Data Mining in Credit Card Frauds: An Overview

Vidhi Khurana*
Ramandeep Kaur**

Abstract

With the increasing awareness of customers amongst plastic money and internet banking, the number of frauds in transactions have also emerged. In order to detect these frauds, various data mining techniques can be applied. Financial Fraud Detection(FFD) has been a major concern among the leading organizations and the banks. Hence a framework has been proposed so as to detect the fraud in the early stages as well as forecast which transactions are prone to fraudulent activities. This paper reviews the previous research conducted by the leading researchers in their areas with a focus on credit card fraud detection and prevention using data mining approaches.

Keywords: Credit Card, Data mining, Financial Fraud Detection, Fraud Prevention

I. Introduction

Data Mining has been a very vibrant and upcoming field in all the prevailing industries. From a small and independent IT firm, banking organizations, convenience stores, to leading industries, the implications of data mining can be felt. It may be defined as the logical process of extraction of hidden and interesting information from the huge databases[1]. It is a methodology of mining of knowledge from the given data sources. Hence may aid in Knowledge discovery.

Data Mining can be categorized into three identifiable steps: (i) Exploration (ii) Pattern Identification and (iii) Deployment. On the basis of the kind of data to be mined, there are two categories of functions involved in Data Mining, viz.,Descriptive and Classification and Prediction[27]. Mined knowledge can be used in various domains like: fraud detection, production control, science exploration and market analysis. Financial Fraud Detection(FFD) is of high priority at present. Data Mining help in detection of financial frauds by analysing patterns hidden in the transaction data [8]. FFD is vital for the prevention of the often devastating consequences of financial

Vidhi Khurana*
Pursuing MCA from Institute of Information Technology & Management
Ramandeep Kaur**
Assistant Professor
Institute of Information Technology & Management

fraud. According to the 2008 Javelin fraud survey report, victims who detected the fraud within 24 hours were defrauded for an average of $428. Victims who did not discover the fraud up to a month later suffered an average loss of $572[6].

Financial Fraud can be classified into various categories as depicted in Table 1.

Bank Frauds are very devastating and have a severe repercussion on the organizations. It comprises of all the fraudulent activities involved in the banking sector. It is broadly classified into two categories: i) External: here the assassin are outside the bank ii) Internal: here bank personnel commits the fraud. Card fraud, mortgage fraud and money laundering are few instances of bank fraud. Insurance Fraud is an activity of obtaining fraudulent outcomes from an insurance company[8]. It can be committed by consumer, broker and agents, insurance company employees and others. Automobile fraud and healthcare fraud are in top category of this classification [2,13]. Securities and commodities fraud is a type of white collar crime that can be committed by individuals. [investopedia] The types of misrepresentation involved in this crime include providing false information, withholding key information, offering bad advice, and offering or acting on inside information. Other related financial frauds include corporate and mass marketing fraud. Mass communication media such as telephones and internets are used in mass market fraud [14]. Mass-marketing fraud schemes generally fall into two broad

