieve a higher level of accuracy in detecting financial fraud. Besides other benefits, researchers can take advantage of knowing the most used methods and in which context so that they can develop a research project to either investigating such method in a different context or suggesting a new innovative method in a similar context.

REFERENCES

- [1] A. Abbasi, C. Albrecht, A. Vance, and J. Hansen, Metafraud: A Meta-Learning Framework for Detecting Financial Fraud, *MIS Quarterly*. 36(4), 2012, 1293–1327.
- [2] A. Gepp, J. H. Wilson, K. Kumar, and S. Bhattacharya, A Comparative Analysis of Decision Trees Vis-a-vis Other Computational Data Mining Techniques in Automotive Insurance Fraud Detection, *Journal of Data Science*. 10, 2012, 537–561.
- [3] A. I. Dimitras, Evaluation of Greek Construction Companies' Securities Using UTADIS Method, *European Research Studies*. 2002, 95–107.
- [4] B. Bai, J. Yen, and X. Yang, False Financial Statements: Characteristics of China's Listed Companies and CART Detecting Approach, *International Journal of Information Technology & Decision Making*. 7(2), 2008, 339–359.
- [5] B. Dikmen, and G. Kùçùkkocaođlu, The Detection of Earnings Manipulation: The Three-Phase Cutting Plane Algorithm Using Mathematical Programming, *Journal of Forecasting*. 29(5), 2010, 442–466.
- [6] Cornell University Law School, White-Collar Crime: An Overview, 2015, <https://www.law.cornell.edu/uscode/text/18/1344>
- C. Gaganis, Classification Techniques for the Identification of Falsified Financial Statements: A Comparative Analysis, *International Journal of Intelligent Systems in Accounting and Finance Management*. 16, 2009, 207–229.
- [7] C. Holton, Identifying Disgruntled Employee Systems Fraud Risk Through Text Mining: A Simple Solution for a Multi-Billion Dollar Problem, *Decision Support Systems*. 46, 2009, 853–864.
- [8] D. G. Whiting, J. V. Hansen, J. B. McDonald, C. Albrecht, and W. S. Albrecht, Machine Learning Methods for Detecting Patterns of Management Fraud, *Computational Intelligence*. 28(4), 2012, 505–527.
- [9] D. Olszewski, Fraud Detection Using Self-Organizing Map Visualizing the User Profiles, *Knowledge-Based Systems*. 70, 2014, 324–334.
- [10] D. Sánchez, M. Vila, L. Cerda, and J. Serrano, Association Rules Applied to Credit Card Fraud Detection, *Expert Systems with Applications*. 36(2), 2009, 3630–3640.
- [11] D. Skillicorn, and L. Purda, Detecting Fraud in Financial Reports, *European Intelligence and Security Informatics Conference*. 2012, 7–13.
- [12] E. Kirkos, C. Spathis, and Y. Manolopoulos, Data Mining Techniques for the Detection of Fraudulent Financial Statements, *Expert Systems with Applications*. 32, 2007, 995–1003.
- [13] E. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, The Application of Data Mining Techniques in Financial Fraud Detection: A Classification Framework and an Academic Review of Literature, *Decision Support Systems*. 50, 2011, 559–569.
- [14] F. H. Glancy, and S. B. Yadav, A Computational Model for Financial Reporting Fraud Detection, *Decision Support Systems*. 50, 2011, 595–601.
- [15] Fox News, Bank of America pays \$16.5 bn to settle financial fraud case, 2014), <http://latino.foxnews.com/latino/news/2014/08/21/bank-america-pays-165-bn-to-settle-financial-fraud-case/>
- [16] IRS, Examples of Corporate Fraud Investigations, 2014, <http://www.irs.gov/uac/Examples-of--Corporate-Fraud-Investigations-Fiscal-Year-2014>
- [17] I. Yeh, and C. Lien, The Comparisons of Data Mining Techniques for the Predictive Accuracy of Probability of Default of Credit Card Clients, *Expert Systems with Applications*. 36(2), 2009, 2473–2480.
- [18] J. Perols, Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms, *Auditing: A Journal of Practice & Theory*. 30(2), 2011, 19–50.
- [19] J. Pinquet, M. Ayuso, and Guill'en, Selection Bias and Auditing Policies for Insurance Claims, *The Journal of Risk and Insurance*. 74(2), 2007, 425–440.
- [20] J. W. Lin, M. I. Hwang, and J. D. Becker, A Fuzzy Neural Network for Assessing the Risk of Fraudulent Financial Reporting, *Managerial Auditing Journal*. 18(8), 2003, 657–665.

