International Journal of Accounting Information Systems 24 (2017) 32–58

Contents lists available at ScienceDirect

International Journal of Accounting Information Systems

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

Data mining applications in accounting: A review of the literature and organizing framework

Farzaneh A. Amani [a], Adam M. Fadlalla [b,*]

[a] Qatar University, Doha, Qatar
[b] Department of Accounting and Information Systems, College of Business and Economics, Qatar University, Doha, Qatar

ARTICLE INFO

Article history:
Received 22 September 2015
Received in revised form 5 November 2016
Accepted 20 December 2016
Available online xxxx

Keywords:
Data mining
Accounting
Literature review
Framework
Prospective
Retrospective

ABSTRACT

This paper explores the applications of data mining techniques in accounting and proposes an organizing framework for these applications. A large body of literature reported on specific uses of the important data mining paradigm in accounting, but research that takes a holistic view of these uses is lacking. To organize the literature on the applications of data mining in accounting, we create a framework that combines the two well-known accounting reporting perspectives (retrospection and prospection), and the three well-accepted goals of data mining (description, prediction, and prescription). The framework encapsulates a taxonomy of four categories (retrospective-descriptive, retrospective-prescriptive, prospective-prescriptive, and prospective-predictive) of data mining applications in accounting. The proposed framework revealed that the area of accounting that benefited the most from data mining is assurance and compliance, including fraud detection, business health and forensic accounting. The clear gaps seem to be in the two prescriptive application categories (retrospective-prescriptive and prospective-prescriptive), indicating opportunities for benefiting from data mining in these application categories. The framework presents a holistic view of the literature and systematically organizes it in a structurally logical and thematically coherent manner.

1. Introduction

In the era of rapidly changing, globalized economies, and highly competitive markets, organizations, to become competitively relevant, need to consider, and, many a times, adopt or implement a wide variety of innovative management philosophies, approaches, and advanced information technologies (Dorsch and Yasin, 1998). In particular, artificial intelligence (AI) is important to the future of the accounting profession (Elliott, 1992), and intelligent systems have empowered many enhancements in multidimensional analytical power and efficiency of the accounting processes (Granlund, 2011). Thus, there are clear calls that AI deserves added attention (Debreceny, 2011), and the existence of opportunities of massive scale for companies to better fully leverage the analytical capability of their enterprise systems (White, 2004). An open question is: could the lack of full utilization of these analytical capabilities be explained by the complexity of these systems as suggested by Kim et al., 2009, or could it be due to other factors such as features specific to data mining techniques, or the nature of the intelligent accounting applications themselves?

* Corresponding author.
E-mail addresses: farzanakhoory@gmail.com (F.A. Amani), fadlalla@qu.edu.qa (A.M. Fadlalla).

http://dx.doi.org/10.1016/j.accinf.2016.12.004

Data mining is one of the most important current paradigms of advanced intelligent business analytics and decision support tools. Such significance is acknowledged by the major accounting professional bodies. The American Institute of Certified Public Accountants (AICPA) has identified data mining as one of the top ten technologies for tomorrow, and the Institute of Internal Auditors (IIA) has listed data mining as one of the four research priorities (Koh and Low, 2004). In addition, the Chartered Global Management Accountants (CGMA) has reported that >50% of corporate leaders rank big data and data mining among the top ten corporate priorities that are fundamental for the data-driven era of business (CGMA, 2013). Data mining has been defined as the process of identifying valid, potentially novel, and ultimately understandable patterns in data (Pujari, 2001). It is also known as the process of extracting or mining knowledge from massive amounts of data (Han et al., 2006) to improve decisions in a particular discipline. The key focus of data mining is, therefore, to leverage the data assets of an organization to derive financial or non-financial benefits. Thus data mining has been applied to almost all non-business as well as business disciplines, including accounting.

Data mining is reported to afford organizations a wide array of benefits and capabilities; including effectively predicting future trends of corporate development, helping managers make better decisions, and raising competitiveness of an enterprise (Xiao et al., 2010; Yigitbasioglu and Velcu, 2012). It can also provide managers with logical and causal connections within a company's figures so that issues can be proactively tackled (Yigitbasioglu and Velcu, 2012). In addition, data mining can contribute towards significantly improving judgment, transaction, and compliance in auditing (Vasarhelyi et al., 2004), improve the quality of evidence supplied to auditors (Brown et al., 2007), and contribute to the efficiency of the overall audit (Chan and Vasarhelyi, 2011). Furthermore, data mining can facilitate electronic (Liang et al., 2001) and continuous (Brown et al., 2007; Vasarhelyi et al., 2012) auditing, and has the potential to radically alter the managerial control systems' role and execution in organizations (Sutton et al., 2011; Granlund et al., 2013). Data mining enables organizations to more easily identify statistical relations among performance measures (Ittner and Larcker, 2001), estimate the likelihood an event will occur, thereby supplementing managers' qualitative judgments (Rezaee et al., 2002), and provide a vehicle of control for both accuracy of the data and legitimacy of data requests. Not least, data mining can help organizations quickly discern patterns in data that would take years to discover using older techniques (Mauldin and Ruchala, 1999), identify disgruntled employees from patterns of their email exchanges (Huerta et al., 2012), and empower regulatory agencies with real-time market surveillance and risk profiling of market players (Williams, 2013).

Accounting is a bedrock of any enterprise and spans a wide range of tasks including internal and external reporting, costing, estimating, evaluating, analyzing, and auditing. Many of these tasks involve a great deal of uncertainty and risk complexities. Accounting has a history of intelligent applications dating back more than three decades (Baldwin et al., 2006), and was one of the earliest business disciplines to utilize data mining to better address these risks and complexities. A large body of research has been published describing applications of data mining in accounting. Although many researchers offered literature reviews of such research, these reviews have generally focused on a specific accounting domain and/or data mining technique (Coakley and Brown, 2000; Yang, 2006; Calderon and Cheh, 2002; Wang, 2010; Ngai et al., 2011).

A more encompassing approach is a review that presents this body of knowledge in a manner that simultaneously takes into consideration the multi-faceted nature of the two underlying disciplines of accounting and data mining. This approach can help in addressing effectively questions such as: what is the current status of the amalgamation of accounting, a fundamental business discipline, and data mining, a top ten future information systems technology? How pervasive is this critical technology in accounting, and is it uniformly used across all branches of accounting or is it limited to some and not others? When used, is it used with similar or varying intensities across the different accounting domains, and what are the plausible explanations if there is variability in usage intensity? How much has accounting adopted of the various powerful capabilities (including goals, tasks, and techniques) that data mining has to offer? These questions cannot be answered by the existing reviews individually. In addition, such an approach may provide a mechanism to organize the research on the applications of data mining in accounting in a structurally logical and thematically coherent manner. It is the purpose of this research to attempt to answer these questions, and to propose an organizing framework for the literature on data mining applications in accounting. In so doing, the paper, contributes to extant literature: first, a better understanding of the intersection of these two important disciplines; second, a macro-level perspective of the current status of research and practice on data mining applications in accounting; and third, a direction to potential opportunities for future research in this important domain. In addition, using a framework that succinctly organizes the literature and summarizes its overall topology reveals the main research themes and patterns, provides deeper insights into the underlying conceptual underpinnings and relationships, thereby leading to a better informed research and practice agenda. Without such a reflective well-organized literature review, one is left with a fragmented landscape without a solid handle on the true topography of the literature. Under such disjointed circumstances, it will be difficult, at best, to ascertain the extent to which the capabilities of a crucial technology of the 21st century have been leveraged in the core business discipline of accounting. The contributions of the paper are relevant to both researchers and practitioners with an interest in the application of data mining in accounting.

The objective of this paper is thus to systematically examine published research on data mining applications in accounting to understand the current status of, discern any central themes in, and offer an organizing framework for, this research. We propose a framework that provides a comprehensive view of what has been accomplished by using data mining in accounting, what areas in the accounting discipline have more and which ones have less utilization of this technology. The paper relies on interpretative research using content analysis to understand the relevant literature. Extant research that describes applications of data mining in accounting served as the primary data for understanding the nature of these applications and for mapping them into the organizing framework. The rest of the paper is organized as follows: section 2 provides a background and literature review, section 3 describes the research methodology, section 4 presents the proposed framework, section 5 presents and discusses results, and section 6 offers conclusions, limitations, and future research directions.

2. Background and literature review

Data mining is the application of specific algorithms for extracting patterns from data. It allows the automated discovery of implicit patterns and interesting knowledge hidden in large amounts of data (Jiawei and Kamber, 2001). Data mining helps organizations to focus on the most important information and knowledge available in their existing databases. But it is only a tool; it does not eliminate the need to know the business, to understand the data, or to understand the analytical methods involved (Jackson, 2002). Data mining has three main goals: description, prediction, and prescription. Whereas description focuses on finding human-interpretable patterns describing the data, prediction involves using some variables or fields in the database to predict unknown or future values of other variables of interest (Fayyad et al., 1996). On the other hand, prescription focuses on providing the best solution for the given problem (Evans, 2013). These goals can be achieved by using many data mining tasks, including classification, clustering, prediction, outlier detection, optimization, and visualization. These tasks differ with the type of problem to be solved as follows:

- Classification focuses on mapping data to predefined qualitative discrete attribute set of classes, which could be binary or multi-class.
- Clustering focuses on segmenting the data to some meaningful classes or groups.
- Prediction focuses on finding a future numerical value (forecasting) or non-numerical value (classification).
- Outlier Detection focuses on finding the data that significantly deviates from the normal.
- Optimization focuses on finding the best solution given some resources.
- Visualization focuses on the visual presentation and understanding of data.
- Regression focuses on estimation of a dependent variable from a set of independent variables.

A wide variety of data mining techniques exist, such as artificial neural networks (NNs), case-based reasoning (CBR), genetic algorithms (GA), decision trees (DT), association rules (AR), support vector machines (SVM), regression, self-organizing maps (SOM), k-nearest neighbor (KNN), naïve Bayes (NB), and fuzzy analysis. Each of these data mining techniques serves a particular purpose, problem, and business need. Additional details on these techniques are readily available from a myriad of references on data mining.

Many researchers have investigated the application of data mining in accounting. However, each of these researchers focused on some specialized aspect of this broader topic, and none, to the best of our knowledge, has provided an all-encompassing overview. One of the early papers was that of Foltin and Garceau (1996), which demonstrated the differences between expert systems and neural networks and the future of neural networks applications in accounting. Coakley and Brown (2000) covered the modeling issues of neural networks in accounting and finance and classified them by research question, type of output (continuous versus discrete), and the parametric nature of the model. Yang (2006) pointed out how data mining is useful in both auditing and fraud detection. More specifically in auditing, Baldwin et al. (2006) highlighted the opportunities for AI in auditing, Calderon and Cheh (2002) provided a roadmap for future neural networks research in auditing and risk assessment, and Koskivaara (2004a) reviewed the use of neural networks in auditing and concluded that the focus was mainly on analytical review procedures. In the area of forensic accounting, Wang (2010) provided a review of data mining-based accounting-fraud detection research and summarized the data structures, algorithms, findings, and model performance evaluation with the aim of helping the accountants in selecting the suitable data and data mining technologies for detecting fraud. Furthermore, Ngai et al. (2011) explored the application of data mining techniques in the detection of financial fraud, and Gray and Debreceny (2014) provided a taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. More specifically, Debreceny and Gray (2011) provided an overview of how data mining techniques can mine emails and how such techniques and applications can be used by auditors as audit evidence. Ravi Kumar and Ravi (2007) provided a review of the application of data mining in bankruptcy prediction in banks and firms during the period 1986–2006. Their review highlighted the techniques applied, sources of data, financial ratios used, country of origin, time line of study, and the comparative performance of techniques in terms of prediction accuracy. More broadly, Fisher et al. (2010) and Chakraborty et al. (2014) applied text and data mining to automatically classify academic articles in accounting and improve understanding of the accounting lexicon. Thus, so far researchers focused their reviews of data mining applications to a specific topical context. We did not find any research that provides a comprehensive view of data mining applications in the broader accounting context, and the paper will attempt to address this gap.

3. Research methodology

Broadly following the methodology of a systematic review outlined in Tranfield et al. (2003) and Khlif and Chalmers (2015), our research methodology consists of seven steps:

Step 1. Scoping of the study: This study focuses on the application of data mining in accounting.

Step 2. Identification of search terms: To frame the scope of the study, we identified keywords that we used as search terms to capture relevant articles. We included accounting-related search terms such as accounting, financial, auditing, costing, fraud, combined with any of the following data mining-related search terms such as data mining, AI, big data, machine learning, clustering, decision tree, genetic algorithm, neural network, self-organizing map, regression, case-based reasoning, nearest neighbor, and Bayes. We consider these keywords and their combinations to represent a reasonably broad set of search terms to unravel the relevant literature.

Step 3. Identification of data sources: Our data sources consist of: (a) leading accounting journals[1], (b) all journals, not overlapping with (a), published by the American Accounting Association (AAA), (c) University Library e-Resources, which include subscription to >80 major electronic databases, (d) OhioLINK, which is an electronic database of >9000 journals from 101 publishers, and (e) Google Scholar (see Fig. 1). We believe, these outlets represent a comprehensive collection of literary sources that will cast a wide enough net to cover research relevant to the scope of the study.

Step 4. Article collection: We searched for literature on data mining applications in accounting using combinations of the search terms specified in Step 2, without time or outlet constraint in the multiple electronic sources (similar to Grabski et al., 2011; and Richardson et al., 2015). We also included articles from OhioLINK's and Google Scholar's "related papers" functionality during the collection process.

Step 5. Article filtering: A manual inspection and filtering process is undertaken by the authors to only include papers that satisfy the following inclusion criteria: (1) describe a specific application of data mining in accounting, (2) explicitly describe what data mining techniques have been utilized, and (3) the data mining goal and tasks are discernible from the paper. All other papers that either did not describe a concrete data mining application, such as interpretive articles, commentaries, and literature reviews, or did not provide enough detail to satisfy the inclusion criteria were excluded. A total of 209 papers satisfied the inclusion criteria.

Step 6. Content evaluation: We utilized a data extraction form to capture an article's:

- bibliographic details (including author(s), publication date, title, journal, volume, issue, pages);
- reporting focus (retrospective or prospective);
- accounting topic (e.g., accounting information systems, financial accounting, managerial accounting, compliance and assurance);
- accounting sub-topic (e.g. financial analysis, financial performance, budgeting, asset management, cost management, auditing cycle, business health, forensic accounting, tax compliance);
- data mining goal (description, prediction, or prescription);
- data mining task (e.g., association, classification, clustering, estimation, exploration, forecasting, optimization); and,
- data mining technique(s) (e.g., regression, neural networks, genetic algorithms, decision trees, support vector machines, case-based reasoning).

Step 7. Synthesis and framework development: section 4 details this step.

Our search goal was to capture literature on as many data mining accounting applications as possible, and identify the nature and the major areas of such applications. Our methodology is by no means without limitations. For example, many other search terms could have been used. However, no search strategy could exhaust all possible relevant terms in either accounting or data mining. We believe we have included the major outlets and search terms to capture the major literature related to the applications of data mining in accounting.

4. Proposed framework

Given the large amount of research that has been produced on the use of modern data mining technology in the field of accounting, an obvious question is: can this research be presented in a structurally logical and thematically coherent manner? In an attempt to answer this question in the affirmative, we propose an organizing framework for the applications of data mining in accounting (Fig. 2). Frameworks that organize literature succinctly summarize the topology of the literature, provide better understandability to complex relationships, and offer a convenient mechanism of mapping research in a given domain. We adopt the methodology of juxtaposing elements from different entities to construct a framework, similar to the approach used by Preece and Rombach (1994) and Richardson et al. (2015). While Preece and Rombach (1994) created a framework by combining measurement approaches from the disciplines of software engineering and human-computer interaction, Richardson et al. (2015) built their framework by linking elements of a professional entity (accountant) with elements of a profession entity (accounting). Our proposed framework combines characteristics of accounting discipline with characteristics of data mining discipline. Specifically, it combines the two well-known major reporting perspectives of accounting (retrospective and prospective) and the well-established three main data mining goals (description, prediction, and prescription).

The retrospective-prospective duality of reporting in accounting is manifested in the work of Birnberg (1980), Carnegie (2012), and Owen (2013). Retrospective reporting deals primarily with reflective reporting of the historical financial position of an organization mainly for financial valuation, decision making, and/or compliance purposes. For example, preparing financial statements provides a retrospective summary of an organization's financial position at a point in time (balance sheet) or profit or loss for a span of time (income statement). Prospective reporting, on the other hand, is future-oriented and includes future financial outlooks, estimations, and projections. For example, when historical information is used for predicting some future aspect of an

[1] Top accounting journal is assigned using 2014 Thomson Reuters journal citation reports with impact factor≥1. These, alphabetically, are: Accounting, Auditing and Accountability Journal, Accounting, Organizations and Society, Contemporary Accounting Research, International Journal of Accounting Information Systems, Journal of Accounting and Economics, Journal of Accounting Research, Management Accounting Research, Review of Accounting Studies, and The Accounting Review.

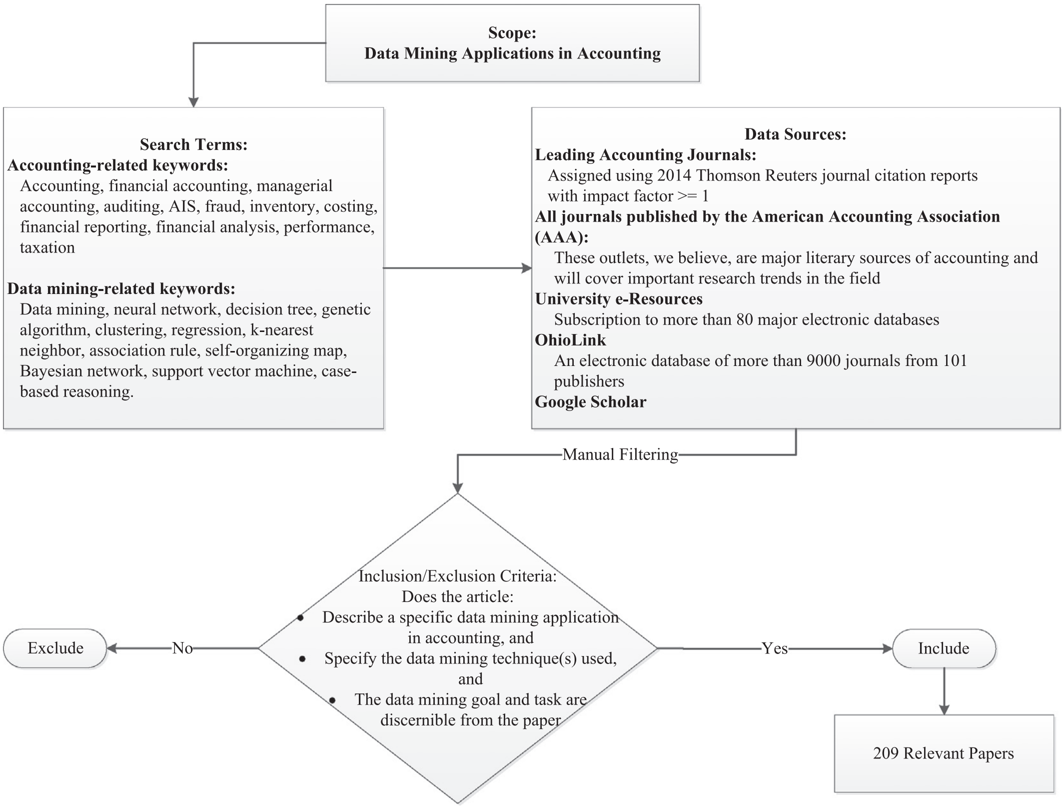

Fig. 1. Search methodology.

organization, such as the direction or the magnitude of its future performance, growth, or any other financial performance or health indicator, the focus of accounting reporting becomes prospective.

