Expert Systems With Applications 90 (2017) 374–393

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Detecting financial restatements using data mining techniques

Ila Dutta[a], Shantanu Dutta[b], Bijan Raahemi[c,*]

[a] University of Ottawa, School of Electrical Engineering and Computer Science and Knowledge Discovery and Data Mining Lab, 800 King Edward Avenue, Ottawa, ON K1N6N5, Canada
[b] University of Ottawa, Telfer School of Management, 55 Laurier Avenue East, Ottawa, ON K1N6N5, Canada
[c] University of Ottawa, Telfer School of Management and Knowledge Discovery and Data Mining Lab, 55 Laurier Avenue East, Ottawa, ON K1N6N5, Canada

A R T I C L E I N F O

Article history:
Received 24 January 2017
Revised 13 August 2017
Accepted 14 August 2017
Available online 16 August 2017

Keywords:
Data mining
Financial restatements
Decision tree (DT)
Artificial neural network (ANN)
Naïve Bayes (NB)
Support vector machine (SVM)

A B S T R A C T

Financial restatements have been a major concern for the regulators, investors and market participants. Most of the previous studies focus only on fraudulent (or intentional) restatements and the literature has largely ignored unintentional restatements. Earlier studies have shown that large scale unintentional restatements can be equally detrimental and may erode investors' confidence. Therefore it is important for us to pay a close to the significant unintentional restatements as well. A lack of focus on unintentional restatements could lead to a more relaxed internal control environment and lessen the efforts for curbing managerial oversights and instances of misreporting. In order to address this research gap, we focus on developing predictive models based on both intentional (fraudulent) and unintentional (erroneous) financial restatements using a comprehensive real dataset that includes 3,513 restatement cases over a period of 2001 to 2014. To the best of our knowledge it is the most comprehensive dataset used in the financial restatement predictive models. Our study also makes contributions to the datamining literature by (i) focussing on various datamining techniques and presenting a comparative analysis, (ii) ensuring the robustness of various predictive models over different time periods. We have employed all widely used data mining techniques in this area, namely, Decision Tree (DT), Artificial Neural Network (ANN), Naïve Bayes (NB), Support Vector Machine (SVM), and Bayesian Belief Network (BBN) Classifier while developing the predictive models. We find that ANN outperforms other data mining algorithms in our empirical setup in terms of accuracy and area under the ROC curve. It is worth noting that our models remain consistent over the full sample period (2001-2014), pre-financial-crisis period (2001-2008), and post-financial-crisis period (2009-2014). We believe this study will benefit academics, regulators, policymakers and investors. In particular, regulators and policymakers can pay a close attention to the suspected firms and investors can take actions in advance to reduce their investment risks. The results can also help improving expert and intelligent systems by providing more insights on both intentional and unintentional financial restatements.

1. Introduction

Financial restatements have been a major concern for the regulators, investors and market participants. In the event of a material inaccuracy, firms need to revise their previously published financial statement(s) which is termed as financial restatements. There are two types of financial restatements: intentional (i.e. frauds) and unintentional (i.e. material errors in financial statements). Most of the data mining predictive models concentrate on fraudulent or intentional restatements for model building (e.g. Cecchini, Aytug, Koehler, & Pathak, 2010; Kim, Baik, & Cho, 2016) and unintentional restatements have received very little attention in the literature. However, focussing only on fraudulent cases has some practical shortcomings.

First, many companies that manipulates earnings, remain undetected and the financial restatements are wrongly categorized as unintentional restatement (Dechow, Ge, Larson, & Sloan, 2011). Detected intentional/fraudulent restatements are relatively rare as compared to the unintentional cases. Our sample shows that unintentional restatements are almost 31 times (3404 unintentional restatement to 109 intentional/ fraudulent restatement) more prevalent than fraudulent (or intentional) restatements. While unintentional restatements are more common in general, this can also be attributed to the evolving nature of financial frauds and an increased level of sophistication to hide fraudulent activities. Therefore, exclusion of unintentional restatements from the sample, may lead to incomplete predictive models.

* Corresponding author.
E-mail addresses: idutt040@uottawa.ca (I. Dutta), shantanu.dutta@telfer.uottawa.ca (S. Dutta), braahemi@uottawa.ca, braahemi@gmail.com, raahemi@telfer.uottawa.ca (B. Raahemi).

http://dx.doi.org/10.1016/j.eswa.2017.08.030
0957-4174/

Second, focussing primarily on fraudulent restatements may create a misleading perception that unintentional restatements are somewhat not detrimental for a firm and its investors. Earlier studies have shown that magnitude of the market reactions to large scale unintentional restatements are quite similar to that of fraudulent restatements and can erode investors' confidence (Bowen, Dutta, & Zhu, 2017; Palmrose, Richardson, & Scholz, 2004). Therefore it is important for us to pay a close to the significant unintentional restatements as well. A lack of focus on unintentional restatements could lead to a more relaxed internal control environment and lessen the efforts for curbing managerial oversights and instances of misreporting.

In order to address this research gap, we focus on developing predictive models based on both intentional (fraudulent) and unintentional (erroneous) financial restatements using a comprehensive real dataset that includes 3513 restatement cases over a period of 2001–2014. To the best of our knowledge it is the most comprehensive dataset used in the financial restatement predictive models. Our study also makes contributions to the datamining literature by (i) focussing on various datamining techniques and presenting a comparative analysis, (ii) ensuring the robustness of various predictive models over different time periods. The literature suggests that there is no consensus on the appropriate datamining techniques, primarily due to the evolving nature of restatement activities and changes in business environment. As Cecchini et al. (2010) suggest, "…future studies should take into account the fact that fraudsters will change their tactics to hide fraud" (Cecchini et al., 2010, p. 1158). It appears that effectiveness of any particular technique depends on the nature of dataset and different time periods. We address this concern in the following ways.

First, we have employed all widely used data mining techniques in this area (Albashrawi, 2016; Ngai, Hu, Wong, Chen, & Sun, 2011), namely, Decision Tree (DT), Artificial Neural Network (ANN), Naïve Bayes (NB), Support Vector Machine (SVM), and Bayesian Belief Network (BBN) Classifier while developing the predictive models. This allows us to explore the most suitable data mining technique – leading to the best possible predictive model in a comprehensive dataset. In order to ensure the robustness of the predictive models, we have considered a comprehensive list of attributes (116 firm- and industry-specific attributes) while building our models.[1] Subsequently, in order to develop more parsimonious models, we performed feature subset selection using stepwise forward selection to reduce less significant or redundant attributes. Our final dataset includes 15 attributes. Another challenge with a financial restatement study is the rarity of restatement events which result into class imbalance problem. In order to address this issue we use an oversample technique, namely Synthetic Minority Oversampling Technique (SMOTE), which increases the instances of minority class and leads to informational gain in the dataset. Our results show that Artificial Neural Network (ANN) leads to the best predictive model with an AUC value of 83%. Decision Tree (DT) and Bayesian Belief network (BBN) algorithms show a similar result while the application of Support Vector machine (SVM) and Naïve Bayes (NB) algorithm lead to weaker predictive models.

Second, there is a concern that restatement activities evolve over time and hence we need to be careful about the stability of predictive models in different time periods. We use the recent financial crisis of 2008 as an external shock to the market and examine whether predictive models remain stable during pre- and post-crisis periods. The financial crisis has exposed a series of scandalous and unethical activities (e.g. subprime mortgage scandal, stock option manipulation, accounting restatements) undertaken by the executives of large U.S. firms. As a result, firms are subjected to more intense scrutiny in the post crisis period. Does a renewed focus on financial integrity have an impact on financial restatement predictive model? In order to address this concern, we develop predictive models for three different time periods: (i) full sample period (i.e. 2001–2014), (ii) pre-crisis period (i.e. till 2008), and (iii) post-crisis period (i.e. post 2008). Our results show consistent performance of various predictive models in different time periods.

The data mining literature further suggests that data quality is quite important for developing reliable predictive models (Banarescu, 2015). We have taken a number of steps to ensure the integrity of data quality. First, we have considered a comprehensive list of restatement cases in developing our predictive models. Only a few studies have focussed on unintentional restatement cases while developing predictive models. However, it appears that number of restatement cases included in those studies are quite limited.[2] Therefore, by employing a more comprehensive dataset on financial restatement in this study, we are able to present more reliable predictive models. Second, some of the restatement cases could be very trivial and may create 'noise' in predictive models. In order to eliminate the trivial restatement instances we retain only the "adverse effect" [3] cases, as identified in Audit Analytics databases. We also exclude the cases that involve clerical errors and are associated with accounting rule change.

Third, it is also quite important to recognize that the choice of restatement data sources may significantly affect the validity of predictive models. Majority of the studies use U.S. GAO database. However, since this database mainly focusses on restatement announcement year and does not include the information on restatement periods, it is quite challenging to use this dataset to develop predictive models (Dechow et al., 2011). The other commonly used database, Accounting and Auditing Enforcement Releases (AAER) database, mainly includes intentional restatement cases. While this database is quite useful for developing predictive models for fraudulent cases, it is less effective for developing a more holistic model that focusses on all 'adverse effect' restatements. In order to overcome these challenges, we use Audit Analytics database that includes a comprehensive set of both intentional and unintentional restatement cases.

Overall, the study makes the following contributions to the datamining literature: (i) unlike earlier studies that focusses only on financial frauds, we consider both intentional (fraud) and unintentional restatements in this study, (ii) we use the most comprehensive restatement dataset in developing our predictive models, (iii) we employ all widely used datamining techniques in this study and present a comparative performance analysis, and (iv) we examine the robustness of various predictive models over different time periods. We believe this study will benefit academics, regulators, policymakers and investors. In particular, regulators and policymakers can pay a close attention to the suspected firms and investors can take actions in advance to reduce their investment

[1] We have primarily used yearly attributes (predictors) in the prediction models as our class variables are based on yearly observation. In our models, we have not used any quarterly attributes. Because quarterly attributes may suffer from seasonality bias.

[2] For example, Kim et al. (2016) include 355 unintentional cases in their study that focusses on predictive model developments; other studies that not necessarily focus on predictive model developments, also use fewer unintentional restatement cases. For example, Hennes et al. (2008) include only 83 unintentional cases, and Hennes et al. (2014) uses approximately 1,500 restatement cases. Both studies focusses on the consequences of restatements and do not employ any data mining technique to develop a predictive model.

[3] Many of the restatement cases are trivial and do not affect shareholder value. We do not consider such cases in our study and focus only on "adverse effect" restatements. "Adverse effect" cases refer to restatement instances that impact the financial statements (e.g. income statement) negatively. These include both unintentional restatements and fraudulent cases. This information is obtained from Audit Analytics database.

risks. The results can also help improving expert and intelligent systems by providing more insights on both intentional and unintentional financial restatements.

In the following section, we review relevant literature. In Section 3, we present the research methodology. In Section 4, we describe different datasets and data pre-processing. Section 5 discusses various components of model building. We then present the results in Section 6. Section 7 concludes this paper.

2. Related works

There is a large body of literature on financial restatements. These studies can be broadly categorized into three groups: (i) market reactions to financial restatements, (ii) actions following financial restatements; (iii) financial restatements/fraud detection models. While our primary interest is to review the usage and effectiveness of various detection models and data mining techniques, we also pay attention to two other categories of literature to address important methodological issues: data sources and data imbalance. Earlier studies show that these two factors can impact the model outcome considerably. We highlight these two issues more elaborately towards the end of this sub-section.

2.1. Market reactions to financial restatements

One of the earlier studies by Turner, Dietrich, Anderson, and Bailey (2001) examined the market reactions to 173 restatement cases from 1997 to 1999 period. They report significant negative abnormal market returns, ranging from −12.3% to −5% for different types of restatement over an eight day event window. Palmrose et al. (2004) identified 403 cases by searching the Lexis-Nexis News Library using key-word searches for restatements over the period of 1995 and 1999. They find an average abnormal return of about −9 percent over a 2-day announcement window for these restatement cases. While these results are significant – both economically and statistically – market reacts even more negatively for the restatement cases involving frauds (−20%). A later study by Files, Swanson, and Tse (2009) used a restatement sample from GAO database and analyzed 381 cases over a period of 1997–2002. They report an abnormal returns of −5.5% over a 3-day (−1 to +1) event window. They further report that market reactions vary according to the prominence of announcements. Overall, these studies report that financial restatements destroy shareholder value quite significantly.

