ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)
Vol. 2, Issue 3 (July - Sept. 2014)
International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014)

Advanced Security Mechanism for Online Financial Transactions Through Data Mining Techniques

P.V.D Prasad
Software Consultant, J M R InfoTech, Beta Building, Sigma Soft Tech Park, Whitefield, Bangalore, India

Abstract
In this paper we provide a detailed summarisation of data mining techniques that it is possible to characterize and detect false transactions over e commerce in a year, depending on the information, historical performance and characteristics, using different types of data mining techniques. As these days, businesses have become more concerned about using transaction data to generate knowledge about the world in which they function to as called "business intelligence." The business intelligence is generated using tools such as data mining and knowledge discovery. Although much of that focus on business intelligence initially was generated about relationships with other firms, such as sales, increasingly there is a focus on internal processes. That focuses on generating business intelligence about internal processes, to facilitate management and monitoring of processes.

Key words
Business Intelligence, KDD, Artificial Intelligence, Data Quality

I. Introduction

Data Mining deals extracting hidden knowledge, unexpected pattern and new rules from large database. Accordingly, the purpose is to investigate and to assess and facilitate data quality, including finding fraudulent activity. It can be implemented by an organizing architecture for the data. The scope of system to investigate the processes of purchasing and accounts payable, within the context of ecommerce process management. As a result, this Paper focuses on the application approaches to facilitate data quality analysis of a process. Particular attention is given to data mining and its ability to ascertain when data seems appropriate or anomalous. This can be useful in facilitating the monitoring process. Trends in data mining include further efforts towards the exploration of new application areas and methods for handling complex data types, algorithm scalability, constraint based data mining and visualization methods.Internet technologies have faultlessly automated interface processes between customers and retailers, retailers and distributors, distributors and factories, and factories and their numerous suppliers. In general, e-commerce and e-business have enabled on-line Payment transactions. Data Mining (DM) is a well honoured field of Computer Science. It emerged in late 80's by using concepts and methods from the fields of Artificial Intelligence, Pattern Recognition, Database Systems and Statistics, Data Mining discovers valid, complex and not obvious hidden information from large amounts of data. Data Mining is also known as Knowledge Discovery in Databases (KDD). Financial data are collected by many organizations like banks, stock exchange authorities, taxation authorities, big accounting and auditor offices specialized data bases, etc and in some cases are publicly available. The application of Data Mining techniques on financial data can contribute to the solution of classification and prediction problems and facilitate the decision making process. Typical examples of financial classification problems are corporate bankrupt, credit risk estimation, going concern reporting, financial distress and corporate performance prediction. The importance of Data Mining in finance and accounting has been recognized by many organizations. The aim of the present study is to provide a state-of-the-art review about current research efforts on applying in Finance and accounting.

Challenges in Web Mining:

As the World Wide Web contains the huge information that provide rich source for data mining. The web possesses great challenges for resource and knowledge discovery based on the following observations:

The web is too big to handle:

The size of the web is very huge and rapidly increasing. This seems that the web is too huge for data warehousing and data mining.

Complexity in Web pages

The web page does not have unifying structure. They are very complex as compared to traditional text document.

Web is a dynamic information source

The information on the web is rapidly updated. The data such as news, stock markets, weather, sports, shopping etc are regularly updated.

Diversity of user communities

The user community on the web is rapidly expanding. These users have different backgrounds, interests, and usage purposes. There are more than 100 million workstations that are connected to the Internet and still rapidly increasing.

Relevancy of Information:

It is considered when a person is searching for small portion of the web, while the rest of the portion of the web contains the information that is not relevant to the user and may swamp desired results.

Data Mining is widely used in diverse areas. There are number of commercial data mining system available today yet there are many challenges in this field.

Financial Data Analysis

The financial data in banking and financial industry is generally reliable and of high quality which facilitates the systematic data analysis and data mining.