Data mining has three main goals: description, prediction, and prescription. The main goal of descriptive data mining is business and data understanding (the what happened), the goal of predictive data mining is using the past to understand the future (the what could happen), and the goal of prescriptive data mining is to achieve the best outcome (the what should happen). Descriptive data mining, the most commonly used and most well understood type, focuses on the use of data to understand the past

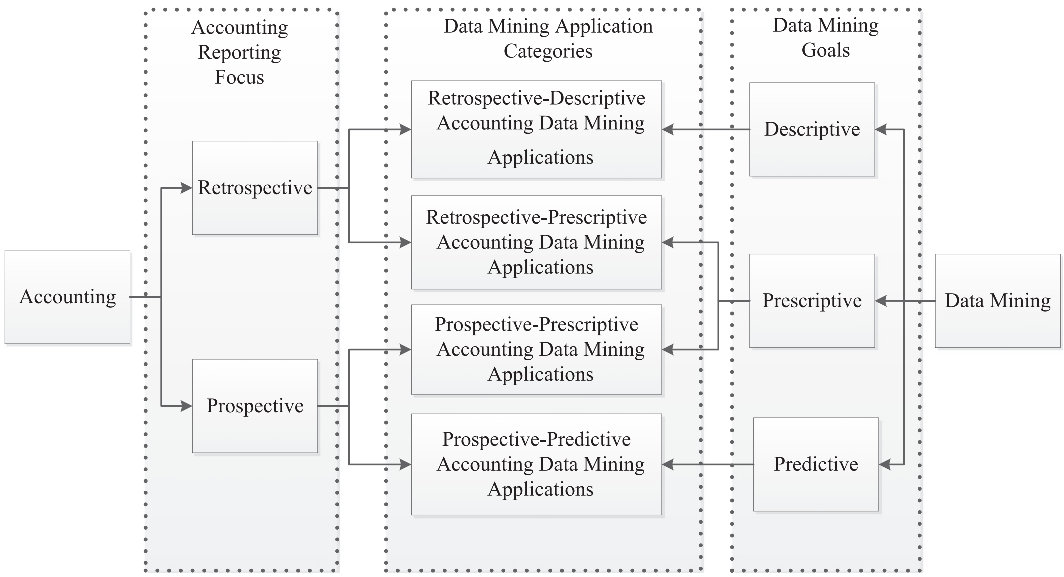

Fig. 2. Proposed framework.

and present and, accordingly, make informed decisions. It uses techniques to categorize, characterize, consolidate, and visualize data to convert it into useful information for the purposes of better data and business understanding. Descriptive data mining enables users to identify patterns and trends in data and discover problems and/or areas of opportunity. On the other hand, predictive data mining analyzes the past in an effort to predict the future by examining historical data, detecting patterns or relationships in these data, and then extrapolating these relationships forward in time. For example, using predictive data mining, a bank system might alert a credit card customer to a potentially fraudulent charge. Predictive data mining primarily informs future-oriented decision making. Finally, prescriptive data mining uses optimization techniques to identify the best alternatives to minimize or maximize some objective function. The mathematical and statistical techniques of predictive data mining can be combined with optimization to make decisions that take into account the uncertainty in the data (Evans, 2013). Whether the goal is retrospective or prospective reporting, prescription (optimization) may be utilized to support maximizing or minimizing these goals in the most resource-efficient manner. Indeed, data mining provides advanced techniques to facilitate descriptive, predictive, and prescriptive modeling, and thus it has a great deal to offer in direct support for the major reporting perspectives of accounting.

Juxtaposing the two major perspectives of accounting reporting and the three main goals of data mining, six combinations result. Only four of these combinations are logically feasible; namely, (1) descriptive data mining in retrospective reporting, (2) prescriptive data mining in retrospective reporting, (3) prescriptive data mining in prospective reporting, and (4) predictive data mining in prospective reporting. These four combinations represent the major groupings of the proposed framework and are used to organize the published research on accounting applications of data mining. We, respectively, refer to the categories formed by these combinations as retrospective-descriptive, retrospective-prescriptive, prospective-prescriptive, and prospective-predictive (Fig. 2).

Retrospective-descriptive applications focus on business and data understanding from a historical viewpoint. The main data mining tasks used by this strand of applications include exploration, clustering, association analysis, visualization, segmentation, and pattern recognition. Retrospective-prescriptive and prospective-prescriptive applications emphasize efficiency, and thus utilize optimization and estimation as data mining tasks, yet they differ on their temporal orientation; where retrospective-prescriptive applications focus on the past and the present, and prospective-prescriptive applications focus on the future. Finally, prospective-predictive applications focus on a future business aspect using historical data, and employ data mining tasks such as classification, forecasting, and estimation. The proposed framework is comprehensive, simple, easy to understand, and empirically verifiable. Section 5 demonstrates how the proposed framework is capable of capturing the research on data mining applications in accounting.

5. Results and discussion

The review of the applications of data mining in accounting reveals various patterns relating to temporal trends; data mining goals, tasks, and techniques used; primary accounting sub-domains and areas covered; current literature macro themes and patterns; results of mapping the literature to the various categories of the proposed framework, and intensity of coverage in each of these categories. These findings are elaborated in the following sub-sections.

5.1. Temporal trends

Our search methodology identified a total of 209 applications (23 described in conference proceeding papers and 186 described in journal articles) of data mining in accounting between 1989 and 2014. There are clear upward trends in the application of data mining in accounting between 1995 and 2001 and 2004–2014, with the highest number of applications in 2013 (Fig. 3). It appears that accounting researchers and professionals have realized benefits of applying data mining to accounting and thus have shown greater propensity to adopt it over time. Of special interest is the quantum leap in the number of such applications since 2010. One possible reason for such a leap is the need for more modeling sophistication in the accounting practice following the major worldwide financial crisis of 2008 and the subsequent business meltdown.

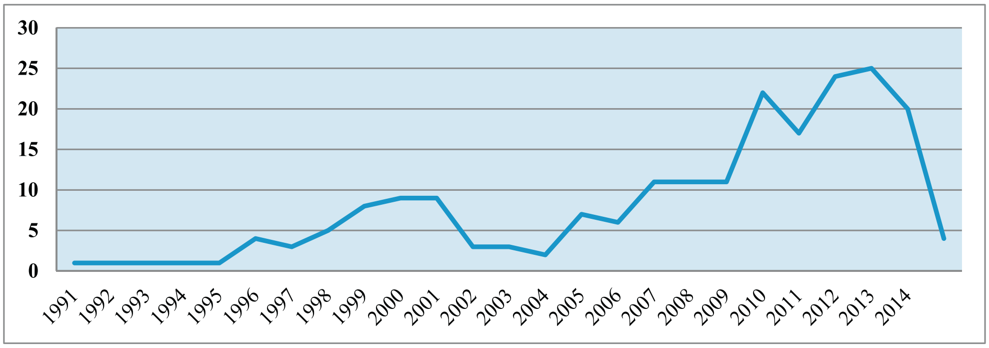

Fig. 3. Number of data mining applications in accounting 1989–2014.

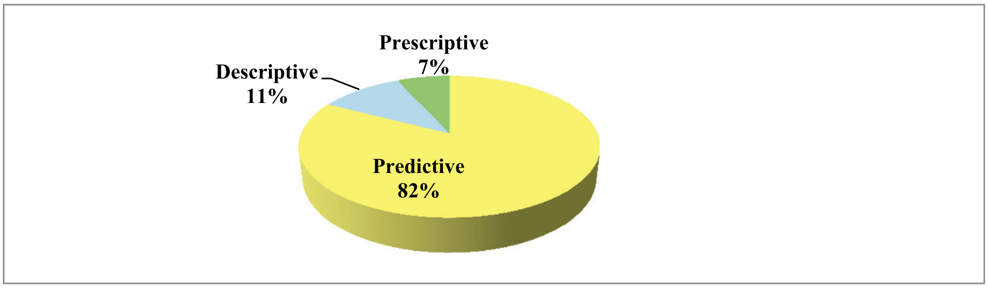

Fig. 4. Applications of data mining in accounting by data mining goals.

5.2. Goals, tasks, and techniques

The analysis shows that the vast majority (82%) of applications focused on predictive data mining, 11% on descriptive, and 7% on prescriptive (Fig. 4). These patterns may be a reflection of the fact that prediction provides more strategic value to accounting decision making as it embodies future outlook, strategic orientation, guidance, and positioning. Hence, the heavier emphasis on prediction, as opposed to description and prescription, in applications of data mining in accounting.

Analysis of the reviewed papers revealed that the classification data mining task is used by the vast majority (67%) of data mining applications, followed by estimation (12%), clustering (6%), and optimization (5%), and that the least used data mining tasks are pattern analysis (<0.5%), exploration (2.5%), and association (2.5%) (Fig. 5). There is a clear trend in the accounting data mining applications to favor binary classification as it seems to fit many of the problems tackled in these applications; for example, the objective being to classify into a binary class of: fraud/no fraud, bankruptcy/no bankruptcy, healthy/not healthy, good performance/poor performance,… etc. It seems, however, that accounting has yet to fully benefit from the many other important data mining tasks such as pattern and association analyses that have shown great benefits in other business disciplines such as marketing (Cil, 2012; Liao and Chen, 2014).

Data mining is a multi-disciplinary approach that uses a variety of techniques from statistics, machine learning, databases, and others. The analysis of the literature shows that neural networks is the most widely used technique (Table 1), and was used by almost half (47%) of the applications. Such dominance of neural networks may be due to the nature of neural networks as a general problem solving technique that can be utilized in all data mining types, tasks, and business problems. Regression, a well-established and accepted method, comes a distant second and was used by 20% of the applications. Following behind are decision trees used by 14%, support vector machines and genetic algorithms each used by 11% of the applications. Other less extensively used techniques include: text mining, self-organizing maps, k-nearest neighbor, discriminant analysis, association rules, case-based reasoning, Bayesian networks, and k-means. There may be a familiarity gap in the accounting community with the more advanced data mining techniques, hence reflected by the low usage of these techniques.

5.3. Application areas

A topical analysis of data mining applications in accounting showed that almost two thirds (64%) of these applications focused on assurance and compliance, one fourth (25%) on managerial accounting, and the remaining on financial accounting and accounting information systems (AIS) (due to the small number of AIS applications, they are sometimes combined with financial

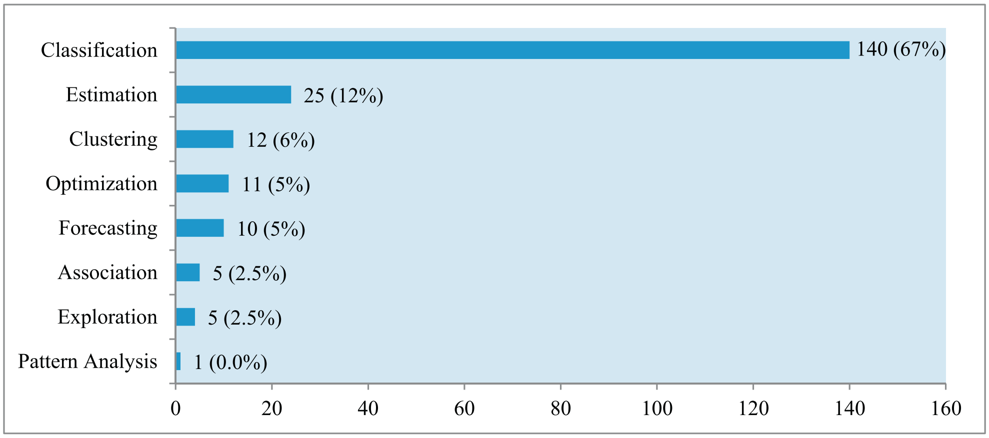

Fig. 5. Applications of data mining in accounting by data mining tasks.

Table 1
Data mining techniques used in accounting applications*.

Data mining technique	Count
Neural networks	99
Regression	41
Decision tree	30
Support vector machines	23
Genetic algorithms	22
Text mining	15
Self-organizing maps	13
Discriminant analysis; K-nearest neighbor, Bayesian networks	9
Association rules; Case-based reasoning	7
K-means; Fuzzy analysis	6
Expert systems	5
Data envelopment analysis	4
Analytic hierarchy process; Principal component analysis; Hybrid	3
Proprietary; Rough sets; Process mining	2
Collocational networks; Digital analysis; OLAP; PMI (pointwise mutual information); Linear programming; Particle swarm optimization	1

* Some applications reported using more than one technique.

accounting applications) (Fig. 6). Fig. 7 further summarizes the number of applications in each accounting topic and sub-topic, and the sections that follow provide more detail on each of these application areas. The varying intensity of using data mining across the various branches of accounting may reflect the intensity of the need for advanced analytics in each of these branches. The well-publicized auditing failures and corresponding bankruptcies as well as the tightening of regulatory legislations and oversight may have necessitated the search for advanced technological support in the domain of assurance and compliance. Similarly, competitive pressures and the pursuit of corporate efficiencies may have created more need for data mining in managerial accounting than in financial accounting.

5.3.1. Data mining in AIS

Few authors examined the application of data mining in the area of AIS. For example, Wang et al. (2009) used self-organizing maps, nominal data analysis, and the concept of entropy in building a chart of accounts structure for enterprise resource planning (ERP) accounting information system. Although Wang et al. (2009) reported savings of effort and time as a result of utilizing data mining techniques in addressing this core accounting problem that represents the initial step in an ERP accounting module implementation, their sample size was very limited (five data sets) and they did not shed any light on the accuracy or reliability of their approach. Zheng (2011) built a resources, events, agents (REA)-based accounting information systems framework that, combined with data warehousing, decision support systems, data mining, and other information technologies was adaptable to e-commerce environments. Zheng (2011) suggested an enabling role of data mining in an accounting information system, particularly in the an e-commerce context, yet there is no clear reason why can't data mining play a similar role in accounting information systems in general.

It is not clear whether this thin research coverage of data mining applications in AIS is due to lack of reporting of such applications or due to true lack of such applications. If the former, it may be, and understandably so, because of the unwillingness to reveal these applications for competitive considerations. If the latter, it is counter-intuitive as one would expect AIS to be a major beneficiary from such key analytics technology, and thus represents a research gap and an opportunity to showcase the power of data mining in AIS. One could also conjecture that AISs are viewed more as technology rather business systems and may thus be tended to by technical staff rather than business people, thus missing the opportunity to leveraging the business dimension of these systems by incorporating advanced data mining applications that aim primarily to deriving business benefits.

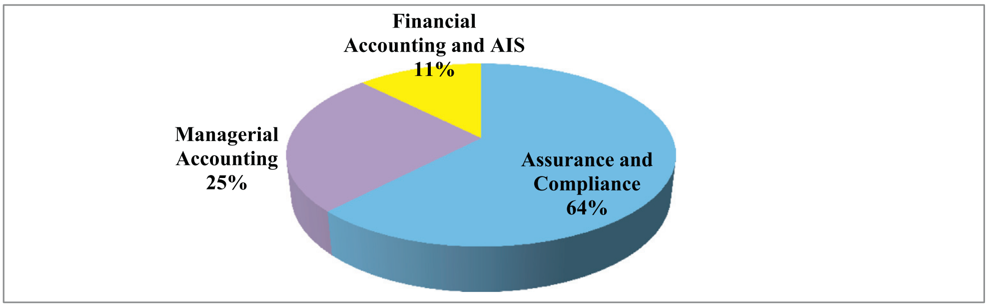

Fig. 6. Data mining applications in primary accounting topics.

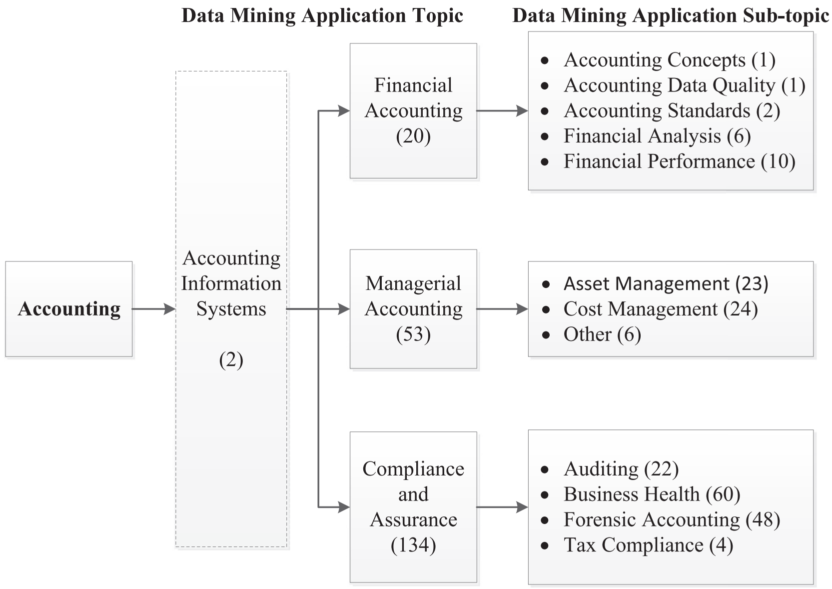

Fig. 7. Data mining applications in accounting topics and sub-topics (numbers in parentheses indicate the number of data mining applications in the corresponding area).