2.2. Actions following financial restatements

Given the significant impact of financial restatements on shareholders' wealth and firm reputation, a number of studies have focused on the actions taken by firms and their boards following such instances. As discussed in Hennes, Leone, and Miller (2014) firms take both short-term and long-term actions following a restatement. In the short-term, relevant studies report that restating firms generally reshape their governance structure and involved committees and replace responsible executives. For example, Farber (2005) find that following a fraudulent restatement, a firm improves its board structure by inducting more independent directors within next three years. Further, fraud firms hold more audit committee meeting compared to their matched firms following the damaging restatements. Desai, Hogan, and Wilkins (2006), Hennes, Leone, and Miller (2008), and Burks (2010) among others find that restating firms are more likely to replace their top executives following significant restatements.

In order to restore the accounting and auditing quality and to prevent future restatement instances, many firms also replace their auditors. Srinivasan (2005) reports that auditor turnover is significantly higher for restating firms than for non-restating firms. Hennes et al. (2014) report that probability of auditor turnover is related to the severity of financial restatements. However, the results are more pronounced for non-Big 4 auditors rather than Big 4 auditors, which can be attributable to the higher switching costs in replacing a Big 4 auditor. Auditor turnover could also be attributed to auditor resignation, as restatements are damaging for auditor reputation. Huang and Scholz (2012) find that due to increased client risk, reputed auditing firms are more likely to resign following restating instances.

2.3. Studies on financial restatements/fraud detection models

As discussed above, financial restatements can have a significant negative effect on shareholders' wealth. In response, firms and auditors take many actions to mitigate reputational damage, restore investor confidence and lower the probability of future restatement incidences. However, it is important to note that there are two types of financial restatements: intentional (i.e. frauds) and unintentional (i.e. material errors in financial statements). Most of the data mining predictive models concentrate on fraudulent or intentional restatements for model building (e.g. Cecchini et al., 2010; Kim et al., 2016) and unintentional restatements have received very little attention in the literature. Further, we find that there is no consensus on the best financial restatement model. It appears that efficiency of detection models depend on data structure and time period. Given the focus of this study, we present a brief overview of data mining studies involving both types of financial restatements. We also summarize the results in Table 1 towards the end of this sub-section.

Existing studies on financial fraud detection mainly explore different classification techniques. In an empirical study, Lin, Chiu, Huang, and Yen (2015) employ logistic regression, decision trees (CART) and artificial neural networks (ANN) to detect financial statement fraud. The authors concluded that ANN and CART provide better accuracy than the logistic model. In another comparative study, Ravisankar, Ravi, Rao, and Bose (2011) apply logistic regression, Support Vector Machine (SVM), Genetic Programing (GP), Probabilistic Neural Network (PNN), multilayer feed forward Neural Network (MLFF) and group method of data handling (GMDH) to predict the financial fraud. The authors used the *t*-statistics to achieve feature selection on the dataset. They conducted the experiment with or without the feature selection on different classifiers. Their study results showed that PNN has outperformed all the other techniques in both cases.

Ata and Seyrek (2009) use decision tree (DT) and neural networks for the detection of fraudulent financial statements. The authors selected the variables that are related to liquidity, financial situation and the profitability of the firms in the selected dataset. The authors concluded that neural network did better prediction of the fraudulent activities than decision tree method. Kirkos, Spathis, and Manolopoulos (2007) explore the performance of Neural Networks, Decision Tree and Bayesian Belief Network to detect financial statement fraud. Among these three models, Bayesian Belief Network showed the best performance in terms of class accuracy. Kotsiantis, Koumanakos, Tzelepis, and Tampakas (2006) examine the performance of Neural Networks, Decision Tree, Bayesian Belief Network, support vector machine and *K*-Nearest Neighbour to detect financial fraud. They proposed and implemented a hybrid decision support system to identify the factors which would be used to determine the probability of fraudulent financial statements. According to the authors, the proposed stacking variant method achieves better performance than other described methods, such as, BestCV, Grading, and Voting. Both studies (Kirkos et al., 2007 and Kotsiantis et al., 2006) used accuracy as performance measure,

Table 1
Summary of data mining studies.

Sl. No	Author	Journal and Year of publication	Paper	Data sample	Database used	Main findings
1	Green and Choi	Auditing: A Journal of Practice and Theory, 1997	Assessing the risk of management fraud through neural-network technology	1982–1990	SEC/AAER, Compustat	Green and Choi develop neural network fraud classification model employing endogenous financial data
2	Feroz et al.	International Journal of Intelligent Systems in Accounting, Finance and Management, 2000	The Efficacy of Red Flags in Predicting the SEC's Targets		SEC/AAER	The authors use ANN and logistic regression to predict possible fraudster and accounting manipulators. The ANN models classify the membership in target (investigated) versus control (non-investigated) firms with an average accuracy of 81%.
3	Bolton et al.	Statistical Science, 2002	Statistical Fraud Detection: A Review	–	–	Bolton and Hand review statistical and data mining techniques used in several subgroups within fraud detection, such as, telecommunications fraud, credit card fraud, medical and scientific fraud, and as well as in other domains such as money laundering and intrusion detection. The authors describe the available tools and techniques available for statistical fraud detection.
4	Kotsiantis et al.	International Journal of Computational Intelligence, 2006	Forecasting Fraudulent Financial Statements using data Mining	2001–2002	Athens Stock Exchange	Kotsiantis et al. (2006) examine the performance of Neural Networks, Decision Tree, Bayesian Belief Network and *K*-Nearest Neighbour to detect financial fraud. They proposed and implemented a hybrid decision support system to identify the factors which would be used to determine the probability of fraudulent financial statements. According to the authors, the proposed stacking variant method achieves better performance than other described methods (BestCV, Grading, and Voting).
5	Kirkos et al.	Expert Systems with Applications, 2007	Data Mining techniques for the detection of fraudulent financial statements	–	Athens financial and taxation databases and Athens Stock Exchange	Kirkos et al. explore the performance of Neural Networks, Decision Tree and Bayesian Belief Network to detect financial statement fraud. Among these three models, Bayesian Belief Network showed the best performance.
6	Ata et al.	The Journal of Faculty of Economics and Administrative Sciences, 2009	The use of data mining techniques in detecting fraudulent financial statements: An application on manufacturing firms	2005	Istanbul Stock Exchange	Ata et al. use decision tree (DT) and neural networks for the detection of fraudulent financial statements. The authors selected the variables that are related to liquidity, financial situation and the profitability of the firms in the selected dataset. The authors conclude that neural network did better prediction of the fraudulent activities than decision tree method.
7	Cecchini et al.	Management Science, 2010	Detecting management fraud in public companies	1991–2003	SEC/ AAER	The authors provide a methodology based on support vector machine and domain specific kernel to implicitly map the financial attributes to ratio and year over year changes of the ratio. Their methodology correctly achieve 80% of fraud cases using publicly available data and also correctly label 90.6% of the non-fraud cases with the same model. Their experiment shows that their model outperforms *F*-score model by Dechow et al. (2011) and the neural network model by Green and Choi (1997).
8	Ravisankar et al.	Decision Support Systems, 2011	Detection of financial statement fraud and feature selection using data mining techniques		Various Chinese stock exchange	The authors apply logistic regression, SVM, Genetic Programing (GP), Probabilistic Neural Network (PNN), multilayer feed forward Neural Network (MLFF) and group method of data handling (GMDH) to predict the financial fraud. The authors use the *t*-statistics to achieve feature selection on the dataset. They conducted the experiment with or without the feature selection on different classifiers. Their study results show that PNN has outperformed all the other techniques in both cases.

(*continued on next page*)

Table 1 (*continued*)

Sl. No	Author	Journal and Year of publication	Paper	Data sample	Database used	Main findings
9	Zhou et al.	Decision Support Systems, 2011	Detecting evolutionary financial statement fraud	–	–	In order to improve the efficiency of the existing data mining framework, Zhou and Kapoor propose an adaptive learning framework. According to Zhou and Kapoor, one way to improve the efficiency of data mining technique is to use adaptive learning framework (ALF). This framework proposes the inclusion of many exogenous variables, governance variables and management choice variables that will contribute towards the adaptive nature of financial fraud detection techniques.
10	Dechow et al.	Contemporary Accounting Research, 2011	Predicting material accounting misstatements	1982–2005	SEC Accounting and Auditing Enforcement Releases (AAER)	Dechow et al. develop a comprehensive database of financial misstatements and develop *F*- score model to predict misstatement.
11	Perols	Auditing: A Journal of Practice and Theory, 2011	Financial Statement Fraud Detection: An Analysis of Statistical and Machine Learning Algorithms	1998–2005	SEC/ AAER	The authors compare the performance of six popular statistical and machine learning models in detecting financial statement fraud under different assumptions of misclassification costs and ratios of fraud firms to non-fraud firms. They show that logistic regression and SVM outperform ANN, bagging, C4.5, and stacking.
12	Ngai et al.	Decision Support Systems, 2011	The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature	–	–	Ngai et al. present a comprehensive literature review on the data mining techniques. They show that Logistic Regression model, NN, the Bayesian belief Network and DT are the main data mining techniques that are commonly used in financial fraud detection.
13	Thiruvadi et al.	Information Technology Journal, 2011	Survey of Data-mining Techniques used in Fraud Detect ion and Prevention	–	–	Thiruvadi and Patel classify the frauds into management fraud, customer fraud, network fraud and computer-based fraud and discuss data mining techniques to fraud detection and prevention.
14	Sharma et al	International Journal of Computer Applications, 2012	A review of financial accounting fraud detection based on data mining techniques	–	–	The authors present a comprehensive review of the exiting literatures on financial fraud detection using the data mining techniques. Their finding shows that regression analysis, DT, Bayesian network, NN are the most commonly used data mining techniques in fraud detection.
15	Abbasi et al.	MIS Quarterly, 2012	MetaFraud: a meta learning framework for detecting financial fraud	1995–2010	SEC Accounting and Auditing Enforcement Releases (AAER)	They propose a meta-learning framework to enhance financial model fraud detection. This framework integrates business intelligence methods into a meta-learning artifact. This framework uses organizational and industry contextual information, quarterly and annual data and more robust classification methods using stacked generalization and adaptive learning.
16	Lin et al.	Knowledge-Based Systems, 2015	Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments	1998–2010	Taiwan Securities and Futures Bureau and the group litigation cases announced by the Securities and Futures Investors Protection Center	The authors employ logistic regression, DT (CART) and ANN to detect financial statement fraud. The authors concluded that ANN and CART provide better accuracy (over 90%) than the logistic model.
17	Aris et al.	The Journal of Applied Business Research, 2015	Fraudulent Financial Statement Detection Using Statistical Techniques: The Case Of Small Medium Automotive Enterprise	2010–2012	A small medium automotive company in Malaysia	The authors evaluate the possibility of fraudulent financial statements using statistical analysis, such as, financial ratio, Altman *Z*-score and Beneish model. They analyzed the financial statement data in a small medium automotive company in Malaysia and suggests further investigation by the management.

(*continued on next page*)

Table 1 (*continued*)

Sl. No	Author	Journal and Year of publication	Paper	Data sample	Database used	Main findings
18	Albashrawi	Journal of Data Science, 2016	Detecting Financial Fraud using Data Mining Techniques: A Decade Review from 2004 to 2015	–	–	Albashrawi present a recent comprehensive literature survey based on 40 articles on data mining applications in corporate fraud, such as, financial statement fraud, automobile insurance fraud, accounting fraud, health insurance fraud and credit card fraud. Their detail study suggest that logistic regression, decision tree, SVM, NN and Bayesian networks have been widely used to detect financial fraud. However, these techniques may not always yield best classification results.
19	Li et al.	Proceedings of the 7th Asian Conference on Machine Learning, Journal of Machine Learning Research Workshops and Conference Proceedings, 2015	Detecting Accounting Frauds in Publicly Traded U.S. Firms: A Machine Learning Approach	1982–2010	AAER, Compustat	The authors evaluate machine learning techniques in detecting accounting fraud. They employ linear and non-linear SVM and ensemble methods to handle class imbalance problem.
20	Kim et al.	Expert Systems with Applications, 2016	Detecting financial misstatements with fraud intention using multi-class cost sensitive learning	1997–2005	GAO, Compustat	Kim et al. develop multi-class financial misstatement detection model to detect the misstatements with fraud intension. They investigate three groups; 1) intentional restatement or fraud, 2) un-intentional restatement or error, and 3) no restatement. The authors develop three multi-class classifier model, multinomial logistic regression, support vector machine and Bayesian networks, using cost-sensitive learning to predict fraudulent intention in financial misstatements.

and ignores the embedded issues of class and cost imbalance in the dataset (Perols, 2011).