- [21] K. A. Kaminski, T. S. Wetzel, and L. Guan, Can Financial Ratios Detect Fraudulent Financial Reporting, *Managerial Auditing Journal*. 19(1), 2004, 15–28.
- [22] L. Bermúdez, J. Pérez, M. Ayuso, E. Gómez, and F. Vázquez, A Bayesian Dichotomous Model with Asymmetric Link for Fraud in Insurance, *Insurance: Mathematics and Economics*. 42(2), 2008, 779–786.
- [23] M. Cecchini, H. Aytug, G. Koehler, and P. Pathak, Detecting Management Fraud in Public Companies, *Management Science*, 56(7), 2010, 1146–1160.
- [24] M. Jans, J.M. Werf, N. Lybaert, and K. Vanhoof, A Business Process Mining Application for Internal Transaction Fraud Mitigation, *Expert Systems with Applications*. 38, 2011, 13351–13359.
- [25] N. Mahmoudi, and E. Duman, Detecting Credit Card Fraud by Modified Fisher Discriminant Analysis, *Expert Systems with Applications*. 42(5), 2015, 2510–2516.
- [26] O. R. Bidder, H.A. Campbell, A. Gómez-Laich, P. Urgé, J. Walker, Y. Cai, and R. P. Wilson, Love The Neighbour: Automatic Animal Behavioural Classification of Acceleration Data Using the K-Nearest Neighbour Algorithm, *PLoS ONE*. 9(2), 2014, 1–7.
- [27] P.C. González, and J. D. Velásquez, Characterization and Detection of Taxpayers with False Invoices Using Data Mining Techniques, *Expert Systems with Applications*. 40, 2013, 1427–1436.
- [28] P. Dechow, W. Ge, C. Larson, and R. Sloan, Predicting Material Accounting Misstatements, *Contemporary Accounting Research*. 28(1), 2011, 1–16.
- [29] P-F. Pai, M-F. Hsu, and M-C. Wang, A Support Vector Machine-Based Model for Detecting Top Management Fraud, *Knowledge-Based Systems*. 24, 2011, 314–321.
- [30] P. Ravisankar, V. Ravi, G. R. Rao, and I. Bose, Detection of Financial Statement Fraud and Feature Selection Using Data Mining Techniques, *Decision Support Systems*. 50, 2011, 491–500.
- [31] R. Chen, T. Chen, and C. Lin, A New Binary Support Vector System for Increasing Detection Rate of Credit Card Fraud, *International Journal of Pattern Recognition and Artificial Intelligence*. 20(2), 2006, 227–239.
- [32] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, Data Mining for Credit Card Fraud: A Comparative Study, *Decision Support Systems*. 50, 2011, 602–613.
- [33] S. Caudill, M. Ayuso, and M. Guillen, Fraud Detection Using A Multinomial Logit Model with Missing Information, *The Journal of Risk and Insurance*. 72(4), 2005, 539–550.
- [34] S. L. Humpherys, K. C. Moffitt, M.B. Burns, J.K. Burgoon, and W. F. Felix, Identification of Fraudulent Financial Statements Using Linguistic Credibility Analysis, *Decision Support Systems*. 50, 2011, 585–594.
- [35] S-H. Li, D. C. Yen, W-H. Lu, and C. Wang, Identifying the Signs of Fraudulent Accounts Using Data Mining Techniques, *Computers in Human Behavior*. 28, 2012, 1002–1013.
- [36] S. Panigrahi, A. Kundu, S. Sural, and A. Majumdar, Credit Card Fraud Detection: A Fusion Approach Using Dempster–Shafer Theory and Bayesian Learning, *Information Fusion*. 10, 2009, 354–363.
- [37] S. Viaene, R. Derrig, and G. Dedene, A Case Study of Applying Boosting Naive Bayes to Claim Fraud Diagnosis, *IEEE Transactions on Knowledge and Data Engineering*. 16 (5), 2004, 612 - 620.
- [38] S. Viaene, M. Ayuso, M. Guillen, D. Gheel, and G. Dedene, Strategies for Detecting Fraudulent Claims in the Automobile Insurance Industry, *European Journal of Operational Research*. 176(1), 2007, 565–583.
- [39] S. Viaenea, G. Dedene, and R. Derrig, Auto Claim Fraud Detection Using Bayesian Learning Neural Networks, *Expert Systems with Applications*. 29(3), 2005, 653–666.
- [40] S. Yang, and L. Wei, Detecting Money Laundering Using Filtering Techniques: A Multiple-Criteria Index, *Journal of Economic Policy Reform*. 13(2), 2010, 159–178.
- [41] S-Y. Huang, R-H. Tsaih, and F. Yu, Topological Pattern Discovery and Feature Extraction for Fraudulent Financial Reporting, *Expert Systems with Applications*. 41, 2014, 4360–4372.
- [42] W. Yang, and S. Hwang, A Process-Mining Framework for the Detection of Healthcare Fraud and Abuse, *Expert Systems with Applications*. 31(1), 2006, 56–68.
- [43] W. Zhou, and G. Kapoor, Detecting Evolutionary Financial Statement Fraud, *Decision Support Systems*. 50(3), 2011, 570–575.
- [44] Y. Jin, R. Rejesus, and B. Little, Binary Choice Models for Rare Events Data: A Crop Insurance Fraud Application, *Applied Economics*. 37(7), 2005, 841–848.
- [45] Y. Sahin, S. Bulkan, and E. Duman, A Cost-Sensitive Decision Tree Approach for Fraud Detection, *Expert Systems with Applications*. 40(15), 2013, 5916–5923.
- [46] Z. Gao, and M. Ye, A Framework for Data Mining-Based Anti-Money Laundering Research, *Journal of Money Laundering Control*. 10(2), 2007, 170–179.