Data Mining and Application in Accounting and Auditing

KeramatOllah Heydari Rostami¹, Saber Samadi¹, Hamed Omrani¹, Ahmad Kazemi Margavi¹, Hamid Asadzadeh¹, Hemad Nazari²
¹University of Tehran, Iran
²University of Tehran Azad Oloom va Tahghighat, Iran
*heydarykeramat@gmail.com

Abstract: Traditional manner for making information from data was laid on handy interpret and analyzing. This kind of analyzing of data is so slow and also expensive and objective. It hopes computer technology helps when data using and analyzing is above humane ability. Knowledge exploration emphasizes on understandable patterns. Totally, it can be said that rising to the upward knowledge is from raw data that are down knowledge.

Key words: Data mining, data warehouse, data mart

1. Introduction

Data mining has several applications in several scopes like investing, finance, accounting, marketing and banking. Several financial analyzing applications use predicting modeling like statistical report and neural networks for optimizing investing portfolios and making financial modeling. It can concentrate on different bank accounts for uncovering fraud. Many companies use data mining for managing all life cycle stages of their companies. This life cycle contains adding new customers, uprising income from existing customers and good support for these customers. By using customers' history and their characters, company can concentrate on the services that they do not use (Tarizadeh et al, 2002).

Data mining, sometimes referred to as data or knowledge discovery, derives its name from searching for valuable information in a large database, data warehouse or data mart. Such information can be used to establish relationships between variables, and then to validate the findings by applying the detected patterns to new subjects of data. The ultimate goal of data mining is prediction (Fadairo et al, 2008). Data mining is a way to gain market intelligence from a huge amount of data. The problem today is not the lack of data, but how to learn from it. In data mining, the data tell the story, but it is up to you how to use that information (Keating, 2008). Data mining is used to search for valuable information from the mounds of data collected over time, which could be used in decision-making. The information may be certain patterns and/or relationships that exist (Keating, 2008). In the scope of data mining, data warehouse and data mart are heard so much, so they are explained here:

What is data warehouse?

- A set of objective, complex, nonvolatile and dependant data that is support for management making decisions.
- A set of several data mart designed for easing processing and querying of employees.
- A data warehouse collects data from several sources and converts them to information, then store with data and place information to make them useful for decision-making.
- A data warehouse is a store of information for analyzing and querying.

What is data mart: It's custom to point data of marts as constant data. Constant means that it is different from the other data like input, output and totally from all data that are transient in origin. In other words, it is said data mart or database is constant because data just can be eliminated when a request of this action is received from data mart management system. Data mart or the same database is set of constant data that are used by applied systems in some institution (Meshkani, 2009).

Data mining: Data mining is debating as a stage of knowledge exploration and obviously is heart of it. Data mining is analysis process of tremendous volume of data for finding useful, obvious and important information, patterns by using automatic and semi automatic developing tools. Such data may be stored
in data warehouse or/and every storage of data. Data mining’s aim for several organizations can prove marketing, sale or operations about customers services by finding better understanding of customers’ needs, knowing unknown patterns, future predicting on the base of past experiences and proven working process (Miller & Han, 2001). By data mining and information that it provides, people can use it as a leverage to make new values in organizations (Wu, 2002).

Data mining is a procedure of seek and find several models, summarizing that has below stages:

- **Hypothesis formulation**: in this stage, one modeler usually points one set of variables for dependency function, in this stage, it can be proposed several hypos for one theorem.
- **Data collecting**: This stage is about making and collecting data. Totally, there are two possibilities. First is when possibility that is data making process under modeler control. Second possibility is shown when professional cannot effect the production in procedure. This method is called visual method.
- **Data preprocessing**: In visual method, data usually collected from data marts in data warehouses and data centers. Data preprocessing usually contains below two stages:
- Omitting unusual data
- Characters, encrypting and selecting
- **Model evaluation**: The principle work of data mining in this stage is selecting suitable techniques. This process is not usually obvious. Its' implementing is upon several models and selecting best of them is an additional role and work.
- **Model interpretation and making conclusion**.

2. Data Mining Framework

Data mining framework contains three layers: First layer is about technology used in data mining e.g. data mart managing, machine learning, statistic, parallel processing, visualization, help in decision making and warehouse. Second layer contains doing work techniques and data mining tools. In third layer, there are several modules e.g. mining distributed databases, mining multimedia data, and mining data on the World Wide Web (Tarizadeh et al., 2002).