In a recent paper, Zhou and Kapoor (2011) has discussed the challenges faced by the traditional data mining techniques and proposed some improvements. They posit that while data mining methodologies and fraud detection models have evolved over the years, yet they face an increase level of difficulty in detecting financial frauds. Because "managements have adapted certain ways to avoid being identified by automated detection systems" (2011). In order to improve the efficiency of the existing data mining framework, Zhou and Kapoor (2011) have proposed an adaptive learning framework. According to Zhou and Kapoor, one way to improve the efficiency of data mining technique is to use adaptive learning framework (ALF). This framework proposes the inclusion of many exogenous variables, governance variables and management choice variables that will contribute towards the adaptive nature of financial fraud detection techniques.

Sharma and Panigrahi (2012) suggest that a fraud detection model based on only financial statement data may not give the best prediction. The authors presented a comprehensive review of the exiting literatures on financial fraud detection using the data mining techniques. Their finding shows that regression analysis, decision tree, Bayesian network, Neural network are the most commonly used data mining techniques in fraud detection. Gupta and Gill (2012) present a data mining framework for financial statement fraud risk reduction. In the framework, classification technique is proposed to successfully identify the fraudulent financial statement reporting. In another survey paper, Phua, Lee, Smith, and Gayler (2010) discuss technical nature of fraud detection methods and their limitations. In a similar study, Ngai et al. (2011) present a comprehensive literature review on the data mining techniques. Their review is based on 49 articles ranging from 1997 to 2008. They show that Logistic Regression model, Neural Networks, the Bayesian belief Network and Decision Tree are the main data mining techniques that are commonly used in financial fraud detection.

In light of Albashrawi (2016), Ngai et al. (2011) present a recent comprehensive literature survey based on 40 articles on data mining applications in corporate fraud, such as, financial statement fraud, automobile insurance fraud, accounting fraud, health insurance fraud and credit card fraud. Their detail study suggests that logistic regression, DT, SVM, NN and BBN have been widely used to detect financial fraud. However, these techniques may not always yield best classification results (Albashrawi, 2016).

Bolton and Hand (2002) review statistical and data mining techniques used in fraud detection with several subgroups, such as, telecommunications fraud, credit card fraud, medical and scientific fraud, and as well as in other domains such as money laundering and intrusion detection. The authors describe the available tools and techniques available for statistical fraud detection. Weatherford (2002) focusses on recurrent neural networks, backpropagation neural networks and artificial immune systems for fraud detection. In a similar study, Thiruvadi and Patel (2011) classify the frauds into management fraud, customer fraud, network fraud and computer-based fraud and discuss data mining techniques in light of fraud detection and prevention.

Green and Choi (1997) developed a neural network fraud classification model for endogenous financial data. Following Green and Calderon's (1995) study, the sample dataset initially consisted of 113 fraudulent company information. The authors used SEC filed financial statements which clearly indicates fraudulent account balances. Further, they used 10-K company annual reports for 1982–1990 from SSEC's Accounting and Auditing Enforcement Releases (SEC/AAER). Their study revealed that the neural network model has significant fraud detection capabilities as a fraud investigation and detection tool.

Hennes et al. (2008) propose a classification procedure to distinguish unintentional restatement (errors) and intentional restatement (irregularities) in financial restatements. In this paper, the authors used 8-K annual reports for 2002 to 2006 with a clear indication of error, irregularity or any investigation on accounting restatements to distinguish between intentional and non-intentional restatement. Many researchers followed their methodology in classifying errors and irregularities or fraud related to financial restatement.

Abbasi, Albrecht, Vance, and Hansen (2012) proposed a metalearning framework to enhance financial model fraud detection. This framework uses organizational and industry contextual information, quarterly and annual data and more robust classification methods using stacked generalization and adaptive learning. Cecchini et al. (2010) provide a methodology based on support vector machine and domain specific kernel. Their methodology correctly detect 80% of fraud cases using publicly available data and also correctly label 90.6% of the non-fraud cases with the same model. In their paper, the authors classify misstatement vs nonmisstatement firms using a graph kernel. Their methodology suggests that using exhaustive combinations of fraud variables has high fraud prediction probability. Further, their experiment shows that their model outperforms *F*-score model by Dechow et al. (2011) and the neural network model by Green and Choi (1997).

Another recent study by Perols (2011) attempts to evaluate the efficiency of different statistical and machine learning models in detecting financial frauds. Based on prior data mining research, the study uses various classification algorithms available in Weka (a popular data mining software) and perform a comparative analysis. In particular, Perols (2011) evaluates following six algorithms from Weka: (1) J48, (2) SMO, (3) MultilayerPerceptron, (4) Logistics, (5) stacking, and (6) bagging.[4] The study finds that logistic regression and support vector machine outperforms other four data mining models.

Aris, Arif, Othman, and Zain (2015) evaluates the possibility of fraudulent financial statements using statistical analysis, such as, financial ratio, Altman *Z*-score and Beneish model. They analyzed the financial statement data in a small medium automotive company in Malaysia and suggest further investigation by the management.

Li, Yu, Zhang, and Ke (2015) evaluates machine learning techniques in detecting accounting fraud. Their dataset initially consists of the fraud samples over the period between 1982 and 2010 from SEC's Accounting and Auditing Enforcement Releases (AAERs). The authors excluded the fraud occurrences prior to 1991 since there was a significant shift in U.S. firm behaviour. For further preprocessing, they also excluded the data samples for year 2006 to 2008 because of low fraud percentages. Thus their final dataset consists of fraud sample ranging from 1997 to 2005. In this study, the authors employ linear and non-linear SVM and ensemble methods and use raw accounting variables. They show that raw accounting variables have a better discriminant ability and help in developing improved predictive models.

As stated earlier, most of the earlier data mining models (in this area) focus on intentional or fraudulent restatements. Models involving both intentional and unintentional restatements are very limited. In a recent paper, Kim et al. (2016) developed a three-class financial misstatement detection model. Inspired by Hennes et al. (2008), they investigate three groups; (1) intentional restatement or fraud, (2) unintentional restatement or error, and (3) no restate-

[4] "J48 is a decision tree learner and Weka's implementation of C4.5 version 8. SMO is a support vector machine (SVM) and Logistics is Weka's logistic regression implementation. Both these classification algorithms are linear functions. MultilayerPerceptron is Weka's backpropagation ANN implementation, and stacking and bagging are two ensemble-based methods" (Perols, 2011, p. 25).

ment. They have used a matched sample approach to identify non-restatement cases to develop their final dataset. The authors then developed three multi-class classifier model using multinomial logistic regression, support vector machine and Bayesian networks to predict fraudulent intention in financial restatements. However, the study has used only a limited set of unintentional restatement cases which can affect the generality of their predictive models. Further, the use of a matched sample methodology could overstate the likelihood of restatement (Burns and Kedia, 2006) and impact the generality of the model (Dechow et al., 2011).

In summary, the existing literature suggests that data mining techniques have been widely used to detect financial fraudulent activities by the practitioners and regulators. These techniques are also well researched by the academics. However, studies involving all types of financial restatements are quite limited. Further, it appears that effectiveness of datamining techniques vary with respect to data sources, data structure and time period. In order to respond to such dynamic environment, it is important to continue with testing various data mining techniques with newer data sets and time period. Especially, given the market disruptions and changes in regulatory focus, it will be quite interesting to develop and examine the efficiency of data mining techniques during the pre- and post-financial crisis periods.

2.4. Various data sources on financial frauds/misstatements

Prior studies have used different data sources in order to identify fraud/ misstatement cases. However, these data sources differ in terms of scope and data coverage. As a result, the data structure and the instances of misstatements could differ quite significantly depending on data sources. Since the fraud/ misstatement cases are relatively rare, any small variation in instance count may affect the outcome of detection models. We provide a brief description of various data sources below.

The Government Accountability Office (GAO) financial restatement database: The GAO database covers a relatively shorter time period (January 1997 to September 2005) but include a large number of restatements (2309 cases). It includes all restatement cases relating to accounting irregularities without differentiating in terms of managerial intent, and materiality. One shortcoming of this database is that it includes information only on the first year of restatement without mentioning the overall restatement periods.

Stanford Law Database on Shareholder Lawsuits: This database includes all shareholder lawsuit instances. However, one major challenge with using this database in the context of developing financial fraud/ misstatement detection model is that lawsuits are initiated not only for financial restatements but also for many other unrelated reasons.

Accounting and Auditing Enforcement Releases (AAER) database: Dechow et al. (2011) present a useful discussion on AAER database, which is based Security Exchange Commission's (SEC) action against a firm or related parties due to rule violation.

> "The SEC takes enforcement actions against firms, managers, auditors, and other parties involved in violations of SEC and federal rules. At the completion of a significant investigation involving accounting and auditing issues, the SEC issues an AAER. The SEC identifies firms for review through anonymous tips and news reports. Another source is the voluntary restatement of the financial results by the firm itself, because restatements are viewed as a red flag by the SEC. The SEC also states that it reviews about one-third of public companies' financial statements each year and checks for compliance with GAAP" (p. 24).

One advantage of using AAER database is that identification of a misstatement does not depend on a researcher's judgement. These are determined by the SEC after a thorough investigation. The empirical results using AAER database, thus, can be easily replicable. However, AAER database does not include many restatement cases, especially the ones that are not investigated by SEC.

Audit Analytics (AA) database: It is a commercial database which includes all restatement cases from 2000 to current period. Karpoff, Koester, Lee and Martin (2014) present a useful discussion on Audit Analytics data collection process:

> "Audit Analytics (AA) defines a restatement as an adjustment to previously issued financial statements as a result of an error, fraud, or GAAP misapplication [and] does not include restatements caused by adoption of new accounting principles or revisions for comparative purposes as a result of mergers and acquisitions. AA extracts its data principally from SEC Form 8-K or required amended periodic reports (Forms 10-K/A, 10-Q/A, 10KSB/A, 20-F/A, and 40-F/A). AA claims to analyze all 8-K and 8-K/A filings that contain "Item 4.02 - Non-Reliance on Previously Issued Financial Statements or a Related Audit Report or Completed Interim Review" (an item required by the SEC since August 2004). In addition, all amended Forms 10-K/A, 10-Q/A, 10KSB/A, 20-F/A, and 40-F/A are reviewed to determine if the amendment is due to a restatement, and all audit opinions are searched for derivatives of the word "restate" with the intent of detecting the so-called "stealth" restatements contained in periodic reports rather than event filings" (Appendix A and B, page 9, 10).

While all four data sources have been widely used in financial restatement studies (Karpoff, Koester, Lee and Martin, 2014), Audit Analytics database is the most suitable one for our study because of its extensive coverage of all restatement cases. In this study, since we focus on all restatement cases (not just the fraudulent cases) that affect income statement or shareholder equity negatively, we need to reply on a comprehensive data sources on restatement cases, such as Audit analytics.

2.5. Addressing data imbalance and cost imbalance issues in financial fraud/restatement detection models

There are significant challenges in developing a reliable restatement detection model due to data imbalance and cost imbalance issues. While it is recognized that one should somewhat balance the restatement and non-restatement cases in the detection models, there is no universally accepted method to address this issue. We discuss relevant issues and options in the following subsections.

2.5.1. Addressing data imbalance issue

Financial restatement instances are relatively rare. A few studies such as Kim et al. (2016) use a matching sample technique (generally using industry and size matching) to make the data structure more balanced.

Dechow et al. (2011) question the appropriateness of such matching technique in developing a restatement detection model. They argue that matching techniques are useful only if we want to find whether a variable is significantly different relative to a control firm. In the process, however, we ignore majority of non-restatement firms in model building, which reduces the inclusiveness/ generality of a model. Dechow et al. (2011) further stress that "it is more difficult when matching to determine Type I and Type II error rates that users will face in an unconditional setting" (p. 23). Another detail study by Burns and Kedia (2006) also argues that a matched control firm approach would overstate the likelihood of restatement.

Other options to mitigate the data imbalance issue include undersampling and oversampling. In undersampling, like a matched sample approach, the instances of majority class are reduced. While a number of studies have used this technique has used this approach (Perols, 2011), there are many challenges with this strategy. First, as we eliminate number of instances from the control group (i.e. majority class), we may lose valuable information on non-restatement firm characteristics (Batista, Prati, & Monard, 2004; Han, Kamber, & Pei, 2011). Second, the resulting predictive model will be less general as it would exclude many firms in the population. Both reasons may also artificially lead to a reduction in the Type I and Type II error rates and improve model performance. Finally, random undersampling may also lead to high variance in the dataset (Wallace, Small, Brodley, & Trikalinos, 2011).

