A METHODOLOGY FOR EVALUATING AND SELECTING DATA MINING SOFTWARE

Ken Collier, Ph.D.[1]
Center for Data Insight
Box 15600, NAU
Flagstaff, AZ 86011-1560
520.523.5412
Ken.Collier@nau.edu

Donald Sautter[2]
Center for Data Insight
Box 15600, NAU
Flagstaff, AZ 86011-1560
520.523.0370
Don.Sautter@nau.edu

Muralidhar Medidi, Ph.D.[3]
Center for Data Insight
Box 15600, NAU
Flagstaff, AZ 86011-1560
520.523.3759
Murali.Medidi@nau.edu

James Morgan, Ph.D.[4]
Center for Data Insight
Box 15600, NAU
Flagstaff, AZ 86011-1560
520.523.7385
James.Morgan@nau.edu

Michael Ratliff, Ph.D.[5]
Center for Data Insight
Box 15600, NAU
Flagstaff, AZ 86011-1560
520.523.6895
Michael.Ratliff@nau.edu

Curt Marjaniemi[6]
Center for Data Insight
Box 15600, NAU
Flagstaff, AZ 86011-1560
520.523.0370
Curt.Marjaniemi@nau.edu

Bernard Carey, Ph.D.[7]
Center for Data Insight
Box 15600, NAU
Flagstaff, AZ 86011-1560
520.523.38867
Bern.Carey@nau.edu

Guy P. Abramo[8]
KPMG Peat Marwick LLP
2001 M Street, N.W.
Washington, DC 20036-3389
gabramo@kpmg.com

1 Associate Professor of Computer Science and Engineering & Center for Data Insight, Technical Director
2 Center for Data Insight, Systems Administrator
3 Assistant Professor of Computer Science and Engineering
4 Professor of Computer Information Systems, Area Coordinator
5 Professor of Mathematics
6 Center for Data Insight, Knowledge Analyst
7 Center for Data Insight, Director
8 Managing Director, Data Warehousing Practice Group, KPMG Peat Marwick LLP

A METHODOLOGY FOR EVALUATING AND SELECTING DATA MINING SOFTWARE

Abstract

As data mining evolves and matures more and more businesses are incorporating this technology into their business practices. However, currently data mining and decision support software is expensive and selection of the wrong tools can be costly in many ways. This paper provides direction and decision-making information to the practicing professional. A framework for evaluating data mining tools is presented and a methodology for applying this framework is described. Finally a case study to demonstrate the method's effectiveness is presented. This methodology represents the first-hand experience using many of the leading data mining tools against real business data at the Center for Data Insight (CDI) at Northern Arizona University (NAU).[9] This is not a comprehensive review of commercial tools but instead provides a method and a point-of-reference for selecting the best software tool for a particular problem. Experience has shown that there is not one best data-mining tool for all purposes. This instrument is designed to accommodate differences in environments and problem domains. It is expected that this methodology will be used to publish tool comparisons and benchmarking results.

INTRODUCTION

Knowledge discovery in databases (KDD) remains a young and evolving field. Although underlying technology and algorithms are advanced largely in research laboratories, it is primarily the commercial developers who advance the application of these technologies to real business and scientific problems. Furthermore, the business users' demands on these tools continue to exceed the available technology. As the KDD field matures in response to these demands, it is relevant to question which data mining software vendors are positioned to dominate the market. Meanwhile business users face the daunting task of deciding which tool best suits their needs and budgets. Currently the dollar cost of these tools is substantial. However, the cost of selecting an improper data-mining tool for a particular application is even more costly in terms of personnel resources, wasted time, and the potential for acting on spurious results.