Table 1: Classification for Financial Fraud based on FBI, 2007

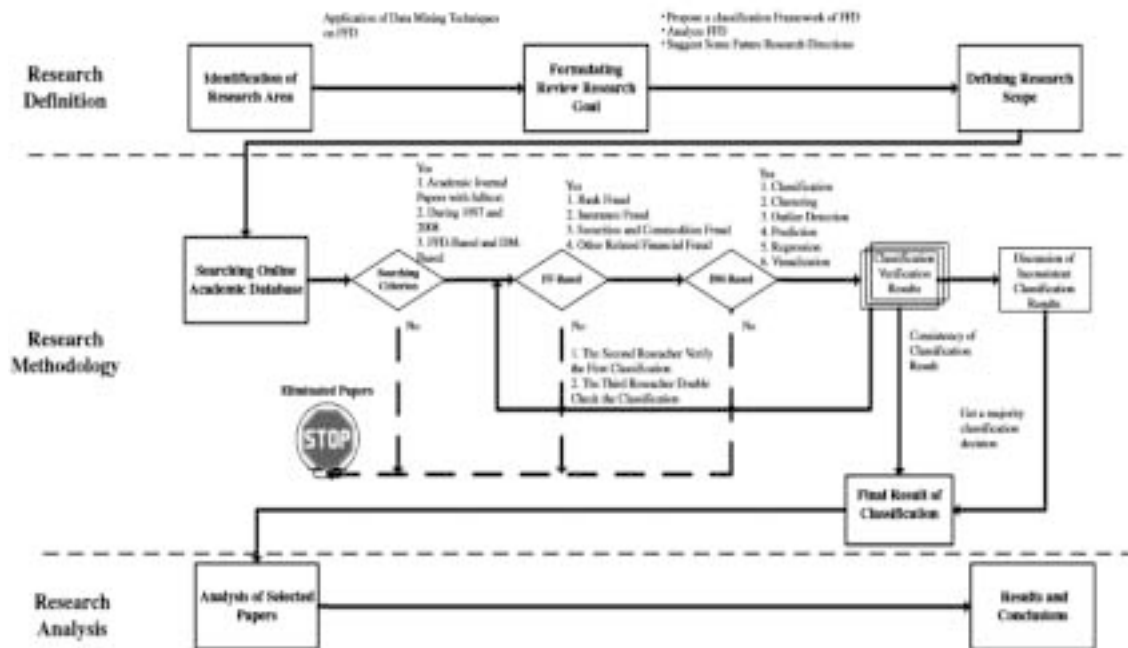| Financial fraud based categories	Fraudulent activities
Bank fraud	Mortgage fraud, Asset forfeiture/money laundering
Insurance fraud	Healthcare fraud, Insurance fraud
Securities and commodities fraud	Securities and commodities fraud
Other related financial fraud	Corporate fraud, Mass marketing fraud

categories: (1) schemes that defraud numerous victims out of comparatively small amounts, such as several hundred dollars, per victim; and (2) schemes that defraud comparatively less numerous victims out of large amounts, such as thousands or millions of dollars per victim.

The objective of this paper is to describe generalized architecture of Financial Fraud detection as well as the techniques of preventing the frauds. Special focus has been laid on Credit Card Financial Frauds. The remainder of the paper is divided in the following sections: Section II deals with a detailed review of literature. Section III deals with a framework for Financial Fraud Detection. Section IV deals with Fraud detection in Credit Cards. Section V gives a concluding remark on the review carried out.

II. Literature Review

Vast research has been carried out in the field of data mining and fraud detection but the challenge in dealing with the increasing number of frauds remains the same. Data mining enables a user to seek valuable information and their interesting relationships [24]. A number of data mining techniques are available such as decision trees, neural networks (NN), Bayesian belief networks, case based reasoning, fuzzy rule-based reasoning, hybrid methods, logistic regression, text mining, feature selection etc. Financial fraud is a serious problem worldwide and more so in fast growing countries like China[21]. According to Kirkos et al. [7], some estimates stated that fraud cost US business more than $400 billion annually. An innovative fraud detection mechanism was developed on the basis of

Fig 1: Methodological Framework for research[8]

Table 2: Research on data mining techniques in FFD[8]

| FF-based categories	Fraudulent activities	Data mining application class	Data mining techniques
Bank fraud	Credit card fraud	Classification	Ada boost algorithm, decision trees, CART, RIPPER, Bayesian Belief Network, Neural networks, discriminant analysis
			K-nearest neighbor, logistic model, discriminant analysis, Naïve Bayes, neural networks, decision trees
			Support vector machine, evolutionary algorithms
		Clustering	Hidden Markov Model
			Self-organizing map
Insurance fraud	Money laundering	Classification	Network analysis
	Crop insurance fraud	Regression	Yield-switching model
			Logistic model, probit model
	Healthcare insurance fraud	Classification	Association rule
			Polymorphous (M-of-N) logic
			Self-organizing map
		Visualization	Visualization
		Outlier detection	Discounting learning algorithm
	Automobile insurance fraud	Classification	Logistic model
			Neural networks
			Principal component analysis of RIDIT(PRIDIT)
			Logistic model
			Logistic model, decision trees, neural networks, support vector machine, K-nearest neighbor, Naïve Bayes, Bayesian belief network
			Fuzzy logic
			Logistic model
			Logistic model, Bayesian belief network
			Self-organizing map
			Naïve Bayes
		Prediction	Evolutionary algorithm
			Logistic model
		Regression	Probit model
			Logistic model
			Probit model
Other related financial fraud	Corporate fraud	Classification	Neural networks, decision trees, Bayesian belief network
			Multicriteria decision aid (MCDA), UTilite's Additives DIscriminantes (UTADIS)
			Evolutionary algorithm
			Fuzzy logic
			Neural networks
			Neural networks, logistic model
			Logistic model
			CART
			Decision trees, neural networks, Bayesian belief network, K-nearest neighbor, RIPPER, support vector machine, stacking variant methodology
		Clustering	Naïve Bayes
		Prediction	Neural networks
		Regression	Logistic model
			Logistic model
			Logistic model

Zipf's Law with a purpose of assisting the auditors in reviewing the bulbous volumes of datasets while at the same time intending to identify any potential fraud records[26]. The study of Bolton and Hand [22] provides a very good summary of literature on fraud detection problems. Some researchers used methods such as ID3 decision tree, Bayesian belief, back-propagation Neural Network to detect and report the financial frauds[7,12]. Fuzzy logic based techniques based on soft computing were also incorporated to deal with the frauds [15, 16]. Panigrahi et. al.[25] suggested a four component fraud detection solution with an idea to determine a set of suspicious transactions and then predict the frauds by running Bayesian learning algorithm. Further, a set of fuzzy association rules were extracted from a data set containing genuine and fraudulent transactions w.r.t credit cards to analyze and compare the frauds. It was

suggested that novel combination of meta-heuristic approaches, namely the genetic algorithms and the scatter search when applied to real time data, may yield fraudulent transactions which are classified correctly[5]. Padhy et al (2012) provided a detailed survey of data mining applications and its feature scope. A number of researchers also discussed the application of data mining in anomaly detection [17, 19, 20, 23].

III. Framework of FFD

Methodological framework for review is a three step process: i) Research Definition ii) Research Methodology and iii) research analysis. Research definition is a phase mining technique.Goal of the research is to create a classification framework for data mining techniques applicable to FFD. Research scope here is the literature comprising application of data
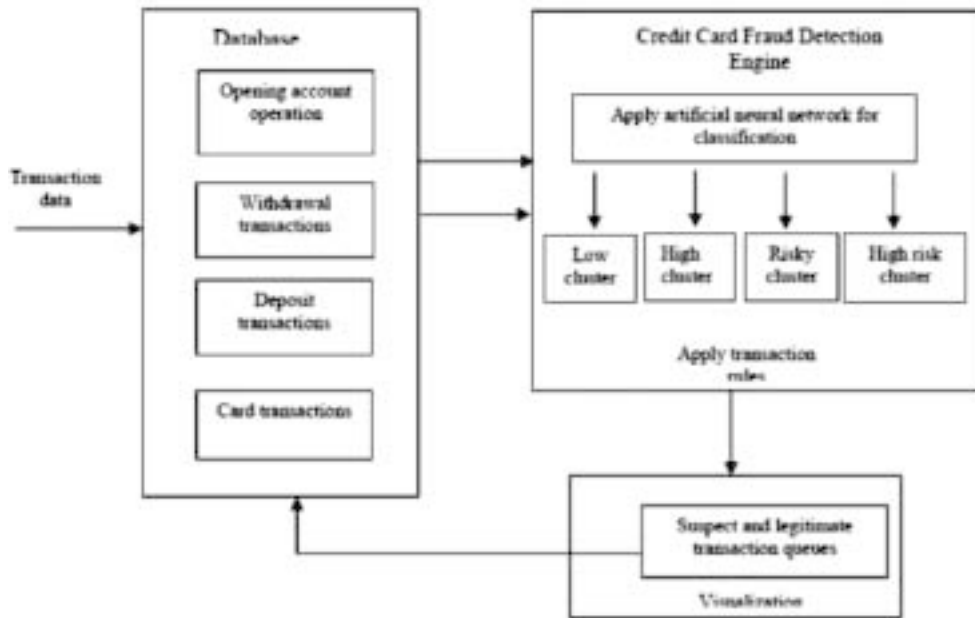
Fig 2: Architecture for Credit Card Fraud Detection[10]

mining techniques on FFD published from 1997 to 2008. Phase to is the research methodology. In this phase the online academic databases are searched for FFD. In each iteration these databses are filtered out to obtain the articles that were published in the academic journals(1997-2008) and should present data mining techniques along with application to FFD. A detailed process for FFD has been depicted in Fig 1. All the obtained articles consistency are verified and final result of classification is passed to third phase of the framework. Research analysis phase includes the analysis of the selected where the topic or area of research is identified for formulating the research goal and definingg the scope of the performed research. Here identified research area: the academic reserch on FFD that applies data papers to formulate conclusion and results based on the analysis of paper[8].

IV. Fraud Detection in Credit Cards

Credit card fraud is sort of identity theft, where an unauthorized person makes fraudulent transactions. It can be classified into: Application fraud and Behaviour fraud. Application fraud occurs when a fraudster gets a credit card issued from companies by providing false information[3]. It is very serious because victim may learn about the fraud too late.

Various data mining techniques used in credit card fraud detection are logistic regression, support vector machine and random forests. Credit card fraud detection scheme scans all the transactions inclusive of fraudulent transactions[10]. Data obtained from the data warehouse is divided into various dataset. Dataset comprises of primary attributes (account number, sale, purchase, date name and many others) and derived attributes (for instance transactions grouped monthly). Derived attributes are not precise, which causes approximation of results and therefore not accurate information. Therefore derived attributes are limitation to the credit card fraud detection scheme. The implemented architecture [Fig2] comprises of database interface subsystem and credit card fraud (CCF) detection engine. The former enables the reading of transactions, i.e. it acts as an interface for banking software.

In the CCF detection subsystem, the host server checks every transaction rendered to it using neural networks and transactions business rules.

V. Conclusion

Data mining gained weightage in the areas where finding the patterns, forecasting, discovery of knowledge etc., is required and becomes obligatory in

different industrial domains. Various techniques and algorithms such as feature selection, classification, memory based reasoning, clustering etc., aids in fraud detection in areanas of insurance, financial frauds etc.. Financial sector has been majory affected ny fradulent activities due to increase in conversion rate of non- internet users to internet users. A detailed review was conducted to understand how these financial frauds can be detected and avoided using data mining techniques. A special reference to Credit card frauds was mentioned to understand the architecture of credit card fraud detection.

References

1. Bose, R.K. Mahapatra, "Business data mining — a machine learning perspective", *Information Management,* vol.39, no.3, pp.211–225, 2001.

2. Coalition against Insurance Fraud, "Learn about fraud," http://www.insurancefraud.org/learn_about_fraud.htm, Last accessed 23 January 2017.

3. Credit Card Fraud: An Overview, Legal Information Institute, web: https://www.law.cornell.edu/wex/credit_card_fraud, Last Accessed: 23 January 2017.

4. D. Sánchez, M. A. Vila, L. Cerda, and J. M. Serrano, "Association rules applied to credit card fraud detection," *Expert Syst. Appl.*, vol. 36, no. 2 PART 2, pp. 3630–3640, 2009.

5. E. Duman and M. H. Ozcelik, "Detecting credit card fraud by genetic algorithm and scatter search," *Expert Syst. Appl.*, vol. 38, no. 10, pp. 13057–13063, 2011.

6. E. Joyner, "Enterprisewide Fraud Management", *Banking, Financial Services and Insurance*, Paper 029, 2011

7. E. Kirkos, C. Spathis and Y. Manolopoulos, "Data mining techniques for the detection of fraudulent financial statement", *Expert Systems with Applications*, vol.32, pp.995–1003, 2007.

8. E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert Syst. Appl.*, vol. 36, no. 2 PART 2, pp. 2592–2602, 2009.

9. FBI, Federal Bureau of Investigation, Financial Crimes Report to the Public Fiscal Year, Department of Justice, United States, 2007, http://www.fbi.gov/publications/financial/fcs_report2007/financial_crime_2007.htm.

10. F. N. Ogwueleka, "Data Mining Application In Credit Card Fraud Detection System", Journal of Engineering Science and Technology, vol. 6, no. 3, pp.311 – 322, 2011.

11. F.N. Ogwueleka, and H.C. Inyiama, "Credit card fraud detection using artificial neural networks with a rule-based component', The IUP Journal of Science and Technology, vol.5, no.1, pp.40-47, 2009.

12. J.E. Sohl and A.R. Venkatachalam, "A neural network approach to forecasting model Selection", *Information & Management*, vol.29, no.6, pp. 297–303, 1995.

13. J.L. Kaminski, "Insurance Fraud", OLR Research Report, http://www.cga.ct.gov/2005/rpt/2005-R-0025.htm. 2004

14. "Mass Marketing Fraud(MMF)", Strategy, Policy & Training Unit, Department of Justice, http://www.justice.gov/criminal-fraud/mass-marketing-fraud, Last Accessed: 23 January 2017.

15. M. Delgado, D. Sa´nchez, and M.A. Vila, "Fuzzy cardinality based evaluation of quantified sentences", *International Journal of Approximate Reasoning*, vol.23, pp.23–66, 2000.

16. M. Delgado, N. Marý´n, D. Sa´nchez, and M.A.Vila, "Fuzzy association rules: General model and applications", *IEEE Transactions on Fuzzy Systems*, vol.11, no.2, pp.214–225, 2003.

17. N. Kaur, "Survey paper on Data Mining techniques of Intrusion Detection", *International Journal of Science, Engineering and Technology Research*, vol. 2, no. 4, pp. 799-804, 2013.

18. N. Padhy, P. Mishra, and R. Panigrahi, "The Survey of Data Mining Applications and Feature Scope", *International Journal of Computer Science, Engineering and Information Technology*, vol. 2, no. 3,pp. 43-58, 2012.

19. P. Dokas, L. Ertoz, V. Kumar, A. Lazarevic, J. Srivastava and P.N.Tan, "Data mining for network intrusion detection", *Proceedings of NSF Workshop on Next Generation Data Mining*, pp. 21-30, 2002.

20. P. Garcia-Teodoro, J. Diaz-Verdejo, G. Maciá-Fernández and E. Vázquez, "Anomaly-based network intrusion detection: Techniques, systems and challenges", *Computers and security*, vol.28, no. 1, pp. 18-28, 2009.

21. P. Ravisankar, V. Ravi, G. Raghava Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," ", *Decision Support Systems*, vol. 50, no. 2, pp. 491–500, 2011.

22. R. Bolton, and D. Hand, 'Statistical fraud detection: A review", *Statistical Science*, vol.17, pp.235–255, 2002.

23. S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," *Procedia Comput. Sci.*, vol. 60, no. 1, pp. 708–713, 2015.

24. S. H. Weiss, and N. Indurkhya, "Predictive Data Mining: A Practical Guide", , *CA: Morgan Kaufmann Publishers*, 1998.

25. S. Panigrahi, A. Kundu, S. Sural, and A. Majumdar, "Credit card fraud detection a fusion approach using Dempster–Shafer theory and bayesian learning", *Information Fusion*, pp.354–363, 2009.

26. S.-M. Huang, D.C. Yen, L.-W. Yang and J.-S. Hua, "An investigation of Zipf's Law for fraud Detection", *Decision Support Systems*, vol.46, no. 1, pp. 70–83, 2008.

27. Tutorialspoint, "Data mining Tasks", http://www.tutorialspoint.com/ data_mining/ dm_tasks.htm, Last Accessed: 24 January 2017.