The other method to address data imbalance issue is oversampling. However, random oversampling that relies on producing minority tuples by sampling with replacement, does not prevent the risk of overfitting (Batista et al., 2004). Chawla, Bowyer, Hall, and Kegelmeyer (2002) proposed a modified oversampling method, namely, Synthetic Minority Oversampling Technique (SMOTE), which does not rely on random oversampling. This process generates new instances of restatements using nearest neighbour algorithm and hence does not induce any systematic bias in the data structure (Chawla et al., 2002). This algorithm does not rely on any specific attribute(s) of restatement or control groups while generating more instances. Accordingly, in this study, we rely on SMOTE methodology to address data imbalance issue.

In SMOTE, synthetic examples are created for each minority class instances using k-nearest neighbors' technique. Thereafter, depending on the oversampling requirement, cases are chosen from k-nearest neighbors. As pointed out by Chawla et al. (2002), "The synthetic examples cause the classifier to create larger and less specific decision, rather than smaller and more specific regions. More general regions are now learned for the minority class samples rather than those being subsumed by the majority class samples around them" (p. 328). Due to this informational gain, SMOTE is likely to lead to better predictive models.

2.5.2. Addressing cost imbalance issue

Since this study deals with a binary classification (i.e. restatement and non-restatement), the detection model leads to four potential outcomes: (i) true positive, a restatement firm is correctly classified as a restatement firm; (ii) false negative, a restatement firm is incorrectly classified as a non-restatement firm; (iii) true negative, a non-restatement firm is correctly classified as a non-restatement firm; and (iv) false positive, a non-restatement firm is incorrectly classified as a restatement firm. False negative and false positive cases are associated with misclassification costs that would reduce the efficiency of a predictive model. Further, in the context of a financial restatement study, cost associated with false negative and false positive classifications are likely to be different. For example, it would be more costly to misclassify a true restatement instance as a non-restatement case. "The classifiers, therefore, have to be both trained using appropriate cost ratios in the training sample and evaluated using relevant cost ratio assumptions" (Perols, 2011, p. 26). In view of these arguments, we need to consider using cost sensitive learning to address misclassification cost to generate a better fit model.

3. Research methodology

Our research methodology follows the Cross Industry Standard Process for Data Mining (CRISP - DM). It consists of four phases. A schematic diagram of the research method that is implemented in this study is shown in Fig. 1. In the first phase (data collection phase) we collect data from various sources, such as Compustat, and Audit analytics. This first phase also includes gathering domain knowledge about the data.

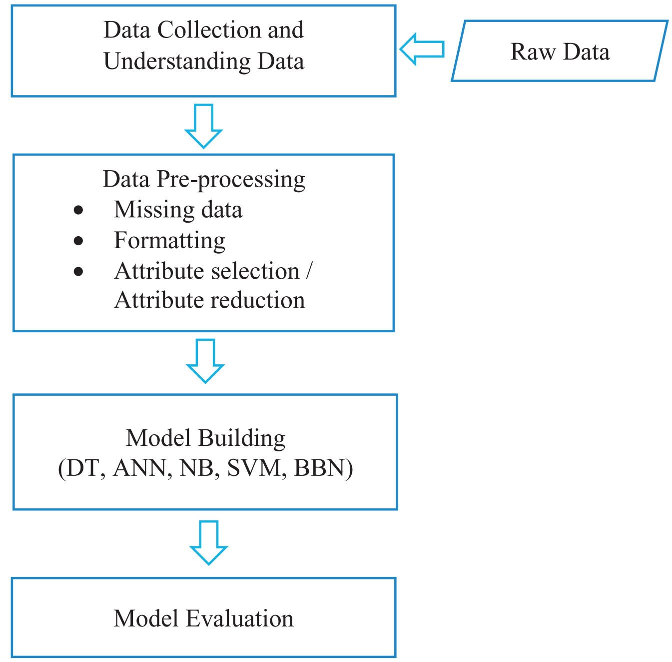

Fig. 1. A schematic diagram of the research methodology: CRISP-DM Approach.

In the second phase or the data pre-processing phase, we first address the missing and duplicate data issue by deleting these instances. Thereafter, we select appropriate attributes by using domain knowledge and following existing research (Dechow et al., 2011; Perols, 2011; Ravishankar et al., 2011). In order to mitigate the outlier effect, we further winsorize the data at 1% and 99% percentiles. We also perform feature subset selection using stepwise forward selection to remove less significant or redundant attributes. This step reduces total attributes from 116 to 15, which help building a more parsimonious model. Next, we prepare different versions of final dataset. These datasets are used to examine the efficiency of different classifiers as employed in the study.

The third phase is model building in which we build, train and run the various models on the test datasets. We use several data mining classification techniques to build a set of predictive models on financial restatement detection. Based on prior research, we select Artificial Neural Network (ANN) (Ata et al., 2009; Green et al. 1997; Feroz, Kwon, Pastena, & Park, 2000; Kirkos et al., 2007; Kotsiantis et al., 2006; Lin et al., 2015; Ngai et al., 2011; Perols, 2011; Ravisankar et al., 2011), Decision Tree (DT) (Ata et al., 2009; Kirkos et al., 2007; Kotsiantis et al., 2006; Lin et al., 2015; Ngai et al., 2011; Perols, 2011), Naïve Bayes (NB) (Sharma et al., 2012), Bayesian Belief Network (BBN) (Albashrawi, 2016; Kim et al., 2016; Kirkos et al., 2007; Kotsiantis et al., 2006; Ngai et al., 2011; Sharma et al., 2012), and Support Vector Machine (SVM) (Albashrawi, 2016; Cecchini et al., 2010; Kim et al., 2016; Perols, 2011; Ravisankar et al., 2011) classification algorithmes. We present a more detail discussion on these data mining techniques and rationale behind using the same in a subsequent section.

We perform ANN, DT, NB, BBN and SVM classification algorithms using Weka, a well-known machine learning software. We further address data imbalance and cost imbalance issues using two different approaches: (i) Synthetic minority oversampling technique (SMOTE), and (ii) cost sensitive learning with SMOTE.

The fourth and final phase is the model evaluation phase where we analyze the models' performance using appropriate performance measures, such as, sensitivity, false positive rate (FP), accuracy, precision, specificity and area under the ROC curve (AUC). Weka presents a number of performance measures that can help

Table 2
Restatement data sources and filtering.

Initially, we download all restatement cases from 1995 to 2014 as reported in Audit Analytics database. The relevant restatement cases are reported in "Stage 1: Year 1995 to 2014 restatement cases (based on restatement begin year)" table. Although, our final sample includes restatement cases for 2001–2014 period, we need to download restatement data from previous years. Many of the restatement cases span over multiple years. Therefore, some of the pre-2011 restatement cases may still have influence in 2011 or later years

Stage 1: Year 1995 to 2014 restatement cases (based on restatement begin year)

	No. of cases
Fraudulent restatement	260
Accounting restatement	15,387
Clerical error restatement	632
Total	16,279

Next stage of data filtering, involves following steps: (i) we expand above restatement cases for all restatement years, (ii) we delete duplicate restatement years, (iii) we merge restatement data with Compustat financial variables using CIK identifier and exclude non-merged cases, (iv) following Hennes et al. (2014), we keep only one restatement year per restatement case and delete FIN 48, SAB No. 101, SAB No. 108 restatement categories. The relevant restatement cases are reported in "Stage 2: Year 2001 to 2014 restatement cases" table. It includes restatement cases from 2001 to 2014.

Stage 2: Year 2001 to 2014 restatement cases

	No. of cases
Fraudulent restatement	170
Accounting restatement	6333
Clerical error restatement	311
Total	6814

Next, we delete firm-year observations, which have missing data for some of the key financial variables, such as market to book value ratio and market excess returns. Following Hennes et al. (2014), we also ignore the restatement cases that are attributed to clerical errors. The relevant restatement cases are reported in "Stage 3: Year 2001 to 2014 restatement cases" table.

Stage 3: Year 2001 to 2014 restatement cases

	No. of cases
Fraudulent restatement	109
Accounting restatement	3404
Total	3513

us in determining the most appropriate learning algorithm. Most of these measures are derived from the confusion matrix.

4. Data collection and pre-processing

4.1. Data sources and related attributes

In this study, we use primarily two different data sources. First, financial restatement incidences are collected from Audit Analytics (AA) database. We restrict our sample from 2001 to 2014 financial year. Second, in order to develop the predictive models it is imperative to gather all relevant financial data. We collect relevant data from COMPUSTAT financial database. Thereafter, by using CIK identifier, we match the cases obtained from AA database with relevant COMPUSTAT entries. Table 2 includes a more detail discussion on data download and filtering process.

We use a comprehensive list of attributes while developing relevant predictive models. Based on the existing literature, we first identify a large set of attributes (116) that are related to fraud/financial restatement detection models (Dechow et al., 2011; Green & Choi, 1997; Lin et al., 2015; Perols, 2011; Ravisankar et al., 2011). A list of the initial set of 116 attributes is presented in Table 3.

Our class variable has two categories; (i) cases without restatement incident in a year – denoted with a value '0′, and cases with restatement incident in a year – denoted with a value '1′. In the final dataset we have 3513 cases with restatement category and 60,720 cases with non-restatement category. It appears that the classes (restatement vs. non-restatement) are highly imbalanced. We address this issue in the dataset pre-processing stage, as discussed in a later section (Section 4.2).

4.2. Data pre-processing

Pre-processing a dataset is a crucial factor in the experimental setup. Since the experiment is conducted on Weka, the data need to be transformed to arff file to be readable to Weka. During pre-processing this dataset, any non-required attributes or noise should be removed to run the classifier successfully. For example, the 'firm ID' and 'ticker symbol' attributes are removed from the initial dataset downloaded from COMPUSTAT database.

It is also very important to address the missing and mismatched values in the dataset. For example, in many instances the Tobin's Q (tob_q) values were missing in the dataset. These cases are replaced by '?' to make it readable for Weka. Similarly, for some of the instances the percentage sign from performance variables (e.g. roa_ebit) is removed by changing data type to keep the consistency with other data.

We then winsorize the data at 1% and 99% percentiles to limit the outliers. Ravisankar et al. (2011) show that we can obtain almost similar efficiency for a predictive model by carefully selecting a smaller set of attributes. Accordingly, we perform feature subset selection using stepwise forward selection to remove less significant or redundant attributes. This step reduces total attributes from 116 to 15, which help building a more parsimonious model. Table 4 includes a list of reduced attribute set.

4.3. Forming different datasets

Earlier studies show that financial misrepresentation could lead to a substantial loss in terms of market value and reputation loss of firms (Karpoff, Lee, & Martin, 2008). A mere look at the financial penalties imposed by the regulatory authorities (e.g. SEC), however, may not give us a holistic perspective. While actual penalty might seem lower, stock price of the firms that are accused of financial misrepresentation dip significantly in the subsequent period. Such instances are not desirable for the investors, employees and other stakeholders alike. Furthermore, Chakravarthy, Haan, and Rajgopal (2014) point out that it requires a significant level of efforts by the firms in rebuilding their reputation during the post financial misrepresentation period.

Therefore, it would be quite helpful for the investors, stakeholders and regulators if they can predict the financial misrepresentation (or, restatement) behaviour of a firm. This will allow the respective parties to take precautionary measures well in advance.

While there have been a number of studies on financial fraud detections, most of the studies worked on different sample period. More importantly, empirical studies on U.S. corporate fraudulent data in the post financial crisis are quite limited. The sample period of this study (2001–2014) will give an opportunity to examine the performance of fraud detection models in the pre- and post-financial crisis period. One implication of this development is that we may be able to come up with better predictive model in the post financial crisis period.

We construct different datasets to compare the restatement prediction for different financial years. In the dataset, the outcome variables (i.e. class) are: (i) restatement in a year, and (ii) no-restatement in a year. We categorize the whole sample into three different types of dataset.

Dataset 1 (2001 to 2014): It contains all the attributes on the full sample period. It includes all financial data obtained from

Table 3
List of attributes.