This paper proposes a methodology for selecting from among the assortment of commercially available data mining software tools. This methodology is based on firsthand experiences in data mining using commercial data sets from a variety of industries. The Center for Data Insight (CDI) at Northern Arizona University (NAU) is uniquely poised to provide a perspective of data mining applications ranging across platforms and across industries. The KDD studies taking place in the CDI involve a wide-variety of commercial software tools applied to business data from a variety of industries as well as scientific data. Analysis of real corporate data introduces issues that are not apparent in many of the sanitized data sets used for demonstration and training, e.g. missing values, interpretation of blanks, inconsistent categorical values, etc. These problems exist to some degree in most real data sets. Hence one must consider data mining tools in light of how well they accommodate real data. Finally, while these tools are widely referred to as "data mining" software, most encompass ancillary activities that surround the actual data mining step of the KDD process (Fayyad, et al, 1996).

The remaining sections of this paper presents a scoring framework, a methodology within which the framework is to be applied, a case study one CDI experience with this methodology, and a plan for future development of the methodology.

AN EVALUATION FRAMEWORK

At the core of the CDI's tool evaluation methodology is a scoring framework. Experience and research suggests four categories of criteria for evaluating data mining tools: performance, functionality, usability, and support of ancillary activities (Morrill, March 1998, Adriaans & Zantinge, 1996). These categories form the basis of the CDI evaluation framework.

---

[9] The CDI is a nonprofit research and development center for Knowledge Discovery in Databases issues. The CDI is embodied in a state-of-the-art laboratory which supports the best-in-class data mining tools with a staff that possesses expertise in both tools and data mining methods. The academicians in the CDI represent a collaborative effort between the colleges of engineering, business, and mathematics at NAU. The CDI is funded by the KPMG Peat Marwick consulting firm and maintains partnerships with many of the software vendors in this field. Visit http://insight.cse.nau.edu for more information.

Performance – As per Table 1 is the ability to handle a variety of data sources in an efficient manner. Hardware configuration has a major impact on tool performance from a computational perspective. Furthermore, some data mining algorithms are inherently more efficient than others (Nakhaeizadeh and Schnabl, 1997). This category focuses on the qualitative aspects of a tool's ability to easily handle data under a variety of circumstances rather than on performance variables that are driven by hardware configurations and/or inherent algorithmic characteristics.

Functionality – As per Table 2 is the inclusion of a variety of capabilities, techniques, and methodologies for data mining. Software functionality helps assess how well the tool will adapt to different data mining problem domains.

Usability – As per Table 3 is accommodation of different levels and types of users without loss of functionality or usefulness. One problem with easy-to-use mining tools is their potential misuse. Not only should a tool be easily learned, it should help guide the user toward proper data mining rather than "data dredging". KDD is a highly iterative process. Practitioners typically adjust modeling variables to generate more valid models. A good tool will provide meaningful diagnostics to help debug problems and improve the output.

Ancillary Task Support – As per Table 4 allows the user to perform the variety of data cleansing, manipulation, transformation, visualization and other tasks that support data mining. These tasks include data selection, cleansing, enrichment, value substitution, data filtering, binning of continuous data, generating derived variables, randomizing, deleting records, etc. (Adriaans & Zantinge, 1996). Since it is rare that a data set is truly clean and ready for mining, the practitioner must be able to easily fine-tune the data for the model building phase of the KDD process.

CRITERIA	DESCRIPTION
Platform Variety	Does the software run on a wide-variety of computer

CRITERIA	DESCRIPTION
	platforms? More importantly, does it run on typical business user platforms?
Software Architecture	Does the software use client-server architecture or a stand-alone architecture? Does the user have a choice of architectures?
Heterogeneous Data Access	How well does the software interface with a variety of data sources (RDBMS, ODBC, CORBA, etc)? Does it require any auxiliary software to do so? Is the interface seamless?
Data Size	How well does the software scale to large data sets? Is performance linear or exponential?
Efficiency	Does the software produce results in a reasonable amount of time relative to the data size, the limitations of the algorithm, and other variables?
Interoperability	Does the tool interface with other KDD support tools easily? If so, does it use a standard architecture such as CORBA or some other proprietary API?
Robustness	Does the tool run consistently without crashing? If the tool cannot handle a data mining analysis, does it fail early or when the analysis appears to be nearly complete? Does the tool require monitoring and intervention or can it be left to run on its own?

