

## SELECTING CLASSIFICATION AND CLUSTERING TOOLS FOR ACADEMIC SUPPORT

Manying Qiu, Virginia State University, [mqiu@vsu.edu](mailto:mqiu@vsu.edu)

---

### ABSTRACT

*Classification and clustering are powerful and popular data mining techniques. Organizations use them to capture information, retain customers, and improve business performance. This paper presents a method for selecting data mining software for an academic environment based on its classification and clustering tools. This research applies the data mining software evaluation framework to evaluate three major commercial data mining software: SAS Enterprise Miner, Clementine from SPSS, and IBM DB2 Intelligent Miner. We added to the framework a criterion that became important in the Internet age. After ranking software on relevant criteria in the framework then purchase the best one that is affordable for academic support.*

**Keywords:** Data mining, classification, clustering, software evaluation

### INTRODUCTION

Data mining techniques have been successfully applied in many different fields including marketing, manufacturing, process control, fraud detection, and network management [1]. Data mining is supported by a variety of software employing basically the same tools. Because implementations are different it could be difficult to select appropriate software for a particular environment. This paper focuses on how to select appropriate data mining software for an academic environment based on the implementations of classification and clustering tools. First we explain why these tools are useful and how they work. Then we use a software evaluation framework to evaluate three major commercial data mining software: SAS Enterprise Miner, Clementine from SPSS, and IBM DB2 Intelligent Miner. SAS and SPSS statistical packages are commonly used for academic support, while IBM DB2 is widely used in industries.

A typical motivation for data mining is to develop an in-depth knowledge of customers and prospects that is essential for businesses to stay competitive in today's marketplace. Companies in every industry are trying to move towards the one-to-one ideal of understanding each individual customer and to use that understanding to make it easier for

the customer to do business with them rather than with a competitor. Companies are learning to look at the lifetime value of each customer. Organizations want to know which customers are worth investing money and effort to hold on to and which ones may be potentials for loss. Also, companies could use knowledge about customers to give each customer individualized attention. For example, a company could allow a customer to create his or her own personalized home page at the company's Website.

To learn about customers a company must gather data from various sources and organize it in a consistent and useful way. Usually a data warehouse stores the tremendous amount of data needed for doing business. This data must be analyzed, understood, and turned into actionable information. Data mining combs through the records to discover patterns, devise rules, come up with new ideas to do business, and make predictions about the future [2]. Data mining employs one or more computer learning techniques to automatically analyze and extract knowledge from data contained within a database [6].

One of the most powerful and popular machine learning methods is classification, which constructs a decision tree from examples. A decision tree (consisting of nodes and branches) represents a collection of rules, with each terminal node (i.e. leaf) corresponding to a specific decision rule. Decision trees are usually constructed beginning with the root of the tree at the top and proceeding down to its leaves at the bottom. This technique may be used directly for predictive or descriptive purposes, and there are a variety of algorithms for building decision trees.

Another popular technique is clustering. Clustering is to divide a heterogeneous population into a number of more homogeneous segments or clusters. Business frequently use clustering to carry out market segmentation study, group customers into clusters with similar buying behaviors and then develop marketing strategies to promote sales best for each cluster. Clustering is undirected knowledge discovery – no target variable is defined. Clustering techniques assign groups of records to the same cluster if they have something in common with the hope that this will make data mining techniques easier to discover meaningful patterns from the dataset. Clustering is often served as

prelude to some other technique of data mining or modeling. Data mining continues to grow in importance and corporations need to use the best products available for their tasks. There is no single classification tool and/or clustering tool that satisfies all user purposes. Instead, we must consider tools with respect to the particular environment and analysis needs. Collier et al [3] provide a framework for evaluating data mining tools. The framework categorizes criteria for evaluating data mining software into four areas: performance, functionality, usability and ancillary task support. Let's assume that a group of university professors are looking for classification and clustering tools for their academic support: teaching, research and consulting activities. We added to the framework a criterion "text miner availability" that became important in the Internet age. The framework will be applied to evaluate classification and clustering techniques in three major commercial data mining software, namely SAS Enterprise Miner, Clementine from SPSS and IBM DB2 Intelligent Miner. Although there are many data mining software which include classification and clustering techniques, these three representative software are selected to investigate how they cover components of classification and clustering process and whether they differ in criteria such as software architecture, data access, algorithm variety, prescribed methodology, visualization, user types and so on. The evaluation process should provide an example to help organizations compare characteristics of classification and clustering tools to select data mining software that meets their needs.

