

# A Study of Fraud Identification in Online Payment Instruments Using Data Mining Techniques

Assistant Professor R. Ramya

Department of Computer Science  
Adthiyaman Arts and Science College for Women  
Uthangarai, Krishnagiri, India  
ramyar295@gmail.com

**Abstract-** The aim of the present work is to survey and analyze the use of electronic payment instruments and fraud detection on banks across the country using statistics and information retrieved from the Central Bank and the data mining techniques. For this purpose, initially, according to the volume of the transactions carried out and using the K-Means algorithm, a label was dedicated to any record. Viewing profile data mining solution that valuable but hidden in mass volume of online transactions, gives valuable information related to this criminal process loses. We analyze some of the recent approaches and architectures where data mining has been applied in the fields of e payment systems. In this study we limit our discussions to data mining in the Context of e-payment systems.

**Keywords-** K means algorithm, clustering technique, kohonen methods, Extrinsic technique, CART decision tree, E-payment, fuzzy computing, POS, Bayes algorithm etc.

## I. INTRODUCTION

In the recent years, information and communication technology (ICT) has made a significant progress. Over time, the role of information technology (IT) in all industries has become more prominent, and the banking industry has not been an exception to this. There are a variety of tools, such as electronic banking, electronic money, payment cards, etc. In general, traditional banking has changed very serious in the meantime, along with various benefits such as speed, efficiency, security, application diversity, anonymity and the use of these tools, there are also faced with challenges Because it has brought the characteristics of these new instruments, a new approach to criminal activity becomes attractive.

Including money laundering, organized crime is due to the complex nature and expertise of those involved in this process, the facilities and opportunities to take the most advantage of these new tools. The Different methods of this generation include abstractions, aggregations, summarizations, and characterizations of data [1].

These forms, in turn, are the results of applying sophisticated modeling techniques from the diverse fields of statistics, artificial intelligence, and database Management and computer graphics. In summary, it is little surprise that e-payment transaction is the killer application for data mining ([3]).

## II. DATA MINING TECHNIQUES USED IN CREDIT CARD FRAUD DETECTION

### 1. Clustering

Among various data mining techniques, Clustering is a data mining technique that makes significant or useful cluster of object(s) that have similar characteristic using automatic technique. Apart from classification, clustering technique also defines the classes and put objects in them, although in classification, object(s) are assigned into predefined classes [4] Farokhi et al. have investigated the application of DM approach in the implementation of the Customer relationship management system. The information gathered from Point of Sales (POS) in one of the Iranian private banks was used in this research work. They used the two methods K Means and Kohonen to detect the most profitable customers, and clustered them into four segments [20]. Fade Ngami & Moat tar have suggested a novel DM model for credit card fraud detection. Their proposed model was to consider the feature selection and the decision cost for accuracy Enhancement of credit card fraud detection.

### 2. Classification of clustering algorithms

Classification may refer as gathering of different types of clustering algorithms. Clustering algorithms may also vary based on whether they produce overlapping or non-overlapping clusters. Non-overlapping clusters can be viewed as Extrinsic Clusters or Intrinsic Clusters. Extrinsic technique/algorithms categorize the items to support in the classification process. Clustering algorithms are the traditional classification supervised

learning algorithms that uses a special input training set. On the other side, intrinsic algorithms/techniques do not use a priori category labels but depend only on the adjacency matrix containing the distance objects.

### 3. Clustering with Neural Networks

Neural Networks (NNs) that use unsupervised learning attempt to find features in the data that characterize the desired output. They look for clusters of like data. These types of NNs are often called Self-Organizing Neural Networks (SONN). There are two types of unsupervised learning: noncompetitive and competitive. With the noncompetitive learning, the weight between two nodes is changed to be proportional to both output values. That is,  $\Delta w = \eta y_1 y_2$  [7] With competitive learning, nodes are allowed to compete. This approach usually assumes a two-layer NN in which all nodes from one layer are connected to all nodes in the other layer. Thus, this provides a grouping of topless together into a cluster.