1. Design and construction of data warehouses for multidimensional data analysis and data mining.
2. Loan payment prediction and customer credit policy analysis.

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

3. Classification and clustering of customers for targeted marketing.
4. Detection of money laundering and other financial crimes.

Retail Industry

Data Mining has its great application in Retail Industry because it collects large amount data from on sales, customer purchasing history, goods transportation, consumption and services. The quantity of data collected will continue to expand rapidly because of increasing ease, availability and popularity of web. The Data Mining in Retail Industry helps in identifying customer buying patterns and trends. That leads to improved quality of customer service and good customer retention and satisfaction.

1. Design and Construction of data warehouses based on benefits of data mining.
2. Multidimensional analysis of sales, customers, products, time and region.
3. Analysis of effectiveness of sales campaigns.
4. Customer Retention.
5. Product recommendation and cross-referencing of item

Predictive patters from quantitative time series analysis will be invented fortunately, a field known as data mining using quantitative analytical techniques is helping to discover previously undetected patterns present in the historic data to determine the buying and selling points of equities.When market beating strategies are discovered via data mining, there are a number of potential problems in making the leap from a back-tested strategy to successfully investing in future real world conditions. The first problem is determining the probability that the relationships are not random at all market conditions. This is done using large historic market data to represent varying conditions and confirming that the time series patterns have statistically significant predictive power for high probability of profitable trades and high profitable returns for the competitive business investment.

Types of Commercial Payment Solution Misuse

Non-compliance with an organization's card usage policies erodes the process efficiencies and cost savings of the payment solution. This out-of-policy spend can result from a lack of knowledge of corporate policies or from intentional abuse Examples of transactions that should trigger audit consideration include those made by cardholders who:

- Purchase with restricted merchants or MCCs (merchant category codes)
- Withdraw cash on the company card with no associated travel or on non-work days (to help create a cash float)
- Use convenience checks excessively to withdraw against the card's limits
- Make out-of-policy travel arrangements
- Transact with only one merchant (the cardholder may be directing funds from the card to a shadow business they have set up)
- Split large purchases into several transactions (may be an attempt to circumvent card limits or expense thresholds)
- Make multiple purchases of even dollar amounts (e.g., $100, $250), which could signal an attempt to exceed cash limits by getting funds from a merchant or financial institution as a charge, rather than from an ATM
- Suddenly begin using their card more frequently or for higher dollar amount purchases than usual

Emerging Compliance Management Tools

Uncovering instances of misuse not easily detected by traditional methods can be a costly and arduous task, compounded by the sheer volume of information that is created with each transaction. Cardholder name, number, merchant name, MCC, purchase amount and date are a mere fraction of the data captured each time a commercial payment solution is used. Multiply this by the number of cardholders and number of transactions each cardholder makes per month and it is clear that intensive data mining is needed to drill down through these volumes of information and sort the data in a meaningful manner.

The marketplace is responding to this need with sophisticated new tools designed to allow organizations to reap full value from their commercial payment solutions. Using cutting-edge data mining techniques, this new class of compliance management applications is being implemented by leading organizations to:

- Detect and deter fraud and waste
- Enforce policies and procedures
- Deliver financial justification for expanding the card program, thereby delivering efficiencies and savings to the organization
- Fulfil government reporting requirements

Key Attributes

While each provider takes a unique approach to compliance management, there are a number of attributes that can be found across many solutions. The following lists the intended benefits these features present.

Ongoing Monitoring (continuous or as needed)

- Uncover unforeseen risks and opportunities for operational improvement
- Accelerate the ability to act on questionable transactions
- Remove the need for time-consuming, costly manual efforts

Transaction Evaluation (all or sample)

- Detect questionable or out-of-policy transactions immediately
- Provide a more complete picture of patterns of use
- Eliminate spot checking (only when all transactions are evaluated); no transaction falls through the cracks

Data Integration

- Pull information from multiple systems and sources
- Provide a more holistic view for greater insight

Standard and Ad-hoc Reporting

- Allow organizations to run reports of the most benefit to them
- Permit customization of report types

Dashboard Visibility

- Give managers a snapshot of program performance metrics
- Provide a central location where users can easily access the level of information they need

Transaction Scoring

- Use algorithms to search for anomalies, including those that

would not typically be detected via known rules
- Leverage interactions between various transactional and cardholder attributes
- Provide a view of usages patterns and trends.