## DATA MINING SOFTWARE EVALUATION

### FRAMEWORK

The data mining software evaluation framework [3] is applied to evaluate classification and clustering tools in different data mining software. The framework employs the following categories, each having criteria that facilitate evaluation in the category and are relevant to academic teaching, research and consulting activities.

1. **Performance** – The ability to handle a variety of data sources in an efficient manner.

Criteria: (1) Software Architecture; (2) Heterogeneous Data Access

2. **Functionality** – the inclusion of a variety of capabilities, techniques, and methodologies for data mining.

Criteria: (1) Algorithmic Variety; (2) Prescribed Methodology

3. **Usability** – accommodation of different levels and types of users without loss of functionality or usefulness.

Criteria: (1) User types; (2) Data visualization

4. **Ancillary Task Support** -- allows the user to perform data cleansing, manipulation, transformation, visualization and other tasks that support data mining.

Criteria: (1) Data filtering; (2) Deriving attributes

We added the following criteria to the framework to include organizational and management issues.

5. **Text Miner Availability** – Data Mining Tools can only access and process structured, basically numerical data to perform decision tree induction and clustering. In the Internet age, there is an increasing volume of unstructured text data which includes document, email, web page, etc. We want to make sure the selected data mining tool will be able to categorize, search or personalize textual data with its text miner.

Criterion: Text Miner Availability

6. **Cost** – cost of licenses for academic server and workstations

Criterion: academic license price

ROI is extremely important to organizations when purchasing software. Because the "return" is usually measured in dollar value, we did not include it in the framework for an academic environment where profit is generally not the goal. However a similar measure of payoff could be designed for an academic environment. One could measure value to academic constituents (e.g. students) by how prevalent is each tool in industry (market share). This prevalence indicates the likelihood students can use their skill without having to learn a different tool when they take an internship or get a job. Professors may want to consider this ROI-like perspective when deciding on a purchase.

### DESCRIPTION OF DM TOOLS

We shall discuss each tool with respect to the framework criteria and estimate the cost for an academic environment.

#### SAS Enterprise Miner

SAS Institute's software started off in the mid 1970s as a 4GL based statistical package for financial and economic analysis. Over the years the SAS system has expanded to become a multi-faceted product providing organizations with a complete information delivery system. Enterprise Miner addresses the entire data mining process—all through an intuitive point-and-click graphical user interface. Combined with SAS data warehousing and on-line analytic processing (OLAP) technologies, it creates a synergistic, end-to-end solution that addresses the full spectrum of knowledge discovery [7].

**Software Architecture** Enterprise Miner sits on top of a large, bundled collection of SAS statistical products. Enterprise Miner is available in a standalone, workstation configuration or in a client/server configuration. In the latter case, you can perform analysis on both the workstation and the server simultaneously [9]. Enterprise Miner is a distributed client/server based system which integrates fully with the rest of SAS software, including OLAP and a tool allowing applications to be deployed via the Internet. The client/server enablement distributes data intensive processing to the most appropriate machine, as well as maintaining a central data source, and allowing access to diverse data sources from DBMS located on different servers [7]. SAS software also supports management by an administrator.

**Heterogeneous Data Access** SAS software has built its reputation on the ability to access, manage and analyze data from any source, and provides a range of drivers that allow full access to data stores. Enterprise Miner is able to access data in warehouse/data mart and over 50 different file structures. Supported data sources include all the major relational databases as well as non-SQL data sources, PC sources and ODBC compliant databases. SAS software has an MDDDB server which provides a platform independent specialized OLAP storage facility for rapid access by end users through various tools [7] [8].

**Algorithmic Variety** Enterprise Miner supports various models and algorithms for classification. With Enterprise Miner, you can access an integrated suite of advanced models and algorithms, including clustering, decision trees, linear and logistic regression to achieve analytical depth. Enterprise Miner supports classification and regression trees, CHAID and C4.5 algorithms for building decision trees, enhanced methods to evaluate a tree based on profit or lift objectives and prune accordingly, interactive growing/pruning of trees; new C-based algorithms and interactive thin client VC++ Windows tree results viewer [8]. Enterprise Miner supports clustering and segmentation of databases, using k-means clustering, self-organizing maps and Kohonen networks techniques [8]. The results of clustering can be passed to other nodes such as Decision Tree Node for explanation of the clusters formed. It can also be passed as a group variable that enables the user to automatically construct separate models for each cluster.