Attribute	Description	Reference
dechw_ab_chg_orderbklg	(OB - OBt-1) / OBt-1 - (SALE - SALEt-1) / SALEt-1	Dechow et al. (2011)
dechw_actual_issuance	IF SSTK>0 or DLTIS>0 THEN 1 ELSE 0	Dechow et al. (2011)
dechw_book_mkt_value	CEQ / (CSHO * PRCC_F)	Dechow et al. (2011)
dechw_rsst_accru	RSST Accruals = (ΔWC + ΔNCO + ΔFIN)/Average total assets, where WC = (ACT - CHE) - (LCT - DLC); NCO = (AT - ACT - IVAO) - (LT - LCT - DLTT); FIN = (IVST + IVAO) - (DLTT + DLC + PSTK)	Dechow et al. (2011)
dechw_chg_free_cf	(IB - RSST Accruals) / Average total assets - (IBt-1 - RSST Accrualst-1) /Average total assetst-1	Dechow et al. (2011)
dechw_chg_invt	(INVT- INVTt-1)/Average total assets	Dechow et al. (2011)
dechw_chg_oper_lease	((MRC1/1.1+ MRC2/1.1^2+ MRC3/1.1^3+ MRC4/1.1^4+ MRC5/1.1^5) - (MRC1t-1/1.1+ MRC2t-1/1.1^2+ MRC3t-1/1.1^3+ MRC4 −1/1.1^4+ MRC5t-1/1.1^5))/Average total assets	Dechow et al. (2011)
dechw_chg_receiv	(RECT- RECTt-1)/Average total assets	Dechow et al. (2011)
dechw_chg_roa	IB / Average total assets - IBt-1 / Average total assetst-1	Dechow et al. (2011)
dechw_def_taxexp	TXDI / ATt-1	Dechow et al. (2011)
dechw_demand_fin	IF ((OANCF-(CAPXt-3+CAPXt-2+ CAPXt-1)/ 3) /(ACT) < −0.5 THEN 1 ELSE 0	Dechow et al. (2011)
dechw_earn_price	IB / (CSHO x PRCC_F)	Dechow et al. (2011)
dechw_exist_oplease	IF (MRC1 > 0 OR MRC2 > 0 OR MRC3 > 0 OR MRC4 > 0 OR MRC5 > 0 THEN 1 ELSE 0	Dechow et al. (2011)
dechw_lev_fin_raised	FINCF / Average total assets	Dechow et al. (2011)
dechw_leverage	DLTT / AT	Dechow et al. (2011)
dechw_per_chg_cashmargin	((1-(COGS+(INVT-INVTt-1))/(SALE-(RECT-RECTt-1)))-(1-(COGSt-1+(INVTt-1-INVTt-2))/(SALEt-1-(RECTt-1-RECTt-2)))) /(1-(COGSt-1+(INVTt-1-INVTt-2))/(SALEt-1-(RECTt-1-RECTt-2)))	Dechow et al. (2011)
dechw_per_chg_cashsale	((SALE - (RECT - RECTt-1)) - (SALEt-1 - (RECTt-1 - RECTt-2))) /(SALEt-1 - (RECTt-1 - RECTt-2))	Dechow et al. (2011)
dechw_softasset	(AT-PPENT-CHE)/Average total assets	Dechow et al. (2011)
dechw_fm_ind_unexp_em_prod	FIRM((SALE/EMP - SALEt-1/EMPt-1)/(SALEt-1/EMPt-1)) - INDUSTRY((SALE/EMP - SALEt-1/EMPt-1)/(SALE t-1/EMPt-1))	Dechow et al. (2011)
dechw_wc_accrual	(((ACT - ACTt-1) - (CHE - CHEt-1)) - ((LCT - LCTt-1) - (DLC - DLCt-1) - (TXP - TXPt-1)) - DP)/Average total assets	Dechow et al. (2011)
perols1	Accounts receivable to sales	Perols (2011)
perols2	Accounts receivable to total assets	Perols (2011), Ravisankar et al. (2011)
perols3	Allowance for doubtful accounts	Perols (2011)
perols4	Allowance for doubtful accounts to accounts receivable	Perols (2011)
perols5	Allowance for doubtful accounts to net sales	Perols (2011)
perols6	Altman Z score	Perols (2011)
perols7	Big four auditor	Perols (2011)
perols8	Current minus prior year inventory to sales	Perols (2011)
perols9	Days in receivables index	Perols (2011)
perols10	Debt to equity	Perols (2011), Ravisankar et al. (2011)
perols11	Declining cash sales dummy	Perols (2011)
perols12	Fixed assets to total assets	Perols (2011)
perols13	Four year geometric sales growth rate	Perols (2011)
perols14	Gross margin	Perols (2011)
perols15	Holding period return in the violation period	Perols (2011)
perols16	Industry ROE minus firm ROE	Perols (2011)
perols17	Inventory to sales	Perols (2011)
perols18	Net sales	Perols (2011)
perols19	Positive accruals dummy	Perols (2011)
perols20	Prior year ROA to total assets current year	Perols (2011)
perols21	Property plant and equipment to total assets	Perols (2011)
perols22	Sales to total assets	Perols (2011)
perols24	Times interest earned	Perols (2011)
perols25	Total accruals to total assets	Perols (2011)
perols26	Total debt to total assets	Perols (2011)
perols27	Total discretionary accrual	Perols (2011)
perols28	Value of issued securities to market value	Perols (2011)
perols29	Whether accounts receivable>1.1 of last year's	Perols (2011)
perols31	Whether gross margin percent > 1.1 of last year's	Perols (2011)
perols32	Whether LIFO	Perols (2011)
perols33	Whether new securities were issued	Perols (2011)
perols34	Whether SIC code larger (smaller) than 2999 (4000)	Perols (2011)
lag1_mkt_excess_ret	Lag of (firm's yearly stock returns minus stock index's (i.e. S&P500) yearly stock returns)	Dechow et al. (2011)
long_term_debt	Total long term debt	In light of Ravisankar et al. (2011)
total_asset	Total asset	Ravisankar et al. (2011)
inv_total_asset	Ratio: Inventory to Total asset	Ravisankar et al. (2011)
accounts_recev	Account receivables	Green and Choi (1997); Lin et al. (2015); Ravisankar et al. (2011)
accnt_rec_sales	Ratio: Account receivable to Total sales	Green and Choi (1997); Feroz et al. (2000); Lin et al. (2015); Kaminski, Wetzel, and Guan (2004)
current_ratio	Ratio: Current asset to Current liability	Ravisankar et al. (2011)

(continued on next page)

Table 3 (*continued*)

Attribute	Description	Reference
inv_current_liab	Ratio: Inventory to Current liability	Ravisankar et al. (2011)
tob_q	Tobin's Q; represents the market value to book value ratio of the firm	Brown and Caylor (2006)
acc_recev_to_lag	Ratio: Account receivable to Last year's account receivables	Ravisankar et al. (2011)
acc_payable_to_lag	Ratio: Account payable to Last year's account payable	In light of Ravisankar et al. (2011)
total_asset_to_lag	Ratio: Total asset to Last year's total asset	Ravisankar et al. (2011)
sale_to_lag	Ratio: Total sales to Last year's total sales	In light of Ravisankar et al. (2011)
fyear	Fiscal year; represents the respective financial year of the firm	Dechow et al. (2011)
trad_current_asset	Current Assets (e.g. inventory, accounts receivables)	In light of Ravisankar et al. (2011)
trad_current_liab	Current Liability (e.g. short term payment obligation)	In light of Ravisankar et al. (2011)
trad_inventory	Total inventory	In light of Ravisankar et al. (2011)
trad_current_total_asset	Ratio: Current asset to Total asset	In light of Ravisankar et al. (2011)
trad_accounts_payable	Account payables	In light of Ravisankar et al. (2011)
compu_ff48	Fama-french 48 sectors	Conventional financial/accounting variable
roe_higher_indavg48	Dummy = 1 if firm return on equity (ROE) >= industry average ROE	Conventional financial/accounting variable
roa_ebitda3yr_higher_indavg48	Dummy = 1 if firm Return on Asset to EBITDA (3-yr average) ratio is >= Industry Return on Asset to EBITDA (3-yr average) ratio; EBITDA = Earnings before interest, tax, depreciation & amortization	Conventional financial/accounting variable
tsr_crsp_higher_indavg48	Dummy = 1 if firm total shareholder return (TSR) >= Industry average total shareholder return (TSR); where, TSR = buy and hold return for last 12-monthly returns.	Conventional financial/accounting variable
tsr_crsp3yr_higher_indavg48	Dummy = 1 if firm total shareholder return (TSR) - 3-yr avg >= Industry average total shareholder return (TSR) - 3-yr avg; where, TSR = buy and hold return for last 12-monthly returns.	Conventional financial/accounting variable
lnsize_tasset	Log of total asset	Conventional financial/accounting variable
lnsize_tasset_3yr_avg	Log of total asset - 3-yr average	Conventional financial/accounting variable
lnsize_mktval	Log of market value of equity	Conventional financial/accounting variable
lnsize_sale	Log of total sale	Conventional financial/accounting variable
mkt_excess_ret	firm's yearly stock returns minus stock index's (i.e. S&P500) yearly stock returns	Conventional financial/accounting variable
std_bhret_48m_incl_curr	Standard deviation of last 48 months stock returns (denotes risk)	Conventional financial/accounting variable
std_bhret_60m_incl_curr	Standard deviation of last 60 months stock returns (denotes risk)	Conventional financial/accounting variable
pct_change_ppe_net_0_1	Percentage change in plant, property and equipment (PPE) net	Conventional financial/accounting variable
pct_change_ppe_net_3yr_avg	3 year arithmetic average of yearly percentage change in PPE	Conventional financial/accounting variable
pct_change_long_ast_0_1	Yearly percentage change in long term asset	Conventional financial/accounting variable
pct_change_long_ast_3yr_avg	3 year arithmetic average of percentage change in long term asset	Conventional financial/accounting variable
pct_change_capex_0_1	Percentage change capex between current and one lagged year	Conventional financial/accounting variable
pct_change_capex_3yr_avg	3 year arithmetic average of percentage change in capital expenditure (capex)	Conventional financial/accounting variable
trad_mkt_val_equity	Market value of equity - calculated as: stock price * outstanding shares	Conventional financial/accounting variable
trad_ltd_to_equity	Ratio: Long term debt to common equity (CEQ)	Conventional financial/accounting variable
trad_ltd_to_asset	Ratio: Total long term debt to Total asset	Conventional financial/accounting variable
trad_ni	Net Income	Conventional financial/accounting variable
trad_roa	Return on asset (calculated as: Net Income / Total asset)	Conventional financial/accounting variable
trad_roa_ebit	Return on asset (calculated as: EBIT / Total asset); EBIT = Earnings before interest & tax	Conventional financial/accounting variable
trad_roa_ebitda	Return on asset (calculated as: EBITDA / Total asset); EBITDA = Earnings before interest, tax, depreciation & amortization	Conventional financial/accounting variable
trad_roe	Return on equity (calculated as: Net income / total market value of equity)	Conventional financial/accounting variable
trad_r_and_d	R&D expenditure	Conventional financial/accounting variable
trad_sale	Total sales	Conventional financial/accounting variable
trad_cash	Total cash holding in a year	Conventional financial/accounting variable
trad_cash_shrt_invest	Total cash holding and short term investment in a year	Conventional financial/accounting variable

(*continued on next page*)

Table 3 (*continued*)

Attribute	Description	Reference
trad_cash_to_tasset	Ratio: Total cash holding to Total asset	Conventional financial/accounting variable
trad_accnt_pay_sales	Ratio: Account payable to Total sales	Conventional financial/accounting variable
trad_intangible_assets	Intangible assets (e.g. goodwill, patent)	Conventional financial/accounting variable
trad_intan_total_asset	Ratio: Intangible asset to Total assets	Conventional financial/accounting variable
trad_ppe_gross	Gross value of Plant, Property and Equipment (Gross PPE)	Conventional financial/accounting variable
trad_ppe_net	Net value of Plant, Property and Equipment (Net PPE)	Conventional financial/accounting variable
trad_ppe_net_total_asset	Ratio: Net value of Plant, Property and Equipment (Net PPE) to Total asset	Conventional financial/accounting variable
trad_auditor_code	auditor identification	Conventional financial/accounting variable
trad_ocf_tasset_new	Ratio: Operating cash flow to Total asset	Conventional financial/accounting variable
trad_ocf_mequity_new	Ratio: Operating cash flow to Total market value of equity	Conventional financial/accounting variable
trad_mkt_to_bookvalue	Ratio: market value of common equity to book value of common equity	Conventional financial/accounting variable
trad_div_total_asset	Ratio: Total dividend to Total asset	Conventional financial/accounting variable
trad_hfi	Herfindahl index; represents the market concentration in that industry	Conventional financial/accounting variable
trad_lag1_net_income	Net income of last year (lagged time period)	Conventional financial/accounting variable
trad_lag1_total_asset	Total asset of last year (lagged time period)	Conventional financial/accounting variable

Table 4
List of reduced attributes.

