

SPECIALIST READING

A Find the answers to these questions in the following text.

- 1 What tool is often used in data mining?
- 2 What AI method is used for the following processes?
 - a Separate data into subsets and then analyse the subsets to divide them into further subsets for a number of levels.
 - b Continually analyse and compare data until patterns emerge.
 - c Divide data into groups based on similar features or limited data ranges.
- 3 What term is used for the patterns found by neural networks?
- 4 When are clusters used in data mining?
- 5 What types of data storage can be used in data mining?
- 6 What can an analyst do to improve the data mining results?
- 7 Name some of the ways in which data mining is currently used.

DATA MINING

Data mining is simply filtering through large amounts of raw data for useful information that gives businesses a competitive edge. This information is made up of meaningful patterns and trends that are already in the data but were previously unseen.

The most popular tool used when mining is artificial intelligence (AI). AI technologies try to work the way the human brain works, by making intelligent guesses, learning by example, and using deductive reasoning. Some of the more popular AI methods used in data mining include neural networks, clustering, and decision trees.

Neural networks look at the rules of using data, which are based on the connections found or on a sample set of data. As a result, the software continually analyses value and compares it to the other factors, and it compares these factors repeatedly until it finds patterns emerging. These patterns are known as rules. The software then looks for other patterns based on these rules or sends out an alarm when a trigger value is hit.

Clustering divides data into groups based on similar features or limited data ranges. Clusters are used when data isn't labelled in a way that is favourable to mining. For instance, an insurance company that wants to find instances of fraud wouldn't have its records labelled as fraudulent or not fraudulent. But after analysing patterns within clusters, the mining software can start to figure out the rules that point to which claims are likely to be false.

Decision trees, like clusters, separate the data into subsets and then analyse the subsets to divide them into further subsets, and so on (for a few more levels). The final subsets are then small enough that the mining process can find interesting patterns and relationships within the data.

Once the data to be mined is identified, it should be cleansed. Cleansing data frees it from duplicate information and erroneous data. Next, the data should be stored in a uniform format within relevant categories or fields. Mining tools can work with all types of data storage, from large data warehouses to smaller desktop databases to flat files. Data warehouses and data