**Prescribed Methodology** Enterprise Miner provides a guiding, yet flexible, framework for conducting decision tree induction encompassing five primary steps (SEMMA)--

- Sample the data by extracting a portion of a data set.
- Explore the data by searching for unanticipated trends and anomalies.
- Modify the data by creating, selecting, and transforming the variables to focus the model selection process.
- Model the data by formalizing the patterns, from which predictions can be made.
- Assess the data by evaluating the usefulness and reliability of the findings from the decision tree induction process and estimate how well it performs.

Enterprise Miner user builds a process flow diagram in the workspace by choosing nodes from the tool bar and menus and using drag and drop to link them together. These objects allow users to choose from a wide range of node types, which represent the operations that users perform to effect specific data mining steps, such as sampling, data extraction and so on. It supports both automatic and interactive training; in other words, the software can build the entire tree automatically, or build the tree node by node with user input; users use their business intuition to specify the variables used in subsequent tree development from a list provided by the software.

**User Types** Enterprise Miner is designed for a combination of beginning, intermediate, and advanced users. A rich set of analysis and modeling capabilities is provided for business users and professional users with some necessary statistical insight and business knowledge.

**Data Visualization** Enterprise Miner supports visual analysis and reporting. In addition to graphs showing outputs from the data mining techniques, there are novel plots which show business users the financial benefits and risks, associated with a particular model. Each node has its own list of properties, parameters and values, which the user can view by clicking on the object, and edit if desired. Data and statistics associated with a decision tree node can be examined by clicking on that object. Graphics include 3D rotating charts and histograms are used to view the desired data. Users can watch a model being built in real time through a window, and can stop the operation any time they wish.

**Data Filtering** Data filtering is handled by the Filter Outliers node. It enables the user to identify and remove outliers from data sets. Users can eliminate rare values in class variables and/or extreme values in interval variables.

**Deriving Attributes** User can create new variables from existing ones in the Transform Variables node. Enterprise Miner provides default functions, such

as squares, inverse, exponential, standardized, square root and logarithms, but allows users to input their own computational formulae, and build expressions.

**Text Miner Availability** Text mining includes clustering algorithms, document categorization and data extraction.

**Cost** An academic server license is from \$40,000 to \$100,000, and a mainframe license is from \$47,000 to \$222,000. Data Mining for the Classroom costs \$8,000 for 100 workstations.

### **Clementine from SPSS**

Clementine was originally developed by Integrated Solutions Ltd (ISL), which was formed in 1989 by Dr. Alan Montgome and five colleges. Clementine was one of the very first products to bring machine learning to business intelligence, while providing a user interface that was intelligible to non-experts [4]. SPSS acquired ISL in 1998. SPSS itself was founded in 1968 and earned its reputation primarily as a provider of statistical software, plus graphical applications to represent those statistics [4].

**Software Architecture** Historically, Clementine was a client-only product in which all processing took place on the client itself. While external databases could be accessed, they were treated by the software as local files; this approach necessarily put a heavy burden on the client platform. In 1999, SPSS introduced a server-based version of the product, with middleware running on a middle-tier server taking the load off the client and using the superior performance of back-end databases to support in situ data mining [4].

**Heterogeneous Data Access** Because Clementine is an SPSS package, importing SPSS data is straightforward, as is data in SAS or CSB format. Clementine can read a variety of different file types including data stored in spreadsheets and databases. Data can be read from ASCII files in either free-field or fixed-field format. Data can be imported from a variety of other packages including Excel, MS Access, dBase, FoxPro and Paradox using the ODBC [10].

**Algorithmic Variety** Rule induction – SPSS supports C5.0 and C&RT decision tree algorithms. SPSS has introduced a graphical decision tree browser so that the user can select the most intuitive way to view decision trees. Rule sets can also be produced from C5.0 while C&RT can be adopted to predict numeric outputs.

Regression modeling - It uses linear regression to generate an equation and a plot. Logistic regression is also available [4]. For clustering, Clementine offers the traditional Kohonen and K-

means techniques along with a new TwoStep technique. TwoStep is a hierarchical clustering algorithm that begins by automatically generating a set of low-level subclusters, then recursively merging them into larger, more generalized clusters until the process can no longer be performed without sacrificing the internal cohesiveness of the higher-level cluster [5].

