

# A Study on Efficient & Secure Data Mining Technique for Detecting Financial Frauds

Vikas Chahar  
(Associate Professor & Head, VIMT – Rohtak)

## Abstract

Data might be one of the most valuable assets of any corporation—but only if it knows how to reveal valuable knowledge hidden in raw data. Data mining allows to extract diamonds of knowledge from the historical data, and predict outcomes of future situations. It helps optimise business decisions, increase the value of each customer and communication, and improve customer satisfaction.

**Keywords:** Data Mining Technique, Financial Frauds.

## 1. Introduction

Data mining is the process of extracting previously unknown information, typically in the form of patterns and associations, from large databases. Today, organisations are realising the numerous advantages that come with data mining. It is a valuable tool—by identifying potentially useful information from the large amounts of data collected, an organisation can gain a clear advantage over its competitors.

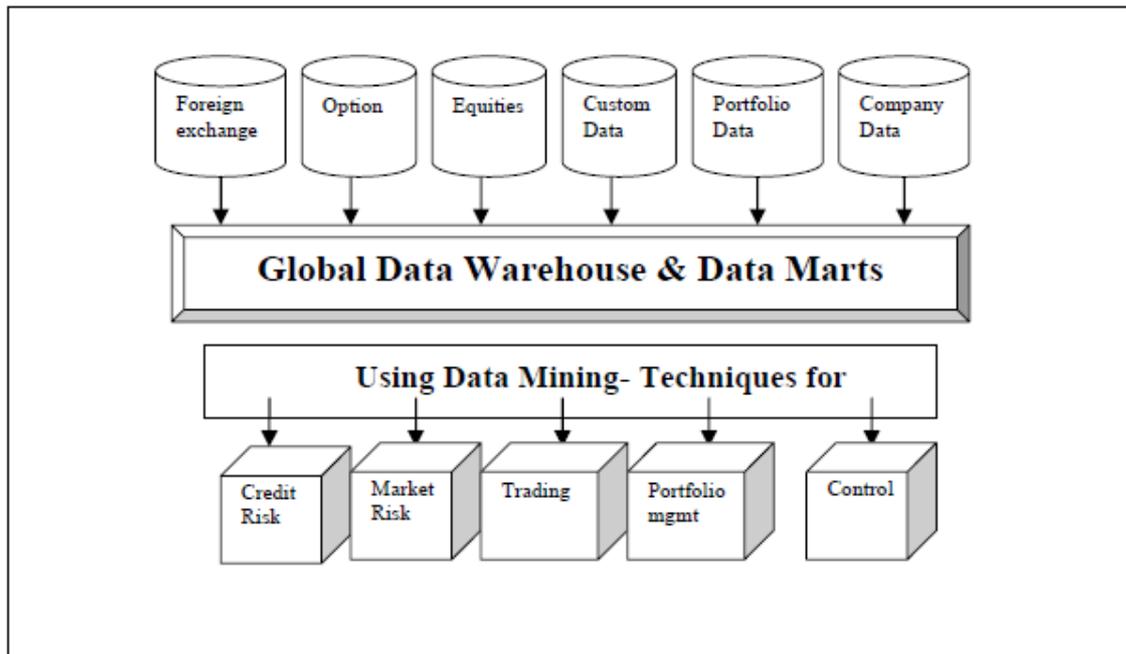
According to *Witten and Frank (2000)*, data mining can be defined as :

*"... The process of discovering patterns in data. The process must be automatic or (more usually) semi-automatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities."*

This definition validates Michalski et al. (1998)'s explanation. If data mining results in discovering meaningful patterns, data turns into information. Information or in this case- patterns that are novel, valid and potentially useful are not merely information, but knowledge. One speaks of discovering knowledge, before hidden in the huge amount of data, but now revealed. This brings us to the term 'Knowledge Discovery', which is usually called in the same breath as 'Data Mining'. The data mining process consists of four basic steps:

- ❖ Question Definition;
- ❖ Data Preparation And Pre-Processing;
- ❖ Data Mining
- ❖ Result Interpretation and Validation.