**Data mining utility**: Data mining raising aim is predicting and description or mixed of prediction and description. The aim of predicting concentrates on attention predicting power, user does not think weather model reflects reality or not (Tarizadeh et al., 2002). Data mining can be used for solving four issues as below:

**Relationship discovery**: Portfolio analysis is the most famous of this application that is called "link analysis”. Patterns analysis is obvious in routing sales: if sales products are the same products customers buy, sales raises (Berry and Linoff, 1997).

**Making choices**: Making decision is choosing one way from several ways. In work and business world, decision makers ought to choose from several trades, expensive or cheap products and even good or bad customers. By using data mining techniques like "classification” can evaluate data for decision-making and preference ways and proposing best way (Berry and Linoff, 1997).

**Making prediction**: Prediction is choosing from several ways about future. Its’ application almost is in prediction of future customer prediction, future market need. Predicting methods use last experience and conditions till predicting future condition.

**Making better work procedure**: Work and business are in complicated system of "providers", communicating structure", and market pressure” and other factors that several of these factors cannot obvious (Berry and Linoff, 1997). By using data, mining can light the several dimensions of procedures that are efficient in prosperity (Delmater, 2001).
3. Data Mining Accounting

Mining for accounting demonstrates how companies, financial institutions, insurance companies, tax authorities and other governmental agencies can employ data in accounting, taxation and auditing-related tasks. Models can be used to detect accounting patterns and irregularities, improper practices, questionable transactions, potential fraud and money laundering. Data mining and data manipulation tools can check the quality and integrity of data, including missing data, multiple ledgers, individual account entries for a single-person supplier and dominant accounting. The identification of such items can lead to correction of errors in standing data, pricing and processing saving and the reduction of duplicate payments. Data mining is one of the important weapons in the fraud examiner's armory. Identifying instances of business fraud is not easy. Data mining possesses its own unique problems due to the requirements of data protection legislation. It is not recommended that data mining be performed without first obtaining legal opinion or other expert advice (Fadairo et al., 2008).

Data mining and data making procedure are tools for evaluating truth and quality of hypos, data and totally can be classify as below frames:

- Disadvantages controlling and control things that can be used for doing fraud.
- Decreasing invoices of one customer or supplier in just on and shorter one.
- Discovering mistakes in tax calculates.
- Discovering unequally in the price mentioned in invoices and paid bills.
- Groth (1997) counts five important points in using data mining:
  - Answering this question that, data mining is used if it has role in investment returns.
  - Data implementation is not simple. Around 60-80 percent of applied expenses of data mining is for implementing of data.
  - Many of works that are done for finding and implementing data are hard to understand.
  - Principle concentration of data mining is for making a supporter decision system. Data should be obvious, acquisitive, surveying and from thinking process.
  - Making applied conclusion. Conclusions of data mining should be applied and operative.

Data mining and find frauds in auditing: Data mining in this stage concentrate on professional doubt and ways of analyzing methods of finding frauds. Data mining is a technique that by using data analyzing softwares as a right work in audit procedure. By the studies done in audit scope recently, it is light that many companies still collect, implement the works and point that auditors mentioned, that by using data mining, it is not a problem anymore (Harding, 2007).

4. Data mining ability: Tools of Data Mining

There are four categories of data mining tools:

- Prediction
- Classification
- Clustering Analysis
- Association Rules Discovery (Keating, 2008).

Prediction tools: They are the methods derived from traditional statistical forecasting for predicting a variable's value. Classification tools: Most commonly used in data mining, classification tools attempt to distinguish different classes of objects or actions. For instance, a particular credit card transaction may be either normal or fraudulent. These tools could classify it as one or the other, thereby saving the credit card company a considerable amount of money. In another instance, an advertiser may want to know which aspect of its Promotion is most appealing to consumers. Is it price, quality, and/or reliability of a product? Maybe it is a special feature that is missing on competitive products. The classifications tools help give such information on all the products, making possible to use the advertising budget in a most effective manner. Clustering analysis tools: These are very powerful tools for clustering products into groups that naturally fall together. These groups are identified by the program and not by the researchers. Most of the clusters discovered may not be useful in business decision.

However, they may find one or two that are extremely important, the ones the company can take advantage of. The most common use for clustering tools is probably in what economists refer to as
“market segmentation”. In market segmentation, a company divides the customer base into segments dependent upon characteristics such as income, wealth, geographic location, lifestyle, and so on. Each segment is then treated with a different marketing approach, one suited precisely to that particular segment. Association Rules Discovery: Here the data mining tools discover associations; e.g., what kinds of books certain groups of people read, what products certain groups of people purchase, what movies certain groups of people watch, etc. Businesses use this information in targeting their markets. Netflix, for example, recommends movies based on movies people have watched and rated in the past. Amazon does the much the same thing in recommending books (Keating, 2008).