Attribute	Description
roa_ebitda3yr_higher_indavg48	Dummy =1 if firm Return on Asset to EBITDA (3-yr average) ratio is >= Industry Return on Asset to EBITDA (3-yr average) ratio; EBITDA = Earnings before interest, tax, depreciation & amortization
std_bhret_60m_incl_curr	St. dev of last 60 months of return (denotes risk)
dechw_actual_issuance	IF SSTK>0 or DLTIS>0 THEN 1 ELSE 0
dechw_chg_oper_lease	((MRC1/1.1+ MRC2/1.1^2+ MRC3/1.1^3+ MRC4/1.1^4+ MRC5/1.1^5) - (MRC1t-1/1.1+ MRC2t-1/1.1^2+ MRC3t-1/1.1^3+ MRC4 −1/1.1^4+ MRC5t-1/1.1^5))/ Average total assets
dechw_leverage	DLTT / AT
dechw_per_chg_cashmargin	((1-(COGS+(INVT-INVTt-1))/(SALE-(RECT-RECTt-1)))- (1-(COGSt-1+(INVTt-1-INVTt-2))/(SALEt-1-(RECTt-1-RECTt-2)))) / (1-(COGSt-1+(INVTt-1-INVTt-2))/(SALEt-1-(RECTt-1-RECTt-2)))
dechw_softasset	(AT-PPENT-CHE)/Average total assets
trad_div_total_asset	Ratio: Total dividend to Total asset
trad_lag1_net_income	Net income of last year (lagged time period)
perols12	Fixed assets to total assets
perols13	Four year geometric sales growth rate
perols20	Prior year ROA to total assets current year
perols21	Property plant and equipment to total assets
perols31	Whether gross margin percent > 1.1 of last year's

COMPUSTAT dataset, industry dummy variable (12 categories), and a measure that denotes the industry concentration based on sales (hfi – Herfindahl Index). This data includes all attributes and all restatement cases on the full sample period (between 2001 and 2014 period). As we are interested in developing a predictive model based on the past financial information, we use lagged data (by one year) for all attributes. For example, if the restatement data is for year 2001, we use all financial data from year 2000. Failing to organize data in this way will induce biases in our analysis.

As the two classes are highly imbalanced in the dataset, we opted to employ SMOTE algorithm to mitigate class imbalance problem. Thereafter, we use the randomized algorithm to shuffle the order of instances passed through it.

Dataset 2 (2001–08): It contains all attributes for the pre- financial crisis period, 2001 to 2008.

Dataset 3 (2009–14): It contains all attributes for the post-financial crisis period 2009 to 2014.

We summarize the dataset characteristics in Table 5.

5. Model building

Based on popularity of the models in prior research on financial fraud detection using data mining techniques, we select three classifiers to build the predictive model. In this study, we employ Artificial Neural Network (ANN), Decision Tree (DT), Naïve Bayes (NB), Bayesian Belief Network (BBN), and Support Vector Machine (SVM) classification algorithms. Further, we use SMOTE to address data imbalance issue.

Table 5
Different dataset characteristics.

Dataset	Dataset period	No. of Instances	No. of restatement Instances	No. of non-restatement instances	No. of attributes (before stepwise forward selection)	No. of attributes (after stepwise forward selection)
1	2001–2014	64,233	3513	60,720	116	15
2	2001–2008	44,076	2513	41,563	116	15
3	2009–2014	20,157	1000	19,157	116	15

Table 5 reports the restatement and non-restatement cases for different datasets used in the study. "Dataset 1'' includes restatement cases from 2001 to 2014. Corresponding financial variables are from 2000 to 2013. "Dataset 2'' includes restatement cases for the pre-crisis period, from 2001 to 2008. Corresponding financial variables are from 2000 to 2007. "Dataset 3'' includes restatement cases for the post-crisis period, from 2009 to 2014. Corresponding financial variables are from 2008 to 2013.

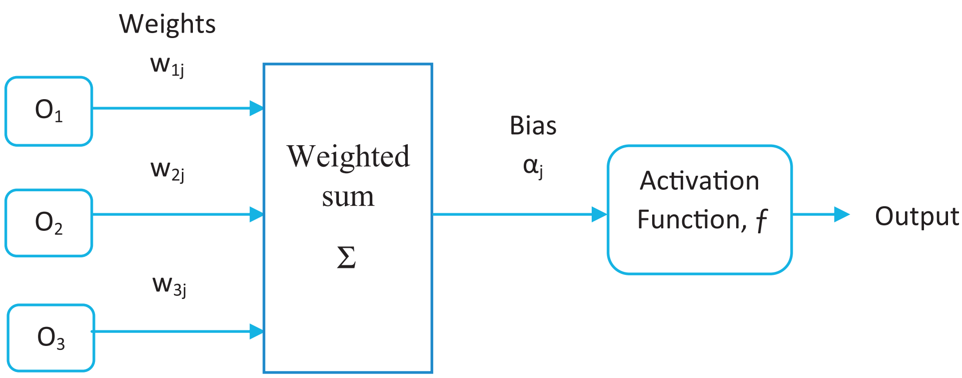

Fig. 2. A schematic diagram of ANN classifier.

5.1. Artificial neural network (ANN)

The neural networks is a supervised machine learning algorithm that is inspired by the neurons in human brain system. A network of node are arranged in three layers; mainly, input layer, hidden layer and output layer. Though ANN is criticized for its poor interoperability, researchers find that neural networks give more accurate result, is highly adaptive and very flexible in noise tolerance. Hence, it can generate a robust model (Ngai et al., 2011). The non-linear statistical data modelling feature of neural network can be used as supervised learning, unsupervised learning and reinforced learning.

The output unit of artificial neural network takes a weighted sum of the outputs from the units in the previous layer. This classification model can be easily modified by changing the weights in the hidden layers. A multi-layer feed forward artificial neural network uses the backpropagation algorithm to modify the weights to minimize the mean- squared error between the network's prediction and the actual target value (Fig. 2).

Fig. 2 shows a model of ANN backpropagation (Han et al., 2012, p. 402). The inputs of a given unit, O_1, O_2, O_3 …. O_n, are either the outputs from the previous hidden layer or the input tuples of the dataset.

$$I_j = \sum_{i=0}^{n} w_{ij} O_i + \alpha_j$$

These inputs are multiplied by their corresponding weights to form a weighted sum. The weighted sum is then added to the bias associated to the unit. An activation function is applied to calculate the final output of the unit, O_j.

$$O_j = \frac{1}{1 + e^{-I_j}}$$

Many researchers (Ata et al., 2009; Kirkos et al., 2007; Kotsiantis et al., 2006; Lin et al., 2015; Ravisankar et al., 2011) have used artificial neural network in their study. This techniques is frequently used in credit card fraud detection, economic and financial forecasting.

5.2. Decision tree (DT)

Decision tree is a supervised learning algorithm that is also commonly used to predict corporate, credit card and other financial fraud (Sharma & Panigrahi, 2012). It uses a tree-like structure to represent the attributes and possible outcomes. The outcomes or predictions are represented by leaves and attributes by branches. One of the advantages of DT classifier(s) is that it does not require any prior domain knowledge to develop the branches and leafs or to develop the predictive model. Based on top-down attribute selection approach, DT algorithm first identifies the initial node. Subsequently, it uses a series of 'if then' steps along with the attribute selection method to complete the predictive model. DT algorithms use various attribute selection measures, such as, information gain, gain ratio or Gini index.

The machine learning algorithms ID3, CART, C4.5 are commonly used as decision tree classification models. ID3 uses information gain as its attribute selection measure. C4.5 uses gain ratio whereas CART uses the Gini index. All these algorithms use simple IF-THEN classification rules to represent the learned knowledge. In our study, we use decision tree C4.5 algorithm.

5.3. Naïve Byes (NB)

A Naïve Byes (NB) classifier is based on Bayes' theorem. One advantage of using Naïve Bayes classifier is that in order to estimate the parameters it only needs small portion of training data. This method is quite useful when input dimensionality is high (Han et al., 2012). This classification technique is widely used in banking and financial fraud detection (Sharma et al., 2012). Phua et al. (2010) find that in the context of financial fraud detection model, NB classifier outperforms other classifiers. It results in lower false negative cases and almost no false positive cases.

This probabilistic model assumes strong independence among the predictor variables. In other words, it assumes that for any given class (y), the value of a predictor (x) does not depend on other predictors. According to the Bayes theorem, the conditional probability $P(y|x)$, can be determined in terms of $P(y), P(x)$, and $P(x|y)$.

$$P(y|x) = \frac{P(x|y)P(y)}{P(x)}$$

$$P(y|x) = P(x_1|y) * P(x_2|y) * P(x_3|y) * \ \dots \ * \ P(x_m|y) * P(y)$$

where, $P(y|x)$ is the conditional probability of a class for a given attribute (predictor); $P(y)$ is the prior probability of a class; $P(x|y)$ is the likelihood which is the probability of predictor given a class; and $P(x)$ is the prior probability of predictor.

Since the Naïve Bayes classifier makes the class independence assumption, if the assumption is valid for a given dataset, the Naïve Bayes classifier can predict the best outcome. In such cases, it can have the best accuracy compare to other classifiers.

5.4. Bayesian Belief network (BBN)

Bayesian Belief networks (BBN) follows Bayes theorem. But unlike the Naïve Bayes classifier, BBN does not make the assumption of conditional independence. In real life problems, dependencies can exist among the attributes. In such cases, BBN is preferable over naïve Bayes classifier.

A BBN is a directed acyclic graph (DAG) where the nodes have one-to one correspondence with the attributes. Any node in the network can be selected as a 'class attribute' node. An arrow drawn to node B from node A represents node A as a parent of node B (Han et al., 2012). Based on Bayes theorem, each attribute or node is conditionally independent of its non-descendants given its parents in the network graph. For each data tuple, $X = (x_{1,} x_{2,} x_{3,} \ldots\ldots x_{n,})$, there is a conditional probability table that represents the conditional probability of the data tuple X for each combination of the values of its parents. Let $P(X)$ be the probability of a data tuple having n attributes.

$$P(X) = \ P(x_{1,} \ x_{2,} \ x_{3,} \ldots\ldots x_{n,})$$

$$P(X) = \ P(x_1|Parents(y_1)) * P(x_2|Parents(y_2)) * \ \dots \ * \times P(x_n|Parents(y_n))$$

where $P(x_i|Parents(y_i))$ is the conditional probability of each combination of tuple values of a parent and $i = 1, \ 2, \ \ldots..n$. This is one of the widely used classification techniques in fraud detection studies (Albashrawi, 2016; Kim et al., 2016; Kirkos et al., 2007; Kotsiantis et al., 2006; Ngai et al., 2011; Sharma et al., 2012).

5.5. Support vector machine (SVM)

Support vector machines (SVM) is one of the most widely used data mining classification techniques (Vapnik, 1995) and may lead to better predictive models in financial domains (Cecchini et al., 2010). However, some of the existing fraud detection studies (Kotsiantis et al., 2006; Ravisankar et al., 2011) find that other data mining techniques outperform SVM. This classifier works with feature space instead of the data space. In a binary classification problem, given labeled training dataset, SVM finds an optimal hyperplane to classify the examples based on the largest minimum distance to the training tuples. This hyperplane is expressed as a 'margin'. In a predictive mode, a trained SVM classifier then classifies the target class based on which side of the margin the test sample falls in.

Let $X = (x_{i,} \ y_{i,})$ be the dataset where $x_{i,}$ is the instances, $y_{i,}$ is the class label and $i = 1, \ 2, \ \ldots..n$.

For a linearly separable dataset, a SVM with linear kernel can be denoted as

$$f(x) = W \ . \ (x_{i,} \ y_{i,}) + b$$

where, W is the weight vector or the co-efficient and b is a scalar value called bias (Han et al., 2012). However, most of the real world problems falls into the category where the dataset is not linearly separable. Therefore, a linear margin does not exists that will clearly separate the positive and negative tuples of the dataset. In such cases, a non-linear mapping is used to transform the dataset into a higher dimensional feature space. Thus the non-linear SVM method are kernelized to define a separating hyperplane. Among the SVM's commonly used kernels, radial basis function (RBF) kernel is the most widely used kernel in the extant literature. A RBF kernel can be denoted as

$$K(x_{i,} \ y_{i,}) = e^{-\frac{||x_{i,} - \ y_i \ ||^2}{2\sigma^2}}$$

where σ is a kernel parameter. Han et al. (2012) pointed out that there is no consensus on administering a specific kernel that will result in the most accurate SVM. According to the authors "the kernel chosen does not generally make a large difference in resulting accuracy. SVM training always finds a global solution" (p. 415). In a recent study on fraud detection, Li et al. (2015) state that "It seems that either financial kernel or expansion is not necessary, as they provide only marginal improvement" (p. 182). Based on the suggestions in existing literatures, we employ SVM RBF kernel – which is a standard feature in data mining software – in predicting the financial restatements in this study.