The broad categories of application of Data Mining and Business Intelligence techniques in the banking and financial industry vertical may be viewed as follows :



The technological improvement that's partly responsible for the increasing trend in fraud, is also part of the solution. Prevention and detection technologies are implemented, tested, customized and commercialized. Software companies sell 'The solution to all your business problems, including fraud'. The term '**Data Mining**' is sold as an expensive, all-problems-solving word. If your business doesn't use data mining, you're not in the game. Starting from this background, *it is interesting to do some research about how to improve existing data mining techniques for detection of Frauds.*

### 1.1 Definition Of Fraud:

Fraud is any dishonest act and behaviour by which one person gains or intends to gain advantage over another person. Fraud causes loss to the victim directly or indirectly. Fraud has not been described or discussed clearly in The Indian Penal Code but sections dealing with cheating, Concealment, forgery, counterfeiting and breach of trust has been discusses which leads to the act of fraud. In Contractual term as described in the Indian Contract Act, Sec 17 suggests that a fraud means and includes any of the acts by a party to a contract or with his connivance or by his agents with the intention to deceive another party or his agent or to induce him to enter in to a contract.

The Concise Oxford Dictionary defines fraud as '*criminal deception; the use of false representations to gain an unjust advantage*'. Fraud is as old as humanity itself, and can take an unlimited variety of different forms. However, in recent years, the development of new technologies (which have made it easier for us to communicate and helped increase our spending power) has also provided yet further ways in which criminals may commit fraud.

The no. of Frauds is increasing every year. "45% of companies worldwide have fallen victim to economic crime in 2004 and 2005. The average damage to the companies from tangible frauds (i.e. asset misappropriation, false pretences, and counterfeiting) was US\$ 1.7 million." according to the 'Global economic crime survey 2005' of PriceWaterhouseCoopers [1].

The number of financial frauds in India is substantial. It is increasing with the passage of time. All the major operational areas in banking represent a good opportunity for fraudsters with growing incidence being reported under deposit, loan and inter-branch accounting transactions, including remittances. Financial Institutions fraud is a big business in today's world. With

more educational qualifications, banking becoming impersonal and increase in banking sector have gave rise to this white collar crime. In a survey made till 1997 bank frauds in nationalised banks was of Rs.497.60 crore.

## 1.2 Types of Financial Fraud

The financial frauds may be of following types:

- ❖ Internet Frauds
- ❖ Check Fraud
- ❖ New Account Fraud
- ❖ Identity Fraud
- ❖ Credit/Debit Card Fraud
- ❖ ATM Transaction Fraud
- ❖ Wire Fraud
- ❖ Loan Fraud

Further, Online banking fraud could be divided into three categories, each of which poses a unique threat to customers and institutions Shirley Inscoe et al. (2003) [2]. The three categories are:

- ❖ Identity Theft
- ❖ Friendly Fraud, or fraud committed by a trusted relative or friend
- ❖ Internal Fraud, that is, fraud perpetrated by a financial institution employee.

## 1.3 What is Fraud Detection?

- ❖ Identify Wrongful Actions :
  - Is Right And Wrong Universal?
  - If So, Why Not Just Prevent Wrong Actions
- ❖ Identify Actions By The Wrong People.
- ❖ Identify *Suspect* Actions :
  - Legal
  - But Probably Not Right

## 2. Literature Survey

Extensive research has been done in the area of "Financial Fraud Detection". This section highlights the work of eminent researchers and

explores the challenges, which still need to be done.

Shamik Sural et al. (2008) [3] proposed an application of HMM in credit card fraud detection. The different steps in credit card transaction processing are represented as the underlying stochastic process of an HMM. We have used the ranges of transaction amount as the observation symbols, whereas the types of item have been considered to be states of the HMM. We have suggested a method for finding the spending profile of cardholders, as well as application of this knowledge in deciding the value of observation symbols and initial estimate of the model parameters. It has also been explained how the HMM can detect whether an incoming transaction is fraudulent or not.

