

# Chapter 19

## A Specialized Evaluation and Comparison of Sample Data Mining Software

**John Wang**

*Montclair State University, USA*

**Xiaohua Hu**

*Drexel University, USA*

**Kimberly Hollister**

*Montclair State University, USA*

**Dan Zhu**

*Iowa State University, USA*

### ABSTRACT

*Finding the right software is often hindered by different criteria as well as by technology changes. The authors performed an AHP (Analytic Hierarchy Process) analysis using Expert Choice to determine which data mining package was most suitable. Deliberating a dozen alternatives and objectives led them to a series of pair-wise comparisons. When further synthesizing the results, Expert Choice helped provide a clear rationale for the decision. The issue is, data mining technology is changing very rapidly. This chapter focused only on the major suppliers typically available in the market place. The method and the process that was used can be easily applied to analyze and compare other data mining software or knowledge management initiatives.*

### INTRODUCTION

Based on the *Knowledge Life Cycle* model, four stages of knowledge creation, knowledge storage/retrieval, knowledge transfer, and knowledge application have been proposed by Alavi and Leidner

(2001) and confirmed by Jennex (2006). “To be effective Knowledge Management Systems, KMS, must support the various knowledge management functions of knowledge capture, storage, search, retrieval, and use” (Jennex, 2006, p.3). Knowledge discovery is generally one of important stages or phases of KM. And while this incorporates identi-

DOI: 10.4018/978-1-60566-954-0.ch019

ifying critical knowledge (this may also be what this stage is called), using data mining to aid in knowledge discovery is appropriate as being a useful KM tool.

Data mining is a promising tool that assists companies to uncover patterns hidden in their data. These patterns may be further used to forecast customer behavior, products and processes. It is important that managers who understand the business, the data, and the general nature of the analytical methods, are involved. Realistic expectation can yield rewarding results across a wide range of applications, from improving revenues to reducing costs (Davenport & Harris, 2007; Porter & Miller, 2001; Wang, 2008). It is crucial to properly collect and prepare the data, and to check the models against the real figures. The best model is often found after managers build models of several different types or by trying different technologies or algorithms. This alone demonstrates the active role managers play in the data mining or other knowledge management processes.

Selecting software is a practical yet very important problem for a company (James, Hakim, Chandras, King, & Variar, 2004). However, not enough attention is given to this critical task. Current literature is quite limited since selecting software is such a complex problem, due to many criteria and frequent technology changes (Elder IV & Abbott, 1998; Giraud-Carrier & Povel, 2003). Haughton, Deichmann, Eshghi, Sayek, Teebagy, & Topi (2003) generally reviewed several computer software packages for data mining including SPSS Clementine, XLMiner, Quadstone, GhostMiner, and SAS Enterprise Miner. Corral, Griffin, & Jennex (2005) examined the potential of knowledge management in data warehousing from an expert's perspective. Jennex (2006) introduced technologies in support of knowledge management systems.

Firstly, this article will take a brief look at data mining today, through describing some of the opportunities, applications and available

technologies. We will then discuss and analyze several of the most powerful data mining software tools available on the market today. Ultimately we will also attempt to provide an analytical analysis and comparison among the brands we have selected. Our selection is based, in part, on our own experience using data mining software as well as writing data mining code, SQL code and our work as relational database administrators. For our analytical comparison we will be using *Expert Choice* (Version 11) advanced decision support software (Wang, Hu, Hollister, & Zhu, 2008).

## **DATA MINING SOFTWARE**

Data mining software analyzes- based on open-ended user queries- relationships and patterns that are stored in transaction data. Available are several types of analytical software: statistical, machine learning and neural networks, decision trees, Naive-Bayes, K-Nearest Neighbor, rule induction, clustering, rules based, linear and logistical regression time sequence, etc. Along the lines of Mena (1998) and Martin (2005), the basic steps of data mining for knowledge discoveries are:

1. Define business problem
2. Build data mining data base
3. Explore data
4. Prepare data for modeling
5. Build model
6. Evaluate model
7. Deploy model
8. Results

Note: Each of these steps contains managerial issues which must be addressed.

The key to knowledge discovery is a true understanding of your data and your business. Without this understanding, no algorithm is going to provide you with a result in which you should confide. Moreover, without this background you will not be able to identify the problems you are

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/specialized-evaluation-comparison-sample-data/41870](http://www.igi-global.com/chapter/specialized-evaluation-comparison-sample-data/41870)

## Related Content

---

### Modeling Operational Robustness and Resiliency with High-Level Petri Nets

Madjid Tavana, Timothy E. Busch and Eleanor L. Davis (2011). *International Journal of Knowledge-Based Organizations* (pp. 17-38).

[www.irma-international.org/article/modeling-operational-robustness-resiliency-high/53460](http://www.irma-international.org/article/modeling-operational-robustness-resiliency-high/53460)

### Knowledge Flow

(2014). *Harnessing Dynamic Knowledge Principles in the Technology-Driven World* (pp. 26-39).

[www.irma-international.org/chapter/knowledge-flow/83669](http://www.irma-international.org/chapter/knowledge-flow/83669)

### Organization of Lessons Learned Knowledge: A Taxonomy and Implementation

Subramanian Rama Iyer, Ramesh Sharda, David Biro, Joyce Lucca and Upton Shimp (2009). *International Journal of Knowledge Management* (pp. 1-20).

[www.irma-international.org/article/organization-lessons-learned-knowledge/4050](http://www.irma-international.org/article/organization-lessons-learned-knowledge/4050)

### Attributive Idea Evaluation: A New Idea Evaluation Method for Corporate Open Innovation Communities

Sven Schwarz and Freimut Bodendorf (2012). *International Journal of Knowledge-Based Organizations* (pp. 77-91).

[www.irma-international.org/article/attribution-idea-evaluation/61429](http://www.irma-international.org/article/attribution-idea-evaluation/61429)

### Exploring the Extent and Impediments of Knowledge Sharing in Chinese Business Enterprise

Wen Bing Su, Xin Li and Chee W. Chow (2012). *Conceptual Models and Outcomes of Advancing Knowledge Management: New Technologies* (pp. 266-290).

[www.irma-international.org/chapter/exploring-extent-impediments-knowledge-sharing/62427](http://www.irma-international.org/chapter/exploring-extent-impediments-knowledge-sharing/62427)