5.6. Synthetic minority oversampling technique (SMOTE)

Chawla et al. (2002) proposed synthetic minority oversampling technique (SMOTE) which generates random synthetic instances of the minority class on the feature space rather than replicating the existing instances of the minority sample. Thus it improves the bias SMOTE algorithm uses k-nearest neighbour technique to create these synthetic minority examples. For each minority class sample X_i, SMOTE finds the k-nearest neighbours, N_i for that instance. Then, it calculates the difference between the feature sample and its nearest neighbours, $\Delta = X_i - N_i$. Next, after multiplying this difference by random number 0 or 1, it adds it to the feature sample; thus creating a set of random points along the line segments between two specific features (Chawla et al., 2002). The random points are generated as

$$F_{new} = \ X_i \ + (X_i \ - \ N_i \ ) \times rand \ (0, 1)$$

where F_{new} is the generated random point and $rand\,(0, \ 1)$ is the random variable generated a random number either 0 or 1.

In order to create a balanced dataset, different amount of synthetic oversampling may be needed for different datasets.

5.7. Performance measures

In the model evaluation phase where we analyze the models' performance using appropriate performance measures, such as, sensitivity, false positive rate (FP), accuracy, precision, specificity and area under the ROC curve (AUC). Weka presents a number of performance measures that can help us in determining the most appropriate learning algorithm. Most of these measures are derived from the confusion matrix. Fig. 3 presents a confusion matrix and its components for a binary classification problem.

Sensitivity or Recall or the true positive (TP) rate is the proportion of correctly identified positive instances whereas false positive (FP) rate is the proportion of falsely classified positive instances. In this study, Sensitivity measures the number of restatements instances that are correctly identified as a restatement by a particular model to the number of actual restatement instances. FP rate measure the number of restatements instances that are incorrectly identified. The accuracy measures the true instances (i.e. correctly classified restatement and non-restatement instances) among the

	Actual Positive	Actual Negative
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)
Total	P = TP + FN	N = TN + FP

Fig. 3. A confusion matrix.

whole population. It can be described as the ratio of correctly classified instances to the total number of instances. Precision is the ratio of correctly identified restatement instances predicted as restatement instances to the total number of actual and predicted restatement instances.

Another measure, specificity or true negative rate is a measure of the proportion of the number of non-restatement instances predicted as non-restatement by a model to the total number of actual non-restatement instances. These measures can be termed as

$$Recall\ or\ Sensitivity = \frac{TP}{TP+FN}$$

$$False\ positive\ rate = \frac{FP}{FP+TN}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Specificity = \frac{TN}{FP+TN} = 1 - false\ positive\ rate$$

where, TP – True positive, TN – True negative, FP – False positive, FN – False negative

Our last evaluation measure is area under the receiver operating characteristic (ROC) curve. An ROC curve is a plot in which the *x*- axis denotes the false positive rate (FPR) and *y*-axis denotes the true positive rate (TPR) of the classifier (Fawcett, 2006). It is a standard technique that denotes a classifier performance on the basis of true positive and false positive error rates (Swets, 1988). As Han et al. (2012) describe:

> "For a two-class problem, an ROC curve allows us to visualize the trade-off between the rate at which the model can accurately recognize positive cases versus the rate at which it mistakenly identifies negative cases as positive for different portions of the test set. Any increase in *TPR* occurs at the cost of an increase in *FPR*. The area under the ROC curve is a measure of the accuracy of the model" (p. 374).

Traditionally, we use area under the curve (AUC) as a performance metric for a ROC curve (Bradley, 1997; Duda, Hart, & Stork, 2001; Lee, 2000; Chawla et al., 2002). AUC shows the relationship between the true positive rate and the false positive rate. The larger the AUC, the better is the classifier performance.

6. Analysis of the result

We present the results in two subsections, based on the treatment of class imbalance and cost sensitive learning analysis. First, we address the class imbalance issue through SMOTE algorithm in Weka. Second, we incorporate cost sensitive learning in addition to SMOTE to address both class and cost imbalance problem, in order to test the effectiveness of cost sensitive learning approach.

We perform classification using five classifiers, Decision tree, ANN, Naïve Byes, SVM and Bayesian belief network using 10- fold cross validation. We summarize the test results and reported all the performance measures (e.g. sensitivity, specificity, precision, AUC etc.). However, as Japkowicz and Shah (2011) discussed there are particular advantages and disadvantages associated with different performance measures. It appears that there is no undisputed performance measure to validate the appropriateness of a learning algorithm. However, As Han et al. (2012) suggest, in case if class imbalance problem, only accuracy rate can't be a good measure of the prediction. Further,

> "Most of the fraud detection studies using supervised algorithms since 2001 have abandoned measurements such as true positive rate and accuracy as a chosen threshold. In fraud detection, misclassification costs (false positive and false negative error costs) are unequal, uncertain, can differ from example to example, and can change over time. In fraud detection, a false negative error is usually more costly than a false positive error" (Phua et al., 2010, p. 4)

Therefore, we have focused on AUC measure while discussing the relevant results.

6.1. SMOTE analysis

We performed the analysis using smote in dataset 1 (with sample year 2001 to 2014) and present the results in Table 6. Based on the results, it appears that for this dataset, decision tree and ANN classifiers lead to good predictive models in terms of AUC and accuracy. In general, ANN showed a higher reliability with an AUC value of approximately 83%. DT and BBN performance is also close with an AUC of 81.8% and 78.6%. It appears that Naïve Bayes and SVM could not generate a reliable model that could have been compared with ANN and DT.

Tables 7 and 8 present the performance measure for dataset 2 and 3. Dataset 2 shows the performance measure prior to financial crisis period (2001 to 2008), whereas dataset 3 shows the performance measure after the financial crisis period (2009 to 2014). Based on the results presented in these tables, we find that for all of these datasets, ANN classifier would generate a reliable predictive model in terms of AUC (approx. 83% value).

In dataset 2, ANN outperformed other models with an accuracy value of 78.06% and a recall value of 78.1%. DT also came close with 77.53% accuracy and 77.5% recall values. Dataset 3 also shows similar results. It is interesting to note that predictive model effectiveness remains quite similar in the pre- and post-crisis periods. It implies that our predictive models are quite stable over the sample periods.

Table 6
Performance measures for the period 2001 to 2014.

Classifier	Recall	FP	Accuracy	Precision	Specificity	*F* - Measure	AUC
DT	0.778	0.312	77.756	0.774	0.688	0.775	0.818
ANN	0.78	0.363	77.9846	0.772	0.637	0.769	0.83
NB	0.602	0.311	60.2067	0.701	0.689	0.614	0.693
SVM	0.656	0.333	65.6005	0.704	0.667	0.668	0.662
BBN	0.726	0.311	72.6155	0.741	0.689	0.732	0.786

Table 7
Performance measures for before financial crisis period 2001 to 2008.

Classifier	Recall	FP	Accuracy	Precision	Specificity	*F* - Measure	AUC
DT	0.775	0.31	77.5293	0.771	0.69	0.772	0.818
ANN	0.781	0.337	78.0628	0.774	0.663	0.773	0.831
NB	0.601	0.306	60.1456	0.702	0.694	0.611	0.7
SVM	0.674	0.311	67.4121	0.717	0.689	0.684	0.682
BBN	0.722	0.299	72.2056	0.741	0.701	0.728	0.787

Table 8
Performance measures for after financial crisis period 2009 to 2014.

Classifier	Recall	FP	Accuracy	Precision	Specificity	*F* - Measure	AUC
DT	0.799	0.311	79.9131	0.794	0.689	0.796	0.821
ANN	0.777	0.349	77.7185	0.77	0.651	0.773	0.826
NB	0.601	0.285	60.1281	0.727	0.715	0.616	0.71
SVM	0.734	0.339	73.447	0.744	0.661	0.739	0.698
BBN	0.751	0.294	75.0709	0.767	0.706	0.756	0.808

Table 9
Two cost matrices.

Classified as –>	Cost matrix 1		Cost matrix 2	
	R	NR	R	NR
Restatement (R)	0	1	0	1
Non-restatement (NR)	5	0	20	0

6.2. Cost sensitive learning with SMOTE analysis

Existing literature shows that financial frauds or restatements can lead to a significant level of shareholder wealth destruction. Hence, financial fraud/ restatement models are quite useful. However, prediction models are not perfect and can lead to false negative and false positive cases. Such outcomes could be costly and should be addressed in the predictive models. Yet, we should recognize that cost of false negative outcome and false positive outcome are different. Perols (2011) observed that “It is more costly to classify a fraud firm as a non-fraud firm than to classify a non-fraud firm as a fraud firm” (p. 20). We address this issue with a cost sensitive learning algorithm included in Weka. We investigate the classification algorithms using cost sensitive learning with SMOTE on the sample datasets so that we could take both class and cost imbalance into account simultaneously. In this context, we create two sets of cost matrices as shown in Table 9. These cost matrices represent two sets of relative error costs (false positive: false negative): (i) 1:5, and (ii) 1:20. In case of a false positive outcome, a non-restatement firm is incorrectly classified as a restatement firm; whereas, in case of a false negative outcome, a restatement firm is incorrectly classified as a non-restatement firm. Given that false negative cases are more damaging, we assign a higher cost weight for this category.

Tables 10–12 present the performance measure for dataset 1 (full sample 2001–2014), dataset 2 (2001–2008), and dataset 3 (2009–2014) respectively using the relative error cost of 1:5. Similarly, Table 13, Table 14 and Table 15 present the performance measure for dataset 1 (full sample 2001–2014), dataset 2 (2001–2008), and dataset 3 (2009–2014) respectively using the relative error cost of 1:20.

Similar to the analyses performed using only SMOTE, it appears that ANN and DT classifiers lead to better predictive models in terms of AUC values for a lower relative error cost matrix (i.e. 1:5). As the cost ratio increases, the performance of the predictive models deteriorates. Overall, we find that an application of cost sensitive learning algorithm does not improve the performance of predictive models. Addressing the class imbalance issue with SMOTE leads to a similar result (as reported in Table 6, 7 and 8).

7. Conclusion and future research

Investors and regulators pay a close attention to the financial restatement events. Earlier studies have shown that financial restatements could erode shareholders wealth quite significantly. Such events can also damage a firm’s reputation considerably. Therefore, an early detection of suspect firms, which may make financial restatements or get involved with financial frauds in the subsequent periods, will be quite beneficial for the investors and regulators. Accordingly, a number of studies have focussed on developing predictive models for financial frauds. However, most of the studies have focussed only on fraudulent cases and largely ignored other financial restatements. A number of studies (for example, Bowen et al., 2017) show that financial restatements are also associated with significant negative market reactions. Moreover, most of the datasets use a partial set of restatement/ fraudulent cases that may induce significant biases in the predictive models.

In this study, we developed predictive models based on all restatement cases (both intentional and unintentional restatements) using a real comprehensive dataset which includes approximately 3513 cases over a period of 2001 to 2014. We built different classifiers (ANN, DT, NB, SVM, and BBN) to test the efficiency of various predictive models. Further, we addressed the issues of class imbalance and cost imbalance by employing SMOTE and cost sensitive learning algorithm. We found that two classifiers (namely, ANN and DT) lead to useful predictive models for restatement

Table 10
Performance measures for the period 2001 to 2014 smote with cost sensitive learning (cost ratio 1:5).

Classifier	Recall	FP	Accuracy	Precision	Specificity	*F* - Measure	AUC
DT	0.692	0.2	69.2099	0.792	0.8	0.702	0.81
ANN	0.644	0.22	64.3981	0.775	0.78	0.652	0.824
NB	0.488	0.292	48.8224	0.708	0.708	0.47	0.693
SVM	0.514	0.258	51.4174	0.75	0.742	0.497	0.628
BBN	0.613	0.23	61.3145	0.767	0.77	0.618	0.786

Table 11
Performance measures for the period 2001 to 2008 smote with cost sensitive learning (cost ratio 1:5).

Classifier	Recall	FP	Accuracy	Precision	Specificity	*F* - Measure	AUC
DT	0.691	0.203	69.0758	0.789	0.797	0.699	0.812
ANN	0.661	0.222	66.058	0.773	0.778	0.668	0.826
NB	0.491	0.296	49.0862	0.705	0.704	0.468	0.7
SVM	0.511	0.268	51.0549	0.743	0.732	0.488	0.621
BBN	0.611	0.236	61.0602	0.762	0.764	0.613	0.787

Table 12
Performance measures for the period 2009 to 2014 smote with cost sensitive learning (cost ratio 1:5).

