

## COMPARISON OF DATA MINING AND AUDITING TOOLS

Simmi Bagga<sup>1</sup> and G.N. Singh<sup>2</sup>

<sup>1</sup>Deptt. of Computer Science, Sant Hira Dass Kanya Mahavidyalaya, Kala Senghian, Distt. Kapurthala, (Punjab)

<sup>2</sup>Deptt. of Physics and Computer Science, Sudarshan Degree College, Lalgaon, Distt. Rewa (M.P.)

### ABSTRACT

Auditing means the review of transaction in order to detect frauds and errors in the transactions. For the process of Auditing Auditor has to analyze a large volume of data that is very difficult to perform manual so we apply some computer based auditing tools. In this paper we describe some computer based auditing tool like IDEA and also data mining tools. Based on the characteristic feature of both IDEA and Data mining tools we have compared both the tools.

## 1. INTRODUCTION

### 1.1 Auditing Tools

In the recent year there has been widespread change in the adoption and utilization of new technologies in business. These days even small business has large number of financial transaction. It's the responsibility of the auditor to analyze these transactions to detect frauds and errors in financial transactions. Due to change in business trends its very difficult and complicated to audit financial transactions by traditional or manual methods. The limitations of these manual auditing can overcome by using computer assisted auditing tools. These days various financial audit software are used to assist auditor to improve the overall audit process. Most of the auditing companies develop their own auditing software or used commercially available software like Audit Command Language(ACL), Top CAATs, Statistical Analysis system (SAS) Interactive Data Extraction and Analysis (IDEA) and Microsoft Data Analyser etc.

IDEA is a Generalized Audit Software. IDEA has many unique features and functions that are not found in other audit software. It enhances the audit capabilities and improves the audit results. IDEA is a versatile tool to interrogate any kind of file. Basic features of IDEA are importing, querying and sorting data according to specifications, mathematical computation and statistical sampling, summarizing, file merging and report data in variety of formats. IDEA has more than 80 built-in functions for arithmetic, financial, text and date criteria. It performs analyses of data including calculation of statistic tests, gap detection, and detection of duplicates, fraud detection, summaries, and aging.

### 1.2 Data Mining

Data mining is the hidden pattern extraction process from large database or warehouse. Data mining works on mathematical analysis to drive pattern and trends that

exist in data. Data mining can be performed on data represented in qualitative, textual or multimedia form. Data mining applications can use a variety of pattern to examine the data. They include association, sequence or path analysis, classification, clustering and forecasting. Data mining involves a collection of tools and techniques for finding useful patterns relating the fields of very large database.

To provide a methodology in which process can operate, we divide data mining in to five stages. The stages begins with the collection of statistically representative data that should have enough information, the methodology make it easy for the decision making by applying statistical and visualization techniques, select and transform the usable variable, model the variable to predict outcomes and the accuracy of the model. The overview of these steps is as follows:

- **Sample:** The first stage is the collect data to construct predictive model, which is used to make predictions about an entire database. The validity of the sampled data is critically important because it the representative of the entire database. Two main characteristics of the sampled data should be maintained i.e size and quality of the sample data. The sample size is relatively easy and understandable. As the size of the sample data grows, it should with increasing clarity reflect the patterns that are present in the entire database.
- **Traverse:** This stage is used to searching for unanticipated anomalies to achieve understanding and ideas. The objective of this stage is to develop the deeper understanding of data and identify areas for further evaluation and analysis. This stage mainly involves data visualization, clustering, factor and correspondence. Data visualization facilitate understanding of data. To better understand a variable, unvaried plots of distribution of values are useful. To understand the relationship among the large

number of variables correlation tables are useful. Recent development in the data visualization, one can effectively view huge qualities of data in a meaningful view.

- **Editing:** This stage consists of editing of data by creating, selecting and transformation of variables to perform the selection process. The raw data in the database often requires modification and refinement before the mining algorithm can make effective use of them. Such modification includes derivation of new variables and other data modification that generally enhance the data set. Once the valid homogeneous sample was extracted, variable reduction methods were used to discard all the variable known to be insignificant, the result of which was to reduce the redundancy.
- **Model:** This stage consists of data modeling and permitting the software for automatic searching to get a combination of data that predicts a desired outcome. The partitioned data set was then sent in parallel into neural network, decision tree and regression algorithms.
- **Validation and Verification:** The last stage consists of this stage consists on assessing the data through evaluation in terms of reliability and usefulness of the findings through data mining and estimate its performance. In this stage we inspect the result make sense to the business experts and also verify that these results can be deployed in actionable way to improve the activities.

