iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** II **Month of publication:** February 2016

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Volume 4 Issue II, February 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Credit Card Fraud Detection Analysis

Bendalam Vijay[1], Jallu Swathi[2]
[1,2]*Assistant Professor, CSE Dept. , AITAM Tekkali.*

Abstract--Due to the rise and rapid growth of E-Commerce, use of credit cards for online purchases has dramatically increased and it caused an explosion in the credit card fraud. As credit card becomes the most popular mode of payment for both online as well as regular purchase, cases of fraud associated with it are also rising. In real life, fraudulent transactions are scattered with genuine transactions and simple pattern matching techniques are not often sufficient to detect those frauds accurately. Implementation of efficient fraud detection systems has thus become imperative for all credit card issuing banks to minimize their losses. Many modern techniques based on Artificial Intelligence, Data mining, Fuzzy logic, Machine learning, Sequence Alignment, Genetic Programming etc., has evolved in detecting various credit card fraudulent transactions. A clear understanding on all these approaches will certainly lead to an efficient credit card fraud detection system. This paper presents a survey of various techniques used in credit card fraud detection mechanisms and evaluates each methodology based on certain design criteria.

I. DETECTING FRAUD

Traditional ways of data analysis have been in use since a long time as a method of detecting fraud. They require complex and time-consuming investigations that deal with different domains of knowledge like financial, economics, business practices and law. Fraud often consists of many instances or incidents involving repeated transgressions using the same method. Fraud instances can be similar in content and appearance but usually are not identical (Palshikar 2002).

Techniques used for fraud detection:

Two techniques used for fraud detection. Statistical techniques and artificial intelligence (Palshikar 2002). Examples of statistical data analysis techniques are:

Data preprocessing techniques for detection, validation, error correction, and filling up of missing or incorrect data.

Calculation of various statistical parameters such as averages, quantiles, performance metrics, probability distributions, and so on. For example, the averages may include average length of call, average number of calls per month and average delays in bill payment.

Models and probability distributions of various business activities either in terms of various parameters or probability distributions.

Computing user profiles.

Time-series analysis of time-dependent data.

Clustering and classification to find patterns and associations among groups of data.

II. SYSTEM ANALYSIS

This chapter gives the information regarding analysis done for the proposed system. System Analysis is done to capture the requirement of the user of the proposed system. It also provides the information regarding the existing system and also the need for the proposed system. The key features of the proposed system and the requirement specifications of the proposed system are discussed below.

A. Existing System

The Traditional detection method mainly depends on database system and the education of customers, which usually are delayed, inaccurate and not in-time. After that methods based on discrimate analysis and regression analysis are widely used which can detect fraud by credit rate for cardholders and credit card transaction. For a large amount of data it is not efficient.

B. Problem Recognition

The high amount of losses due to fraud and the awareness of the relation between loss and the available limit has to be reduced. The fraud has to be deducted in real time and the number of false alert has to be minimized.

C. Proposed System

The proposed system overcomes the above mentioned issue in an efficient way. Using genetic algorithm the fraud is detected and the

false alert is minimized and it produces an optimized result. The fraud is detected based on the customers behavior. A new classification problem which has a variable misclassification cost is introduced. Here the genetic algorithms is made where a set of interval valued parameters are optimized.

III. SYSTEM DESIGN

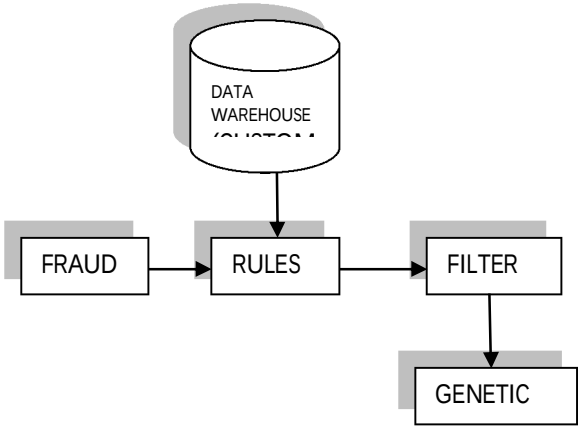
The process of design involves "conceiving and planning out in mind and making a drawing, pattern or a sketch". The system design transforms a logical representation of what a given system is required to do into the physical reality during development. Important design factors such as reliability, response time, throughput of the system, maintainability, expandability etc., should be taken into account. Design constraints like cost, hardware limitations, standard compliance etc should also be dealt with. The task of system design is to take the description and associate with it a specific set of facilities-men, machines (computing and other), accommodation, etc., to provide complete specifications of a workable system.