Automated Alerts
- Proactively flag non-compliant transactions
- Give card managers timely access to exception reports

Behind-the-firewall versus SaaS Solutions

Today's class of compliance management solutions is divided between traditional delivery methods and the Software-as-a-Service (SaaS) model.

Behind-the-firewall applications can offer a number of benefits, including:
- Flexibility to customize to each client's requirements and incorporate requests for enhancements sooner.

An Overview of Visa IntelliLink Compliance Management

Visa IntelliLink Compliance Management is a web-based solution that provides intelligent assistance for optimal card program management. It converts transaction data into information program managers can use to minimize and deter misuse and abuse and maximize savings..

Visa IntelliLink Compliance Management's monitoring of card usage in minute detail also helps card program managers ensure that their commercial payment programs are achieving optimal savings and benefits by operating according to their organization's policies and regulations. The wide range of services includes:
- Analytics and investigative reporting, including drill-down dashboard capabilities to measure program performance
- Misuse and abuse detection, including a neural-network transaction scoring system that can detect questionable behaviour not easily apparent to the human brain
- Reporting that enables program and regulatory compliance
- Self-service administration

Visa IntelliLink Compliance Management has been validated by a third party as meeting the requirements of Section 508 of the Americans with Disabilities Act and the Payment Card Industry Data Security Standards (PCIDSS).

II. Clustering

Using clustering, it can give us a general idea of the insight of the data when there are too many input attributes. This can help to clarify the healthy status of the system at different levels, such as: system daily, system hourly, and district daily and district hourly level. By notifying this result, one can use this information to perform a simple query on the raw data to find out cells within this range to detect the problematic cells causing high

Traffic Analysis by Classification

By day, By week, By month, By year: We start by aggregating weekly data on weekday basis and try to find the distribution of several parameters over each day of week. By gaining access to this type of information, we can better allocate our resources and equipments in keys dates such as days of week with high traffic.

Evolutionary steps in data mining

Data mining techniques are the result of a long process of research and have gone through various steps of evolution. Such evolution began when business data was first stored on computers and generated technologies to allow users for navigating their data in real time. Data mining algorithms have existed for at least ten years, but have recently been implemented as reliable and understandable tools. Now it is supported by further technologies that are sufficiently mature for navigation to prospective and proactive information delivery.

Financial, banking and credit or risk scoring

Data mining can assist financial institutions in various ways, such as credit reporting, credit rating, loan or credit card approval by predicting good customers, risk on sanctioning loan, mode of service delivery and customer and others. A credit card company can leverage its vast warehouse of customer transaction data to identify customers most likely to be interested in a new credit product. In addition, data mining can also assist credit card issuers in detecting potentially fraudulent credit card transaction. In general, data mining methods such as neural networks and decision trees can be a useful addition to the techniques available to the financial analyst.

Procedure of Data Patterns Discovery in Financial Institutions

In this method, The Data bases stores the data for every financial transaction over a organisation day by day, We are implementing the clustering method for storing of data storage in the data warehouse. The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. A cluster of data objects can be treated collectively as one group and so may be considered as a form of data compression. Although classification is an effective means for distinguishing groups or classes of objects, it requires the often costly collection and labelling of a large set of training tuples or patterns, which the classifier uses to model each group. It is often more desirable to proceed in the reverse direction: First partition the set of data into groups based on data similarity and then assign labels to the relatively small number of groups. Additional advantages of such a clustering-based process are that it is adaptable to changes and helps single out useful features that distinguish different groups.