## 2. CHARACTERISTICS OF IDEA (INTERACTIVE DATA EXTRACTION AND ANALYSIS)

IDEA is a powerful and one of the fast auditing tools. The IDEA process involves three main processes involves Import, Analysis and Review Result. It import data from any source in any format and then in Analysis process contains Sort, Search and performs various calculations and statistical tests and then review results in the form of pivot table, reports, charts etc. The main characteristics of Idea are:

- **Improve Audit Work :** IDEA has the ability to analyze large amounts of data. For better performance of auditing requires testing the interrelations among data elements. To fulfill this requirement IDEA has variety of inbuilt functions. It provides powerful reporting utilities to extract the audit data into meaningful reports. With IDEA software auditor can improve performance and extend their capabilities.
- **Less technical skill required:** Due to user friendly interface it is used by financial auditors and IT auditors without any technical skill. To use IDEA software for auditing a very little training is required. IDEA's HTML-based Help, Informative User Guide assist users to perform various audit functions.

- **Speed Up Auditing Process:** IDEA process large amount of transactions with high speed. It creates custom views of the data and creates reports quickly. It imports data from almost any source so there is no need of converting format of the data. That speed up the whole process.
- **Detect and Handle Frauds easily:** IDEA is an effective and efficient way of identifying non-compliance with policies and procedures and fraud in transactions. Once such issues have been discovered and the control this can be used to quickly identify similar incidents throughout the company and apply corrective measures.
- **Professional Judgments required:** IDEA has various inbuilt function for calculation and to perform statistical tests. Auditing through IDEA software does not require any technical training but still require professional auditor to observe and analyze results.

IDEA lowers the cost of auditing and also speed up the work. But still professional judgments require observing and evaluating the results.

## 3. CHARACTERISTICS OF DATA MINING TOOLS

Data Mining tools are user friendly interface to carrying out automated data analysis task. The main features of data mining tools are:

- **Ability to Handle complicated problem:** The objective of Data Mining Software is automatically discovers useful information even from the complex data set. Data Mining Algorithms allows to perform knowledge discovery and used for prediction and searching data patterns even from the complex data easily.
- **Automated discover unknown patterns:** Data Mining automates the process of finding predictive patterns from large databases. Pattern discovery helps to find fraud detection and errors in the transaction that is the main task of auditing.
- **Scalability:** Data mining tools can handle large amount of data that makes the scalability is one of the important feature of it. For the Audit process this feature serves as key point.
- **Relatively high cost:** Data mining software is cheaper but still somewhat expensive than other software. Because in data mining users have to incur overhead costs like data preparation, analyzing and training costs which is relatively high.
- **Technical skill required:** Technical skill is required for Data Mining software users. User must have knowledge of various Data Mining Algorithms to choose appropriate algorithm according to the task requirements. Skills are also required to finding patterns of interest and to evaluate the results of findings.

Data Mining method has various features but there is lack of features as compared to other auditing tools. But still this tool is used to improve the efficiency of the professionals.

#### 4. COMPARISON OF DATA MINING TOOL WITH THE IDEA

These days Data Mining tools are now becoming used in Auditing process but due to cost effective feature of its other Auditing tools are more popular. The IDEA and Data Mining tools both have the Auditing capabilities but still IDEA is more interactively used as compared to the Data Mining tools. One of the main reasons is its cost Data Mining tools are relatively costly that the other auditing tools. IDEA has many inbuilt functions that makes auditing task easier. To operate IDEA user need a very little technical knowledge on the other hand Data Mining tools required technically skilled staff. IDEA is much more user friendly a compared to Data Mining tools because IDEA display statistics and other calculations in graphical form that explains the trend of data or any type of fraud and error easily. With the development of Data Mining tools for the auditing, it is possible to replace other popular auditing tools and also able to replace expertise required in some auditing processes.

#### 5. SUMMARY

If auditors come to know the pattern for transactions and expected error then Data Mining tool can be used to improve the efficiency of the professionals. The integration of Data Mining tools with the Auditing tools is relatively a new concept. If Data Mining tools for Auditing are developed then it make the auditing process fast, cheaper and relatively much more efficient.

#### REFERENCE

- [1] Fayyad U.M., "Data Mining and Knowledge Discovery: Making Sense Out of Data", *IEEE Expert*, **11 (5)**, 1996.
- [2] Fayyad U.M., Piatetsky-Shapiro, G., Smyth P., and Uthurusamy R., 1996. "Advances in Knowledge Discovery and Data Mining". Menlo Park, Calif.: AAAI Press.
- [3] Dunham, M.H. (2003). "Data Mining Introductory and Advanced Topics". *Upper Saddle River, NJ*: Pearson Education, Inc.
- [4] Yao Y.Y., "A Step Towards the Foundations of Data Mining", in: *Data Mining and Knowledge Discovery: Theory, Tools, and Technology V*, Dasarathy, B.V. (ed.), *The International Society for Optical Engineering*.
- [5] Alvin A., and James K., (2000). "Auditing: An Integrated Approach, New Jersey: Prentice-Hall.
- [6] Elder John F. IV, and Daryl Pregibon, (1996). "Advances in Knowledge Discovery and Data Mining", *A Statistical Perspective on KDD*.