This new system must provide for all of the essential data processing and it may also do some of those tasks identified during the work of analysis as optional extras. It must work within the imposed constraints and show improvement over the existing system.. At the outset of design a choice must be made between the main approaches. Talks of 'preliminary design" concerned with identification analysis and selections of the major design options are available for development and implementation of a system. These options are most readily distinguished in terms of the physical facilities to be used for the processing who or what does the work.

A. Architectural Design

Describing the overall features of the software is concerned with defining the requirements and establishing the high level of the system. During architectural design, the various web pages and their interconnections are identified and designed. The major software components are identified and decomposed into processing modules and conceptual data structures and the interconnections among the modules are identified. The following modules are identified in the proposed system.

B. System Architecture



The above architecture describes the work structure of the system.

The customer data in the data warehouse is subjected to the rules engine which consists of the fraud rule set.

The filter and priority module sets the priority for the data and then sends it to the genetic algorithm which performs its functions and generates the output.

IV. DETAILED SYSTEM DESIGN

Detailed design deals with the various modules in detail explaining them with appropriate Diagrams and notations. The Use case diagram is designed to see the working logic of the proposed system. The sequence diagram is designed to describe, how the client and the server interacts with each other when processing content. The flow of the proposed system is described with the activity diagram. We know where the application starts and when it ends after processing the keywords and the current URL link. This will help the programmers to implement the internal logic for the module in the given specification.

In this part of design phase, the design is carried out using the top-down strategy. First the major modules are identified. Then they are divided into sub modules so that each module at the lowest level would address a single function of the whole system. Each module design is explained detail.

This chapter tells us how the input module is design in getting the users requirements. The detailed input design provides as information regarding what are tools used in getting inputs and send to the server.

Output design is gives the user with good interacting option on the screen. The information delivered to the users through the information system. Useful output is essential to ensure the use and acceptance of the information system. Users often judge the merit of a system based upon its output. Productive output can only be achieved via close interaction with users. The output is designed in attractive and effective way that user can access them with a problem. The application starts and when it ends after processing the keywords and the current URL link. This will help the programmers to implement the internal logic for the module in the given specification.

In this part of design phase, the design is carried out using the top-down strategy. First the major modules are identified. Then they are divided into sub modules so that each module at the lowest level would address a single function of the whole system. Each module design is explained detail.

This chapter tells us how the input module is design in getting the users requirements. The detailed input design provides as information regarding what are tools used in getting inputs and send to the server.

Output design is gives the user with good interacting option on the screen. The information delivered to the users through the information system. Useful output is essential to ensure the use and acceptance of the information system. Users often judge the merit of a system based upon its output. Productive output can only be achieved via close interaction with users. The output is designed in attractive and effective way that user can access them with a problem.

A. Machine Learning and Data Mining

Early data analysis techniques were oriented toward extracting quantitative and statistical data characteristics. These techniques facilitate useful data interpretations and can help to get better insights into the processes behind the data. Although the traditional data analysis techniques can indirectly lead us to knowledge, it is still created by human analysts (Michal ski et al. 1998).

To go beyond, a data analysis system has to be equipped with a substantial amount of background knowledge, and be able to perform reasoning tasks involving that knowledge and the data provided (Michal ski et al. 1998). In effort to meet this goal, researchers have turned to ideas from the machine learning field. This is a natural source of ideas, since the machine learning task can be described as turning background knowledge and examples (input) into knowledge (output).

If data mining results in discovering meaningful patterns, data turns into information. Information or 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.

B. Supervised and Unsupervised Learning

The machine learning and artificial intelligence solutions may be classified into two categories: 'supervised' and 'unsupervised' learning. In supervised learning, samples of both fraudulent and non-fraudulent records are used. This means that all the records available are labeled as 'fraudulent' or 'non-fraudulent'. After building a model using these training data, new cases can be classified as fraudulent or legal (Jans et al.).

Furthermore, this method is only able to detect frauds of a type which has previously occurred. In contrast, unsupervised methods don't make use of labeled records. These methods seek for accounts, customers, suppliers, etc. that behave 'unusual' in order to output suspicion scores, rules or visual anomalies, depending on the method (Bolton and Hand 2002).