Clustering is also called data segmentation in some applications because clustering partitions large data sets into groups according to their similarity. Clustering can also be used for outlier detection, where outliers (values that are "far away" from any cluster) may be more interesting than common cases. Applications of outlier detection include the detection of credit card fraud and the monitoring of criminal activities in electronic commerce. For example, exceptional cases in credit card transactions, such as very expensive and frequent purchases, may be of interest as possible fraudulent activity.

Benford's Law

One metric that can be traced and monitored to expectations is Benford's law which states that the first significant digit d ($d \in \{1... b-1\}$) in base b ($b \geq 2$) occurs with probability proportional to logb (d + 1) - logb(d). As a result, Benford's law establishes a set of expectations for the distribution of numbers. For many numeric generating processes, the first digit (or first and second,

ISSN : 2347 - 8446 (Online)
ISSN : 2347 - 9817 (Print)

etc.) can be analyzed to see if it meets expectations. If it does not, then that can indicate an anomaly and that an investigation should be conducted to determine if there is some fundamental problem. It is possible to extend the law to digits beyond the first. In particular, the probability of encountering a number starting with the string of digits n is given by:

$$\log_{10}(n+1) - \log_{10}(n) = \log_{10}\left(1+\frac{1}{n}\right)$$

(For example, the probability that a number starts with the digits 3, 1, 4 is log10(1 + 1/314) ≈ 0.0014.) This result can be used to find the probability that a particular digit occurs at a given position within a number. For instance, the probability that a "2" is encountered as the second digit is

$$\log_{10}\left(1+\frac{1}{12}\right)+\log_{10}\left(1+\frac{1}{22}\right)+\cdots+\log_{10}\left(1+\frac{1}{92}\right)\approx 0.109$$

And the probability that d (d = 0, 1... 9) is encountered as the n-th (n > 1) digit is

$$\sum_{k=10^{n-2}}^{10^{n-1}-1} \log_{10}\left(1+\frac{1}{10k+d}\right)$$

The distribution of the n-th digit, as n increases, rapidly approaches a uniform distribution with 10% for each of the ten digits. Four digits is often enough to assume a uniform distribution of 10% as '0' appears 10.0176% of the time in the fourth digit while '9' appears 9.9824% of the time.

How it works?

Knowledge Discover: Comparison to expectations

Although preventive controls are critical to ensuring that data quality is high, an important approach to ensuring data quality is to compare data to "expectations" to see if the data meets those expectations. There are a number of bases of comparison, including Benford's law and other comparisons.

1. Same, Same, Same

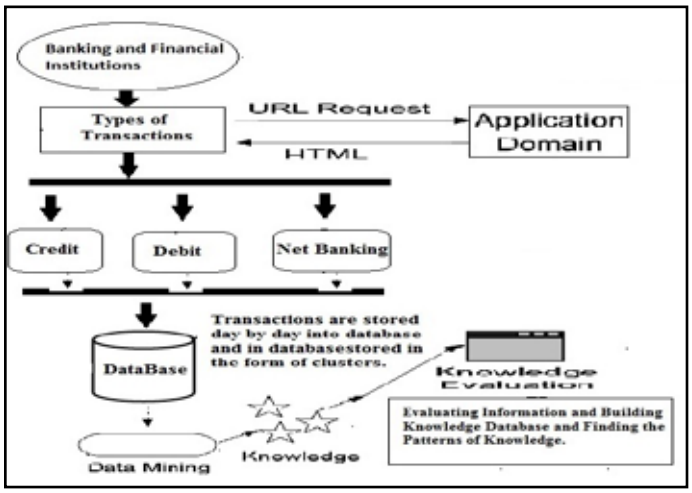
An important test of the quality of the accounts payable data is for duplicate payment of the same invoice to the same vendor .In this situation, the data is investigated for the same invoice number, same amount, and same vendor. Such duplicate payments can occur if the vendor provides multiple copies at different times of the same invoice, whether as part of normal business practice or as part of a fraudulent approach. As part of the analysis, the accounts payable for the match must be determined so that it can be ascertained if there is a systematic problem.