5.3.2. Data mining in financial accounting

Financial accounting applications mainly examined financial performance and analysis. One of the earliest applications of data mining in this area was that of Callen et al. (1996), in which they built a neural networks model to forecast quarterly accounting earnings. This work benchmarked neural networks against linear time series forecasting models, and reported that the linear time series models yielded better quarterly earnings forecasts than an artificial neural network model. The research that followed this early work, reports, in the majority, the opposite of this finding on various problems, perhaps replicability of Callen et al.'s (1996) experiment was difficult due to the lack of exact specification of their neural network model. Back et al. (2001) used self-organizing maps to compare company performance extracted from numerical information versus that extracted from textual information in annual reports. Similarly, but mainly focusing on the future, Kloptchenko et al. (2004) and Magnusson et al. (2005), used data mining techniques to analyze quantitative and qualitative contents of financial reports to predict future financial performance; both concluding that while textual content is more informative of future performance, quantitative content is more informative of past performance. Although the primary strength of these studies is in leveraging the power of textual information, in addition to that of quantitative information, their scope is very narrow: focusing on international pulp and paper industry. Thus the question of generalizability of their findings to other industries is still open.

To enhance financial performance analysis, and taking into consideration both homogeneity of size and sector in their experiment, Hofmann and Lampe (2013) used clustering on balance sheet structure of logistics service providers. This study focused only on the balance sheet and on macro-level variables, but financial performance goes far beyond the balance sheet and spans all other financial statements. Hence Hofmann and Lampe (2013) analysis is bound to have missed important variables relevant to financial performance, but fall outside the scope of the balance sheet. More specifically, to improve financial ratio analysis, Landajo et al. (2007) developed a robust neural networks model for the cross-sectional analysis of accounting information, and Eklund et al. (2008) used self-organizing maps to identify a set of ratios for deriving a financial benchmarking model. The strengths of these studies are that they took into consideration the error cost dimension in evaluating different models and selected financial ratios based on their empirical reliability and validity in international comparisons. Perhaps feature selection techniques could have been a more direct data mining approach to selecting ratios relevant to a specific task – such as financial analysis.

Huang and Li (2011), using advanced text mining features (but not considering interdependence between risk factors), developed a multi-label text classification k-nearest neighbor algorithm to identify risk factors of annual reports. Koskivaara (2004b), an early adopter of SOMs and visualization for financial analysis – although in the context of a single-medium-sized company, used self-organizing maps for classifying and clustering accounting data for signaling unexpected fluctuations. Focusing more on discovering patterns of data quality issues, Alpar and Winkelsträter (2014) applied association rules to accounting transactions data. Their results showed that not including error cost considerations in classification models evaluation could lead to economically bad decisions. However, although their procedure may be useable in other companies, the discovered association rules are company-specific.

Many authors used data mining in accounting at a more macro level. Spear and Leis (1997) developed well-specified, yet too-specific, multiple supervised neural network models to improve the choice of accounting method (full cost vs. successful effort) for oil and gas producing companies. Beaudoin et al. (2010), using logistic regression, examined the potential effects of accounting policy Statement No. 158 of Financial Accounting Standards on management actions. This is one of the few works that used a balanced matched-sample experimental design to contribute to the debate on costs and benefits of accounting regulation. Lodhia and Martin (2011), uniquely focusing on the use of data mining for carbon accounting and reporting, explored the use of text mining in environmental accounting, and whether broader climate change issues were addressed in submissions made by corporations and other stakeholders to regulatory agencies. Garnsey (2006) used clustering to derive related accounting concepts to improve access to, and retrieval of, financial accounting material. Other authors (such as Henry, 2006; Cho et al., 2010; Li, 2010; Davis and Tama-Sweet, 2012; Huang et al., 2013; Huang et al., 2014) focused on the use of text mining in analyzing the narrative across different disclosure outlets to examine the relationship between the tone of the disclosure, future performance, investor, and market reactions. Their overall consensus is that inclusion of predictor variables capturing verbal content and writing style of earnings-press releases results in more accurate predictions of market response earnings announcements. In addition, the language and verbal tone used in corporate environmental disclosures, in addition to their amount and thematic content, should be considered when investigating the relation between corporate disclosure and performance.

Data mining in financial accounting has primarily focused on financial performance and ratio analysis; such as forecasting quarterly accounting earnings, comparing informational value of numerical versus textual data for performance measurement, financial performance benchmarking, identifying risk factors in annual reports, visualization of patterns in accounting data, assessment of quality of accounting data underlying financial reports, and the impact of management announcements and their tones on market response, among others. These applications have primarily focused on description and prediction as goals, and used clustering and classification as data mining tasks. Neural networks and text mining are the most prevalent techniques of these applications. Future research opportunities include more utilization of textual components of the financial reports in predicting financial performance, importance of domain expertise in data mining applications, paying more attention to data quality issues, importance of benchmarking of applications, going beyond financial ratios to capture relevant inputs for better prediction financial performance, sensitivity analysis of derived models to data characteristics, and the importance of considering variable relationships in the analysis.

5.3.3. Data mining in managerial accounting

Managerial accounting applications focused on major areas such as cost management, asset management, and budgeting and pricing management.

5.3.3.1. Cost management. Data mining has been applied in the area of cost management at various costing levels: equipment, process, construction, product, and project. At the *equipment* level, data mining has been used for estimation of equipment manufacturing cost (Chou et al., 2010; Chou et al., 2011), for improving the accuracy of equipment inspection and repair (Chou and Tsai, 2012) and for tracing equipment replacement costs (Dessureault and Benito, 2012). The application of data mining to cost management strand of research demonstrated many strengths and weaknesses. Strengths include well-defined accuracy measures (Chou et al., 2010), use of hybrid paradigms (Chou et al., 2011; Kostakis et al., 2008), utilization of hierarchical analysis approaches (Chou and Tsai, 2012), and highlighting of the importance of data understanding (Dessureault and Benito, 2012). Weaknesses include a very limited scope and sample size (Chou et al., 2010; Chou et al., 2011, and Chou and Tsai, 2012), as well as use of limited variables (Dessureault and Benito, 2012), or simulated data (Kostakis et al., 2008). At the business *process* level, data mining has been used for defining cost drivers in activity-based costing and improving production process routing, (Kostakis et al., 2008; Liu et al., 2012), as well as for intelligent transfer price decision making (Kirsch et al., 1991). At the *construction* level, authors (Yu et al., 2006; Shi and Li, 2008; Migliaccio et al., 2011; Vouk et al., 2011) focused on the application of data mining to construction cost management, creating a neural networks system for simple, fast, and adequately accurate estimation of total or unit cost of construction, operation, and maintenance. Some of these studies are not replicable because the variables used are not specified (e.g. Yu et al., 2006), and some are not integrated into the existing operational systems (e.g. Shi and Li, 2008). Many authors applied data mining to *product* costing; namely, for forecasting product unit cost (Chang et al., 2012), estimating product life-cycle cost (Seo et al., 2002; Yeh and Deng, 2012), estimating project design cost (Deng and Yeh, 2010), and estimating product manufacturing cost (Deng and Yeh, 2011). These studies utilized the power of hybrid data mining modeling techniques to report their results, yet they are narrowly-focused on a single industry, company, and/or product, and thus can hardly be considered of wide applicability. At the *project* level, data mining has been used to develop a project-level cost control system (Zhao and Ding, 2009; Ji et al., 2010, 2011: Kaluzny et al., 2011; Petroutsatou et al., 2011), and to develop a project-level cost estimation system (Shan and He, 2012). These cost estimation applications are not limited to tangible products or projects, but also extend to estimation of cost for intangible projects such as software projects (Huang et al., 2007; Khalifelu and Gharehchopogh, 2012).

5.3.3.2. Asset management. Inventory management, including inventory classification, costing, optimization, and controlling is a key factor that influences company competitiveness. In the area of inventory control, neural networks were used to optimize inventory level (Bansal et al., 1998a, 1998b; Reyes-Aldasoro et al., 1999), reporting a 50% reduction in inventory cost (from over a billion dollars to about half-a-billion dollars) while maintaining the same level of probability that a particular customer's demand will be satisfied. In addition, these papers describe the use of traditional statistical techniques to help determine the best neural network type for a particular application. To better manage inventory, many data mining techniques were used, including decision trees (Braglia et al., 2004), fuzzy neural networks (Li and Kuo, 2008), and genetic algorithms (Zeng et al., 2006). These researches reported improved management processes and, consequently, reduced inventory holding costs by using hybrid data mining models, including combinations of fuzzy systems and neural networks, integrating neural networks, analytic hierarchy process, and genetic algorithms.

The predictive accuracy of data mining techniques on the LIFO/FIFO and ABC inventory classification has been an active area of research to minimize total inventory costs, including ordering cost, holding cost, purchase cost and transportation cost (Liang et al., 1992; Altay Guvenir and Erel, 1998; Partovi and Anandarajan, 2002; Šimunović et al., 2009; Yu, 2011; Kabir and Hasin, 2013). An interesting finding in these studies is that that there was no significant difference between the backpropagation and the genetic algorithms learning methods on the predictive accuracy of neural networks to classify inventory. Data mining has also been used to improve inventory management in e-commerce environments (Chodak and Suchacka, 2012), resulting in better recommender systems that take into consideration product cost, which results in moving inventory items that otherwise remain dormant in e-stores. Many authors (Gaafar and Choueiki, 2000; Megala and Jawahar, 2006) addressed the material requirements planning (MRP) lot-sizing problem using various data mining techniques, genetic algorithms (Lee et al., 2013) to develop an integrated model for lot-sizing with supplier selection and quantity discount, neural networks and genetic algorithms (Zhou et al., 2009) to optimize the multi-objective function of selecting materials for a product, similarly (Wu and Hsu, 2008) for designing bill of material configuration for reducing logistic costs for spare parts inventory, stochastic neuro-fuzzy (Gumus et al., 2010) for inventory management in a multi-echelon environment, and neural networks (Wang, 2011) for classification of inventory risk level. The overall focus of these authors is to achieve high solution quality at acceptable computational time, and the common finding is that data mining techniques, used individually or integrated in hybrid models, are capable of solving the static or dynamic lot-sizing problem with notable consistency and reasonable accuracy.

In addition, data mining techniques have been used to improve the accuracy and efficiency of asset evaluation (Liu and Ren, 2009), for identifying important factors affecting intangible assets value (Tsai et al., 2012), and for improving prediction of cash flow (Cheng and Roy, 2011). Accuracy of physical or intangible asset valuation is important to both investors and creditors, especially in the context of knowledge-based economies where assets are becoming more and more intangible knowledge as opposed to physical assets. Thus the need for novel approaches to valuation of such intangible assets as reported in Tsai et al., 2012 who used feature selection, an important data-preprocessing step in data mining, to identify important and representative factors affecting intangible assets, concluding that their model is simple and feasible, and improve the valuation accuracy and efficiency.

5.3.3.3. Other. In addition to the broad categories of cost and asset management, this study revealed application of data mining to other areas of managerial accounting. For *budgeting* (Chou, 2009; Tang, 2009), developed a web-based case-based reasoning system for early cost budgeting to assist decision makers in project screening, and applied fuzzy analytic hierarchy process for multi-criteria decision-making to improve budget allocation decisions. In the area of *revenue management* Ragothaman and Lavin (2008) used neural networks to curb improper revenue recognition practices by predicting firms that will restate their revenue. Their results show that the neural network model has superior predictive power for predicting revenue restatement firms compared to the Multiple Discriminant Analysis (MDA) and Logit models; although the Logit and MDA models predict nonrevenue restatement firms better. Moreover, when misclassification costs are taken into consideration, the neural network model performs the best with the lowest relative misclassification costs. On the other hand, to improve revenue underpayment recovery process in a healthcare organization, Hennigan and Chowlera (2011) developed a proprietary data mining algorithm to recover more than $20 million in less twenty months of implementation, and to increase staff productivity by 100% in less than six months, allowing auditors to resolve 40 to 50 claims per day versus 20 per day prior to their systems implementation. For account *reconciliation*, Chew and Robinson (2012) explored application of natural language processing to achieve 100% precision and recall on a real-life dataset, suggesting that their approach is highly reliable and eliminated most of the manual work for their test problem, suggesting the possibility of highly desirable improvements in information technology controls to reduce the cost of external audit work. In the area of *mergers and acquisitions*, Shawver (2005) used neural networks for accurately predicting bank merger premiums to better price mergers, accrue competitive advantage in pricing merger offers, and enhance the possibility that the merger will achieve its intended financial, strategic, and/or operational synergy.

Data mining applications in managerial accounting mainly addressed cost management at different levels (product, equipment, process, and project levels), asset (mainly inventory) management, among others with less emphasis. These applications cover many specific implementations such as classifying, selecting, predicting, and optimizing inventory management, defining cost drivers, estimating and forecasting project and product cost, developing cost estimation models for product, equipment, and project, developing budgeting systems, predicting cash flows, etc. The primary goals in these applications involve prediction and prescription. The main tasks are estimation and optimization, and the predominant technique is neural networks. Future research opportunities in this domain include: exposure of these applications using web services, better identification of relevant variables, more sensitivity analysis of generated models, improve model parsimony, improve data handling, clearly differentiating between causation and association, addressing issues of data availability, and cross-industry model validation.

5.3.4. Data mining in assurance and compliance

5.3.4.1. Auditing. The accounting transactions are becoming more complicated and easier to manipulate with the increasing use of online systems and the proliferation of smart devices and the internet of things. This necessitates a more sophisticated auditing profession, including an increasing use of the advanced techniques of data mining. Needless to ignore, the important role of information technology has to play in improving the efficiency of the monitoring and controlling process (Daigle and Lampe

2005). Data mining has been applied throughout the *auditing cycle*: planning (such as engagement, risk assessment, design of audit plan), conducting (mainly performing substantive audit tests), and reporting (audit report). Data mining has also been applied post the audit cycle, including impact and consequences of auditor's opinion.

In the *engagement* phase, data mining has been used to predict auditor selection (Kirkos et al., 2008, 2010) and switching (Kirkos, 2012), to find the optimal match between the audit project characteristics and auditor expertise in public construction projects (Wang and Kong, 2012), and to classify the level of corporate audit costs and variation in audit fees (Curry and Peel, 1998; Beynon et al., 2004). In today's information-rich environments, *risk assessment* involves recognizing patterns in the data, such as complex data anomalies and discrepancies that may conceal one or more error or hazard conditions (Ramamoorti et al., 1999). Calderon (1999) and Ramamoorti et al. (1999) studied the ability of neural networks to enhance auditors' risk assessment and reported that neural network modeling is invaluable in directing internal auditor attention to those aspects of financial, operating, and compliance data most informative of high-risk audit areas, and thus enhances audit efficiency and effectiveness. Similarly, Davis et al. (1997) and Hwang et al. (2004) developed neural networks models to support auditors in conducting control risk assessment; concluding that neural networks provide auditors an effective way to recognize patterns in the large number of control variable inter-relationships that even experienced auditors cannot express. Likewise, Issa and Kogan (2014) proposed a predictive logistic regression model as a tool for quality review of control risk assessments, and thus improve audit efficiency by focusing on the concept of audit by exception. For audit *planning*, Ragothaman et al. (1995) developed a rule-based system that assists auditors at the planning stage in the design of subsequent substantive tests, when material errors and irregularities in the financial statements are probable; demonstrating that this system outperforms a model based on discriminant analysis in classifying firms into error and non-error categories. Unfortunately, the sample size used in their study limits the generalizability of the generated rules.

Some of the interesting findings in the application of data mining in the audit engagement phase is that the level of debt is a factor that influences the auditor choice decision, gross profit is the variable that best predicts auditor switching, and that adoption of an effective auditor procurement process increases the likelihood that a company will engage and match the right auditor at a fair price (Kirkos et al., 2008, 2010; Kirkos, 2012; Wang and Kong, 2012). In addition, Curry and Peel, 1998 report that neural network models exhibit superior forecasting accuracy to their ordinary least squares counterparts in predicting the cross sectional variation in corporate audit fees, although this differential reduces when the models are tested out of sample.

In the audit *conducting* phase, Argyrou and Andreev (2011) proposed a semi-supervised tool for clustering accounting database as an internal control procedure through usage of self-organizing maps to supplement internal controls, verify processing of accounting transactions, and assess accuracy of financial statements. Their empirical results suggest that the proposed tool can compress a large number of accounting transactions, generating homogeneous, well-separated, and interpretable clusters. In *performing substantive tests*, Coakley and Brown (1993) and Koskivaara (2000a) used neural networks in predicting patterns in auditing monthly balances as part of auditors' analytical review process, and suggest that neural networks recognize patterns within financial accounts as well as the dynamics and the relationships between these accounts more effectively than did financial ratio and regression methods. Koskivaara (2000b), focusing on the pre-processing of the data, investigated the ability of neural networks for recognizing the dynamics and the relationships between financial accounts values in order to detect unexpected fluctuations. His findings indicate that the best results were achieved when all the data were preprocessed by scaling them either linearly or linearly on a yearly basis, yet no further elucidation is provided for why this is the case – although the author cautioned about the stability of the proposed model. Along the same lines, Coakley (1995) suggested the use of neural networks in pattern recognition of the investigation signals generated by analytical procedures; demonstrating that the use of neural networks provides a more reliable indication of the presence of material errors than either traditional analytic procedures or pattern analysis, and also provides insight into the plausible causes of these errors. Their results suggest that the use of an ANN to analyze patterns of related fluctuations across numerous financial ratios provides a more reliable indication of the presence of material errors than either traditional analytic procedures or pattern analysis, offer improved performance in recognizing material misstatements within the financial accounts, and, not less importantly, provide insight to the plausible causes of the error. Koskivaara and Back (2007) proposed a neural networks model for analytical review for continuous auditing of financial data; namely, estimating the future revenues and expenses of an organization, concluding that the neural networks model is most successful for such estimates.

In the area of post auditing cycle, the informational content conveyed by the auditor's going concern opinion has substantial impact on a firm's current and future standing. Jones (1996) examined the abnormal stock returns surrounding the release of the auditor's going concern report using ordinary least squares regression, and found that ordinary least squares regression tests indicated that mean abnormal returns surrounding the release of the auditor's report were lower for going concern opinions than for clean opinions and that the magnitude of the abnormal returns depended on the extent to which the opinion type was unexpected by investors. Bhimani et al. (2009) examined the influence of the release on the firm ability to continue and the possibilities of subsequent default, revealing that the likelihood of default for firms that received going concern opinion is 2.8 times that of firms that received a clean opinion.

