Data mining is one of the new “buzz” words for the auditor, but in fact, data mining has been around as long as auditing has been around. Data mining can be described as the process of sorting through data and extracting relevant information and performing various analytical functions on the data prior to the actual on site audit. The information extracted would be used in various aspects of decision making, depending on the data being analyzed. The reason why this has become so popular is the sophistication of the software available. With this sophistication has come a change in our approach to the auditing process in the 21st century.

The Development of Data Mining in the Audit Process

Data mining has assumed increased importance in the audit process. Previously, detailed analysis was a time consuming process, with constraints based on the number of records that could be stored in available software. Databases gave us the ability to look at a far greater number of records, but are cumbersome to work with and are limited based on the source of the data. Today’s software allows the auditor to analyze large amounts of data from various sources (i.e., different hardware and software configurations), changing the way audits are performed. Audits today have become much more focused and allow the auditor to perform many tests on the data prior to going on site making the entire audit process more efficient and cost effective. The benefit of data mining has never been more evident than on the transactional audit.

In the late 1980s and early 1990s, the focus of an inspection of records was geared towards the overall administration of underwriting and claims. The accounting section of the audit was usually limited to conducting an audit trail on a limited basis, involving premiums and claims from a few files.

Companies seldom supplied electronic data. Often, data was provided in printed form, such as a printout from a mainframe. The detail would be basic at best, with little more than inception-to-date premium information or (more commonly) claims.

Samples of documents were selected based on materiality (e.g., large claims and/or large premium), plus a sample of other files based on various criteria such as every hundredth file or any files with premium/claims over a certain amount.

Audits with an accounting emphasis were usually carried out as part of a legal process, such as litigation or arbitration. These early audits normally involved the reconstruction of the accounts of ceding companies at certain dates. The objective was to verify the accuracy of the premium and claims representations made by the cedant to the reinsurer around the time of the renewal of a treaty, and to make sure that the ceding company had not understated the reserves at the date of the representation.
These audits would often take many weeks to complete, and would often require the full time input of an IT professional and several support staff, as well as the audit team.

Data mining has evolved rapidly in the last few years, due to the technological advances in key areas such as:

- The ability to store more data into smaller spaces; the information stored on a mainframe computer, can now be housed in much smaller computers and servers.
- The cost effectiveness of investment in computer technology; as a result, larger volumes of data can be handled relatively inexpensively.
- Faster processing and the increased sophistication of data mining software; data mining has become more user-friendly; algorithms have become more reliable, and are increasingly replacing older statistical methods.

**Benefits of Data Mining in the Audit Process**

The effect of these technological advances has been profound. Although the auditing of underwriting and claims functions is still important, a review of the accuracy and timeliness of accounts has become equally important. These days, an auditor can ask for access to an entire database, or request certain key elements from the database. The data mining software does not alter the structure of the database; however, it does allow the development of an exception analysis by converting the data from different systems to a common platform without manual intervention. Following the analysis of the data, the relevant items can be copied into an electronic file.

To an extent, analyses are usually time-constrained or subject to budgetary limits. In the early days, claims errors and under-reserving were much easier to determine than accounting errors, therefore the focus was much narrower. Technological advances in data mining mean that the analyses that used to take several weeks can now be done in a matter of hours. The software can usually be implemented by one or two experienced auditors, without the need for full time input from the IT professional and support staff. Analysis can now be done in much greater detail than before. Instead of reviewing summaries, data mining software can drill down to the transactional level and pick out single transactions from a population of several million.

Certain information may be kept separately, such as premium and claim information. As long as the databases contain at least one section in common (usually the policy number), the data can be merged. In the past, data anomalies were usually identified by accident where zero or negative items were identified by reviewing 10’s to 100’s of pages of mainframe reports, and in no instance could all such anomalies be identified. Now with the advances in data mining software the auditor can have a confidence that 100% of all such instances are identified in minutes. Obviously, these are basic tests but more complex functions, such as aggregate losses or the allocation of premiums based on Increased Limits Factors (ILF’s) would also be easier to calculate.
The ability to detect certain patterns becomes much easier when a database can be fully analyzed. For a reconstruction of accounts exercise, such as the early accounting audit described above, statistical information at certain dates can now be calculated easily.

Summary data, which was often provided in the early days, would not show the possible suppression of reserves at key dates, therefore the selection of large claims would not necessarily be relevant. The data mining software would be able to concentrate on certain files, where reserves had been raised around a particular date. In this way, the selection of files for testing would be more relevant.

There is a much greater ability to detect fraudulent activity. For example, insurers can detect whether certain expense billings are excessive or have been billed multiple times.

Data mining allows certain functions to be automated. Again, the construction of macros to automate queries used to be done exclusively by the IT professional. Although professional help may be required to automate the most complex data mining exercises, many functions can be done by the auditor.

Improvements in data mining have allowed the scope of audits to become much wider. These days, audits are not just carried out on ceding companies on behalf of reinsurers, but companies can now review their own data to determine such things as aging of claims, duplicate payments, and workload of claims staff.

In summary, the audit function has evolved as a result of the improvements in computer hardware and data mining software, and the variety of applications to the insurance industry continues to grow. This new technology has helped usher in a new era in reinsurance auditing.

-Don Wustrow & John Kelly for AIRROC Matters in 2011