### 4. Clustering Large Databases

Clustering techniques should be able to adapt as the database changes. A clustering algorithm should have:

- Require no more than one scan of the database.
- It should have the ability to provide status. This is sometimes referred to as the ability to be online.
- It should be suspend able, stoppable and resume able.
- It should process each tuple only once.

## III. CLASSIFICATION

Classification is the task of assigning objects to their respective categories. Classification is a supervised method in DM. It means that the purpose of classification is clear, and there are Variables whose values are predictable from the values of other variables and data. The goal of supervised predictive models is to find a model or mapping that will correctly associate the inputs with the targets. In this research work, the classification algorithm was applied to analyze and investigate the factors affecting the number and amount of transactions conducted using the E payment instruments.

For this purpose, the CART decision tree was selected. The CART decision tree is a binary recursive partitioning procedure capable of processing Continuous and nominal attributes as targets and predictors. The data is handled in its raw form; no binning is required or recommended. Beginning with the root node, the data is split into two Children and each child are, in turn, split into grandchildren.

Trees are grown to a maximal size Without the use of a stopping rule, essentially the tree-growing process stops when no further splits are possible due to the lack of

data. The proposed process for specifying the factors Affecting transactions conducted with the E-payment.

## IV. ROLE OF STATISTICS IN DATA MINING

Extracting causal information from data is often one of the principal goals of data mining and more generally of statistical inference. Data for decades; thus DM has actually existed from the time large-scale statistical modeling has been made possible.

Statisticians consider the causal relationship between the dependent variables and independent variables as proposed by the user (usually the domain expert), and try to capture the degree and nature of dependence between the variables. Modeling methods include Simple linear regression, multiple regressions, and nonlinear regression. Such models are often parameter driven and are arrived at after solving attendant Optimization models.

## V. E-PAYMENTS AND DATA MINING

In this section, we survey articles that are very specific to DM implementations in e-payment Systems.

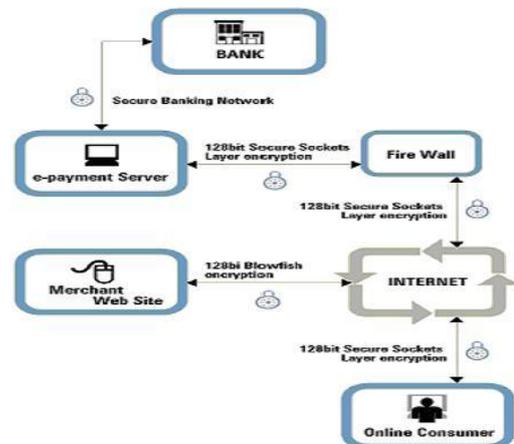


Fig.1 E-Payment System.

The salient applications of DM techniques are presented first. Later in this section, architecture and data collection issues are discussed.

The following are the factors which affect the user acceptability of e-payment-

- Awareness.
- Education / knowledge.
- Accessibility.
- Convenience.
- Security issue.
- Atomicity.

### 1. Money Laundering

Washing the proceeds of crime by banks or financial institutions for offenders is a major success US federal agency stated that more than 300 million dollars annually worldwide money laundering is done. Between 40 and 80 million dollars of this amount is carried out by drug trafficking in the United States. FATF to combat money laundering, which is estimated at 85 million dollars annually from drug trafficking to money laundering United Nations Office on Drugs and Crime said in its report that in 2009 money-laundering the proceeds of criminal activities was about 1.6 trillion \$(The Financial Action Task Force, 2014).

### 2. DM in Recommendation Systems

Systems have also been developed to keep the customers automatically informed of important Events of interest to them. The article by [9] discusses an intelligent framework called PENS that has the ability to not only notify customers of events, but also to predict events and event classes that are likely to be triggered by customers. The event notification system in PENS has the following components Event manager, event channel manager, registries, and proxy manager.