The auditing profession has become highly litigious in recent years. Misclassification of a potential future bankruptcy candidate as healthy is referred to as an audit failure, and may result in substantial litigation costs. For example, Ernst & Young was compelled to pay $400 million and KPMG paid $186.5 million for audit failures (Anandarajan et al., 2001). Blacconiere and DeFond (1997) investigated the independent audit opinions of publicly-traded savings and loans that subsequently failed and the resulting substantial independent auditor litigation, concluding that the only variable consistently related to independent auditor litigation was client size - possibly because failures of larger firms are more costly to regulators than the failure of smaller firms; in which case, regulators may be more likely to initiate lawsuits against auditors of larger firms. Chen et al. (2009a, 2009b) looked to the ability of data mining techniques, mainly neural networks, in predicting fraud litigation for assisting accountants on devising an audit strategy and found that neural networks are highly capable of identifying potential lawsuits.

In addition, data mining has been applied to improve business processes. For example, Jans et al. (2013) explored the value-added by process mining to audit practice, and Mueller-Wickop and Schultz (2013) demonstrated the benefits of process mining in audit domain through an algorithm that determines an activity sequence from accounting data to construct considerably improved business process models. These studies concluded that process mining, among its many benefits, allows the auditor to conduct analyses not possible with existing audit tools, such as discovering the ways in which business processes are actually being carried out in practice, and to identify social relationships between individuals.

5.3.4.2. Business health. In the audit conducting phase, the main goal is to assess the financial position of a firm in order to decide on the auditor opinion for the final reporting phase. In business health, researchers focused on three major areas in the application of data mining: financial viability, bankruptcy, and going concern. *Financial viability* or business failure can be defined as a situation that a firm cannot pay lenders, preferred stock shareholders, suppliers, or other creditors, or the firm goes into bankruptcy according to the law (Dimitras et al., 1996). The are many focus areas of using data mining in this area: supporting auditor's judgment about a client's continued financial viability (Etheridge et al., 2000), predicting business failure (Ahn et al., 2000; Chakraborty and Sharma, 2007; Tang and Chi, 2005; Huang et al., 2008; Youn and Gu, 2010; Benhayoun et al., 2013; Chen, 2013; Chen et al., 2013; Li et al., 2013), classifying, predicting, and preventing bank failures (Alam et al., 2000; Tung et al., 2004; Boyacioglu et al., 2009; Quek et al., 2009). A common feature of these researches is that almost all of them used a hybrid data mining modeling approach. Some of their main findings include: using overall error rate metric, a probabilistic neural network is the most reliable tool for predicting financial viability, but when the estimated relative costs of misclassification are considered, the best such predictor is the categorical learning neural network model. While neural networks and regression demonstrate comparable Type I errors, neural networks show lower Type II errors for both in-sample and hold-out sample predictions. Additionally, interest coverage is the most important signal of business failure (Youn and Gu, 2010). Chen, 2013 reported that different neural networks learning techniques have different accuracy of prediction across time horizons. Tang and Chi, 2005 investigated the influences of network architecture, variable selection, sample mixture of training and testing subsets on neural network models' learning and prediction capability.

Bankruptcy prediction is a critical topic that has been studied extensively and persistently in the accounting and finance literature. Many authors used data mining techniques for bankruptcy prediction (Jo et al., 1997; O'Leary, 1998; Yang et al., 1999; Zhang et al., 1999; Charalambous et al., 2000; Lee et al., 2005; Min and Lee, 2005; McKee, 2007; Tsai and Wu, 2008; Chen et al., 2009a, 2009b; Mokhatab Rafiei et al., 2011; Shirata et al., 2011; Olson et al., 2012; Fedorova et al., 2013; Kasgari et al., 2013; Korol, 2013; Serrano-Cinca and Gutiérrez-Nieto, 2013; Tinoco and Wilson, 2013). A somewhat surprising results are that of Yang et al., 1999 where back-propagation was reported to have failed to discriminate between bankrupt and non-bankrupt firms, and the superiority of linear discriminant analysis over probabilistic neural networks. On the other hand, Zhang et al., 1999 reported that neural networks are robust to sampling variations in overall classification performance. Shirata et al., 2011 work demonstrated the effectiveness of text mining bankruptcy prediction, in that certain combinations of terms were effective in distinguishing between bankrupt and non-bankrupt companies. More specifically, Pompe and Bilderbeek (2005) examined factors that influence bankruptcy prediction, noting that models generated from the final annual report published prior to bankruptcy were less successful in the timely prediction of failure, and economic decline coincided with the deterioration of a model's performance. While all these authors use only quantitative measures, mainly financial ratios, in their bankruptcy prediction modeling, Anandarajan et al. (2001) used both qualitative and quantitative measures. Whereas Cho et al. (2009) developed an integrated model combining statistical and AI techniques for bankruptcy prediction, others focused on the performance accuracy of bankruptcy prediction models (Tseng and Hu, 2010; Kim and Kang, 2010; du Jardin, 2010; Tseng and Hu, 2010; Kim and Kang, 2010; du Jardin, 2010) with non-conclusive agreement on which modeling technique offers the best predictive power. This conclusion is not surprising given the many different ways each technique can be parametrized and the specifics of the problem addressed. In a nutshell, there is no evidence that one data mining technique outperforms the others under all circumstances.

Auditing standards, namely SAS No. 126 in 2012, address the auditor's responsibilities in an audit of financial statements with respect to evaluating whether there is substantial doubt about the entity's ability to continue as a *going concern*. Determining the going concern status of a company is not an easy task. For predicting going-concern qualification, Peel (1989) used logistic regression and found that high gearing, low profitability, and low ownership concentration were consistently associated with the auditor's decision to issue a going-concern qualification. Koh and Tan (1999), Lenard et al. (1995), Koh and Low (2004) used logistic regression as well as neural networks and decision trees for predicting firm going concern status. The consensus of the first two studies is that neural networks are proposed as a robust model for auditors to support their assessment of going concern opinion. Koh and Low (2004) reported the superiority of decision tree as a predictive model for going concern over neural networks and logistic regression. Kleinman and Anandarajan (1999) used non-financial variables as predictors of an auditor's going concern opinion, and for supporting an auditor's going-concern assessment decision, highlighting the power of qualitative data in predicting going-concern qualification. Lenard et al. (2000) developed a going concern evaluation decision model based on fuzzy clustering and a hybrid model of a statistical model and an expert system to identify categories of firms with particular characteristics that may indicate whether or not the audit report of the firms requires a going concern modification. Whereas Lenard et al. (2001) examined decision making capabilities of a hybrid rule-based expert system and discriminant analysis to provide insight into the characteristics of firms that experience problems, but do not necessarily receive a going-concern modification, Martens et al. (2008) constructed a an effective going concern predictive system using support vector machines and rule-based classifiers. Shirata and Sakagami (2008) used text mining for clarifying the difference between going-concern and non-going-concern companies by analyzing the nonfinancial (qualitative) information disclosed in financial reports. Doumpos et al. (2005) developed a support vector machine model combining publically available financial information and credit-risk rating indicators in explaining qualifications in audit reports. Their major conclusion is that linear and non-linear support vector machines models are capable of distinguishing between qualified and unqualified financial statements, at a point in time as well as over time, with satisfactory accuracy. Spathis (2003) used logistic regression and ordinary least squares regression to test the extent to which combinations of financial and non-financial information can be used to enhance the ability to discriminate between the choices of a qualified or unqualified audit report. The qualification decision is associated with financial information such as financial distress and with non-financial information such as firm litigation. Salterio (1996) used case-based reasoning in investigating whether precedents and the client's preferred accounting policy affect auditors' accounting policy judgments. Kirkos et al. (2007a), Gaganis et al. (2007a), and Gaganis et al. (2007b) focused on using data mining for identifying qualified auditors' opinions, and more specifically, Anandarajan and Anandarajan (1999) compared the predictive ability of neural networks, expert systems, and multiple discriminant analysis in determining what type (modified or disclaimer) of going concern report should be issued. The findings of these studies indicate that the inclusion of credit rating in the models results in a considerable increase both in terms of goodness of fit and classification accuracy, and that the results are mixed concerning the accuracy of industry-specific models, as opposed to general models. Furthermore, financial distress and profitability are reported strongly related to qualified opinions, yet liquidity and auditor's characteristics seem to be irrelevant to identifying qualified auditors' opinions.

5.3.4.3. Forensic accounting. AICPA explicitly acknowledges the responsibility of auditors in *fraud detection* (Cullinan and Sutton, 2002). Detection of manipulated financial statements by using normal audit procedures becomes an incredibly difficult task (Dikmen and Küçükkocaoğlu, 2010). “Fraud risk assessment is a highly complex process that is a part of every audit engagement. Over time, regulatory requirements have steadily increased the amount of time and effort required of the auditor to assess fraud. It follows, therefore, that fraud risk assessment presents an ideal opportunity for technological assistance” (Comunale et al., 2010). The review of literature showed prevalent usage of data mining by researchers and practitioners to detect fraud. Researchers tackled different levels and areas of fraud. Some focused on detecting fraud risk at the more macro level of audit engagement level (Comunale et al., 2010), and others focused on detecting fraud at the more micro level of business transactions (Debreceny and Gray, 2010; Bella et al., 2009; Tackett, 2013). Whereas Debreceny and Gray (2010) researched the journal entry fraud using digit analysis and found that the distribution of first digits of journal dollar amounts differed from that expected by Benford's Law, Bella et al. (2009) developed a four-stage self-organizing map fraud detection architecture of electronic billing records, and Tackett (2013) suggested the use of association rules in detecting fraud through finding patterns and relationships when examining a company's digital records. On the other hand, Bay et al. (2006) focused on identifying the irregularities at the general ledger level, and (Jans et al., 2010; Jans et al., 2011; Owusu-Ansah et al., 2002) focused on detecting fraud at the business cycle or process level. While Jans et al. (2010) used descriptive data mining techniques for detecting and reducing risk of internal fraud at the procurement cycle level, Jans et al. (2011) examined the effectiveness of fraud detection audit procedures at the stock and warehousing cycle level, and Owusu-Ansah et al. (2002) employed business process mining for mitigating internal transactions fraud in the procurement processes. These authors found that size of the audit firm, auditor's position tenure, and auditor's years of experience are statistically significant predictors of fraud. Using a combination of Benford's Law and neural networks, Busta and Weinberg (1998) focused on detecting manipulated financial data in analytical review procedures, and Kim and Vasarhelyi (2012) used data mining for detecting company level internal fraud.

Management fraud is a type of fraud that adversely affects stakeholders through misleading or fraudulent financial statements (FFS) (Elliott and Willingham, 1980), thus many researchers focused on detecting FFS with the help of data mining at different levels: detection of top management fraud (Fanning and Cogger, 1998; Pai et al., 2011), detection of fraud based on prediction of company future performance (Virdhagriswaran and Dakin, 2006), and detection of fraud in financial reports (Kirkos et al., 2007b; Ata and Seyrek, 2009; Deng, 2009; Zhou and Kapoor, 2011). Other researchers (Cerullo and Cerullo, 2006; Feroz et al., 2000; Green and Choi, 1997; Hoogs et al., 2007; Huang et al., 2012; Jie and Wei, 2009; Krambia-Kapardis et al., 2010; Ogut et al., 2009; Ravisankar et al., 2011; Tsaih et al., 2009; Yue et al., 2009; Zouboulidis and Kotsiantis, 2012; Kotsiantis et al., 2006; and Perols, 2011) used data mining for predicting FFS. Important findings of these authors include: the ability of neural networks models to classify membership in SEC investigated versus non-investigated firms with high accuracy. One explanation for such relative success of neural networks is their ability to use adaptive learning processes to determine what is important to distinguish true signal from noise. The researches also explored the effectiveness of combining financial and governance indicators, exogenous and endogenous factors, and feature selection to detect fraudulent financial statements. Along the same lines, the work of Gaganis (2009) involved the use of data mining classification techniques combining both financial and nonfinancial data for the identification of FFS and concluded that classification accuracy depends on the way the data is pre-processed, the objective function, and the search strategy of the model. Alden et al. (2012) used genetic algorithms in detecting patterns of FFS and concluded that GAs and estimation of distribution algorithm demonstrate a better ability to classify patterns of financial statement fraud than those the traditional logistic regression model. More specifically, Lin et al. (2003) developed an integrated fuzzy neural networks model to assess the risk of FFS. The fuzzy neural network model of Lin et al. (2003) outperformed most statistical models and artificial neural networks reported in prior studies, and its performance compared favorably with a baseline Logit model. Liou (2008) explored the differences and similarities between falsified financial reporting detection and business failure prediction models using logistic regression, neural networks, and decision tree and found that the financial factors used to detect fraudulent reporting are helpful for predicting business failure. Welch et al. (1998) developed a data mining-based classifier system for modeling auditor decision when estimating the likelihood of fraud by contractors developing bids for government contracts, and reported that, in classification decision models involving simultaneous processing, genetic algorithms represent an innovative heuristic approach that may produce improved models when compared to traditional mathematical approaches. Kochetova-kozloski et al. (2011) used data mining for improving auditors judgments on fraudulent management events.

The search for indicators of financial reporting fraud is not limited to using the numeric part of financial statements, which was the focus of many previous researchers, but it extends to the evaluation of the qualitative part. For example, Humpherys et al. (2011) and Purda and Skillicorn (2014) evaluated and classified the management's discussion and analysis section of Form 10-K using linguistic credibility analysis and data mining techniques. Their findings indicate that writers of fraudulent disclosures may write more to appear credible while communicating less in actual content, and support the usefulness of linguistic analyses by auditors to flag questionable financial disclosures and to assess fraud risk. Similarly, Yu et al. (2013) constructed models based on data mining techniques in detecting and classifying the violations to the accounting information disclosure by listed companies. Goel et al. (2010) and Goel and Gangolly (2012) examined qualitative textual content in annual reports to predict fraud, and found that textual information provides valuable clues pertaining to fraud prediction. Similarly, Gupta and Gill (2012a) used support vector machines in detecting fraud in the qualitative part of financial statements, and Gupta and Gill (2012b) examined the efficacy of decision trees, Naïve Bayes and genetic algorithms for preventing and detecting FFS. Larcker and Zakolyukina (2012) focused on detecting financial statement manipulations in CEO and CFO narratives during quarterly earnings conference calls. These researchers also reached similar conclusions as those of Humpherys et al. (2011) and Purda and Skillicorn (2014) in that employment of linguistic features is an effective means for detecting fraud, and result in significant improvement in the accuracy of detecting fraud in financial reports.

In the area of predicting *earnings management*, Tsai and Chiou (2009) developed neural networks and decision tree models to be used by investors in predicting the level of earnings management in advance, and whether earnings are managed upward or downward. The results of Tsai and Chiou (2009) indicate that using data mining techniques significantly improved prediction of earnings management and generated decision rules that help in detecting earnings management. On the other hand, Ezazi et al. (2013) examined the usefulness of various data mining techniques in predicting earnings management, clearly questioning the assumption of linearity for modeling the accrual process, and concluded that a non-linear approach to predicting earnings management is more effective than a linear approach. Focusing on detecting earnings management, Hoglund (2012) assessed the performance of different data mining techniques, more specifically Hoglund (2013a) assessed the performance of the cross-sectional Jones (1991) accrual model using a genetic algorithm. The results indicated the superiority of genetic algorithms compared to other grouping methods. To overcome the problem of data availability in estimating time series, Hoglund (2013b) found that fuzzy linear regression-based Jones model outperforms the regression-based Jones model in detecting simulated earnings management when the estimation time series is short. Song et al. (2013) examined the association between earnings management and assets misappropriation and found that misappropriation of assets has a significant positive association with discretionary accruals.

Data mining applications in assurance and compliance focused primarily on three main topics: auditing (including engagement, planning, conducting, and post-auditing phases), business health (including financial viability, bankruptcy, and going concern), and forensic accounting (including fraud detection and earnings management). Specific applications include: auditor selection and switching, auditor fee determination, supporting auditor in the level of substantive tests, identifying patterns in accounting data and analytical procedures, bankruptcy prediction, detection of fraudulent financial statements and reports, detection of earnings management, etc. The main goal of applications in this domain is prediction, and the primary task is classification. The predominant techniques are neural networks and regression. Future research opportunities include: enriching input with variables related to managerial characteristics, testing different approaches to classifiers aggregation, testing of different learning algorithms and model architectures, exploring different time granularities and data preprocessing approaches, expand the scope of model development to multiple firms and business types, more careful selection of input variables, elongating the prediction horizon, including non-financial variables and more visual analysis, pay more attention to model comparison, combing data and text mining in predicting financial fraud, etc.

5.3.4.4. Tax compliance. Data mining has also been used in taxation, such as the work of Denton et al. (1995), who used neural networks in classifying employees for tax purposes. For tax compliance at the company level, Wu (1994) used neural networks and Kallio and Back (2011) used a self-organizing map, for successfully identifying companies that require further tax audit investigation. Wu et al. (2012) applied a data mining technique in enhancing tax evasion detection performance by using data mining to develop a screening framework to filter possible non-compliant tax reports that may be subject to further auditing. The results of Wu et al. (2012) show that their proposed data mining technique enhances the detection of tax evasion, and therefore can be employed to effectively reduce or minimize losses from tax evasion.

5.3.5. Overall summary of findings of data mining application areas

In summary, it is clear that the overwhelming focus of using data mining in accounting, regardless of proposed framework category, is in the area of compliance and assurance. Thus, the "disconnect between the application domain of auditing and assurance and the technology domain of AI" that was reported by Baldwin et al. (2006) is not supported by these findings. One could argue that the well-publicized financial scandals over the last two decades, such as that of Enron and WorldCom, presented the greatest challenge to the compliance and assurance functions of the accounting profession. This may have consequently expanded the reach of the accounting researchers to incorporate modern technology paradigms such as data mining to improve

Table 2
Strengths, weaknesses, and recommendations for data mining applications in accounting.