Whether supervised or unsupervised methods are used, note that the output gives us only an indication of fraud likelihood. No stand alone statistical analysis can assure that a particular object is a fraudulent one. It can only indicate that this object is more likely to be fraudulent than other objects (Jans et al.).

1) Supervised Methods: The field of neural networks has been extensively explored as a supervised method. Jans et al. mention the studies of Barson, Field, Davey, McAskie, and Frank (Barson et al.) and Green and Choi (1997) all use neural network technology for detecting respectively fraud in mobile phone networks (Barson et al.) and financial statement fraud. Lin et al. (2003) apply a fuzzy neural net, also in the domain of fraudulent financial reporting. Both Brause et al. (1999) and Estevez et al. (2006) use a combination of neural nets and rules. Bayesian learning neural network is implemented for credit card fraud detection by Maes et al. (2002) for telecommunications fraud by Ezawa and Norton (1996) and for auto claim fraud detection by Viaene et al. (2005). In the same field as Viaene et al. (2005), insurance fraud, Major and Riedinger (2002) presented a tool for the detection of medical insurance fraud. They proposed a hybrid knowledge/statistical-based system, where expert knowledge is integrated with statistical

power. Another example of combining different techniques can be found in Fawcett and Provost (1997). A series of data mining techniques for the purpose of detecting cellular clone fraud is used. Specifically, a rule-learning program to uncover indicators of fraudulent behavior from a large database of customer transactions is implemented. Fawcett and Provost (1999) the Activity Monitoring is introduced as a separate problem class within data mining with a unique framework. Stolfo et al. and Lee et al. delivered some interesting work on intrusion detection. They provided a framework, MADAM ID, for Mining Audit Data for Automated models for Intrusion Detection. Next to this, the results of the JAM project are discussed. Cahill et al. (2000) design a fraud signature, based on data of fraudulent calls, to detect telecommunications fraud. For scoring a call for fraud its probability under the account signature is compared to its probability under a fraud signature. The fraud signature is updated sequentially, enabling event-driven fraud detection. Link analysis comprehends a different approach. It relates known fraudsters to other individuals, using record linkage and social network methods (Wasserman and Faust 1998). Cortes et al. (2002) proposed a solution to fraud detection in this field (Phua, 2005).