**Prescribed Methodology** SPSS (and ISL before it) has espoused the use of a formal methodology for data mining and it has been a member of the CRISP-DM (Cross Industry Standard Process for Data Mining) group since its foundation in 1996. This methodology defines a six stage process for data mining projects [4]:

- Business understanding – A number of Clementine Application Templates (CATs) are available as add-on modules that encapsulate, at least some extent, business understanding within specific sectors. Some CATs available are: crime, fraud, microarray, CRM, Web mining and Telco.
- Data Understanding – There are two aspects to data understanding: the extraction of data and the examination of that data for its utility in the proposed data mining operation.
- Data Preparation – Data preparation (and retrieval) is performed through ‘data manipulation’ operations.
- Modeling – Modeling (and visualization) is at the heart of Clementine. Clementine is a fully graphical end user tool based on a simple paradigm of selecting and connecting icons from a palette to form what SPSS calls a ‘stream’ to analyze the data under consideration.
- Evaluation – Evaluation is about visualizing the results of the data mining process to understand the data.
- Deployment – Output data, together with the relevant predictions, can be written to files or exported to ODBC compliant databases as new tables or appended to existing ones.

**User Types** From the outset, the aim of ISL was to build an integrated environment for data mining operations that could be used and understood by business people without the need to rely on technical experts [4].

**Data Visualization** A list of generic visualization capabilities includes tables, distribution display, plots and multiplots, histograms, webs in which different line thicknesses show the strength of a connection. Some visualization techniques are available in 3D as well as 2D and the user can overlay multiple attributes onto a single chart in order to try to get more information. For example, the user might have a 3D scatter diagram and then

add color, shape, transparency or animation options to make patterns more easily discernible.

**Data Filtering** The Filter node is used to rename fields or remove unwanted fields from analysis, for example those having invariant values or those having a high proportion of missing information.

**Deriving Attributes** The Derive node allows modifying data values or creating new fields as functions of others. Clementine Language for Expression Manipulation (CLEM) allows the user to derive new values

**Tax Miner Availability** Text Mining for Clementine uses the LexiQuest solution's linguistic extraction technology to access and process unstructured data. The text mining for Clementine supports extraction of concepts from text data stored in a database, determine frequencies and categories, and link results to structured data and display the document or document selected through text mining.

**Cost** A teaching license for 100 users costs \$4,000 per year.

### **IBM DB2 Intelligent Miner**

IBM proposed many of the principles that now constitute mainstream data warehouse technology. It introduced the term Information Warehouse in 1991, and developed data mining software in various forms over the years [11]. IBM believes that doing business in the New Economy requires personalization and timing—getting the right messages to the right people at the most opportune time. The business analysis results must be applicable in real-time. With Intelligent Miner, the user can have an end-to-end solution in which the results of data mining are driven back into operational applications [12].

**Software Architecture** IBM sees data mining as a key component of an information warehouse framework. It therefore normally implements data mining in an organization as part of a data warehousing architecture. The main processing component, the Mining Kernel, is a client/server system [11].

**Heterogeneous Data Access** Intelligent Miner is not specific to IBM data warehouses or databases. It can process information stored in DB2 or flat files on AS/400, AIX or MVS. If the data to be mined has been downloaded to a flat file, the data mining processor is capable of accessing this file directly. Decision tree induction operations can also be performed directly on Oracle or Terradata databases. In addition, a high-speed extract is

provided to import data into DB2 Universal Database from Oracle, Sybase, or DB2 for OS/390 databases [11] [13].

**Prescribed Methodology** IBM provides minimum user guidance. There is no single set way of using intelligent Miner. Different users utilize the tool set differently, using the operations alone or in combination, that best meet the needs of the business. However, a custom interface to select pre-defined subsets of the function is designed for business analysts.

**Algorithmic Variety** IBM offers two techniques for decision tree induction:

- Classification explores synergies between two different types of entity (e.g. between products and occupation of purchasers). It enables user to profile customers based on a desired outcome, such as propensity to buy high-end clothing. Tree induction creates models that are represented either as decision trees, or as IF THEN rules.
- Prediction explores changes in established patterns (often used in the identification of opportunities for cross selling). Linear regression is used for value prediction.

IBM DB2 Intelligent Miner Demographic Clustering automatically determines the number of clusters based on the measure representing how similar the records within the individual clusters should be.

**User Types** Users include expert mining analysts and business analysts. IBM has provided a set of technologies that allows developers who have an understanding of the issues facing decision tree induction and clustering to build their own solutions, based on the algorithms and functions that IBM provides. Typically new users are assisted by IBM consultants. IBM expects this software to appeal particularly to independent software vendors and large customers [11].