Philip K. Chan et al. (1999) [4] explained large-scale data-mining techniques can improve on the state of the art in commercial practice. Scalable techniques to analyze massive amounts of transaction data that efficiently compute fraud detectors in a timely manner is an important problem, especially for e-commerce. Besides scalability and efficiency, the fraud-detection task exhibits technical problems that include skewed distributions of training data and nonuniform cost per error, both of which have not been widely studied in the knowledge-discovery and data mining community.

Min-Jung Kim et al. (2002) [5] described that even using salient features and advanced classification techniques, the detection of credit card frauds is a very complicated problem. Two major characteristics in the credit card transaction data may explain this difficulty. The one is the skewed distribution of data. Fraudulent transactions happen very rarely as compared with legitimate transactions and the percentage of fraudulent transaction is usually considered to be 0.1% or less . And the other is that a large part of fraudulent transactions are overlapped with the legitimate ones. It is not easy to extract frauds from non-frauds efficiently. So achieving high fraud detection rate always inevitably accompanies lots of mis-detections.

Emin Aleskerov et al. (1997) [6] presented CARDWATCH, a database mining system used for credit card fraud detection. The system is based on a neural network learning module,

provides an interface to a variety of commercial databases and has a comfortable graphical user interface. Test results obtained for synthetically generated credit card data and an auto associative neural network model show very successful fraud detection rates. The system is easily extensible and able to work directly on a large variety of commercial databases. The current version of the system was tested on synthetically generated data using an autoassociator with very promising results: a fraud detection rate of 85% and a legal transaction identification rate of 100% were achieved.

Tao Guo et al. (2008) [7] presented work and demonstrate the advantages of the data mining techniques including neural networks, and calculation of confidence. They modeled the sequence of operations in credit card transaction processing using a confidence-based neural network. Receiver operating characteristic (ROC) analysis technology is also introduced to ensure the accuracy and effectiveness of fraud detection. A neural network is initially trained with synthetic data. If an incoming credit card transaction is not accepted by the trained neural network model (NNM) with sufficiently low confidence, it is considered to be fraudulent. They described how confidence value, neural network algorithm and ROC can be combined successfully to perform credit card fraud detection.

Vasilis Aggelis et al. (2006) [8] described that offline internet banking fraud detection system offers many benefits to both bank and customers. Fraud detection system gives added value to e-banking. Especially, nowadays, where fraudsters' attacks are increased considerably in our country, such system differentiate bank owner from other bank competitors. Bank takes lead. Such in-house system implementations, which are set up for customer benefit, are infrequent in local market. Fraud detection system indicates quality of ebanking services. Quality depends on user friendly interface, on a full of electronic transactions portfolio, but also depends on user protection and guarantee.

Hongjun Lu et al. (1996) [9] explained that ONE of the data mining problems is classification. Various classification algorithms have been designed to tackle the problem by researchers in different fields such as mathematical

programming, machine learning, and statistics. Recently, there is a surge of data mining research in the database community. The classification problem is re-examined in the context of large databases. Unlike researchers in other fields, database researchers pay more attention to the issues related to the volume of data. They are also concerned with the effective use of the available database techniques, such as efficient data retrieval mechanisms.

Osama Dandash et al. (2007) [10] stated that internet banking fraud can be performed internally by genuine staff or externally by customers or suppliers. Here, they explained a security analysis of the proposed Internet banking model compared with that of the current existing models used in fraudulent Internet payments detection and prevention. Several modern models in preventing and detecting fraud are evolving and being applied to many banking systems. However, they have no effective detection mechanism to identify legitimate users and trace their unlawful activities. Also they are not secure enough to prevent fraudulent users from performing fraudulent transactions over the Internet. The proposed model facilitates Internet banking Fraud Detection and Prevention (FDP) by applying two new secure mechanisms, Dynamic Key Generation (DKG) and Group Key (GK).]

Yufeng Kou et al. (2007) [11] described that due to the dramatic increase of fraud which results in loss of billions of dollars worldwide each year; several modern techniques in detecting fraud are continually evolved and applied to many business fields. Fraud detection involves monitoring the behavior of populations of users in order to estimate, detect, or avoid undesirable behaviour. Undesirable behavior is a broad term including delinquenc , fraud, intrusion, and account defaulting. This paper presents a survey of cuwent techniques used in credit card fraud detection, telecommunication fraud detection, and computer intrusion detection.