Table 1Computational Performance Criteria

CRITERIA	DESCRIPTION

CRITERIA	DESCRIPTION
Algorithmic Variety	Does the software provide an adequate variety of mining techniques and algorithms including neural networks, rule induction, decision trees, clustering, etc.?
Prescribed Methodology	Does the software aid the user by presenting a sound, step-by-step mining methodology to help avoid spurious results?
Model Validation	Does the tool support model validation in addition to model creation? Does the tool encourage validation as part of the methodology?
Data Type Flexibility	Does the implementation of the supported algorithms handle a wide-variety of data types, continuous data without binning, etc.?
Algorithm Modifiability	Does the user have the ability to modify and fine-tune the modeling algorithms?
Data Sampling	Does the tool allow random sampling of data for predictive modeling?
Reporting	Are the results of a mining analysis reported in a variety of ways? Does the tool provide summary results as well as detailed results? Does the tool select actual data records that fit a target profile?
Model Exporting	After a model is validated does the tool provide a variety of ways to export the tool for ongoing use (e.g., C program, SQL, etc.)?

Table 2 Functionality Criteria

CRITERIA	DESCRIPTION
User Interface	Is the user interface easy to navigate and uncomplicated? Does the interface present results in a meaningful way?
Learning Curve	Is the tool easy to learn? Is the tool easy to use correctly?
User Types	Is the tool designed for beginning, intermediate, advanced users or a combination of user types? How well suited is the tool for its target user type? How easy is the tool for analysts to use? How easy is the tool for business (end) users to use?
Data Visualization	How well does the tool present the data? How well does the tool present the modeling results? Are there a vari ety of graphical methods used to communicate information?
Error Reporting	How meaningful is the error reporting? How well do error messages help the user debug problems? How well does the tool accommodate errors or spurious model building?
Action History	Does the tool maintain a history of actions taken in the mining process? Can the user modify parts of this history and re-execute the script?
Domain Variety	Can the tool be used in a variety of different industries to help solve a variety of different kinds of business problems? How well does the tool focus on one problem domain? How well does it focus on a variety of domains?

Table 3 Usability Criteria

CRITERIA	DESCRIPTION
Data Cleansing	How well does the tool allow the user to modify spurious

CRITERIA	DESCRIPTION
	values in the data set or perform other data cleansing operations?
Value Substitution	Does the tool allow global substitution of one data value with another (e.g., replacing ‘M’ or ‘F’ with 1 or 0 for uniformity)?
Data Filtering	Does the tool allow the selection of subsets of the data based on user-defined selection criteria?
Binning	Does the tool allow the binning of continuous data to improve modeling efficiency? Does the tool require continuous data to be binned or is this decision left to user discretion?
Deriving Attributes	Does the tool allow the creation of derived attributes based on the inherent attributes? Is there a wide-variety of methods available for deriving attributes (e.g. statistical functions, mathematical functions, boolean functions, etc.)?
Randomization	Does the tool allow randomization of data prior to model building? How effective is the randomization? How efficient is the randomization?
Record Deletion	Does the tool allow the deletion of entire records which may be incomplete or may bias the modeling results in some way? Does the tool allow the deletion of records from entire segments of the population? If so, does the tool allow these records to be easily reintroduced later if necessary?

CRITERIA	DESCRIPTION
Handling Blanks	Does the tool handle blanks well? Does the tool allow blanks to be substituted with a variety of derived values (e.g., mean, median, etc.)? Does the tool allow blanks to be substituted with a user-defined value? If so, can this be done globally as well as value-by-value?
Metadata Manipulation	Does the tool present the user with data descriptions, types, categorical codes, formulae for deriving attributes, etc.? If so, does the tool allow the user to manipulate this metadata?
Result Feedback	Does the tool allow the results from a mining analysis to be fed back into another analysis for further model building?