5. Limitations of Data Mining

While data mining can be immensely useful, it carries limitations. Data mining only identifies anomalies, exceptions, irregularities, patterns or trends. It does not assign meaning or value to that information. This may result in "false positive" findings, findings that prompt initial concern, but for which there are sound explanations. Data mining activities can also have an adverse impact on IT operations and data integrity. Data mining requires considerable IT processing power, especially if it involves complex queries that examine very large data sets. Using live files for data mining also subjects the underlying data manipulation. If information must be used. In addition to recognizing the potential for false positive findings and the needs to segregate data mining from live IT activities, auditors and managers also need to remember that data mining is only an aid for decision-making, not a substitute for sound judgment (Martin, 2009).

The relevance and success of data mining for the going concern decision is driven by a number of factors. First of all, much data of previously audited firms is available, a prerequisite for any data mining application. Secondly, the going concern decision is a complex task with widespread consequences to both the company being audited and the auditor, for which decision support systems are more than welcome. This has prompted the development of numerous models to predict the issuance of a going concern opinion in the past. Finally, recent accounting debacles only stress the importance of good auditing practices, increasing the relevance of such predictive data mining models even further. In the existing literature body, the automated prediction of such opinions is commonly done with logistic regression. Although more advanced data mining techniques —, which have been widely, researched and applied in domains such as credit scoring, bio-informatics and marketing — were large lymissing from the audit domain. Decision support tools can be very helpful, though user friendliness is a key requirement as auditors are often rather skeptical to the use of statistical, rather incomprehensible models. An intuitive decision table on the other hand, can very easily be incorporated into the auditor’s guidelines, assuring that going concern opinions are expressed more consistently (Martens et al. 2008).

The history of knowledge discovery in databases (KDD), now known as data mining (DATAMINING), is not much different, at least so far. In the early 1990s, when the KDD processing term was first coined (Piatetsky and Frawley, 1991), there was a rush to develop DATAMINING algorithms that were capable of solving all a company’s problems of searching for knowledge in large volumes of data. Apart from developing algorithms, tools (Clementine (1995), IBM Intelligent Miner Weka DB Miner (1995)) were also developed to simplify the application of DATAMINING algorithms and provide some sort of support for all the activities involved in the KDD process. From the viewpoint of DATAMINING process models, the year 2000 marked the most important milestone, as this was when the first standard and tool-independent DATAMINING process model was published. This standard is known as CRISP- DATAMINING (Cross-Industry Standard Process for DATAMINING) (Chapman et al., 2000).

A process model can be defined as the set of tasks to be performed to develop a particular element, as well as the elements that are produced in each task (outputs) and the elements that are necessary to do a task (inputs). The ultimate goal of a process model is to make the process repeatable, manageable and measurable (to be able to get metrics). A good process model should be (Richards, 1997 & Tyrrell, 2000): Effective: An effective process should help us produce the right product.

- Maintainable: So we can quickly and easily find and remedy faults or work out where to make changes.
- Predictable: Any new product development needs to be planned, and those plans are used as the basis for allocating resources: both time and people. A good process will help us do this. The process helps lay out the steps of development.
• Repeatable: If a process is found to work, it should be replicated in future projects. Ad hoc processes are rarely replicable unless the same team is working on the new project. Even with the same team, it is difficult to keep things exactly the same.
• Quality: Quality in this case may be defined as the product’s fitness for its purpose.
• Improvable: No one would expect their process to reach perfection and need no further improvement itself. Even if we were as good as we could be now, both development environments and requested products are changing so quickly that our processes will always be running to catch up.
• Traceable: A defined process should allow the project staff to follow the status of a project. (Marba’n, 2009)

6. Concluding Remarks

Data mining is discovering predictable hidden information in tremendous data marts and availability possibility of important information in data marts to their owners. Data mining tools can predict future behavior of customers, stockholders and totally stakeholders and raise decision power. The power of data mining is more powerful of other tools. These tools can solve problems that were very time consumer in short time. These tools seek patterns for exploring hidden predictable information in data marts.

References

Wu, J. (2002). Business Intelligence: the value in Mining Data published in DM Review Online.