Llew Mason et al. (2001) [12] proposed an architecture that successfully integrates data mining with an e-commerce system. The proposed architecture consists of three main components: Business Data Definition, Customer Interaction, and Analysis, which are connected using data transfer bridges. This

integration effectively solves several major problems associated with horizontal data mining tools including the enormous effort required in pre-processing of the data before it can be used for mining, and making the results of mining actionable. The tight integration between the three components of the architecture allows for automated construction of a data warehouse within the Analysis component. The shared metadata across the three components further simplifies this construction, and, coupled with the rich set of mining algorithms and analysis tools (like visualization, reporting and OLAP) also increases the efficiency of the knowledge discovery process. The tight integration and shared metadata also make it easy to deploy results, effectively closing the loop. Finally presented several challenging problems that need to be addressed for further enhancement of this architecture.

Dr. Bhasin et al. (2006) [13] described that Data Mining is a Competitive Tool in the Banking and Retail Industries. He explained that Data mining, as is evident from its key elements, typically involves the use of predictive modelling, forecasting and descriptive modelling techniques. By using these techniques, an organisation can proactively manage customer retention, identify cross-sell and up-sell opportunities, profile and segment customers, set optimal pricing policies, and objectively measure and rank which suppliers are best suited for their needs.

Monkol Lek et al. (2001) [14] presented research in progress on the development of a data mining prototype for detecting fraud patterns and irregularities in electronic commerce transactions. These transactions are vulnerable to fraud since they are done remotely and often it is quite difficult to verify the legitimacy of customers. This fraud detection software prototype provides auditors and IS management with the appropriate tools for detecting, managing, and controlling fraud patterns, as well as applying the rules derived from these patterns to commercial e-commerce databases. The software prototype provides a graphical user interface for the merchants' fraud investigators to determine the data to be considered in the rule generation process. The following sections provide an outline of the theoretical foundation for fraud pattern detection and continuous

auditing then discuss the design and implementation of the e-commerce fraud detection prototype software and its usefulness to the business, auditors and IS management.

Rajanish Dass (Indian Institute of Management Ahmedabad) [15] explained that data mining can contribute to solving business problems in banking and finance by finding patterns, causalities, and correlations in business information and market prices that are not immediately apparent to managers because the volume data is too large or is generated too quickly to screen by experts. The managers of the Financial Institutions may go a step further to find the sequences, episodes and periodicity of the transaction behaviour of their customers which may help them in actually better segmenting, targeting, acquiring, retaining and maintaining a profitable customer base. Business Intelligence and data mining techniques can also help them in identifying various classes of customers and come up with a class based product and/or pricing approach that may garner better revenue management as well.

Philip Matkovsky et al. (1998) [16] here, evaluated five of the most highly acclaimed data mining tools are so compared on a fraud detection application, with descriptions of their distinctive strengths and weaknesses. Further, he explained that the data mining tool market has become more crowded in recent years, with more than 50 commercial data mining tools, for example, listed at the KDNuggets web site (<http://www.kdnuggets.com>). Rapid introduction of new and upgraded tools is an exciting development, but does create difficulties for potential purchasers trying to assess the capabilities of off-the-shelf tools. Dominant vendors have yet to emerge (though consolidation is widely forecast). According to Data Quest, in 1997, IBM was the data mining software market leader with a 15% share of license revenue, Information Discovery was second with 10%, Unica was third with 9%, and Silicon Graphics was fourth with 6% .

Philip K. Chan et al. (1999) [17] described that large-scale data-mining techniques can improve on the state of the art in commercial practice. Scalable techniques to analyze massive amounts of transaction data that efficiently compute fraud

detectors in a timely manner is an important problem, especially for e-commerce. Besides scalability and efficiency, the fraud-detection task exhibits technical problems that include skewed distributions of training data and non uniform cost per error, both of which have not been widely studied in the knowledge-discovery and data mining community.