The event-prediction system is based on association rule-mining and clustering algorithms. The PENS system is used to actively help an e-commerce service provider to forecast the demand of product categories better. Data mining has also been applied in detecting how customers may respond to promotional offers made by a credit Card Company ([11]). Techniques including fuzzy computing and interval computing are used to generate if-then-else rules.

### 3. DM Applied to Retail Payment System

They share their experience in terms of lessons that they learnt. They classify the important issues in practical studies, into two categories

- Business-related.
- Technology related.

We now summarize their findings on the technical issues here. Collecting data at the right level of abstraction is very important. Web server Logs were originally meant for debugging the server software. Hence they convey very little useful information on customer-related transactions.

Approaches including seasoning the web logs may yield better results. A preferred alternative would be having the application server itself log the user related activities. This is certainly going to be richer in semantics compared to the state less web logs, and is easier to maintain Compared to state-full web logs.

### 4. POS Data Transfer and Audit using online Payment

POS Data Transfer and Audit refers to the collection of point-of-sale (POS) transaction data from connected stores. POS transaction data is received by the repository in the form of transaction logs (TLOGs), which are processed by the POS Inbound Processing Engine (PIPE). Once the POS transactions are received, PIPE performs several processing activities, such as master data verification, sales auditing, POS transaction data summaries, and sending of POS transaction data to follow-on applications. Additionally, POS transaction data is made available to consuming applications through virtual data models and individual SAP HANA views.

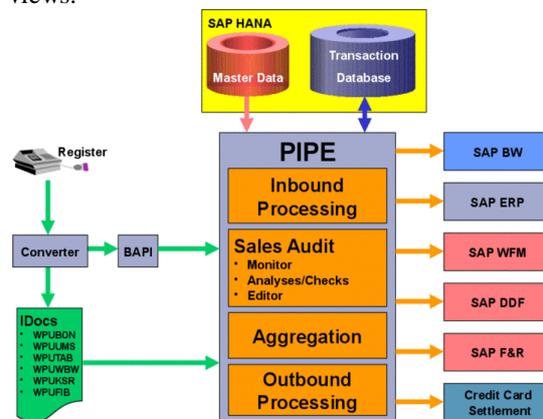


Fig.2 Online Payment process between two process.

## VI. FRAUD DETECTION TOOLS

Fraud detection can be classified as ‘supervised’ or ‘unsupervised’. Supervised methods uses database of known fraudulent/legitimate cases from which model has been constructed which yields a suspicion score for new cases when some different score for which past behavior has been analyzed. (Han, 1981; McLachlan, 1992)demonstrated traditional statistical classification methods such as linear discriminate analysis and logistic discrimination that have proved to be effective tools for many applications but more powerful tools (Ripley, 1996;Hand, 1997; Webb, 1999) such as neural networks that have also been extensively applied. Supervised learning algorithms are Rule-based methods that produce classifiers using rules of the form.

If {certain conditions}, Then {a consequent} Some of the examples of such algorithms include BAYES (Clark and Niblett, 1989), FOIL (Quinlan 1990) and RIPPER (Cohen 1995), Tree-based algorithms such as CART (Brieman et al, 1984) and C4.5 (Quinlan 1993) that produce classifiers of a similar form. Combinations of some or all of these algorithms can be used as meta-

learning algorithms which improve prediction in fraud detection that is, Chan et al (1999).

When building a supervised tool for fraud detection major consideration includes those of uneven class sizes and different costs for different types of misclassification [13]. On the other side, there must be some consideration for the costs of investigating observations and corresponding benefits of identifying fraud. Unsupervised methods, are used when there are no prior sets of legitimate and fraudulent observations available. Some techniques that are being employed are usually a combination of both profiling and outlier detection methods.

There exists model in which baseline distribution that represents normal behavior, then attempt to detect remark that show greatest different behavior from this existing norm. Benford's law (Hill 1996) said that the distribution of the first significant digits of numbers drawn from a vast range of arbitrary distributions would have some certain form. Until currently, this law was regarded as merely a mathematical curiosity with no noticeable useful application. Nevertheless, Nigrini and Mittermaier (1997) and Nigrini (1999) show that Benford's law can be used to detect fraud in accounting data.