**Data Visualization** The main client window is an Explorer-like view, showing a directory of the data available by type. After the data has been mined, IBM provides a range of viewing options including charts and graphs, histograms and tree diagrams. Intelligent Miner grows its decision trees interactively with the user, asking the 'best question' at each node.

**Data Filtering** Data derived from some sources, such as operational systems, may not be of suitable quality for classification and clustering. Sometimes data needs to be pre-treated. Additionally, some databases may hold information which is not of a type which is useful for decision tree induction and clustering. IBM has developed

techniques that allow data to be filtered and reduced, for example, filtering variables out of input records, filtering records out of the input database and filtering records using a value set.

**Deriving Attributes** One may derive new variables from original input data using SQL expressions. For example one may derive aggregate values using SQL column functions such as AVG, SUM, and COUNT. One may calculate new values using SQL expressions.

**Tax Miner Availability** IBM Intelligent Miner for text includes a set of tools that enrich business intelligence solutions, including text analysis, full text search, Web crawler, and Web search.

**Cost** Academic users can obtain this software at no cost through the IBM Scholars program.

each category are assigned weights with respect to target needs and the total of these weights within each category equals 1.00 or 100%. Each category is also assigned a weight with the same respect and the total of these weights equals 1.00 or 100%. Next the tools can be scored for comparison. We select SAS Enterprise Miner as reference software because the professors are using other components of the SAS statistical package for their teaching, research and consulting activities. Enterprise Miner receives a score of 3 for each criterion, and other two types of software are rated against Enterprise Miner for each criterion using the following discrete rating scale:

Relative Performance	Rating
Much worse than Enterprise Miner	1
Worse than Enterprise Miner	2
Same as Enterprise Miner	3
Better than Enterprise Miner	4
Much better than Enterprise Miner	5

**COMPARATIVE ANALYSIS**

The comparative analysis of software will focus on the relevant evaluation criteria. The criteria within **Scoring by University Professors**

**Classification and Clustering Tools Evaluation**

Criteria	Weight	Enterprise Miner (reference)		Clementine		Intelligent Miner	
		Rating	Score	Rating	Score	Rating	Score
<i>Performance (0.20)</i>							
Software Architecture	0.5	3	1.5	3	1.5	2	1.0
Heterogeneous data access	0.5	3	1.5	2	1.0	2	1.0
<b>Category Score</b>			<b>3.0</b>		<b>2.5</b>		<b>2.0</b>
<i>Functionality (0.20)</i>							
Algorithmic Variety	0.5	3	1.5	3	1.5	2	1.0
Prescribed Methodology	0.5	3	1.5	4	2.0	2	1.0
<b>Category Score</b>			<b>3.0</b>		<b>3.5</b>		<b>2.0</b>
<i>Usability (0.20)</i>							
User Types	0.7	3	2.1	2	1.4	1	0.7
Data Visualization	0.3	3	0.9	4	1.2	3	0.9
<b>Category Score</b>			<b>3.0</b>		<b>2.6</b>		<b>1.6</b>
<i>Ancillary Task Support (0.15)</i>							
Data Filtering	0.6	3	1.8	3	1.8	3	1.8
Deriving Attributes	0.4	3	1.2	2	0.8	2	0.8
<b>Category Score</b>			<b>3.0</b>		<b>2.6</b>		<b>2.6</b>
<i>Text Miner Availability (0.25)</i>							
	1.0	3	3.0	3	3.0	3	3.0
<b>Category Score</b>			<b>3.0</b>		<b>3.0</b>		<b>3.0</b>
<b>Weighted Average</b>			<b>3.0</b>		<b>2.9</b>		<b>2.3</b>

**Software Architecture** IBM Intelligent Miner gets 2 for software architecture because it only uses

client-server architecture. Although almost all the organizations have implemented computer

networking, a stand-alone software architecture certainly benefits students or professors who want to do analysis on their own stand alone computers. Like SAS Enterprise Miner, Clementine allows users a choice between standard-alone and client-server versions. Therefore, Clementine gets 3.

**Heterogeneous Data Access** Because Clementine's access to files other than SPSS files is not straightforward Clementine gets 2. Intelligent Miner's access to data sources other than DB2 is not direct so it also gets 2 for this criterion.

**Algorithmic Variety** Like Enterprise Miner Clementine supports multiple models and algorithms for classification and clustering, so Clementine scores a 3. IM does not support as many algorithms as EM so it gets a 2.