Qi Luo et al. (2008) [18] explained that successful Knowledge Discovery and Data Mining applications play an important role in data that have clearly grown to surpass raw human processing abilities. The challenges facing advances in this field are formidable. Some of these challenges include as follows:

- ❖ Develop new mining algorithms for classification, clustering, dependency analysis, and change and deviation detection that scale to large databases.
- ❖ Develop effective means for data sampling, data reduction.
- ❖ Develop schemes capable of mining over non homogenous data sets (including mixtures of multimedia, video, and text modalities).
- ❖ Develop new mining and search algorithms capable of extracting more complex relationships between fields and able to account for structure over the fields (hierarchies, sparse relations).

Xiaoyan Jiang et al. (2009) [19] described e-commerce, related concept of data mining technologies and the realization process of data mining techniques are described. Combine data mining and guide features of e-shopping site, study intelligent shopping guide system based on data mining. Further he explained the use of data mining provide users with intelligent guide and make users easy in the range of goods which not only save user browsing net and considering time, but also provide users with a good proposal to allow the user to get the appropriate selection of goods, is a Kind of high-quality services. Shopping intelligent guide system Based on data mining, is a kind of shopping guide application of e-commerce, intelligence can help users find the appropriate gifts, or can provide constructive suggestions. The ultimate goal is shopping.

Jian-qiang Li et al. (2008) [20] described that due to the development of the information

techniques, the data in electric database increase dramatically and abundantly knowledge is difficult to be acquired in traditional methods due to mass of data, high dimension and strong coupling in electric industrial process. So the data mining techniques are introduced into the power system. The basic concepts and systematic structure of data mining are described. Based on the characteristic of electric data, the data mining technique is introduced into the electric power industry and the feasibility and necessity are discussed. The fault diagnosis and operation optimization based on data mining is researched in detailed. The application of data mining in electric industry can guide the optimal operation based on historical data and improve the economic efficient in power plant.

### 3. Challenges in Detecting Financial Fraud

Many tools and techniques have been developed to fight back Financial Institutions frauds. The current tools and techniques used in financial institutions are static in nature, independent, simple and are based on mental model, experiences and tradition. With the fraudsters using a variety of new methods to commit fraud, heterogeneous profiles of cardholders, fraudsters, merchants and financial institutions and their changing pattern of operations pose increasing challenges in detection of frauds.

So, in my dissertation work, I will develop a predictive model, which will provide an aid to detect financial frauds and will help in making "Financial System" more robust.

### 4. Objectives Of The Study

Objectives are most important part of every research. Without objectives it is like a direction less ship moving without a compass. The research regarding the proposal is going to be done two fold – Business Issues and Technical research.

This work during the research period aims to address few of the listed issues by achieving the following objectives:

- ❖ To identify the various Factors influencing the “Financial Frauds Detection” .
- ❖ To develop a Predictive Model, which will provide an aid to detect Financial Frauds.
- ❖ Perform Evaluation of the proposed model by comparing with that of the current existing models.
- ❖ A Comprehensive Survey of Data Mining-Based “Financial Fraud” Detection Research

## 5. Research Methodology

### Research:

Research is a word that refers to a search of knowledge. One can also define research as a scientific and systematic search of pertinent information. In fact, Research is an art of scientific investigation.

### Research Design:

It is the blue print of the course of action on which the research study will be based. The basic approach for this research study will be:

- ❖ Literature Survey
- ❖ To identify the various factors influencing the “Financial Frauds Detection” .
- ❖ To develop a Predictive Model, which will provide an aid to detect financial frauds.
- ❖ Perform Evaluation of the proposed model by comparing with that of the current existing models.
- ❖ A Comprehensive Survey of Data Mining-Based “Financial Fraud” Detection Research.
- ❖ Verification and Validation of results.
- ❖ Publication of technical papers.
- ❖ Improvement by the feedback of survey and technical publications.
- ❖ Compilation of final result, the theories developed/evolved & final preparation of thesis.

## 6. Tools/Techniques Used

Existing Data Mining technique like Clustering, Classification etc. will be used.