CRISP-DM phase	Strengths	Weaknesses	Recommendations
Business understanding	Technical justifications are well specified – e.g. compare performance of a data mining technique or techniques in addressing a given business problem.	Consideration of business impact is generally lacking.	Business impact is as important as technical justification since the latter does not necessarily translate into the former. More attention needs to be given to questions such as: if some measure of performance (such as accuracy) is improved, how does this translate into business benefit? Such attention could lead to more acceptance to data mining applications in real business contexts, and thus reduce reliance on simulations and synthetic data.
Data understanding	Many studies used benchmarked variables that are available in publically published financial statements and reports.	Sometimes clearly important variables are missing, for example in the case of using only balance sheet variables for predicting firm performance and ignoring variables in other financial statements.	Domain knowledge of the problem being addressed and its business dimensions is important to understand the most pertinent data, appropriate unit of analysis (for example, firm vs. firm-year), and temporal dimension (e.g. point-in-time vs. cross-sectional). Research conducted from a purely technical perspective or not using domain expertise may miss important data understanding considerations
Data preparation	Use of variable normalization to remove scale impact and variable derivation to enrich the attribute set.	Data preparation is rarely discussed, and little attention is paid to variable selection techniques.	Including all relevant variables is key in developing properly-specified models. Missing such variables may result in an inadequate model that does not capture all important components of the problem being solved. As important is not including irrelevant and redundant variables. Strategies to include the proper set of attributes in the analysis include using feature selection techniques, variable clustering, and including in the analysis relationships between variables as well.
Model development	Building on models specified in previous research.	Rarely is the model specification complete.	Complete model specification is necessary for future replicability and benchmarking. Every data mining modeling technique provides multiple architectures and requires setting of many parameters. For example, a neural network model can be specified in terms of topology (Multi-layer perceptron, Radial Basis, etc.), information flow (feedforward vs. recurrent), learning function (most notably backpropagation), number of hidden layers, etc. Not providing all the specifics of a model makes it hard to replicate and benchmark the proposed model
Model assessment	Use of out-of-sample and k-fold cross validation techniques.	Scant attention is paid to considerations of classification error costs and hence rarely a mention of financial impact of classification decision outcomes. In addition, very few utilized a well-formed experimental design in which an experimental group is matched with a control group. Many researches used accuracy to assess model performance.	It is key to take into account the impact of the cost of a false positive and that of a false negative when assessing models – as these costs are in many business situations are far from equally weighted. Ignoring such considerations can lead to bad decisions. Although accuracy is one of the most popular model assessment measures, it is important to recognize that accuracy is not always an appropriate assessment metric – especially when samples are imbalanced. For example, if the cases of interest represent a small (<10%) proportion of the total sample, a model will report >90% accuracy without capturing any of the cases of interest – this is especially true in the case of fraud, auditing, and forensic accounting.
Model deployment	Modeling for real-world business applications.	Rarely a mention of model recalibration and variability of business environment.	Businesses operate in a continually changing environment and thus deployed models should be re-calibrated on regular basis to validate their relevance and performance under the new business realities.

capabilities of combating illegal reporting practices. Similar to what was reported in (Gupta and Gill, 2012b), the major focus of data mining in forensics appears to still be on detection; with less emphasis on other, equally important, dimensions of fraud such as prevention and mitigation. The managerial accounting also received a significant number of data mining applications, perhaps due to the increasing desire to improve overall enterprise efficiencies in light of the ramifications of the 2008 financial crisis and the subsequent business meltdown. Notable in the managerial accounting domain, is the near absence of the application of data mining to management control systems, a major tool to guard against risk and vulnerability. In the financial accounting domain, the most focus was on financial and performance analysis, but little has been reported on the use of data mining in the key functions of recording and validating transactions to improve the accuracy of both data and reporting. The application of data mining in accounting also seems to be thin in the areas of process and text mining. Accounting functions have many embedded processes, such as ordering, billing, paying, and reconciling that may provide rich input for process mining. However, lack of easy access to business process information might be one reason that process mining applications are limited. Similarly, although accounting is numeric in the majority, it has an abundance of textual content that can benefit from text mining. Yet, text mining applications of accounting are relatively few, possibly due to the fact that text mining is itself a relatively newer, less familiar, branch of data mining. These limitations may be aggravated by limited coverage, if at all, of accounting curriculum of such advanced techniques.

We use the well-known Cross Industry Standard Protocol – Data Mining (CRISP-DM) framework as a logical presentation structure to present the strengths, weaknesses, and recommendations for data mining applications in accounting (Table 2). CRISP-DM is a widely accepted industry-, discipline-, technology-, and technique-agnostic guiding methodological standard for implementing data mining projects. CRISP-DM is based on six phases: business understanding, data understanding, data preparation, modeling, testing, and deployment. The reported data mining applications in accounting manifest many strengths as well as some weaknesses. Strengths include: well- reasoned technical justifications, use of publically available variables, use of variable normalization and derivation, incremental modeling approaches, unbiased model testing, and modeling of real-world problems. Some of the weaknesses manifested throughout the previous research include: lacking consideration of business impact, missing inclusion of important relevant variables, thin consideration of data preprocessing and preparation, limited information on model specification, rare consideration of cost of classification errors, and limited dialogue on model calibration to account for the changing business environment. Recommendations for more effective applications of data mining in accounting include: highlighting business impact of the application, inclusion of domain experts in the model development which helps mitigate missing important variables relevant to the modeling situation, complete data and model specification to facilitate ease of replicability, consideration of classification error cost in assessing and comparing models, and regular model recalibration to ensure validity of the model over time. Table 2 provides more details of these strengths, weaknesses, and recommendations within the CRISP-DM framework.

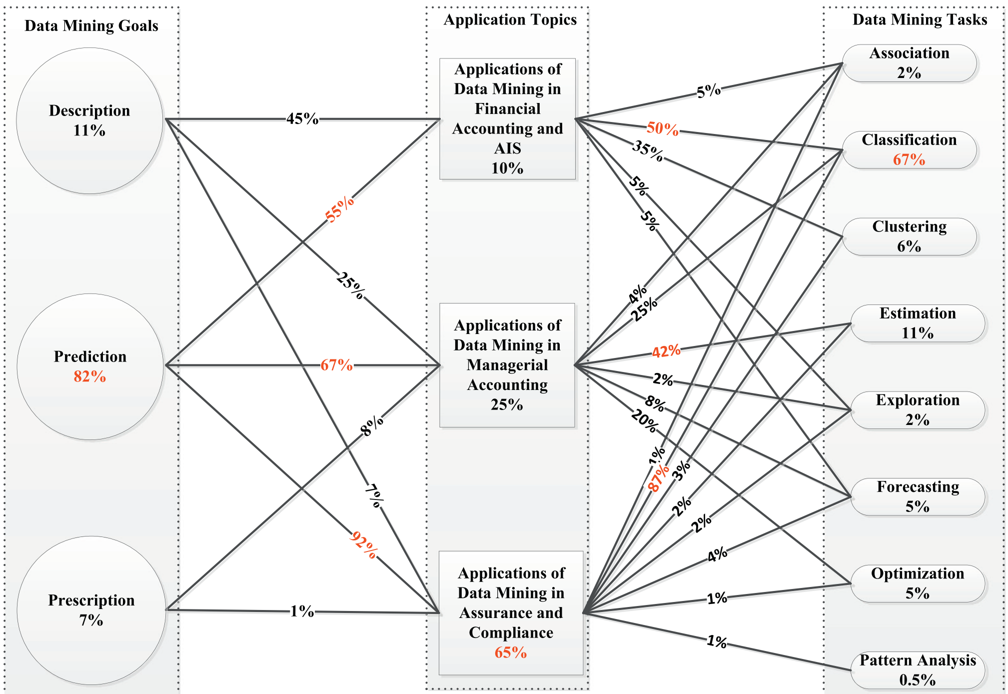

Fig. 8. Data mining applications in accounting by accounting topic, data mining goal, and data mining task. Numbers on the lines indicate the percent of applications for a given goal or a given task. Numbers inside a shape indicate the overall percent of applications for the specific goal, topic, or task. Red numbers indicate the highest percent; e.g., applications of data mining in financial accounting and AIS are 55% predictive, and 50% of them use classification as a task). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.4. Macro themes and patterns (Fig. 8)

5.4.1. Overall application topic, goal, and task patterns

A number of macro patterns of data mining applications in accounting emerged from the literature review. Fig. 8 provides a summary of applications by accounting topic, data mining goal, and data mining task. The analysis reveals that the intensity of data mining usage, goals, and tasks varies across the various branches of accounting. In terms of *usage intensity*, the overwhelming majority of data mining applications in accounting are in the domain of assurance and compliance, representing 65% of the total applications, followed by applications in managerial accounting (25%), and then applications in financial accounting (10%). In terms of overall *goal intensity*, prediction is the undisputable overall data mining goal as is indicated by 82% of the applications, followed by description (11%), and then prescription (7%). At the application topic level, description goal intensity is most concentrated in financial accounting, prediction goal intensity is most concentrated in assurance and compliance, and prescription goal intensity is most concentrated in managerial accounting. In terms of overall *task intensity*, classification is by far the most common task undertaken by 67% of the data mining applications in accounting, distantly followed by estimation in 11% of the applications, and clustering in 6% of the applications.

5.4.2. Financial accounting and AIS applications patterns

It is worth noting that the majority (55%) of the data mining applications in financial accounting and AIS are predictive and the remaining are descriptive. The absence of prescriptive data mining in financial accounting and AIS is somewhat perplexing, and we think is worthy of future investigation. Furthermore, 85% of the data mining applications in financial accounting and AIS focused on only two data mining tasks, classification (50%) and clustering (35%), highlighting a potential opportunity for more utilization of the many other data mining tasks, such as visualization (a form of exploration), in the applications of data mining in financial accounting and AIS.

5.4.3. Managerial accounting applications patterns

Two-thirds (67%) of the applications of data mining in managerial accounting are predictive, 25% are prescriptive, and the remaining are descriptive. Estimation is the most common task used in (42%) of managerial accounting applications of data mining, followed by classification (25%), optimization (20%), forecasting (8%), association (4%), and exploration (2%). Clearly applications of data mining in managerial accounting assumed more variety of data mining tasks compared to financial accounting. This may be due to the varied analytics needs in managerial accounting compared to those in financial accounting.

5.4.4. Assurance and compliance applications patterns

The foremost data mining goal in the application of data mining in assurance and compliance is prediction (92%), followed by description (7%), and prescription (1%). This seems to reflect the vast desire for discrimination among classes in this accounting domain. For example, discrimination between companies that pass the audit from those that do not, companies that are healthy from those that are not, companies that are committing fraud versus those that are not, …, etc. The most common task in data mining applications in assurance and compliance is classification, used in 87% of the applications. Assurance and compliance data mining applications manifested the most task variety, followed, in this respect, by managerial accounting applications, and then financial accounting applications.

Although different branches of accounting assumed similar data mining tasks, the intensity of the tasks is not uniform and varies widely across application types. For example, assurance and compliance, managerial accounting, and financial accounting all performed the classification task; however, assurance and compliance performed it with the highest intensity (87%), followed by financial accounting (50%) intensity, and managerial accounting (25%). In a nutshell, data mining applications in accounting can be characterized as predominantly having an assurance and compliance profile, with prediction as their major goal, classification as their foremost task, and neural networks as their primary technique.

5.5. Mapping data mining applications in accounting to the proposed framework

Published research on data mining applications in accounting can be mapped to the proposed framework as in Fig. 9 and Tables 3 and 4. The prospective-predictive category focuses on the application of predictive data mining in prospective reporting, and has the highest number of applications (82%). This seems logical, since prediction leads to informing future decisions, and if such predictions are accurate enough, the business will reap significant benefits. This is also consistent with the goal profile of data mining applications in other business disciplines (for example, Köksal et al., 2011). This prospective-predictive category spans different branches of accounting. For example, financial accounting prospective topics such as forecasting earnings, and predicting choice of accounting method; managerial accounting prospective topics such as improving cost variance analysis of an activity-based standard costing management system, developing cost estimation models; and compliance and assurance prospective topics such as predicting firm going concern status, predicting auditor selection, switching, and opinion, predicting business health, bankruptcy, and failure, and identifying fraudulent disclosure and irregularities in the general ledger. The topics in this category mainly focus on future financial performance, efficiency considerations, and business health.

The retrospective-descriptive category was the second highest (11%) domain of data mining applications in accounting. It covers topics that span AIS, financial accounting, managerial accounting, and compliance and assurance. For example, in AIS, designing a chart of accounts structure for an ERP accounting information system is very time consuming, laborious, and costly. Data

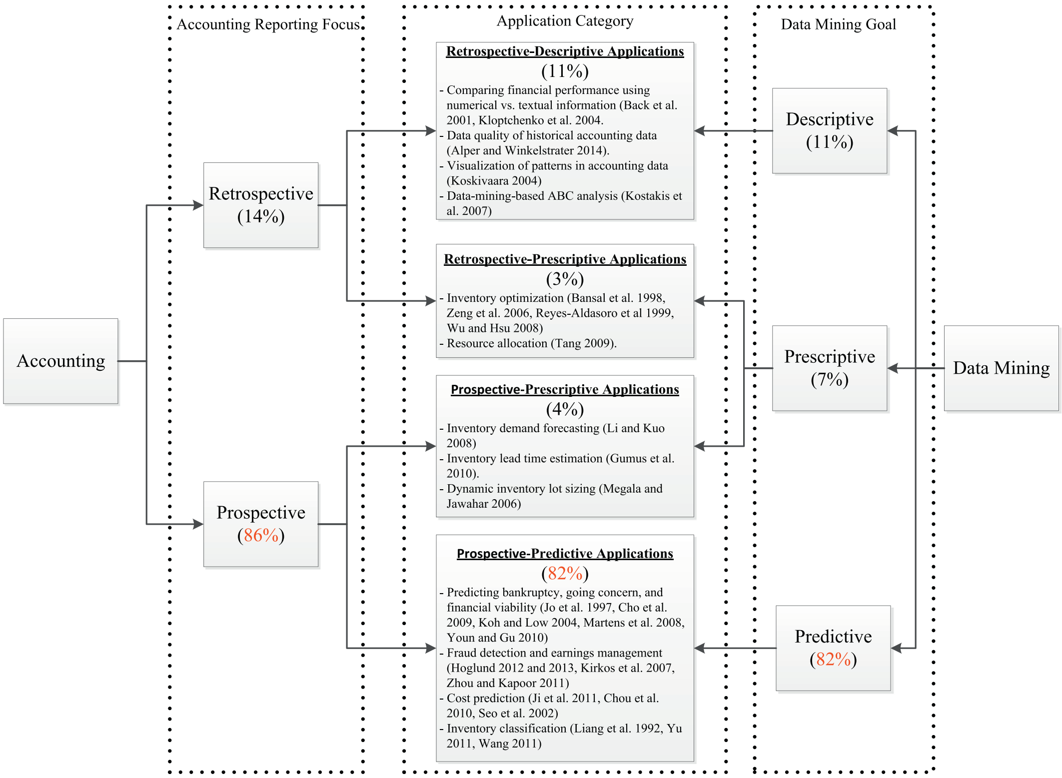

Fig. 9. Mapping accounting data mining applications to the proposed framework.

Table 3
Descriptive and prescriptive data mining applications in retrospective reporting.

Retrospective-descriptive applications	Retrospective-prescriptive applications
AIS applications	
Developing chart of accounts structure for enterprise resource planning (ERP) accounting system	
Developing AIS adaptable to the e-commerce environment	
Financial accounting applications	
Comparing performance using numerical data and text information from annual reports	
Determining firm financial status from language of quarterly reports	
Discovering patterns in accounting data for signaling unexpected fluctuations	
Examining accounting data quality and handling issues	
Extracting meaning from textual part of financial reports	
Financial analysis of logistics firms	
Managerial accounting applications	
Defining the cost derivers	Developing an inventory optimization system
Improving inventory management	Minimizing level and cost of inventory
Managing project costs	Inventory optimization
Assurance and compliance applications	
Clustering accounting database as an internal control procedure	Optimally assigning auditors to projects
Developing a framework for exploring and reducing internal fraud	
Evaluating firm financial statements for creditors	
Explaining qualifications in audit reports	
Exploring fraud based on pattern and relationships	
Exploring fraud in journal entries	
Exploring fraud using fraud management system architecture in next generation network	
Exploring tax evasion	
Grouping employees for tax purposes	
Identifying firms that need to be further investigated and tax audited	
Identifying pattern signals in the analytical procedures	
Using process mining to improve audit practice	

Table 4
Prescriptive and predictive data mining applications in prospective reporting.

Prospective-prescriptive applications	Prospective-predictive applications
Financial accounting	
	Classifying risk factors in 10-k forms
	Forecasting accounting earnings
	Modeling and predicting general relationships between accounting ratios
	Predicting choice of accounting method
	Predicting the effects of accounting policy on management actions
Managerial accounting applications	
Developing a decision support tool for revising and validating the inventory policy for spare parts	Automating account reconciliation using data mining techniques
	Classifying inventory types
Developing an inventory purchasing optimization system	Developing cost estimation model
	Estimating project design, development, and manufacturing cost
Estimating inventory lead time	Forecasting equipment and product unit cost
Forecasting inventory demand	Forecasting inventory risks
Optimally solving lot-sizing inventory problem	Identifying important and representative factors affecting intangible assets
Resource/budget allocation	Improving accuracy of cost predictions for inspection and repair of equipment
Selecting optimal bill of material	Improving cost variance analysis of an activity-based standard costing management systems
	Improving estimation of total or unit cost of construction, operation, and maintenance of projects
	Predicting cash flow
	Predicting choice of inventory method
	Predicting product life cycle cost
	Predicting restatement of financial statement due to revenue recognition
	Predicting transfer pricing decisions
	Solving lot-sizing inventory problem
Assurance and compliance applications	
	Assessing fraud risks; internal controls risks; audit environment risks
	Classifying auditor's going concern opinion
	Classifying qualifications in audit reports
	Detecting fraud based on the prediction of the firm's future performance
	Detecting fraudulent bids; fraudulent transactions; earning management; disclosure violations by listed companies
	Detecting fraudulent financial statements
	Distinguishing the investigated from non-investigated firms by Security Exchange Commission
	Effect of firm financial (distress) and non-financial (litigation) information on the auditor opinion
	Identifying fraudulent disclosure
	Identifying irregularities in the general ledger
	Investigating audit going concern opinion and subsequent litigation
	Modeling the dynamics and the relationships between account values for finding the unexpected fluctuations
	Predicting firm going concern status
	Predicting auditor selection, switching, and opinion
	Predicting business health; business bankruptcy; business failure; bank solvency
	Predicting cross sectional variation of audit fees
	Predicting fraud-lawsuit
	Predicting going concern opinion from qualitative variables
	Predicting material misstatements during the analytical review process using ratio analysis
	Predicting patterns in auditing monthly balances
	Predicting relationship between earning management and asset misappropriation
	Predicting the behavior of stock return surrounding the going concern opinion
	Predicting the effect of auditor going concern opinion on the firm default
	Predicting the type of going concern audit report

mining has been used to address this problem and represent an accounting system in the least possible number of account segments. In financial accounting, data mining has been used to bring to light the value of textual, not just numerical, information in assessing financial performance, and to examine and handle accounting data quality issues, hence enabling more accurate reporting. In the managerial accounting domain, data mining has been used to identify the important cost drivers to improve the overall costing process. In the realm of compliance and assurance, data mining applications have contributed more internal control procedures through clustering of accounting database, more effective fraud detection through exploration of transactional patterns and relationships, more understanding of qualifications in audit reports, and more tax compliance through clustering of employees for such purposes. The main thrust of this category is to achieve better business understanding, better data quality, and better overall accounting system by taking a rear-view perspective.