**Prescribed Methodology** Enterprise Miner's SEMMA (sample, explore, modify, model and assess) provides a methodology that clarifies the data mining process. Clementine adopts CRISP-DM (Cross Industry Standard Process for Data Mining) method that has more components than SEMMA, so Clementine gets 4 in this criterion. Intelligent Miner's method has fewer components so it scores 2 for this criterion.

**User Types** Enterprise Miner can be utilized by a wide range of users having different skill levels. Clementine has a restricted focus on business, non-technical users, so it gets 2. Intelligent Miner aims at technical users so it gets 1 for this criterion.

**Data Visualization** Clementine is a fully graphical end user tool based on a simple paradigm of selecting and connecting icons from a palette to form a 'stream'. A stream may consist of any number of different attempts to analyze the data under consideration and the 'model', so Clementine gets 4. Intelligent Miner is as good as Enterprise Miner so it scores 3.

**Data Filtering** All three products have similar capabilities and are rated 3.

**Deriving Attributes** Clementine and IM do not have as many as predefined methods available for deriving attributes (e.g. statistical functions, mathematical functions, Financial functions, etc.) so both of them are rated 2.

**Text Miner Availability** Because all three data mining software have text miner available we score each a 3.

## CONCLUSION

According to weighted average scores SAS Enterprise Miner is the best (3.0) closely followed by Clementine (2.9) and Intelligent Miner (2.3). One could use this ranking to select the best software whose cost is within the budget. IBM Intelligent Miner is the cheapest, followed by Clementine and SAS Enterprise Miner.

The data mining framework is a valuable instrument to evaluate data mining software based on their classification and clustering tools and the framework could be tailored to meet specific mining needs. For example, in the academic and environments integration of numerical and textual data mining becomes increasingly important at the Internet age. Also one must carefully interpret the criteria so that tools can be evaluated properly against these criteria. The evaluation results in this paper do not necessarily apply to different environments. Other environments may require different criteria in the framework. For example, some issues such as software security may not be so important in an academic environment, while in a business software security is very important. Defining the user environment is a very important step before scoring the tools.

## REFERENCES

1. Barbara, D & Jajodia S. Applications of Data Mining in Computer Security, Boston, MA: Kluwer Academic Publishers, 2002.
2. Berry, M.J.A. & Linoff, G.S., Mastering Data Mining, New York, NY: Wiley Computer Publishing, 2000.
3. Collier, K, Carey, B, Sautter, D. and Marjaniemi, C. "A Methodology for Evaluating and Selecting data Mining Software," Proceedings of the 32<sup>nd</sup> Hawaii International Conference on System Sciences, 1999, IEEE.
4. Howard, P "Clementine from SPSS." Bloor Research, June 2003. <http://www.it-director.com/research.php?productid=660>
5. James, G. "Analysts' Darling," Intelligent Enterprise, Aug 31, 2001, [http://www.intelligententerprise.com/010831/413products1\\_1.jhtml](http://www.intelligententerprise.com/010831/413products1_1.jhtml)
6. Roiger, R.J. and Geatz, M.W. Data Mining: A Tutorial-Based Primer, Boston, MA: Anderson Wesley Publishing, 2003.
7. "SAS Enterprise Miner" (1998) [http://www.bloor-research.com/research\\_library.php?pid=397](http://www.bloor-research.com/research_library.php?pid=397)
8. "SAS Enterprise Miner: Unearthing the truth – profitable data mining results with less time and effort",

- <http://www.sas.com/technologies/analytics/data-mining.html>
9. "Hard Core Mining (SAS Institute's Enterprise Miner 4.1) (Evaluation)" *Intelligent Enterprise*, Oct 4, 2001 4(5) p 46, <http://80web6.infotrac.galegroup.com.proxy.library.vcu.edu/itw/intomark/796/504/38066.html>
  10. "Introduction to Clementine," SPSS Inc., Training Department <http://homepages.wmich.edu/~c1luk/ClemBase5.pdf>
  11. "IBM Intelligent Miner," (1998) [http://www.bloor-research.com/research\\_library.php?pid=399](http://www.bloor-research.com/research_library.php?pid=399)
  12. "IBM DB2 Intelligent Miner," <http://www-3.ibm.com/cgi-bin/software/>
  13. "DB2 Intelligent Miner for Data: Technical Detail," <http://www-3.ibm.com/cgi-bin/software/webtools/print/print.pl>