Classifier	Recall	FP	Accuracy	Precision	Specificity	*F* - Measure	AUC
DT	0.738	0.183	73.8263	0.812	0.817	0.75	0.819
ANN	0.658	0.206	65.7584	0.787	0.794	0.67	0.817
NB	0.481	0.278	48.0944	0.72	0.722	0.47	0.71
SVM	0.532	0.249	53.1502	0.751	0.751	0.531	0.641
BBN	0.663	0.214	66.3439	0.783	0.786	0.677	0.809

Table 13
Performance measures for the period 2001 to 2014 smote with cost sensitive learning (cost ratio 1:20).

Classifier	Recall	FP	Accuracy	Precision	Specificity	*F* - Measure	AUC
DT	0.493	0.258	49.2938	0.77	0.742	0.457	0.768
ANN	0.432	0.271	43.1753	0.771	0.729	0.369	0.8
NB	0.411	0.305	41.1004	0.693	0.695	0.35	0.693
SVM	0.346	0.307	34.5762	0.738	0.693	0.215	0.519
BBN	0.52	0.243	51.9938	0.772	0.757	0.5	0.786

Table 14
Performance measures for the period 2001 to 2008 smote with cost sensitive learning (cost ratio 1:20).

Classifier	Recall	FP	Accuracy	Precision	Specificity	*F* - Measure	AUC
DT	0.493	0.258	49.2938	0.77	0.742	0.457	0.768
ANN	0.447	0.276	44.7403	0.767	0.724	0.387	0.819
NB	0.412	0.313	41.2263	0.688	0.687	0.344	0.7
SVM	0.359	0.314	35.909	0.74	0.686	0.231	0.523
BBN	0.521	0.252	52.1105	0.766	0.748	0.498	0.785

Table 15
Performance measures for the period 2009 to 2014 smote with cost sensitive learning (cost ratio 1:20).

Classifier	Recall	FP	Accuracy	Precision	Specificity	*F* - Measure	AUC
DT	0.508	0.227	50.8193	0.782	0.773	0.494	0.777
ANN	0.476	0.235	47.5789	0.782	0.765	0.45	0.788
NB	0.414	0.289	41.4074	0.706	0.711	0.372	0.71
SVM	0.332	0.286	33.1811	0.742	0.714	0.213	0.523
BBN	0.557	0.22	55.6763	0.78	0.78	0.557	0.805

instances. Furthermore, different time period analyses showed that our models were consistent in both the pre- and post-crisis periods.

The results of this study show a promising outcome and presents reliable predictive models that can guide the investors, regulators and auditors to detect a firm with the intention of financial fraudulent activity or the probability of material errors in financial statements. The results are quite consistent for both pre- and post-crisis periods.

While this study makes some unique contributions by (i) using a comprehensive dataset, (ii) focussing on both intentional and unintentional restatement cases, and (iii) developing predictive models employing all conventional data mining techniques with a comprehensive list of financial and firm specific attributes, it has some limitations. For example, there are many other classification techniques (e.g. random forest, logistic regression, CART) that are not tested in this study. We also did not use any custom built financial kernel while employing SVM technique in the predictive models. As Li et al. (2015) posit that financial kernel or any expansion may not lead to improved results, we did not test this conjecture explicitly. Some of the studies use association rules in financial fraud detection (e.g. credit card fraud detection)

(Sanchez, Vila, Cerda, & Serrano, 2009). While this is an alternative way to explore the probability of financial frauds, it is out of scope for this study and we do not venture into this area. Further, this study is confined to only a specific type of financial irregularities, that is, financial restatement. There are other financial irregularities (e.g. credit card fraud, insurance frauds) that are not considered in this study. Also, we have used a particular oversampling technique (i.e. SMOTE) in order to address class imbalance issue in this study. We recognize that extant literature presents various oversampling techniques that are not examined in this study.

This study can be extended in different ways. A natural extension would be the inclusion of managerial variables (e.g. CEO characteristics, board characteristics) while developing predictive models. Finance and accounting research shows that managerial characteristics play an important role in corporate actions and decisions. Inclusion of managerial variables may lead to better predictive models. In a future study, one can explore a three-class model (intentional, unintentional and no restatement) with a comprehensive dataset for a better understanding of managerial intent and a firm's financial competency. Another avenue of future research is to include social media perception in the model building process. Discussions on social media may reveal useful information about a firm's ethical behaviour and financial irregularities. This in turn could be a valuable input in the development of predictive models. Similarly, text mining algorithms could be employed to perform sentiment analysis on financial restatement documents and use it as an additional attribute. Also, we can explore other complementary techniques in the model building process. For example, in this study we have used single classifier in each model building process. In addition we can use hybrid data mining techniques that combine more than one classifier and study relevant impacts. Further, one can explore different undersampling and oversampling techniques to address class imbalance issue for the better prediction of financial restatements. It appears that there is no consensus in the literature on how to address this issue, which is quite common in financial restatement/ fraud datasets. This needs more in-depth attention. Finally, we can extent the methodology and research framework of this study to other areas of financial irregularities (e.g. credit card fraud, insurance fraud) and significant financial events such as bankruptcy, mergers and acquisitions, strategic alliances, secondary equity offering, and corporate innovation to name a few.

References

Abbasi, A., Albrecht, C., Vance, A., & Hansen, J. (2012). Metafraud: A meta-learning framework for detecting financial fraud. *Mis Quarterly, 36*(4), 1293–1327.

Albashrawi, M. (2016). Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015. *Journal of Data Science, 14*(3), 553–569.

Aris, N. A., Arif, S. M. M., Othman, R., & Zain, M. M. (2015). Fraudulent financial statement detection using statistical techniques: The case of small medium automotive enterprise. *The Journal of Applied Business Research, 14*(4), 1469–1477.

Ata, H. A., & Seyrek, I. H. (2009). The use of data mining techniques in detecting fraudulent financial statements: An application on manufacturing firms. *The Journal of Faculty of Economics and Administrative Science, 14*(2), 157–170.

Banarescu, A., & Baloi, A. (2015). Preventing and detecting fraud through data analytics in auto insurance field. *Romanian Journal of Economics, 40, 1*(49), 89–114.

Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKD Explorations Newsletter - Special Issue on Learning from Imbalanced Datasets, 6*(1), 20–29.

Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical Science, 17*(3), 235–255.

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

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition, 30*(7), 1145–1159.

Brown, L. D., & Caylor, M. L. (2006). Corporate governance and firm valuation. *Journal of Accounting and Public Policy, 25*, 409–434.

Burks, J. (2010). Disciplinary measures in response to restatements after Sarbanes-Oxley. *Journal of Accounting and Public Policy, 29*(3), 195–225.

Burns, N., & Kedia, S. (2006). The impact of performance-based compensation on misreporting. *Journal of Financial Economics, 79*, 35–67.

Cecchini, M., Aytug, H., Koehler, G. J., & Pathak, P. (2010). Detecting management fraud in public companies. *Management Science, 56*(7), 1146–1160.

Chakravarthy, J., Haan, E. D., & Rajgopal, S. (2014). Reputation repair after a serious restatement. *The Accounting Review, 89*(4), 1329–1363.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357.

Dechow, P. M., Ge, W., Larson, C. R., & Sloan, R. G. (2011). Predicting material accounting misstatements. *Contemporary Accounting Research, 28*(1), 17–82.

Desai, H., Hogan, C., & Wilkins, M. (2006). The reputational penalty for aggressive accounting: Earnings restatements and management turnover. *The Accounting Review, 81*(1), 83–112.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed). John Wiley & Sons.

Farber, D. (2005). Restoring trust after fraud: Does corporate governance matter?. *The Accounting Review, 80*(2), 539–561.

Fawcett, T. (2006). An Introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861–874.

Feroz, E. H., Kwon, T. M., Pastena, V., & Park, K. J. (2000). The efficacy of red flags in predicting the SEC's targets: An artificial neural networks approach. *International Journal of Intelligent Systems in Accounting, Finance, and Management, 9*(3), 145–157.

Files, R., Swanson, E., & Tse, S. (2009). Stealth disclosure of accounting restatements. *The Accounting Review, 84*(5), 1495–1520.

GAO. (2002). *Financial statement restatements: Trends, market impacts, regulatory responses, and remaining challenges*. US General Accounting Office, http://www.gao.gov/special.pubs/gao-06-1079sp/.

Green, B. P., & Calderon, T. G. (1995). Analytical procedures and auditor's capacity to detect fraud. *Accounting Enquiries: A Research Journal, 5*(1), 1–47.

Green, B. P., & Choi, J. H. (1997). Assessing the risk of management fraud through neural network technology. *Auditing, 16*(1), 14.

Gupta, R., N. S., & Gill, N. S. (2012). A data mining framework for prevention and detection on financial statement fraud. *International Journal of Computer Applications, 50*(8), 7–14.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques: Concepts and techniques (3rd ed.)*. MA: Morgan Kaufmann: Elsevier.

Hennes, K. M., Leone, A. J., & Miller, B. P. (2008). The importance of distinguishing errors from irregularities in restatement research: The case of restatements and ceo/cfo turnover. *The Accounting Review, 83*(6), 1487–1519.

Hennes, K. M., Leone, A. J., & Miller, B. P. (2014). Determinants and market consequences of auditor dismissals after accounting restatements. *The Accounting Review, 89*(3), 1051–1082.

Huang, Y., & Scholz, S. (2012). Evidence on the association between financial restatements and auditor resignations. *Accounting Horizons, 26*(3), 439–464.

Japkowicz, N., & Shah, M. (2011). *Evaluating learning algorithms: A classification perspective*. New York, NY: Cambridge University Press 343.

Kaminski, K., Wetzel, T., & Guan, L. (2004). Can financial ratios detect fraudulent financial reporting. *Managerial Auditing Journal, 19*(1), 15–28.

Karpoff, J. M., Lee, D. S., & Martin, G. S. (2008). The cost to firms of cooking the books. *Journal of Financial and Quantitative Analysis, 43*(3), 581–612.

Karpoff, J. M., Koester, A., Lee, D. S., & Martin, G. S. (2014). *Database Challenges in Financial Misconduct Research, University of Washington Working Paper*. WA, USA: Seattle.

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.

Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications, 32*(4), 995–1003.

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.

Lee, S. (2000). Noisy Replication in Skewed Binary Classification. *Computational Statistics and Data Analysis, 34*(2), 165–191.

Li, B., Yu, J., Zhang, J., & Ke, B. (2015). Detecting accounting frauds in publicly traded U.S. firms: A machine learning approach. In *Proceedings of the 7th asian conference on machine learning*.

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.

Ngai, E., Hu, Y., Wong, Y., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems, 50*(3), 559–569.

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

Perols, J. (2011). Financial statement fraud detection: An analysis of statistical and machine learning algorithms. *Auditing: A Journal of Practice and Theory, 30*(2), 19–50.

Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *Computing Research Repository, abs/1009.6119*, 1–14.

Ravisankar, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support System, 50*, 491–500.

Sanchez, D., Vila, M. A., Cerda, L., & Serrano, J. M. (2009). Association rules applied to credit card fraud detection. *Expert Systems with Applications, 36*, 3630–3640.
Sharma, A., & Panigrahi, P. K. (2012). A review of financial accounting fraud detection based on data mining techniques. *International Journal of Computer Applications, 39*(1), 37–47.
Srinivasan, S. (2005). Consequences of financial reporting failure for outside directors: Evidence from accounting restatements and audit committee members. *Journal of Accounting Research, 43*(2), 291–334.
Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science, 240*, 1285–1293.
Thiruvadi, S., & Patel, S. C. (2011). Survey of data-mining techniques used in Fraud detection and prevention. *Information Technology Journal, 10*, 710–716.
Turner, L., Dietrich, J. R., Anderson, K., & Bailey, A. J., 2001, Accounting restatements. Ohio State University Working paper (February 2001).
Vapnik, V., & Cortes, C. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.
Wallace, B. C., Small, K., Brodley, C. E., & Trikalinos, T. A. (2011). Class imbalance, redux. In *Proceedings of the 11th international conference on data mining,* (pp. 754–763). IEEE.
Weatherford, M. (2002). Mining for fraud. *IEEE Intelligent Systems, 17*(4), 4–6.
Zhou, W., & Kapoor, G. (2011). Detecting evolutionary financial statement fraud. *Decision Support Systems, 50*(3), 570–575.