## References

- [1] Mieke, “Data Mining for Fraud Detection: Toward an Improvement on Internal Control Systems? ”,proceeding of International Research Symposium on Accounting Information Systems in Milwaukee, USA , (2006).
- [2] Shirley Inscoc, “Fraud Prevention Strategies For Internet Banking”, retrieved from the Publication Of Bits Fraud Reduction Steering Committee (U.S.A.) , (2003).
- [3] Shamik Sural ,“Credit Card Fraud Detection Using Hidden Markov Model” appears in IEEE Transactions On Dependable And Secure Computing, Vol. - 5, pp. 37-48 (2008).
- [4] Philip K. Chan, “Distributed Data Mining in Credit Card Fraud Detection” , appears in “Intelligent Systems and their Applications, IEEE”, Vol. -14, pp. 67-74 , (1999).
- [5] M.J. Kim , “A Neural Classifier with Fraud Density Map for Effective Credit Card Fraud Detection,” Proc. Int’l Conf. Intelligent Data Eng. and Automated Learning, pp. 378-383, (2002).
- [6] E. Aleskerov, “CARDWATCH: A Neural Network Based Database Mining System for Credit Card Fraud Detection,” Proceedings of IEEE/IAFE: Computational Intelligence for Financial Eng., pp. 220-226, (1997).
- [7] Tao Guo, “Neural Data Mining For Credit Card Fraud Detection” , proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, Vol. - 7, pp. 3630-3634 (2008).
- [8] Vasilis Aggelis , “Offline Internet Banking Fraud Detection” proceedings of the IEEE First International Conference on Availability, Reliability and Security (ARES’06) , pp. 2 (2006).
- [9] Hongjun Lu, “Effective Data Mining Using Neural Networks”, appears in IEEE Transactions On Knowledge And Data Engineering, Vol. - 8, pp. 957-961 ( 1996).
- [10] Osama Dandash , “Security Analysis for Internet Banking Models”, Proceedings of the Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking & Parallel/Distributed Computing, Vol. – 3, pp. 1141-1146 ( 2007).

- [11] Yufeng Kou, "Survey of Fraud Detection Techniques", Proceedings of the IEEE International Conference on Networking, Sensing & Control Taipei, Taiwan, Vol. – 2, pp. 749-754 (2004).
- [12] Llew Mason, "Integrating E-Commerce and Datamining: Architecture and Challenges", proceedings of IEEE International Conference on Integrating E-Commerce and Datamining , pp. 27-34 (2001)
- [13] Dr. Bhasin , "Data Mining: A Competitive Tool in the Banking and Retail Industries" retrieved from [icai.org/resource\\_file/9935588-594.pdf](http://icai.org/resource_file/9935588-594.pdf), pp. 588-594 (2006).
- [14] Monkol Lek, "Data Mining Prototype For Detecting E-Commerce Fraud" ,Proceeding of European Conference on Information Systems, Bled Slovenia ( 2001).
- [15] Rajanish Dass, "Data Mining In Banking And Finance: A Note For Bankers" retrieved from <http://www.iimahd.ernet.in> (2008)
- [16] Philip Matkovsky , " An Evaluation of High-End Data Mining Tools for Fraud Detection" , Proceedings of the 1998 IEEE International Conference on Systems, Man, and Cybernetics, Vol. – 3 , pp. 2836-2841 (1998).
- [17] Philip K. Chan , "Distributed Data Mining in Credit Card Fraud Detection", published in IEEE Intelligent Systems, Vol. - 14, pp. 67-74 (1999).
- [18] Qi Luo , "Advancing Knowledge Discovery and Data Mining", appears in First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008), Adelaide, Australia, pp. 6-7 ( 2008).
- [19] Xiaoyan Jiang, "Study on Intelligent E-Shopping System Based on Data Mining " , proceedings of International Conference on e-Business and Information system Security (EBISS 2009) in Wuhan, China, pp. 1-4 (2009).
- [20] Jian-qiang Li , "Research and Application of Data Mining Technique in Power Plant" , proceedings of International Symposium on Computational Intelligence and Design (ISCID 2008) in Wuhan, China, Vol. – 2, pp. 250-253 (2008).