The retrospective-prescriptive and the prospective-prescriptive categories seem to be the least prevalent types of applications of data mining in accounting, 3% and 4%, respectively. The retrospective-prescriptive category includes applications such as minimizing the level and cost of inventory, optimizing resource allocation, and optimally assigning auditors to projects. Examples of applications covered in the prospective-prescriptive category include developing a decision support tool for revising and validating inventory management policy, developing a purchasing optimization system, estimating optimal inventory lead time, forecasting demand, and optimizing future budget allocation. These two application categories seem to be relatively unexplored by accounting researchers, and might represent a promising potential for accounting to benefit from data mining. The low coverage of these categories might be due to the fact that prescriptive analysis requires skill sets different from those that are traditionally covered in the accounting curriculum, and thus might require expertise and/or knowledge that are not readily available to a typical accounting professional.

6. Conclusions, future directions, and limitations

The paper reviewed the literature on data mining applications in accounting, and proposed an organizing framework. The review reveals that the current status of the amalgamation of accounting, a fundamental business discipline, and data mining, a top ten future information systems technology is not at an all-encompassing stage. Data mining applications in accounting have primarily focused, with varying intensities, on three branches of accounting: assurance and compliance, managerial accounting, and financial accounting. Data mining application is most intense in assurance and compliance, followed by managerial accounting, and then financial accounting. Most of these applications focused overwhelmingly on prediction as a goal, classification as a task, and neural networks as a technique. The other two goals of data mining, description and prescription, have not received similar attention in the application of data mining in accounting. Although prediction fits well with many accounting problems, embodies future outlook, strategic orientation, guidance, and position, one cannot argue that other data mining goals that received less emphasis do not provide strategic value to accounting decision making. The low usage of prescriptive applications is a clear gap that needs to be more actively addressed to realize the benefits of optimization in decision making. The overwhelming focus on classification as a data mining task highlights the potential for using other data mining tasks such as pattern and association analysis that have shown great benefits in other business disciplines. Although the predominance of neural networks as the technique of choice reflects the power of neural networks as a general modeling tool, yet, this reliance on a technique that is known for its black-box depiction leads to many models that are lacking in explainability. In short, current data mining applications in accounting can be characterized as predominantly having an assurance and compliance focus, with prediction as their major goal, classification as their foremost task, and neural networks as their primary technique. Under such characterization, it is hard to say that accounting has leveraged enough data mining power and capabilities.

The proposed framework combined the two major accounting reporting foci (retrospection and prospection) and the three data mining goals (description, prediction, and prescription), resulting in four main feasible categories of accounting data mining applications: retrospective description, retrospective prescription, prospective prescription, and prospective prediction. The concentration of applications varies markedly among these categories, with the highest concentration being in the prospective prediction category, followed by the retrospective description category, the prospective prescription category, and lastly the retrospective prescription category. The analysis of the literature mapping to the framework revealed the variation of intensity of usage, goal, task, and technique in these applications, and enabled us to identify not only the current patterns of such applications, but also the potential areas where data mining can be of potential benefit to accounting.

Researchers have not yet systematically assessed the value of, and the benefits from, assimilating data mining techniques into accounting. Notably missing is the application of data mining to major topics in accounting such as corporate governance, integrated reporting, accounting education, enterprise accounting systems, XBRL, and adoption of accounting standards. It seems AIS might benefit greatly from incorporating data mining capabilities, and improve the accuracy of recording and/or accuracy of reporting. Furthermore, the majority of data mining applications in auditing focused primarily on prediction; however, many aspects of risk management such as, risk prevention, incident detection, and incident mitigation, may benefit as well from such technology. Of particular promise might be research into developing models that distinguish between intentional misstatements in financial reporting and unintentional mistakes. For example, detecting whether the true cause of a reporting error is incorrect judgment, improper interpretation of accounting standards, or intentional lack of full disclosure; and whether such errors are due to managerial or procedural reasons are all areas that can potentially benefit from data mining. Further research opportunities exist into the application of data mining to improve enterprise risk management, real time reporting and auditing, transactional accuracy, transactional security, intrusion detection, and lean accounting. Increasing use of data mining visualization capabilities may also accrue great benefits to decision making in accounting, due to the old adage of “a picture is worth a thousand words”. A future research direction might also be to zoom into a specific accounting sub-domain, and apply this organizing framework to derive insights at a more micro level.

Examining applications of data mining in accounting through lens of well-known Cross Industry Standard Protocol – Data Mining (CRISP-DM) framework revealed many practical recommendations may be pertinent:

1. Considering the business impact as technical justification and paying more attention to questions such as: if some measure of performance (such as accuracy) is improved, how does this translate into business benefit?
2. Considering the domain knowledge of the problem being addressed and its business dimensions to understand the most pertinent data, appropriate unit of analysis (for example, firm vs. firm-year) and temporal dimension (e.g. point-in-time vs. cross-sectional).

3. Including all relevant variables in developing properly-specified models through use of feature selection techniques, variable clustering, and including in the analysis relationships between variables as well.
4. Completing model specification for future replicability and benchmarking.
5. Taking into account the impact of the cost of a false positive and that of a false negative when assessing models.
6. Using accuracy as a model assessment and comparison metric judiciously. Although accuracy is one of the most popular model assessment measures, it is important to recognize that accuracy is not always an appropriate assessment metric – especially when samples are imbalanced. For example, if the cases of interest represent a small (< 10%) proportion of the total sample, a model will report >90% accuracy without capturing any of the cases of interest – this is especially true in the case of fraud, auditing, and forensic accounting.
7. Re-calibrating of models deployed on regular basis to validate their relevance and performance under the new business realities.

The research is not without limitations. The methodology is by no means claimed to be perfect. For example, many other search terms could have been used to identify the relevant literature. However, no search strategy could exhaust all possible relevant terms in either accounting or data mining. We believe we have included the main terms and outlets to enable us to capture the most important literature on the applications of data mining in accounting.

References

Ahn, B.S., Cho, S.S., Kim, C.Y., 2000. The integrated methodology of rough set theory and artificial neural network for business failure prediction. Expert Syst. Appl. 18 (2), 65–74.

Alam, P., Booth, D., Lee, K., Thordarson, T., 2000. The use of fuzzy clustering algorithm and self-organizing neural networks for identifying potentially failing banks: an experimental study. Expert Syst. Appl. 18 (3), 185–199.

Alden, M.E., Bryan, D.M., Lessley, B.J., Tripathy, A., 2012. Detection of financial statement fraud using evolutionary algorithms. Journal of Emerging Technologies in Accounting 9 (1), 71–94.

Alpar, P., Winkelsträter, S., 2014. Assessment of data quality in accounting data with association rules. Expert Syst. Appl. 41 (5), 2259–2268.

Altay Guvenir, H., Erel, E., 1998. Multicriteria inventory classification using a genetic algorithm. Eur. J. Oper. Res. 105 (1), 29–37.

Anandarajan, M., Anandarajan, A., 1999. A comparison of machine learning techniques with a qualitative response model for auditor's going concern reporting. Expert Syst. Appl. 16 (4), 385–392.

Anandarajan, M., Lee, P., Anandarajan, A., 2001. Bankruptcy prediction of financially stressed firms: an examination of the predictive accuracy of artificial neural networks. Intelligent Systems in Accounting, Finance and Management 10 (2), 69–81.

Argyrou, A., Andreev, A., 2011. A semi-supervised tool for clustering accounting databases with applications to internal controls. Expert Syst. Appl. 38 (9), 11176–11181.

Ata, H.A., Seyrek, I.H., 2009. The use of data mining techniques in detecting fraudulent financial statements: an application on manufacturing firms. Suleyman Demirel University Journal of Faculty of Economics and Administrative Sciences 14 (2).

Back, B., Toivonen, J., Vanharanta, H., Visa, A., 2001. Comparing numerical data and text information from annual reports using self-organizing maps. Int. J. Account. Inf. Syst. 2 (4), 249–269.

Baldwin, A.A., Brown, C.E., Trinkle, B.S., 2006. Opportunities for artificial intelligence development in the accounting domain: the case for auditing. Intelligent Systems in Accounting, Finance and Management 14 (3), 77–86.

Bansal, K., Vadhavkar, S., Gupta, A., 1998a. Brief application description: neural networks based forecasting techniques for inventory control applications. Data Min. Knowl. Disc. 2 (1), 97–102.

Bansal, K., Vadhavkar, S., Gupta, A., 1998b. Neural networks based data mining applications for medical inventory problems. Int. J. Agil. Manuf. 1 (2), 187–200.

Bay, S., Kumaraswamy, K., Anderle, M.G., Kumar, R., Steier, D.M., 2006. Large scale detection of irregularities in accounting data. Proceedings of the 6th IEEE International Conference on Data Mining (ICDM'06), Hong Kong, HK (December 75–86).

Beaudoin, C., Chandar, N., Werner, E.M., 2010. Are potential effects of SFAS 158 associated with firms' decisions to freeze their defined benefit pension plans? Review of Accounting and Finance 9 (4), 424–451.

Bella, B., Eloff, J.H., Olivier, M.S., 2009. A fraud management system architecture for next-generation networks. Forensic Sci. Int. 185 (1), 51–58.

Benhayoun, N., Chairi, I., El Gonnouni, A., Lyhyaoui, A., 2013. Financial intelligence in prediction of firm's creditworthiness risk: evidence from support vector machine approach. Procedia Economics and Finance 5, 103–112.

Beynon, M.J., Peel, M.J., Tang, Y.C., 2004. The application of fuzzy decision tree analysis in an exposition of the antecedents of audit fees. Omega 32 (3), 231–244.

Bhimani, A., Gulamhussen, M.A., Lopes, S., 2009. The effectiveness of the auditor's going-concern evaluation as an external governance mechanism: evidence from loan defaults. Int. J. Account. 44 (3), 239–255.

Birnberg, J.G., 1980. The role of accounting in financial disclosure. Acc. Organ. Soc. 5 (1), 71–80.

Blacconiere, W.G., DeFond, M.L., 1997. An investigation of independent audit opinions and subsequent independent auditor litigation of publicly-traded failed savings and loans. J. Account. Public Policy 16 (4), 415–454.

Boyacioglu, M.A., Kara, Y., Baykan, O.K., 2009. Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: a comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey. Expert Syst. Appl. 36 (2), 3355–3366.

Braglia, M., Grassi, A., Montanari, R., 2004. Multi-attribute classification method for spare parts inventory management. J. Qual. Maint. Eng. 10 (1), 55–65.

Brown, C.E., Wong, J.A., Baldwin, A.A., 2007. A review and analysis of the existing research streams in continuous auditing. Journal of Emerging Technologies in Accounting 4 (1), 1–28.

Busta, B., Weinberg, R., 1998. Using Benford's law and neural networks as a review procedure. Manag. Audit. J. 13 (6), 356–366.

Calderon, T.G., 1999. Neural networks and preliminary information risk assessment in an auditing environment. Account. Enq. 8, 245–290.

Calderon, T.G., Cheh, J.J., 2002. A roadmap for future neural networks research in auditing and risk assessment. Int. J. Account. Inf. Syst. 3 (4), 203–236.

Callen, J.L., Kwan, C.C., Yip, P.C., Yuan, Y., 1996. Neural network forecasting of quarterly accounting earnings. Int. J. Forecast. 12 (4), 475–482.

Carnegie, G.D., 2012. The special issue: AAAJ and research innovation. Account. Audit. Account. J. 25 (2), 216–227.

Cerullo, M.J., Cerullo, M.V., 2006. Using neural network software as a forensic accounting tool. Information Systems Control Journal 2, 33.

Chakraborty, S., Sharma, S.K., 2007. Prediction of corporate financial health by artificial neural network. International Journal of Electronic Finance 1 (4), 442–459.

Chakraborty, V., Chiu, V., Vasarhelyi, M., 2014. Automatic classification of accounting literature. Int. J. Account. Inf. Syst. 15 (2), 122–148.

Chan, D.Y., Vasarhelyi, M.A., 2011. Innovation and practice of continuous auditing. Int. J. Account. Inf. Syst. 12 (2), 152–160.

Chang, P.C., Lin, J.J., Dzan, W.Y., 2012. Forecasting of manufacturing cost in mobile phone products by case-based reasoning and artificial neural network models. J. Intell. Manuf. 23 (3), 517–531.

Charalambous, C., Charitou, A., Kaourou, F., 2000. Comparative analysis of artificial neural network models: application in bankruptcy prediction. Ann. Oper. Res. 99 (1/4), 403–425.

Chartered Global Management Accountant (CGMA), 2013. Report From Insight to Impact: Unlocking Opportunities in Big Data. Available at:. http://www.cgma.org/Resources/Reports/DownloadableDocuments/From_insight_to_impact-unlocking_the_opportunities_in_big_data.pdf.
Chen, M.Y., 2013. A hybrid ANFIS model for business failure prediction utilizing particle swarm optimization and subtractive clustering. Inf. Sci. 220, 180–195.
Chen, H.J., Huang, S.Y., Kuo, C.L., 2009a. Using the artificial neural network to predict fraud litigation: some empirical evidence from emerging markets. Expert Syst. Appl. 36 (2), 1478–1484.
Chen, H.J., Huang, S.Y., Lin, C.S., 2009b. Alternative diagnosis of corporate bankruptcy: a neuro fuzzy approach. Expert Syst. Appl. 36 (4), 7710–7720.
Chen, Y., Zhang, L., Zhang, L., 2013. Financial distress prediction for Chinese listed manufacturing companies. Procedia Computer Science 17, 678–686.
Cheng, M.Y., Roy, A.F., 2011. Evolutionary fuzzy decision model for cash flow prediction using time-dependent support vector machines. Int. J. Proj. Manag. 29 (1), 56–65.
Chew, P.A., Robinson, D.G., 2012. Automated account reconciliation using probabilistic and statistical techniques. Int. J. Account. Inf. Manag. 20 (4), 322–334.
Cho, S., Kim, J., Bae, J.K., 2009. An integrative model with subject weight based on neural network learning for bankruptcy prediction. Expert Syst. Appl. 36 (1), 403–410.
Cho, C.H., Roberts, R.W., Patten, D.M., 2010. The language of US corporate environmental disclosure. Acc. Organ. Soc. 35 (4), 431–443.
Chodak, G., Suchacka, G., 2012. Cost-oriented recommendation model for e-commerce. Proceeding of the 19th International Conference on Computer Networks in Communications in Computer and Information Science, Szczyrk, Poland :pp. 421–429 June. 10.1007/978-3-642-31217-5_44.
Chou, J.S., 2009. Web-based CBR system applied to early cost budgeting for pavement maintenance project. Expert Syst. Appl. 36 (2), 2947–2960.
Chou, J.S., Tsai, C.F., 2012. Preliminary cost estimates for thin-film transistor liquid–crystal display inspection and repair equipment: a hybrid hierarchical approach. Comput. Ind. Eng. 62 (2), 661–669.
Chou, J.S., Tai, Y., Chang, L.J., 2010. Predicting the development cost of TFT-LCD manufacturing equipment with artificial intelligence models. Int. J. Prod. Econ. 128 (1), 339–350.
Chou, J.S., Cheng, M.Y., Wu, Y.W., Tai, Y., 2011. Predicting high-tech equipment fabrication cost with a novel evolutionary SVM inference model. Expert Syst. Appl. 38 (7), 8571–8579.
Cil, I., 2012. Consumption universes based supermarket layout through association rule mining and multidimensional scaling. Expert Syst. Appl. 39 (10), 8611–8625.
Coakley, J.R., 1995. Using pattern analysis methods to supplement attention directing analytical procedures. Expert Syst. Appl. 9 (4), 513–528.
Coakley, J.R., Brown, C.E., 1993. Artificial neural networks applied to ratio analysis in the analytical review process. International Journal of Intelligent Systems in Accounting and Finance Management 2 (1), 19–39.
Coakley, J.R., Brown, C.E., 2000. Artificial neural networks in accounting and finance: modeling issues. International Journal of Intelligent Systems in Accounting, Finance & Management 9 (2), 119–144.
Comunale, C.L., Rosner, R.L., Sexton, T.R., 2010. The auditor's assessment of fraud risk: a fuzzy logic approach. Journal of Forensic and Investigative Accounting 3 (1), 149–194.
Cullinan, C.P., Sutton, S.G., 2002. Defrauding the public interest: a critical examination of reengineered audit processes and the likelihood of detecting fraud. Crit. Perspect. Account. 13 (3), 297–310.
Curry, B., Peel, M.J., 1998. Neural networks and business forecasting: an application to cross-sectional audit fee data. Int. J. Commer. Manag. 8 (2), 94–120.
Daigle, R.J., Lampe, J.C., 2005. The level of assurance precision and associated cost demanded when providing continuous online assurance in an environment open to assurance competition. Int. J. Account. Inf. Syst. 6 (2), 129–156.
Davis, A.K., Tama-Sweet, I.S.H.O., 2012. Managers' use of language across alternative disclosure outlets: earnings press releases versus MD&a. Contemp. Account. Res. 29 (3), 804–837.
Davis, J.T., Massey, A.P., Lovell II, R.E., 1997. Supporting a complex audit judgment task: an expert network approach. Eur. J. Oper. Res. 103 (2), 350–372.
Debreceny, R.S., 2011. Betwixt and between? Bringing information systems and accounting systems research together. J. Inf. Syst. 25 (2), 1–9.
Debreceny, R.S., Gray, G.L., 2010. Data mining journal entries for fraud detection: an exploratory study. Int. J. Account. Inf. Syst. 11 (3), 157–181.
Debreceny, R.S., Gray, G.L., 2011. Data mining of electronic mail and auditing: a research agenda. J. Inf. Syst. 25 (2), 195–226.
Deng, Q., 2009. Application of support vector machine in the detection of fraudulent financial statements. Proceeding of the 4th IEEE International Conference in Computer Science and Education (ICCSE'09), Nanchang, China, pp. 1056–1059 July.
Deng, S., Yeh, T.H., 2010. Applying least squares support vector machines to the airframe wing-box structural design cost estimation. Expert Syst. Appl. 37 (12), 8417–8423.
Deng, S., Yeh, T.H., 2011. Using least squares support vector machines for the airframe structures manufacturing cost estimation. Int. J. Prod. Econ. 131 (2), 701–708.
Denton, J.W., Sayeed, L., Perkins, N.D., Moorman, A.H., 1995. Neural networks to classify employees for tax purposes. Account. Manag. Inf. Technol. 5 (2), 123–138.
Dessureault, S., Benito, R.O., 2012. Data mining and activity based costing for equipment replacement decisions part 1—establishing the information infrastructure. Min. Technol. 121 (2), 73–82.
Dikmen, B., Küçükkocaoğlu, G., 2010. The detection of earnings manipulation: the three-phase cutting plane algorithm using mathematical programming. J. Forecast. 29 (5), 442–466.
Dimitras, A.I., Zanakis, S.H., Zopounidis, C., 1996. A survey of business failures with an emphasis on prediction methods and industrial applications. Eur. J. Oper. Res. 90 (3), 487–513.
Dorsch, J.J., Yasin, M.Y., 1998. A framework for benchmarking in the public sector: literature review and directions for future research. International Journal of Public Sector Management 11 (2/3), 91–115.
Doumpos, M., Gaganis, C., Pasiouras, F., 2005. Explaining qualifications in audit reports using a support vector machine methodology. Intelligent Systems in Accounting, Finance and Management 13 (4), 197–215.
du Jardin, P., 2010. Predicting bankruptcy using neural networks and other classification methods: the influence of variable selection techniques on model accuracy. Neurocomputing 73 (10), 2047–2060.
Eklund, T., Back, B., Vanharanta, H., Visa, A., 2008. A face validation of a SOM-based financial benchmarking model. Journal of Emerging Technologies in Accounting 5 (1), 109–127.
Elliott, R.K., 1992. The third wave breaks on the shores of accounting. Account. Horiz. 6 (2), 61–85.
Elliott, R.K., Willingham, J.J., 1980. Management Fraud: Detection and Deterrence. Petrocelli Books, New York, pp. 35–46.
Etheridge, H.L., Sriram, R.S., Hsu, H.Y., 2000. A comparison of selected artificial neural networks that help auditors evaluate client financial viability. Decis. Sci. 31 (2), 531–550.
Evans, J.R., 2013. Business Analytics: Methods, Models, and Decisions. Prentice-Hall, Boston, MA.
Ezazi, M.E., Ghotbi, F.S., Ghotbi, S.F., 2013. Predicting earning management using RBF, ICA, and SVM in firms listed in Tehran security exchange. Asian Journal of Management Research 4 (1), 208–220.
Fanning, K.M., Cogger, K.O., 1998. Neural network detection of management fraud using published financial data. International Journal of Intelligent Systems in Accounting, Finance and Management 7 (1), 21–41.
Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. AI Mag. 17 (3), 37.
Fedorova, E., Gilenko, E., Dovzhenko, S., 2013. Bankruptcy prediction for Russian companies: application of combined classifiers. Expert Syst. Appl. 40 (18), 7285–7293.
Feroz, E.H., Kwon, T.M., Pastena, V.S., Park, K., 2000. The efficacy of red flags in predicting the SECs targets: an artificial neural networks approach. International Journal of Intelligent Systems in Accounting, Finance and Management 9 (3), 145–157.
Fisher, I.E., Garnsey, M.R., Goel, S., Tam, K., 2010. The role of text analytics and information retrieval in the accounting domain. Journal of Emerging Technologies in Accounting 7 (1), 1–24.
Foltin, C., Garceau, L., 1996. Beyond expert systems: neural networks in accounting. National Public Accountant 41 (6), 26–32.
Gaafar, L.K., Choueiki, M.H., 2000. A neural network model for solving the lot-sizing problem. Omega 28 (2), 175–184.
Gaganis, C., 2009. Classification techniques for the identification of falsified financial statements: a comparative analysis. Intelligent Systems in Accounting, Finance and Management 16 (3), 207–229.