2. Same, Same, Different

One test of the quality of accounts payable data is the "same, same, different" test. The purpose of the test is to compare different accounts payable entries to determine if they are the same, and as a result, a bill has been paid more than once or if the wrong vendor has been paid. An invoice might be paid twice in the situation where the invoice was paid to the wrong vendor and then the correct vendor. The wrong vendor may have been paid, either purposely or by accident, such as an incorrect keying of the data. As part of the analysis, the accounts payable, the match must be determined so that it can be ascertained if there is a systematic problem.

3. Same, Different, Different

Another test of accounts payable data is the reuse of a purchase order number for other amounts or vendors. In a system that requires a purchase order number, a fraudulent entry could "reuse" a purchase order number to meet the need of providing a purchase order number with each entry. This test would allow detection of such reuse.The Systematic of Data Classification and storage is represented as below:



Data Quality-Based Data Mining

Purchasing and accounts payable systems depend on the underlying data in the system being "good data" to begin with. However, that assumption may not be true. One approach to analyzing data quality is to investigate the data using data mining, in order to determine if the basic data set contains any anomalies, indicating problems with the underlying data. For example, vendors may be fraudulent or goods may be bogus, in which case any transactions involving those vendors or goods would be suspect.

Conclusion

This paper approaches to ensure and analyze data quality in a financial payment Process in the context of business process management. Further, three different scenarios of how the purchasing and accounts payable processes would be generated and were analyzed, Data Mining provides a way of managing processes, by integrating technology. This has not been aimed at those activities but has focused more on managing cash flows in the organisations. This lays out some metrics to monitor process data, helping how knowledge discovery could be used to determine if data is meeting expectations.

References

[1]. *Financial Stock Market forecast using DataMining Techniques-K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik and P. Arumugam.*

[2]. *Data Mining In Finance And Accounting: A Review Of Current Research Trends Efstathios Kirkos Yannis Manolopoulos.Worldwide Software as a Service 2010-2014 Forecast: Software Will Never Be the Same, International Data Corporation, June 2010. http : Visa.com/intellilink*

[3]. *Supporting Mobile Payment QOS by Data Mining GSM Network Traffic Edison Lai, Simon Fong, Yang Hang--University of Macau,*

[4]. *Role of Data Mining in E-Payment systems Sabyasachi Pattanaik,Partha Pratim Ghosh*

FM University, Balasore.
[5]. Usefulness and applications of data mining in extracting information from different perspectives- Jiban K Pal
[6]. DM Concepts & Techniques _ Han&Kamber
[7]. Monitoring and Managing Data and Process Quality –BPM
[8]. Chen, A. "Hasbro Plays to Win with BPM," eWeek.com, August 2, 2004.
[9]. Cognos, "Cognos Financial Analytics," http://www.cognos.com/pdfs/whitepapers/wp_cognos_financial_analytics.pdf
[10]. Coderre, D. "Global Technology Audit Guide Continuous Auditing," The Institute of Internal Auditors, 2005.
[11]. Hill, T. "The first digit phenomenon," American Scientist 86 (July–August 1998), p. 358. http://www.americanscientist.org/template/ AssetDetail/assetid/ 15660;jsessionid=baa6gWCz81?fulltext=true
[12]. Lombardi Software, "Accounts Payable," http://www.lombardisoftware.com/bpmaccounts-payable.php#
[13]. Metastorm, "The Louisiana Department of Social Services," 2006, www.metastorm.com/customers/lodss/Louisiana%20DSS%20Success%20Story.pdfAU8522_C010.
[14]. http://en.wikipedia.org/wiki/Benford's_law