The assertion behind fraud detection using tools such as Benford's law is that fabricating data which is conventional to Benford's law is difficult [14]. Fraudster(s) familiarize new prevention and detection measures as well so that fraud detection needs to be adaptive over time. Still, legitimate account users may regularly change their behavior over a longer period of time and it is significant to avoid false alarms.

## VII. CONCLUSION

Efficient and well-organized credit card fraud detection system is an greatest requirement for any card issuing bank. Credit card fraud detection has drawn quite a lot of interest from the research community and a number of techniques have been proposed to counter/identify credit card fraud. The Fuzzy Darwinian fraud detection systems improve the system accuracy, while neural network improve the method time to detect particular fraud termed as suspicious activity. Since the Fraud detection rate of Fuzzy Darwinian fraud detection systems in terms of true positive is 100% and shows good results in detecting fraudulent transactions on the other side, the neural network based CARDWATCH shows good accuracy in fraud detection and Processing Speed is also high but it is limited to one-network per customer. The Fraud detection rate of using Clustering is very compare to other methods. As usage of credit cards become more and more popular in every field of

the daily life, credit card fraud has become much more rampant. Therefore, there is a need for improving security of the financial transaction systems in an automatic and effective way, by building an accurate and efficient credit card fraud detection system. As, it is the key task for the financial institutions. In this study, we gather various methods that were used to build fraud detecting models. Currently, due to the security issues, only a few approaches for credit card detection are available in public. In between them, neural networks approach is a very popular tool. Though, it is difficult to implement because of lack of available data set.

## REFERENCES

- [1]. Fadaei Noghani, F., & Moattar, M. (2017). Ensemble classification and extended feature selection for credit card fraud detection. *Journal of AI and Data Mining*, vol. 5, no. 2, pp. 235-243.
- [2]. Kudyba, S. & Lawrence, K. (2008). Enhancing information management through data mining analytics to increase product sales in an e-commerce platform. *Int. J. Electronic Marketing and Retailing*, vol. 2, no. 2, pp. 97-104.
- [3]. Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining. *Procedia Computer Science*, vol. 85, no. Cms, pp. 78-85. <http://dx.doi.org/10.1016/j.procs.2016.05.180>
- [4]. Dorronsoro, Ginel, Sgnchez and Cruz. Neural fraud detection in credit card operations. *Neural Networks, IEEE Transactions*. Volume: 8, Issue: 4: 827-834, 1997.
- [5]. Mirjana Pejic-Bach, Profiling intelligent systems applications in fraud detection and prevention: survey of research articles, 2010
- [6]. International Conference on Intelligent Systems, Modeling and Simulation. Prabin Kumar Panigrahi, a Framework for Discovering Internal Financial Fraud using Analytics, International Conference on Communication Systems and Network Technologies 2011
- [7]. Raghavendra Patidar, Lokesh Sharma, "Credit Card Fraud Detection Using Neural Network", *International Journal of Soft Computing and*
- [8]. Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-NCAI2011, and June 2011 1S. Benson Edwin Raj, 2A. Annie Portia Analysis on Credit Card Fraud Detection Methods ICCET2011, 18th & 19th March, 2011 978-1-4244-9394-4/11/\$26.00 ©2011 IEEE] 152
- [9]. Sahin, Y., Duman, E.: An overview of business domains where fraud can take place, and a survey

of various fraud detection techniques. In: Proceedings of the 1st International Symposium on Computing in Science and Engineering, Aydin, Turkey (2010).

- [10].Tej Paul Bhatla, Vikram Prabhu & Amit Dua “Understanding Credit Card Frauds,” 2003.
- [11].V. Filippov L. Mukhanov B. Shchukin Credit Card Fraud Detection System.
- [12].Y. Sahin, E. Duman “Detecting Credit Card Fraud by ANN and Logistic Regression” 2011