Gaganis, C., Pasiouras, F., Doumpos, M., 2007a. Probabilistic neural networks for the identification of qualified audit opinions. Expert Syst. Appl. 32 (1), 114–124.
Gaganis, C., Pasiouras, F., Spathis, C., Zopounidis, C., 2007b. A comparison of nearest neighbours, discriminant and logit models for auditing decisions. Intelligent Systems in Accounting, Finance and Management 15 (1/2), 23–40.
Garnsey, M.R., 2006. Automatic classification of financial accounting concepts. Journal of Emerging Technologies in Accounting 3 (1), 21–39.
Goel, S., Gangolly, J., 2012. Beyond the numbers: mining the annual reports for hidden cues indicative of financial statement fraud. Intelligent Systems in Accounting, Finance and Management 19 (2), 75–89.
Goel, S., Gangolly, J., Faerman, S.R., Uzuner, O., 2010. Can linguistic predictors detect fraudulent financial filings? Journal of Emerging Technologies in Accounting 7 (1), 25–46.
Grabski, S.V., Leech, S.A., Schmidt, P.J., 2011. A review of ERP research: a future agenda for accounting information systems. J. Inf. Syst. 25 (1), 37–78.
Granlund, M., 2011. Extending AIS research to management accounting and control issues: a research note. Int. J. Account. Inf. Syst. 12 (1), 3–19.
Granlund, M., Mouritsen, J., Vaassen, E., 2013. On the relations between modern information technology, decision making and management control. Int. J. Account. Inf. Syst. 4 (14), 275–277.
Gray, G.L., Debreceny, R.S., 2014. A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. Int. J. Account. Inf. Syst. 15 (4), 357–380.
Green, B.P., Choi, J.H., 1997. Assessing the risk of management fraud through neural network technology. Audit. J. Pract. Theory 16 (1), 14.
Gumus, A.T., Guneri, A.F., Ulengin, F., 2010. A new methodology for multi-echelon inventory management in stochastic and neuro-fuzzy environments. Int. J. Prod. Econ. 128 (1), 248–260.
Gupta, R., Gill, N.S., 2012a. Financial statement fraud detection using text mining. Int. J. Adv. Comput. Sci. Appl. 3 (12).
Gupta, R., Gill, N.S., 2012b. Prevention and detection of financial statement fraud—an implementation of data mining framework. Int. J. Adv. Comput. Sci. Appl. 3 (8).
Han, J., Kamber, M., Pei, J., 2006. Data Mining: Concepts and Techniques. Morgan Kaufmann.
Hennigan, D., Chowlera, S., 2011. Adopting a data-driven approach to underpayment recovery. Healthcare Financial Management: Journal of the Healthcare Financial Management Association 65 (2), 58.
Henry, E., 2006. Market reaction to verbal components of earnings press releases: event study using a predictive algorithm. Journal of Emerging Technologies in Accounting 3 (1), 1–19.
Hofmann, E., Lampe, K., 2013. Financial statement analysis of logistics service providers: ways of enhancing performance. Int. J. Phys. Distrib. Logist. Manag. 43 (4), 4.
Hoglund, H., 2012. Detecting earnings management with neural networks. Expert Syst. Appl. 39 (10), 9564–9570.
Hoglund, H., 2013a. Estimating discretionary accruals using a grouping genetic algorithm. Expert Syst. Appl. 40 (7), 2366–2372.
Hoglund, H., 2013b. Fuzzy linear regression-based detection of earnings management. Expert Syst. Appl. 40 (15), 6166–6172.
Hoogs, B., Kiehl, T., Lacomb, C., Senturk, D., 2007. A genetic algorithm approach to detecting temporal patterns indicative of financial statement fraud. Intelligent Systems in Accounting, Finance and Management 15 (1/2), 41–56.
Huang, K.W., Li, Z., 2011. A multi-label text classification algorithm for labeling risk factors in SEC form 10-K. ACM Transactions on Management Information Systems 2 (3), 18.
Huang, X., Ho, D., Ren, J., Capretz, L.F., 2007. Improving the COCOMO model using a neuro-fuzzy approach. Appl. Soft Comput. 7 (1), 29–40.
Huang, S.M., Tsai, C.F., Yen, D.C., Cheng, Y.L., 2008. A hybrid financial analysis model for business failure prediction. Expert Syst. Appl. 35 (3), 1034–1040.
Huang, S.Y., Tsaih, R.H., Lin, W.L., 2012. Unsupervised neural networks approach for understanding fraudulent financial reporting. Ind. Manag. Data Syst. 112 (2), 224–244.
Huang, X., Teoh, S.H., Zhang, Y., 2013. Tone management. Account. Rev. 89 (3), 1083–1113.
Huang, A.H., Zang, A.Y., Zheng, R., 2014. Evidence on the information content of text in analyst reports. Account. Rev. 89 (6), 2151–2180.
Huerta, E., Glandon, T., Petrides, Y., 2012. Framing, decision-aid systems, and culture: exploring influences on fraud investigations. Int. J. Account. Inf. Syst. 13 (4), 316–333.
Humpherys, S.L., Moffitt, K.C., Burns, M.B., Burgoon, J.K., Felix, W.F., 2011. Identification of fraudulent financial statements using linguistic credibility analysis. Decis. Support. Syst. 50 (3), 585–594.
Hwang, S.S., Shin, T., Han, I., 2004. CRAS-CBR: internal control risk assessment system using case-based reasoning. Expert. Syst. 21 (1), 22–33.
Issa, H., Kogan, A., 2014. A predictive ordered logistic regression model as a tool for quality review of control risk assessments. J. Inf. Syst. 28 (2), 209–229.
Ittner, C.D., Larcker, D.F., 2001. Assessing empirical research in managerial accounting: a value-based management perspective. J. Account. Econ. 32 (1), 349–410.
Jackson, J. 2002. Data mining: a conceptual overview. Communications of the Association for Information Systems 8 (1): 19.m
Jans, M., Lybaert, N., Vanhoof, K., 2010. Internal fraud risk reduction: results of a data mining case study. Int. J. Account. Inf. Syst. 11 (1), 17–41.
Jans, M., van der Werf, J.M., Lybaert, N., Vanhoof, K., 2011. A business process mining application for internal transaction fraud mitigation. Expert Syst. Appl. 38 (10), 13351–13359.
Jans, M., Alles, M., Vasarhelyi, M., 2013. The case for process mining in auditing: sources of value added and areas of application. Int. J. Account. Inf. Syst. 14 (1), 1–20.
Ji, S.H., Park, M., Lee, H.S., 2010. Data preprocessing–based parametric cost model for building projects: case studies of Korean construction projects. J. Constr. Eng. Manag. 136 (8), 844–853.
Ji, S.H., Park, M., Lee, H.S., 2011. Cost estimation model for building projects using case-based reasoning. Can. J. Civ. Eng. 38 (5), 570–581.
Jiawei, Han, Kamber, M., 2001. Data Mining: Concepts and Techniques. Morgan Kaufmann, San Francisco, CA.
Jie, L., Wei, L., 2009. Research on detecting technique of financial statement fraud based on fuzzy genetic algorithms BPN. Proceeding of the IEEE International Conference in Management Science and Engineering ICMSE, Moscow, pp. 1462–1468 September.
Jo, H., Han, I., Lee, H., 1997. Bankruptcy prediction using case-based reasoning, neural networks, and discriminant analysis. Expert Syst. Appl. 13 (2), 97–108.
Jones, J.J., 1991. Earnings management during import relief investigations. J. Account. Res. 193–228.
Jones, F.L., 1996. The information content of the auditor's going concern evaluation. J. Account. Public Policy 15 (1), 1–27.
Kabir, G., Hasin, M.A.A., 2013. Multi-criteria inventory classification through integration of fuzzy analytic hierarchy process and artificial neural network. Int. J. Ind. Syst. Eng. 14 (1), 74–103.
Kallio, M., Back, B., 2011. The self-organizing map in selecting companies for tax audit. In Emerging Themes in Information Systems and Organization Studies:347–358 http://dx.doi.org/10.1007/978-3-7908-2739-2_27.
Kaluzny, B.L., Barbici, S., Berg, G., Chiomento, R., Derpanis, D., Jonsson, U., ... Ramaroson, F., 2011. An application of data mining algorithms for shipbuilding cost estimation. Journal of Cost Analysis and Parametrics 4 (1), 2–30.
Kasgari, A.A., Divsalar, M., Javid, M.R., Ebrahimian, S.J., 2013. Prediction of bankruptcy Iranian corporations through artificial neural network and probit-based analyses. Neural Comput. & Applic. 23 (3/4), 927–936.
Khalifelu, Z.A., Gharehchopogh, F.S., 2012. Comparison and evaluation of data mining techniques with algorithmic models in software cost estimation. Procedia Technology 1, 65–71.
Khlif, H., Chalmers, K., 2015. A review of meta-analytic research in accounting. J. Account. Lit. 35, 1–27.
Kim, M.J., Kang, D.K., 2010. Ensemble with neural networks for bankruptcy prediction. Expert Syst. Appl. 37 (4), 3373–3379.
Kim, Y., Vasarhelyi, M.A., 2012. A model to detect potentially fraudulent/abnormal wires of an insurance company: an unsupervised rule-based approach. Journal of Emerging Technologies in Accounting 9 (1), 95–110.
Kim, H.J., Mannino, M., Nieschwietz, R.J., 2009. Information technology acceptance in the internal audit profession: impact of technology features and complexity. Int. J. Account. Inf. Syst. 10 (4), 214–228.
Kirkos, E., 2012. Predicting auditor switches by applying data mining. Journal of Applied Economic Sciences 3 (21), 246–261.
Kirkos, E., Spathis, C., Nanopoulos, A., Manolopoulos, Y., 2007a. Identifying qualified auditors' opinions: a data mining approach. Journal of Emerging Technologies in Accounting 4 (1), 183–197.
Kirkos, E., Spathis, C., Manolopoulos, Y., 2007b. Data mining techniques for the detection of fraudulent financial statements. Expert Syst. Appl. 32 (4), 995–1003.
Kirkos, E., Spathis, C., Manolopoulos, Y., 2008. Support vector machines, decision trees and neural networks for auditor selection. Journal of Computational Methods in Science and Engineering 8 (3), 213–224.

Kirkos, E., Spathis, C., Manolopoulos, Y., 2010. Audit-firm group appointment: an artificial intelligence approach. Intelligent Systems in Accounting, Finance and Management 17 (1), 1–17.
Kirsch, R.J., Strouble, D.D., Johnson, W.A., Aungst, S.G., 1991. Transfer: an expert system approach to transfer pricing. Expert Syst. Appl. 3 (4), 481–487.
Kleinman, G., Anandarajan, A., 1999. The usefulness of off-balance sheet variables as predictors of auditors' going concern opinions: an empirical analysis. Manag. Audit. J. 14 (6), 273–285.
Kloptchenko, A., Eklund, T., Karlsson, J., Back, B., Vanharanta, H., Visa, A., 2004. Combining data and text mining techniques for analyzing financial reports. Intelligent Systems in Accounting, Finance and Management 12 (1), 29–41.
Kochetova-kozloski, N., Messier Jr., W.F., Eilifsen, A., 2011. Improving auditors' fraud judgments using a frequency response mode. Contemp. Account. Res. 28 (3), 837–858.
Koh, H.C., Low, C.K., 2004. Going concern prediction using data mining techniques. Manag. Audit. J. 19 (3), 462–476.
Koh, H.C., Tan, S.S., 1999. A neural network approach to the prediction of going concern status. Account. Bus. Res. 29 (3), 211–216.
Köksal, G., Batmaz, İ., Testik, M.C., 2011. A review of data mining applications for quality improvement in manufacturing industry. Expert Syst. Appl. 38 (10), 13448–13467.
Korol, T., 2013. Early warning models against bankruptcy risk for Central European and Latin American enterprises. Econ. Model. 31, 22–30.
Koskivaara, E., 2000a. Artificial neural network models for predicting patterns in auditing monthly balances. J. Oper. Res. Soc. 51 (9), 1060–1069.
Koskivaara, E., 2000b. Different pre-processing models for financial accounts when using neural networks for auditing. Proceeding of the European Conference on Information Systems (ECIS 2000). Paper 3.
Koskivaara, E., 2004a. Artificial neural networks in analytical review procedures. Manag. Audit. J. 19 (2), 191–223.
Koskivaara, E., 2004b. Visualization of patterns in accounting data with self-organizing maps. In Business Intelligence Techniques:133–147 http://dx.doi.org/10.1007/978-3-540-24700-5_8.
Koskivaara, E., Back, B., 2007. Artificial neural network assistant (ANNA) for continuous auditing and monitoring of financial data. Journal of Emerging Technologies in Accounting 4 (1), 29–45.
Kostakis, H., Sarigiannidis, C., Boutsinas, B., Varvakis, K., Tampakas, V., 2008. Integrating activity-based costing with simulation and data mining. Int. J. Account. Inf. Manag. 16 (1), 25–35.
Kotsiantis, S., Koumanakos, E., Tzelepis, D., Tampakas, V., 2006. Forecasting fraudulent financial statements using data mining. Int. J. Comput. Intell. 3 (2), 104–110.
Krambia-Kapardis, M., Christodoulou, C., Agathocleous, M., 2010. Neural networks: the panacea in fraud detection? Manag. Audit. J. 25 (7), 659–678.
Landajo, M., de Andrés, J., Lorca, P., 2007. Robust neural modeling for the cross-sectional analysis of accounting information. Eur. J. Oper. Res. 177 (2), 1232–1252.
Larcker, D.F., Zakolyukina, A.A., 2012. Detecting deceptive discussions in conference calls. J. Account. Res. 50 (2), 495–540.
Lee, K., Booth, D., Alam, P., 2005. A comparison of supervised and unsupervised neural networks in predicting bankruptcy of Korean firms. Expert Syst. Appl. 29 (1), 1–16.
Lee, A.H., Kang, H.Y., Lai, C.M., Hong, W.Y., 2013. An integrated model for lot sizing with supplier selection and quantity discounts. Appl. Math. Model. 37 (7), 4733–4746.
Lenard, M.J., Alam, P., Madey, G.R., 1995. The application of neural networks and a qualitative response model to the auditor's going concern uncertainty decision. Decis. Sci. 26 (2), 209–227.
Lenard, M.J., Alam, P., Booth, D., 2000. An analysis of fuzzy clustering and a hybrid model for the auditor's going concern assessment. Decis. Sci. 31 (4), 861–884.
Lenard, M.J., Alam, P., Booth, D., Madey, G., 2001. Decision-making capabilities of a hybrid system applied to the auditor's going-concern assessment. Intelligent Systems in Accounting, Finance and Management 10 (1), 1–23.
Li, F., 2010. The information content of forward-looking statements in corporate filings—a naïve Bayesian machine learning approach. J. Account. Res. 48 (5), 1049–1102.
Li, S.G., Kuo, X., 2008. The inventory management system for automobile spare parts in a central warehouse. Expert Syst. Appl. 34 (2), 1144–1153.
Li, H., Sun, J., Li, J.C., Yan, X.Y., 2013. Forecasting business failure using two-stage ensemble of multivariate discriminant analysis and logistic regression. Expert. Syst. 30 (5), 385–397.
Liang, T.P., Chandler, J.S., Han, I., Roan, J., 1992. An empirical investigation of some data effects on the classification accuracy of probit, ID3, and neural networks. Contemp. Account. Res. 9 (1), 306–328.
Liang, D., Lin, F., Wu, S., 2001. Electronically auditing EDP systems: with the support of emerging information technologies. Int. J. Account. Inf. Syst. 2 (2), 130–147.
Liao, S.H., Chen, Y.J., 2014. A rough set-based association rule approach implemented on exploring beverages product spectrum. Appl. Intell. 40 (3), 464–478.
Lin, J.W., Hwang, M.I., Becker, J.D., 2003. A fuzzy neural network for assessing the risk of fraudulent financial reporting. Manag. Audit. J. 18 (8), 657–665.
Liou, F.M., 2008. Fraudulent financial reporting detection and business failure prediction models: a comparison. Manag. Audit. J. 23 (7), 650–662.
Liu, Z., Ren, Y., 2009. The CPVVA evaluating model based on improved DEA model optimized by CO algorithm. Proceeding of the 2nd International IEEE Workshop on Knowledge Discovery and Data Mining, WKDD 2009, Moscow January 19–22. 10.1109/WKDD.2009.29.
Liu, X.B., Zhou, S.K., Meng, Q.N., Bo, H.G., Yang, J.P., 2012. Activity-based standard cost variance analysis. Comput. Integr. Manuf. Syst. 18 (8), 1881–1893.
Lodhia, S., Martin, N., 2011. Stakeholder responses to the National Greenhouse and energy reporting act: an agenda setting perspective. Account. Audit. Account. J. 25 (1), 126–145.
Magnusson, C., Arppe, A., Eklund, T., Back, B., Vanharanta, H., Visa, A., 2005. The language of quarterly reports as an indicator of change in the company's financial status. Inf. Manag. 42 (4), 561–574.
Martens, D., Bruynseels, L., Baesens, B., Willekens, M., Vanthienen, J., 2008. Predicting going concern opinion with data mining. Decis. Support. Syst. 45 (4), 765–777.
Mauldin, E.G., Ruchala, L.V., 1999. Towards a meta-theory of accounting information systems. Acc. Organ. Soc. 24 (4), 317–331.
McKee, T.E., 2007. Altman's 1968 bankruptcy prediction model revisited via genetic programming: new wine from an old bottle or a better fermentation process? Journal of Emerging Technologies in Accounting 4 (1), 87–101.
Megala, N., Jawahar, N., 2006. Genetic algorithm and Hopfield neural network for a dynamic lot sizing problem. Int. J. Adv. Manuf. Technol. 27 (11/12), 1178–1191.
Migliaccio, G.C., Guindani, M., Zhang, S., Ghorai, S., 2011. Regression-based prediction methods for adjusting construction cost estimates by project location. Proceeding of the 3rd International Construction Specialty Conference, Ottawa, Ontario June 14–17.
Min, J.H., Lee, Y.C., 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Syst. Appl. 28 (4), 603–614.
Mokhatab Rafiei, F., Manzari, S.M., Bostanian, S., 2011. Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence. Expert Syst. Appl. 38 (8), 10210–10217.
Mueller-Wickop, N., Schultz, M., 2013. ERP event log preprocessing: timestamps vs. accounting logic. Proceeding of the 8th International on Design Science at the Intersection of Physical and Virtual Design (DESRIST), Helsinki, Finland June 105–119. 10.1007/978-3-642-38827-9_8.
Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y., Sun, X., 2011. The application of data mining techniques in financial fraud detection: a classification framework and an academic review of literature. Decis. Support. Syst. 50 (3), 559–569.
Ogut, H., Aktaş, R., Alp, A., Doganay, M.M., 2009. Prediction of financial information manipulation by using support vector machine and probabilistic neural network. Expert Syst. Appl. 36 (3), 5419–5423.
O'Leary, D.E., 1998. Using neural networks to predict corporate failure. International Journal of Intelligent Systems in Accounting, Finance and Management 7 (3), 187–197.
Olson, D.L., Delen, D., Meng, Y., 2012. Comparative analysis of data mining methods for bankruptcy prediction. Decis. Support. Syst. 52 (2), 464–473.
Owen, G., 2013. Integrated reporting: a review of developments and their implications for the accounting curriculum. Acc. Educ. 22 (4), 340–356.
Owusu-Ansah, S., Moyes, G.D., Oyelere, P.B., Hay, D., 2002. An empirical analysis of the likelihood of detecting fraud in New Zealand. Manag. Audit. J. 17 (4), 192–204.
Pai, P.F., Hsu, M.F., Wang, M.C., 2011. A support vector machine-based model for detecting top management fraud. Knowl.-Based Syst. 24 (2), 314–321.
Partovi, F.Y., Anandarajan, M., 2002. Classifying inventory using an artificial neural network approach. Comput. Ind. Eng. 41 (4), 389–404.
Peel, M.J., 1989. The going-concern qualification debate: some UK evidence. Br. Account. Rev. 21 (4), 329–350.

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.
Petroutsatou, K., Georgopoulos, E., Lambropoulos, S., Pantouvakis, J.P., 2011. Early cost estimating of road tunnel construction using neural networks. J. Constr. Eng. Manag. 138 (6) 679–6.
Pompe, P.P., Bilderbeek, J., 2005. Bankruptcy prediction: the influence of the year prior to failure selected for model building and the effects in a period of economic decline. Intelligent Systems in Accounting, Finance and Management 13 (2), 95–112.
Preece, J., Rombach, H.D., 1994. A taxonomy for combining software engineering and human-computer interaction measurement approaches: towards a common framework. International Journal of Human-Computer Studies 41 (4), 553–583.
Pujari, A.K., 2001. Data Mining Techniques. Universities press.
Purda, L., Skillicorn, D., 2014. Accounting variables, deception, and a bag of words: assessing the tools of fraud detection. Contemp. Account. Res. 1–32.
Quek, C., Zhou, R.W., Lee, C.H., 2009. A novel fuzzy neural approach to data reconstruction and failure prediction. Intelligent Systems in Accounting, Finance and Management 16 (1/2), 165–187.
Ragothaman, S., Lavin, A., 2008. Restatements due to improper revenue recognition: a neural networks perspective. Journal of Emerging Technologies in Accounting 5 (1), 129–142.
Ragothaman, S., Carpenter, J., Buttars, T., 1995. Using rule induction for knowledge acquisition: an expert systems approach to evaluating material errors and irregularities. Expert Syst. Appl. 9 (4), 483–490.
Ramamoorti, S., Bailey, J.A., Traver, R.O., 1999. Risk assessment in internal auditing: a neural network approach. International Journal of Intelligent Systems in Accounting, Finance and Management 8 (3), 159–180.
Ravi Kumar, P., Ravi, V., 2007. Bankruptcy prediction in banks and firms via statistical and intelligent techniques–a review. Eur. J. Oper. Res. 180 (1), 1–28.
Ravisankar, P., Ravi, V., Rao, G.R., Bose, I., 2011. Detection of financial statement fraud and feature selection using data mining techniques. Decis. Support. Syst. 50 (2), 491–500.
Reyes-Aldasoro, C.C., Ganguly, A.R., Lemus, G., Gupta, A., 1999. A hybrid model based on dynamic programming, neural networks, and surrogate value for inventory optimization applications. J. Oper. Res. Soc. 85–94.
Rezaee, Z., Sharbatoghlie, A., Elam, R., McMickle, P.L., 2002. Continuous auditing: building automated auditing capability. Audit. J. Pract. Theory 21 (1), 147–163.
Richardson, P., Dellaportas, S., Perera, L., Richardson, B., 2015. Towards a conceptual framework on the categorization of stereotypical perceptions in accounting. J. Account. Lit. 35, 28–46.
Salterio, S., 1996. The effects of precedents and client position on auditors' financial accounting policy judgment. Acc. Organ. Soc. 21 (5), 467–486.
Seo, K.K., Park, J.H., Jang, D.S., Wallace, D., 2002. Prediction of the life cycle cost using statistical and artificial neural network methods in conceptual product design. Int. J. Comput. Integr. Manuf. 15 (6), 541–554.
Serrano-Cinca, C., Gutiérrez-Nieto, B., 2013. Partial least square discriminant analysis for bankruptcy prediction. Decis. Support. Syst. 54 (3), 1245–1255.
Shan, C., He, J., 2012. Research on cost management system of distribution network construction projects based on data mining. Proceeding of the China International Conference on Electricity Distribution (CICED), Shanghai, China September 1–7. 10.1109/CICED.2012.6508454.
Shawver, T.J., 2005. Merger premium predictions using a neural network approach. Journal of Emerging Technologies in Accounting 2 (1), 61–72.
Shi, H., Li, W., 2008. The integrated methodology of rough set theory and artificial neural-network for construction project cost prediction. Proceeding of the 2nd International Symposium in Intelligent Information Technology Application (IITA'08), Shanghai, China December 60–64. 10.1109/IITA.2008.238.
Shirata, C.Y., Sakagami, M., 2008. An analysis of the "Going Concern Assumption": text mining from Japanese financial reports. Journal of Emerging Technologies in Accounting 5 (1), 1–16.
Shirata, C.Y., Takeuchi, H., Ogino, S., Watanabe, H., 2011. Extracting key phrases as predictors of corporate bankruptcy: empirical analysis of annual reports by text mining. Journal of Emerging Technologies in Accounting 8 (1), 31–44.
Šimunović, K., Šimunović, G., Šarić, T., 2009. Application of artificial neural networks to multiple criteria inventory classification. Strojarstvo: časopis za teoriju i praksu u strojarstvu 51 (4), 313–321.
Song, D.B., Lee, H.Y., Cho, E.J., 2013. The association between earnings management and asset misappropriation. Manag. Audit. J. 28 (6), 542–567.
Spathis, C.T., 2003. Audit qualification, firm litigation, and financial information: an empirical analysis in Greece. Int. J. Audit. 7 (1), 71–85.
Spear, N.A., Leis, M., 1997. Artificial neural networks and the accounting method choice in the oil and gas industry. Account. Manag. Inf. Technol. 7 (3), 169–181.
Sutton, S.G., Reinking, J., Arnold, V., 2011. On the use of grounded theory as a basis for research on strategic and emerging technologies in accounting. Journal of Emerging Technologies in Accounting 8 (1), 45–63.
Tackett, J.A., 2013. Association rules for fraud detection. Journal of Corporate Accounting and Finance 24 (4), 15–22.
Tang, Y.C., 2009. An approach to budget allocation for an aerospace company—fuzzy analytic hierarchy process and artificial neural network. Neurocomputing 72 (16), 3477–3489.
Tang, T.C., Chi, L.C., 2005. Neural networks analysis in business failure prediction of Chinese importers: a between-countries approach. Expert Syst. Appl. 29 (2), 244–255.
Tinoco, M.H., Wilson, N., 2013. Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. Int. Rev. Financ. Anal. 30, 394–419.
Tranfield, D., Denyer, D., Smart, P., 2003. Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br. J. Manag. 14 (3), 207–222.
Tsai, C.F., Chiou, Y.J., 2009. Earnings management prediction: a pilot study of combining neural networks and decision trees. Expert Syst. Appl. 36 (3), 7183–7191.
Tsai, C.F., Wu, J.W., 2008. Using neural network ensembles for bankruptcy prediction and credit scoring. Expert Syst. Appl. 34 (4), 2639–2649.
Tsai, C.F., Lu, Y.H., Yen, D.C., 2012. Determinants of intangible assets value: the data mining approach. Knowl.-Based Syst. 31, 67–77.
Tsaih, R.H., Lin, W.Y., Huang, S.Y., 2009. Exploring fraudulent financial reporting with GHSOM. Proceeding of Pacific Asia Workshop (PAISI) In Intelligence and Security Informatics, Bangkok, Thailand April 31–41. 10.1007/978-3-642-01393-5_5.
Tseng, F.M., Hu, Y.C., 2010. Comparing four bankruptcy prediction models: logit, quadratic interval logit, neural and fuzzy neural networks. Expert Syst. Appl. 37 (3), 1846–1853.
Tung, W.L., Quek, C., Cheng, P., 2004. GenSo-EWS: a novel neural-fuzzy based early warning system for predicting bank failures. Neural Netw. 17 (4), 567–588.
Vasarhelyi, M.A., Alles, M.G., Kogan, A., 2004. Principles of analytic monitoring for continuous assurance. Journal of Emerging Technologies in Accounting 1 (1), 1–21.
Vasarhelyi, M.A., Alles, M., Kuenkaikaew, S., Littley, J., 2012. The acceptance and adoption of continuous auditing by internal auditors: a microanalysis. Int. J. Account. Inf. Syst. 13 (3), 267–281.
Virdhagriswaran, S., Dakin, G., 2006. Camouflaged fraud detection in domains with complex relationships. Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Philadelphia, PA August 941–947.
Vouk, D., Malus, D., Halkijevic, I., 2011. Neural networks in economic analyses of wastewater systems. Expert Syst. Appl. 38 (8), 10031–10035.
Wang, S., 2010. A comprehensive survey of data mining-based accounting-fraud detection research. Proceeding of the International IEEE Conference on Intelligent Computation Technology and Automation (ICICTA), Changsha, China May 50–53. 10.1109/ICICTA.2010.831.
Wang, W., 2011. Model of the forecasting of inventory risk management of spare parts: an analysis framework based on neural network. Energy Procedia 13, 6412–6417.
Wang, Y.R., Kong, S.L., 2012. Applying genetic algorithms for construction quality auditor assignment in public construction projects. Autom. Constr. 22, 459–467.
Wang, P.Y., Chow, T.W., Chiu, C.W., 2009. Computational accounting in determining chart of accounts using nominal data analysis and concept of entropy. Expert Syst. Appl. 36 (3), 6966–6977.
Welch, O.J., Reeves, T.E., Welch, S.T., 1998. Using a genetic algorithm-based classifier system for modeling auditor decision behavior in a fraud setting. International Journal of Intelligent Systems in Accounting, Finance and Management 7 (3), 173–186.
White, L., 2004. Why look at German cost management. Strategic Finance 86 (3), 6–25.
Williams, J.W., 2013. Regulatory technologies, risky subjects, and financial boundaries: governing 'fraud' in the financial markets. Acc. Organ. Soc. 38 (6), 544–558.
Wu, R.C.F., 1994. Integrating neurocomputing and auditing expertise. Manag. Audit. J. 9 (3), 20–26.

Wu, M.C., Hsu, Y.K., 2008. Design of BOM configuration for reducing spare parts logistic costs. Expert Syst. Appl. 34 (4), 2417–2423.
Wu, R.S., Ou, C.S., Lin, H.Y., Chang, S.I., Yen, D.C., 2012. Using data mining technique to enhance tax evasion detection performance. Expert Syst. Appl. 39 (10), 8769–8777.
Xiao, M., Xiaoli, H., Gaojin, L., 2010. Research on application of data mining technology in financial decision support system. Proceeding of the International IEEE Conference on Information Management, Innovation Management and Industrial Engineering (ICIII), Kunming, China November 381–384. 10.1109/ICIII.2010.572.
Yang, J.G.S., 2006. Data mining techniques for auditing attest function and fraud detection. Journal of Forensic Investigative Accounting 1 (1), 4–10.
Yang, Z.R., Platt, M.B., Platt, H.D., 1999. Probabilistic neural networks in bankruptcy prediction. J. Bus. Res. 44 (2), 67–74.
Yeh, T.H., Deng, S., 2012. Application of machine learning methods to cost estimation of product life cycle. Int. J. Comput. Integr. Manuf. 25 (4/5), 340–352.
Yigitbasioglu, O.M., Velcu, O., 2012. A review of dashboards in performance management: implications for design and research. Int. J. Account. Inf. Syst. 13 (1), 41–59.
Youn, H., Gu, Z., 2010. Predicting Korean lodging firm failures: an artificial neural network model along with a logistic regression model. Int. J. Hosp. Manag. 29 (1), 120–127.
Yu, M.C., 2011. Multi-criteria ABC analysis using artificial-intelligence-based classification techniques. Expert Syst. Appl. 38 (4), 3416–3421.
Yu, W.D., Lai, C.C., Lee, W.L., 2006. A WICE approach to real-time construction cost estimation. Autom. Constr. 15 (1), 12–19.
Yu, Z., Guang, Y., Zi-qi, J., 2013. Violations detection of listed companies based on decision tree and K-nearest neighbor. Proceeding of the International IEEE Conference on Management Science and Engineering (ICMSE), Harbin, China :pp. 1671–1676 July. 10.1109/ICMSE.2013.6586490.
Yue, D., Wu, X., Shen, N., Chu, C.H., 2009. Logistic regression for detecting fraudulent financial statement of listed companies in China. Proceeding of the International IEEE Conference on Artificial Intelligence and Computational Intelligence (AICI'09), Shanghai, China :pp. 104–108 November. 10.1109/AICI.2009.421.
Zeng, Y., Wang, L., Chen, T., Lu, Y., 2006. A hybrid intelligent decision support system for spare parts inventory control using neural network and gene algorithms approach. Proceeding of the Conference on Power, Energy, and Applications, Gaborone, Botswana Available at:. http://www.actapress.com/Abstract.aspx?paperId=28205.
Zhang, G., Hu, M.Y., Eddy-Patuwo, B., Indro, D.C., 1999. Artificial neural networks in bankruptcy prediction: general framework and cross-validation analysis. Eur. J. Oper. Res. 116 (1), 16–32.
Zhao, L., Ding, L., 2009. Project cost control system based on data mining. Proceeding of International Forum on Information Technology and Applications (IFITA'09), Chengdu, China :pp. 535–538 May. 10.1109/IFITA.2009.306.
Zheng, C., 2011. A study on the construction of accounting information system based on REA model. Proceeding of the International Conference on Uncertainty Reasoning and Knowledge Engineering, Bali, Indonesia :pp. 165–168 August. 10.1109/URKE.2011.6007935.
Zhou, W., Kapoor, G., 2011. Detecting evolutionary financial statement fraud. Decis. Support. Syst. 50 (3), 570–575.
Zhou, C.C., Yin, G.F., Hu, X.B., 2009. Multi-objective optimization of material selection for sustainable products: artificial neural networks and genetic algorithm approach. Mater. Des. 30 (4), 1209–1215.
Zouboulidis, E., Kotsiantis, S., 2012. Forecasting fraudulent financial statements with committee of cost-sensitive decision tree classifiers. Proceeding of the 7th Hellenic Conference on Artificial Intelligence: Theories and Applications, Lamia, Greece :pp. 57–64 May. 10.1007/978-3-642-30448-4_8.