CHAPTER 1

INTRODUCTION

1.1 Overview
1.2 Objectives
1.3 Scope
1.4 Significance
1.5 Thesis Organisation
1.6 Summary

This chapter provides the foundation for current research work by unfolding the problem being investigated. It also presents the objectives, scope and significance of the research work. Finally, it provides an overview of the subsequent chapters of thesis followed by chapter summary.

1.1 OVERVIEW

Data mining aims to discover hidden knowledge, unknown patterns, and new rules from large databases that are potentially useful and ultimately understandable for making crucial decisions. It applies data analysis and knowledge discovery techniques under acceptable computational efficiency limitations, and produces a particular enumeration of patterns over the data. The insights obtained via a higher level of understanding of data can help iteratively improve business practice.

Data mining is able to uncover unknown patterns and predict future trends and behaviours in financial markets. It creates opportunities for companies to make proactive and knowledge-driven decisions in order to gain a competitive advantage. Data mining has been applied to a number of financial applications, including development of trading models, investment selection, loan assessment, portfolio optimization, fraud detection, bankruptcy prediction, real-estate assessment, and so on. The competitive advantages achieved by data mining include increased revenue, reduced cost, and much improved marketplace responsiveness and awareness.
Fraud detection is one of the most important applications of Data Mining. It is widely accepted by both researchers' community and practitioners that there is a requirement of analytical procedures and data mining techniques along with traditional auditing procedures for prevention and detection of financial statement fraud. Auditing firms and procedures are not capable enough to prevent and detect financial statement fraud, since detection of fraud is not their primary objective and auditors have a very little knowledge about the management of the organization. Moreover, standard auditing procedures may prove insufficient because auditors use sampling technique and do not examine each and every transaction. Researchers are working hard for designing and implementing new data mining methodologies for prevention and detection of fraudulent financial reporting. Enormous research work has been already done by many researchers in this area. However, due to upgrading and changing technology and new methods used by management in perpetrating financial statement fraud, there is always a necessity of more research work in this area. This motivated us to work in this area.

This research work emphasis on analysis and design of data mining techniques for prevention and detection of financial statement fraud. Brief descriptions of these are as follows:

1.1.1 Prevention of financial statement fraud

Data Mining plays a very important role in prevention of financial statement fraud because the aim of data mining is to find out potential knowledge, unknown patterns and unsuspected relationship from a large set of data. This capability of data mining has been utilised in this research for preventing fraudulent financial reporting.

Data mining tasks can be divided in two subgroups: predictive tasks and descriptive tasks. With predictive tasks, the objective is to predict the value of one attribute, based on the values of other attributes. Due to this nature, predictive data mining along with machine learning is best suited for fraud detection. Predictive tasks make a prediction for every observation. Descriptive tasks however, describe the data set as a whole. It aims to describe the underlying relationships in the data set. This fact accounts for the use of descriptive data mining instead of predictive data mining for fraud prevention. An advantage of the use of descriptive data mining techniques is that it is easier to apply it on unsupervised data. Therefore, this research recommends the use of descriptive data mining techniques for prevention of financial statement fraud.
Descriptive data mining techniques namely association rules, clustering and anomaly detection are suggested as appropriate candidates for prevention of financial statement fraud in this research.

1.1.2 Detection of financial statement fraud

Financial statement fraud is a deliberate misstatement of material facts by the management in the books of accounts of a company with the aim of deceiving investors and creditors. This illegitimate task performed by management has a severe impact on the economy throughout the world because it significantly dampens the confidence of investors. Despite the presence of strong internal control and various internal as well as external audit committees, detecting fraudulent financial reporting fraud is a difficult task when using normal audit procedures due to the following reasons. First, there is a shortage of knowledge concerning the characteristics of financial statement fraud. Secondly, given its infrequency, most auditors lack the experience necessary to detect it. Finally, managers deliberately try to deceive auditors. For such managers, who comprehend the limitations of any audit, standard auditing procedures may prove insufficient. It has also been noted that the increased emphasis on system assessment is at odds with the profession’s position regarding fraud detection, since most material frauds originate at the top levels of the organization, where controls and systems are least prevalent and effective. These limitations suggest that there is a need for additional analytical procedures for the effective detection of financial statement fraud.

Cost of financial statement fraud is very high both in terms of finance as well as the goodwill of the organisation and related country. Therefore, this research work discusses various causes and consequences of financial statement fraud.

Detection of financial statement fraud comes in to play only if the prevention mechanism has failed to stop the management in perpetrating the fraud. Therefore, this research work proposes a data mining framework for prevention of financial statement fraud at the first place and detecting it in case of failure of prevention methods.

Data mining methods could possibly assist auditors in detection of fraud, because data mining can use previous instances of fraud to build models in order to identify and detect the risk of fraud. Detection of financial statement fraud is an instance of classification and decision problem and the efficacy of the detection depends on the classification algorithms and the
fraud predictors used and how they are combined. Therefore, this research work identifies various financial ratios / variables as input vector to the classification algorithms and further implements three classification algorithms i.e. Decision Tree (CART), Naïve Bayesian Classifier and Genetic Programming for identification and detection of financial statement fraud.

1.2 OBJECTIVES

Objective of this research work is to study, introduce, design, apply and evaluate the use of data mining techniques by mainly focusing on two aspects of financial statement fraud – prevention and detection.

Following seven specific objectives are decided and achieved:

**Objective – 1:** To study the current use of data mining techniques in prevention and detection of financial statement fraud. This includes the study of the nature of data mining techniques used along with data specifications and empirical results obtained by the use of these techniques.

To achieve this objective existing data mining techniques proposed and implemented by various researchers are comprehensively studied and explored. A review is conducted which includes the analytical and empirical results of various data mining techniques. Studies helped in understanding the applicability, advantages and shortcomings of data mining methods for prevention and detection of financial statement fraud. This also provided path for further research work.

**Objective – 2:** To analyse various factors responsible for financial statement fraud. This includes the study of causes and consequences of fraudulent financial reporting.

To accomplish this objective first concept of financial statement fraud is explored then for better understanding various factors responsible for financial statement fraud are studied followed by the study of effects of fraudulent financial reporting on investors, creditors and a country as a whole.

**Objective – 3:** To perform the comparative study of extensively used data mining techniques in identification and detection of financial statement fraud.
To achieve this objective the four commonly used data mining techniques - Neural Networks, Decision Trees, Genetic Algorithm (GA) and Bayesian Belief Network (BBN) are compared in terms of their performance on the basis of eight parameters on a five point scale ranging from low to very high.

**Objective – 4:** To identify financial ratios, variables for prevention and detection of financial statement fraud.

To accomplish this objective, three financial statements namely balance sheet, income statement and cash flow statements are comprehensively studied and analysed. In order to identify financial variables, behavioural characteristics of an organisation along with profitability, liquidity, safety and efficiency are considered. Sixty two financial variables are identified as key input variables to the proposed framework for prevention and detection of financial statement fraud.

**Objective – 5:** To suggest descriptive data mining techniques for prevention of fraudulent financial reporting.

To realize this objective three descriptive data mining techniques are suggested to prevent financial statement fraud. In this regard, conventional methods of prevention of financial statement fraud are explored followed by the introduction of three data mining techniques namely, Association Rule, Cluster Analysis and Anomaly Detection.

**Objective – 6:** To propose a data mining framework for prevention and detection of financial statement fraud.

To accomplish this objective, a new data mining framework is proposed in order to prevent fraudulent financial reporting at the first place and detecting it if the management of an organisation is able to perpetrate even in presence of anti-fraud environment. This framework suggests the use of descriptive data mining for prevention and predictive data mining techniques for detection of financial statement fraud.

**Objective – 7:** To design association rules for prevention and decision rules for detection of financial statement fraud.

To realize this objective, the proposed framework is implemented by collecting data of 114 organisations. The proposed framework is implemented by using Association Rules as descriptive data mining and Decision Tree (CART), Naïve Bayesian Classifier and Genetic
CHAPTER 1

INTRODUCTION

Programming as predictive techniques for prevention and detection of financial statement fraud. This implementation resulted in seven association rules for preventing fraudulent financial reporting at the first place followed by five decision rules for detection of fraud in case of failure of the prevention mechanism.

1.3 SCOPE

This research work focuses on the design and applicability of data mining framework for prevention and detection of financial statement fraud. Data used for implementation of the data mining framework are extracted from publically available financial statements namely balance sheet, income statement and cash flow statements of 114 organisations. Financial variables used in this research are confined to financial ratios available from financial statements; characteristics such as number of outside board members and composition of administrative board are not in the scope of this research work. This research proposes the use of descriptive data mining techniques as opposed to the use of predictive data mining techniques for prevention of financial statement fraud. Text mining algorithms for sentiment analysis of the textual description of financial statements are not in the scope of this research work. The proposed framework considers the use of individual data mining techniques only. Hybrid systems which integrate two or more data mining techniques are not taken in the scope of this research work.

1.4 SIGNIFICANCE

In addition to contribution towards better understanding and applicability of data mining techniques for prevention and detection of financial statement fraud, following are the specific significant benefits of this research work:

First, the proposed research provides a theoretical framework to the concept of data mining; applicability of data mining methods in prevention and detection of fraud, insight of existing use of data mining methods for prevention and detection of fraudulent financial reporting. This is helpful for designing appropriate research studies, to understand the objectives and make decisions based on these objectives for prevention and detection of financial statement fraud.

Second, it provides complete list of the reasons behind fraudulent financial reporting. These reasons further helps in identifying the methods and the consequences of financial statement fraud.
Third, this research work measures the efficiency of four extensively used data mining techniques by comparing them on the basis of eight parameters. This is helpful in selecting best data mining techniques for detection of financial statement fraud.

Fourth, this research comprehensively analyse the three financial statements namely balance sheet, income statement and cash flow statement and identifies sixty two financial ratios to be used as input vector to the proposed data mining framework for prevention and detection of financial statement fraud. These financial variables are based on the performance measures of an organisation.

Fifth, this research work suggests three descriptive data mining techniques for prevention of financial statement fraud. Three techniques namely Association Rule, Cluster Analysis and Anomaly Detection are explained and investigated. This is useful for selecting a data mining technique for preventing fraudulent financial reporting.

Sixth, this research work proposes a data mining framework for prevention and detection of financial statement fraud. The proposed framework follows the conventional flow of data mining and suggests the use of descriptive data mining methods for prevention and predictive data mining methods for detection of financial statement fraud. This is helpful in better prevention and detection because detection mechanism of the proposed framework is complemented by the prevention mechanism.

Seventh, this research work presents seven association rules for prevention and five decision rules for detection of financial statement fraud. These results are going to be of great help to both researchers and practitioners in preventing fraudulent financial reporting at the first place and detecting it in case of failure of prevention methods.

1.5 THESIS ORGANISATION

This research work is based on analysis and design of data mining techniques by focusing on prevention and detection of financial statement fraud. The whole work is presented in eight chapters which are briefly described as follows:
Chapter 1 introduces the theme of this research work by providing the overview of the research work. Objectives, Scope and Significance of the research work are also given in this chapter.

Chapter 2 provides study and review of literature in order to create an adequate framework for conducting this research work. This chapter is divided into five sections. Section 2.1 provides the introduction of the chapter. Section 2.2 defines the fundamental terminology and concepts of Data Mining. Section 2.3 provides an overview of different applications of data mining. Section 2.4 presents the review of existing use of data mining techniques for prevention and detection of financial statement fraud by focusing on the nature of data mining technique used, data set and experimental results. Section 2.5 summarizes the chapter.

Chapter 3 discusses the financial statement fraud, its causes, methods and consequences. This chapter is divided into ten sections. Section 3.1 provides the introduction of this chapter. In order to understand the concept, section 3.2 defines fraud. Section 3.3 explains different classification of fraud. Section 3.4 introduces the concept of financial statement fraud. Section 3.5 discusses the reasons behind financial statement fraud by explaining different constituents of fraud triangle. This section further elaborates the tricks adopted by the management of an organisation in order to publish fraudulent financial statement. Section 3.6 explores the effects of financial statement fraud. Section 3.7 elucidates the flowchart of financial statement fraud. Section 3.8 summarizes the chapter.

Chapter 4 performs the comparative analyses of four extensively used data mining techniques for detection of fraudulent financial reporting. These techniques are compared on the basis of eight parameters. These performance criteria include classification accuracy, ease of problem encoding, flexibility, computation complexity, interpretability, optimization capability, scalability and accessibility.

Chapter 5 is concerned with identification of financial variables required for prevention and detection of financial statement fraud. It discusses the behavioural characteristics for preventing fraudulent financial reporting and identifies ten input variables. This chapter also identifies fifty two variables for identification and detection of financial statement fraud.
Chapter – 6 suggests three descriptive data mining techniques namely Association Rule, Naïve Bayesian Classifier and Anomaly Detection for financial statement fraud prevention. This chapter also explains the conventional methods for preventing fraudulent financial reporting.

Chapter – 7 proposes and implements a data mining framework for prevention and detection of financial statement fraud. This chapter is divided in two parts. First part explains the proposed framework and explores the issues related with anti – fraud environment. Second part implements the newly designed framework by using one descriptive data mining technique namely association rule and three predictive data mining methods i.e. Decision Tree (CART), Naïve Bayesian Classifier and Genetic Programming. This implementation results in seven association rules for preventing the fraud and five decision rules for financial fraud detection.

Chapter – 8 summarizes research contribution. Objectives of research work are reviewed in this chapter. It also provides research findings and significance. Finally, limitations of current study along with future direction are outlined in this chapter.

1.6 SUMMARY

The whole research work is divided mainly into three parts – analysis of data mining techniques, identification of financial variables and design of new data mining framework for prevention and detection of financial statement fraud.

First part focuses on the analysis of data mining techniques used for prevention and detection of fraud. The reasons and methods of financial statement fraud are also investigated in this part. The efficiency of four commonly used data mining techniques is judged by comparing these techniques on the basis of eight performance criteria.

Second part identifies financial variables for prevention and detection of fraud. This part also suggests three descriptive data mining techniques for prevention of financial statement fraud.

Third part proposes a new data mining framework for prevention and detection of financial statement fraud. This part further implements the proposed framework by applying one descriptive data mining method and three predictive data mining techniques. This
implementation results in seven association rules and five decision rules. Rules generated in this research are the major milestones of this research work.

Outline of the whole research work is provided in this chapter which briefly describes the research work conducted and its motivation. Seven objectives are discussed in this chapter which are fulfilled and have been detailed out in subsequent chapters. Organisation of the thesis presents thesis structure and readers’ guide.