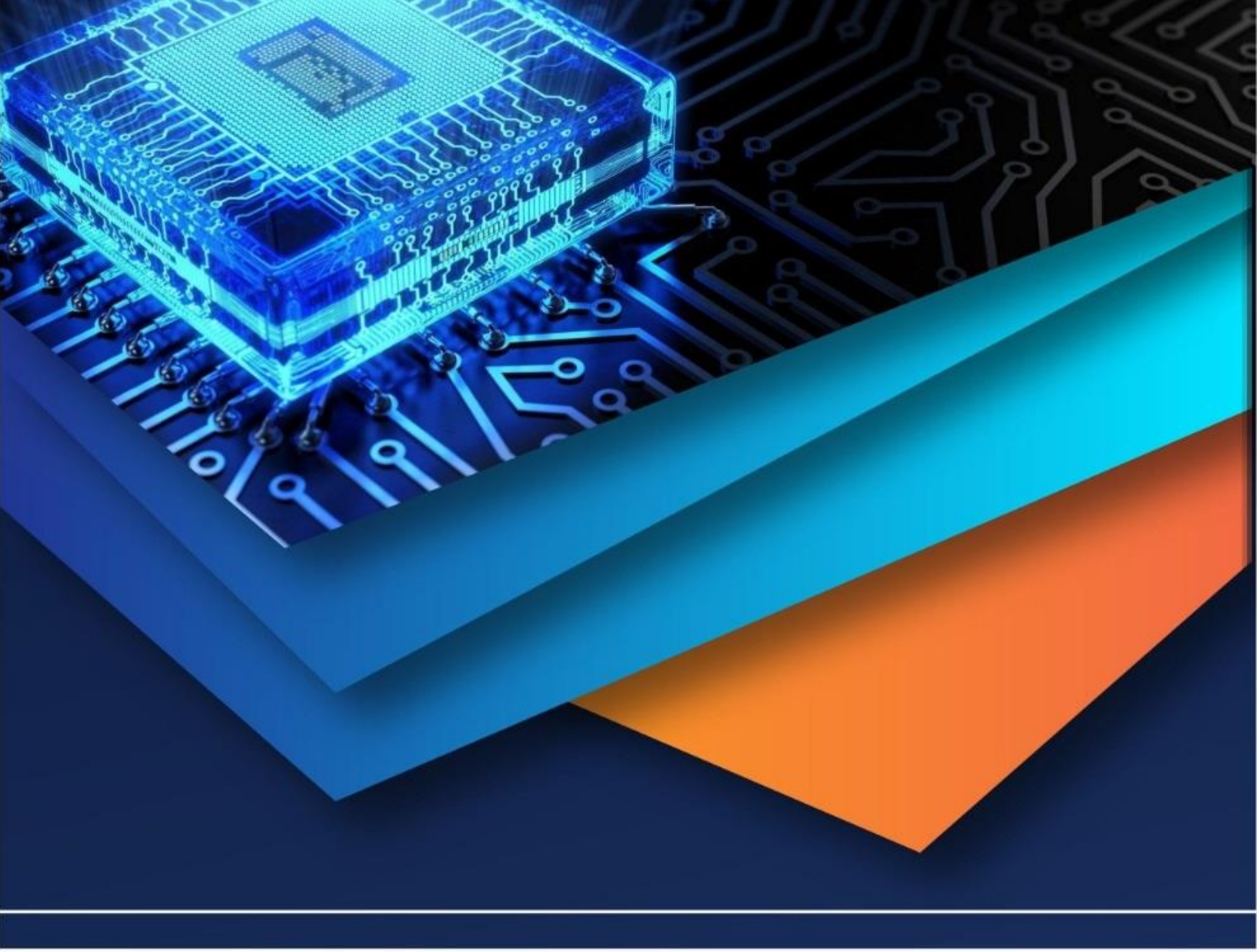
2) Unsupervised Methods: Some important studies with unsupervised learning with respect to fraud detection should be mentioned. For example, Bolton and Hand use Peer Group Analysis and Break Point Analysis applied on spending behavior in credit card accounts. Peer Group Analysis detects individual objects that begin to behave in a way different from objects to which they had previously been similar. Another tool Bolton and Hand develop for behavioral fraud detection is Break Point Analysis. Unlike Peer Group Analysis, Break Point Analysis operates on the account level. A break point is an observation where anomalous behavior for a particular account is detected. Both the tools are applied on spending behavior in credit card accounts. Also Murad and Pinkas (1999) focus on behavioral changes for the purpose of fraud detection and present three-level-profiling. As the Break Point Analysis from Bolton and Hand, the three-level-profiling method operates at the account level and it points any significant deviation from an account's normal behavior as a potential fraud. In order to do this, 'normal' profiles are created based on data without fraudulent records (semi supervised). To test the method, the three-level-profiling is applied in the area of telecommunication fraud. In the same field, also Burge and Shawe-Taylor (2001) use behavior profiling for the purpose of fraud detection. However, using a recurrent neural network for prototyping calling behavior, unsupervised learning is applied.] Cox et al. (1997) combines human pattern recognition skills with automated data algorithms. In their work, information is presented visually by domain-specific interfaces, combining human pattern recognition skills with automated data algorithms.

V. CONCLUSION

This method proves accurate in deducting fraudulent transaction and minimizing the number of false alert. Genetic algorithm is a novel one in this literature in terms of application domain. If this algorithm is applied into bank credit card fraud detection system, the probability of fraud transactions can be predicted soon after credit card transactions. And a series of anti-fraud strategies can be adopted to prevent banks from great losses and reduce risks. The objective of the study was taken differently than the typical classification problems in that we had a variable misclassification cost. As the standard data mining algorithms does not fit well with this situation we decided to use multi population genetic algorithm to obtain an optimized parameter.

REFERENCES

[1] PricewaterhouseCoopers LLP (2009). "2009 Global Economic Crime Survey". http://www.pwc.com/gx/en/economic-crime-survey. Retrieved June 29, 2011.
[2] Nigrini, Mark (June, 2011). "Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations". Hoboken, NJ: John Wiley & Sons Inc.. ISBN 978-0-470-89046-2. http://www.wiley.com/WileyCDA/WileyTitle/productCd-0470890460.html.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY