ABSTRACT

Data mining techniques are the most commonly used techniques for prevention and detection of financial frauds. The implementation of data mining techniques for fraud detection follows the traditional information flow of data mining, which begins with feature selection followed by representation, data collection and management, pre-processing, data mining, post-processing, and performance evaluation. Data mining methods have the capability of detecting fraud because these techniques can use past cases of fraud to build models, which identify and detect the risk of fraud.

Financial statement fraud, one of the financial frauds, has reached the epidemic proportion globally. Collapses of high profile companies have left a dirty smear on the effectiveness of corporate governance, quality of financial reports, and credibility of audit functions. Financial statement fraud has become a critical issue for the organisations around the world.

Detection of Financial Statement Fraud is one of the key application areas of Data Mining, because data mining methods are capable of discovering the reasons behind fraudulent financial reporting. Traditional auditing procedures are available but auditors are supposed to provide information, whether the financial statement is according to GAAP (Generally Accepted Accounting Principles) or not. They cannot provide absolute assurance that all material misstatements are detected and identified. Consequently large number of data mining techniques have been proposed and implemented by researchers’ community to give more effective methods of preventing and detecting financial statement fraud.

This research work conducts extensive literature survey of existing use of data mining methods in prevention and detection of financial statement fraud, by focusing on nature of data mining techniques, data specifications and experimental results. This survey gives in –
depth view of existing techniques and provides a direction to rest of the study. By focusing on analysis and design of data mining techniques for prevention and detection of financial statement fraud, seven objectives are realised in this research work. Entire research work is divided into three parts to achieve these objectives.

First part deals with analysis of data mining techniques for detection of financial statement fraud. Four commonly used data mining techniques have been investigated in this part. Data mining techniques namely Neural Networks, Decision Trees, Genetic Algorithm and Bayesian Belief Networks are compared on the basis of eight performance criteria. These eight criteria includes: classification accuracy, ease of problem encoding, flexibility, computation complexity, interpretability, optimization capability, scalability and accessibility. In order to have better understanding about financial statement fraud, concept of fraud and financial statement fraud along with its causes and consequences are also elaborated in this part.

Second part of this research work identifies sixty two financial variables from publically available financial statements namely Balance Sheet, Income Statement and Cash Flow Statement. These financial variables have been used as input vector to the proposed data mining framework for prevention and detection of financial statement fraud. This part also suggests three descriptive data mining techniques for preventing fraudulent financial reporting. These techniques are Association Rule, Cluster Analysis and Anomaly Detection. Furthermore, these techniques are explained in terms of prevention of financial statement fraud.

Third part proposes a new data mining framework for preventing fraudulent financial reporting at the first place and detecting it once top level managers are able to commit fraud by deceiving the prevention methods. Data mining framework proposed in this research follows the conventional flow of data mining and proposes the use of descriptive data mining
methods for prevention of financial statement fraud and predictive data mining methods for
detection in case of failure of prevention methods. This part of the research also implements
the proposed framework by collecting sixty two financial variables decided in part two from
financial statements of 114 organisations. Association Rules have been generated by the Rule
Engine module of the framework. These rules are further analysed by Rule Monitor module
of the framework for raising an alarm regarding financial statement fraud. In case of failure
of prevention mechanism, three predictive mining methods have been applied for detecting
the fraud. The efficacy of these techniques has been evaluated by using sensitivity and
specificity as the measures of performance evaluation.

The association rules generated in this study are going to be of great importance for both
researchers and practitioners in preventing fraudulent financial reporting. Decision rules
produced in this research complements the prevention mechanism by detecting financial
statement fraud.

Hence, the framework designed and implemented in this research will boast up the prevention
mechanism of an organisation by preventing fraudulent financial reporting and detect it if
management of the organization is capable of perpetrating financial statement fraud despite
the presence of anti-fraud environment.

Overall this research work is helpful in understanding the existing use of data mining
methods for prevention and detection of financial statement fraud, preventing financial
statement fraud at the first place and detecting it in case of failure of prevention mechanism.
It further helps in identifying financial variables responsible for fraud from publically
available financial statements, suggesting data mining methods for prevention of fraudulent
financial reporting and selecting a best data mining method for detection of fraud.