Table 4 Ancillary Task Support Criteria

METHODICAL APPLICATION OF THE FRAMEWORK

Application of the basic framework works best within a larger assessment model. Using standard decision matrix concepts (Ulrich & Eppinger 1995), this model consists of the following phases:

1) Tool prescreening
2) Identify Additional Selection Criteria
3) Weight Selection Criteria
4) Tool Scoring
5) Score Evaluation
6) Tool Selection

The following sections describe the methods applied at each phase and show how the framework is used to support the methodology. During these steps a selection matrix is developed to aid in scoring and selecting the best tool.

Step 1: Tool Prescreening

The goal of step 1 is to reduce the set of tools being considered to a manageable number. Eliminating tools that clearly will not be selected due to rigid constraints of the organization or the tool vendor does this. As an example, if the organization has already made the decision that KDD software will run on a Unix server, then any tools that are not supported on a Unix platform can be eliminated. In our experience this is a simple, but valuable, step in the selection process. Infeasible tools that are not eliminated early simply serve to clutter the selection process unnecessarily.

Step 2: Identify Additional Selection Criteria

Unfortunately evaluating data mining tools is not simply a matter of selecting the best tool for all purposes. Instead a data mining organization must consider the tools with respect to their particular environment, and analysis needs. While the evaluation framework provides most of the technical criteria for selection, the aim of this step is to identify any additional criteria that are specific to a particular organization. Software cost is usually considered during this step in addition to such things as platform restrictions, end-user abilities, specific data mining projects, etc. Additionally, it is during this step that framework criteria are examined and irrelevant items are discarded if necessary.

Step 3: Weight Selection Criteria

Following step 2 the evaluator has five categories of selection criteria. These include the four groups represented by the framework (performance, functionality, usability, and ancillary task support) plus an additional group of organization specific criteria identified in step 2. During step 3 the criteria within each category are assigned weights so that the total weight within each categor y equals 1.00 or 100%. An example of this is provided for the Performance category in Table 5. This weighting must be conducted with respect to the intended use of the software. Consider an organization whose data warehouse is centrally located on a Windows NT server, and whose local area network consists exclusively of Windows NT workstations. Such an organization will probably assign a low weight to platform support since any other platforms on which the tool is supported do not matter.

Criteria	Weight	Tool A	Tool B	Tool C
Performance				
Platform Variety	.05			
Software Architecture	.05			
Heterogeneous Data Access	.10			
Data Size	.40			
Efficiency	.15			
Interoperability	.05			
Robustness	.20			

Table 5 Identifying and weighting selection criteria

Step 4: Tool Scoring

Once the criteria have been weighted with respect to a set of targeted needs, the tools can now be scored for comparison. Rather than scoring on some artificially absolute scale, scor ing is done relative to a reference tool. Generally the evaluator is predisposed toward a favorite tool for a variety of subjective reasons. This "favorite" should be selected as the *reference tool*. Any tool may be selected in the absence of a favorite. Th e reference tool receives a rating of 3 for each criterion.

Other tools are then rated against the reference tool for each criterion using the following discrete rating scale[10]:

Relative Performance	**Rating**
Much worse than the reference tool	1
Worse than the reference tool	2
Same as the reference tool	3
Better than the reference tool	4
Much better than the reference tool	5

Using this scheme a score is calculated for every criterion for each tool. These scores are then totaled to produce a score for each category. Finally, the categorical scores are combined in a weighted-average to calculate an overall tool score. Be default each criteria category recieves a weight of .20. However, adjusting these weights allows the evaluator to emphasize or de -emphasize particular categories of criteria. See Table 6 for a partial example of tool scoring. A more complete example is presented in the case study that follows.

Criteria	**Weight**	**Tool A (reference)**		**Tool B**		**Tool C**	
Performance (.30)		Rating	Score	Rating	Score	Rating	Score
Platform Variety	.05	3	.15	3	.15	4	.20
Software Architecture	.05	3	.15	3	.15	5	.25
Heterogeneous Data Access	.10	3	.30	4	.40	4	.40
Data Size	.40	3	1.2	2	.80	4	1.6
Efficiency	.15	3	.45	2	.30	3	.45
Interoperability	.05	3	.15	3	.15	4	.20
Robustness	.20	3	.60	1	.20	5	1.00
Performance Score		**3.0**		**2.15**		**4.1**	
Functionality (.20)		Rating	Score	Rating	Score	Rating	Score
Mining Techniques	.15	3	.45	4	.60	3	.45
...							

[10] Note that one may use a continuous rather than discrete scale to further increase the precision of the scoring. However, we have discovered that little is gained by doing so.

Model Exporting	.00	3	.00	1	.00	2	.00
Functionality Score		**3.0**		**3.8**		**1.85**	
Usability (.30)		Rating	Score	Rating	Score	Rating	Score
User Interface	.00	3	.00	2	.00	3	.00
...							
Domain Variety	.25	3	.75	3	.75	5	1.25
Usability Score		**3.0**		**1.8**		**3.95**	
Ancillary Task Support (.10)		Rating	Score	Rating	Score	Rating	Score
Data Cleansing	.15	3	.45	4	.60	5	.75
...							
Result Feedback	.05	3	.15	3	.15	4	.20
Ancillary Task Score		**3.0**		**4.7**		**4.25**	
Other Criteria (.10) ...							
Weighted Average		**3.0**		**4.52**		**3.51**	

Table 6 Tool Scoring Example

Step 5: Scoring Evaluation

Occasionally tool scores do not match the evaluator's intuition or subjective assessment. Although this evaluation methodology is designed to objectify an inherently subjective process, intuition should not be ignored altogether. Discrepancies between scores and intuition are generally due to incorrect weightings of criteria. If such a discrepancy exists, step 5 involves reviewing the weightings assigned to selection criteria and adjusting them if necessary. Through an iterative application of this methodology, we have seen that tool comparisons within this framework generally evolve to a justifiable tool comparison.

FRAMEWORK AUTOMATION

Using a simple spreadsheet application this framework can easily be automated. Data mining vendors currently participating in the Center for Data Insight include Angoss, Cognos, DataMind, ISL Decision Systems, SAS Institute, Silicon Graphics Inc., Thinking Machines, TriVida, and Unica. We have developed nine separate spreadsheet templates, each of which uses a different tool as the reference tool. In each template the criteria weighting defaults to an even distribution within each category. . When a client uses the CDI to evaluate tools a template can be selected and the weightings modified to suit their environment and data mining needs. Furthermore, we have designed the weighting so that the user simple ranks the importance of each criterion within a category on a scale of one to ten. These rankings are then converted into the corresponding percentage based on a simple mathematical formula.

CASE STUDY

Northern Arizona University's Office of Financial Aid routinely must determine how to disburse funding to prospective new students. Funding takes the form of grants, loans, scholarships, and gifts. One objective in disbursing these funds is to increase enrollment by offering monies to prospective students who would not otherwise choose to attend NAU. Prospective students fall into one of three categories: those who will attend NAU regardless of funding, those who will not attend NAU regardless of funding, and those who will attend NAU only if they receive financial support. The Office of Financial Aid is exploring data mining as a technology to help identify students who tend to fit in the third category. Similar to target marketing, success on this project will enable them to improve recruitment results through targeted funding.

At the time of this writing NAU is developing an enterprise data warehouse and a collection of data marts using Sun Microsystems hardware and Oracle Corp. software. However, financial aid data has been collected into a Microsoft Access database and student tracking and retention data has been collected into an older IDMS database and subsequently dumped to tape backup in flat-file format.

The Office of Financial Aid has a limited budget to purchase data mining software. Furthermore, its current operational data store resides on a modest Windows NT computer that is accessed by a small number of financial aid staff. This office already has a license to Cognos' on-line analytical processing (OLAP) software, Powerplay, so they have a natural predisposition toward Cognos' Scenario data mining tool.

The CDI was approached by the NAU Financial Aid office with a request to help them determine if data mining was a suitable solution to the ongoing funds disbursement problem and to determine if Cognos Scenario was truly the best tool for their needs and environment. Using the methodology presented in this paper, the CDI was able to help them conduct an object ive tool assessment. Here we briefly describe the steps in the process and show the resulting matrix.

Step 1: Although Financial Aid thought highly of SGI's Mineset tool, it was screened out since the purchase of SGI hardware was presumed to be prohibitive.[11] Cognos Scenario was also eliminated since its CHAID algorithm only handles categorical data and did not meet the needs of the client. Finally, since TriVida was still in beta-test at the time of this evaluation, it was omitted as well.

Step 2: The only criterion that the client wished to add to the framework was tool cost. Unfortunately the cost of each tool is not well-publicized so ballpark estimates based on discussions with sales reps was used to rate this criterion.

Steps 3 & 4: By combining the CDI staff's tool knowledge with the domain knowledge of a Financial Aid Office representative, the criteria were weighted and the remaining tools were scored. These scores are presented in Table 7.

Step 5: The first set of scores fav ored a tool that had many runtime errors when used against the Financial Aid sample data set. Upon reviewing the weightings, it was determined that too little emphasis was placed on performance, especially robustness. An adjustment of the weights resulted in the final matrix seen in Table 7.

<u>DATA MINING TOOL EVALUATION SCORING FOR NAU FINANCIAL AID</u>

Criteria	**Weight**	**Knowledge Seeker**		**Data Mind**		**Model 1**		**Clementine**		**Darwin**	
Performance	*0.15*	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**

[11] The CDI has since analyzed tool costs and has discovered that the cost of a reasonably powerful SGI server together with Mineset is very competitive with other data mining software costs with the added ability to use the server for other purposes.

Platform Variety	0.2	3	0.6	4	0.8	3	0.6	3	0.6	2	0.4
Software Architecture	0	3	0	5	0	3	0	3	0	2	0
Heterogeneous Data Access	0.15	3	0.45	3	0.45	3	0.45	3	0.45	2	0.3
Data Size	0.1	3	0.3	2	0.2	3	0.3	3	0.3	3	0.3
Efficiency	0.1	3	0.3	3	0.3	3	0.3	2	0.2	2	0.2
Interoperability	0.1	3	0.3	3	0.3	3	0.3	3	0.3	3	0.3
Robustness	0.35	3	1.05	3	1.05	3	1.05	1	0.35	2	0.7
Category Score			**3**		**3.1**		**3**		**2.2**		**2.2**
Functionality	*0.2*	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**
Algorithmic Variety	0.2	3	0.6	3	0.6	5	1	5	1	4	0.8
Prescribed Methodology	0.15	3	0.45	4	0.6	4	0.6	4	0.6	3	0.45
Model Validation	0.2	3	0.6	4	0.8	4	0.8	4	0.8	4	0.8
Data Type Flexibility	0.15	3	0.45	3	0.45	3	0.45	2	0.3	3	0.45
Algorithm Modifiability	0.05	3	0.15	3	0.15	4	0.2	4	0.2	4	0.2
Data Sampling	0.05	3	0.15	3	0.15	3	0.15	3	0.15	3	0.15
Reporting	0.2	3	0.6	4	0.8	5	1	4	0.8	4	0.8
Model Exporting	0	3	0	2	0	5	0	4	0	5	0
Category Score			**3**		**3.55**		**4.2**		**3.85**		**3.65**
Usability	*0.2*	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**
User Interface	0.2	3	0.6	3	0.6	3	0.6	4	0.8	2	0.4
Learning Curve	0.15	3	0.45	3	0.45	2	0.3	1	0.15	1	0.15
User Types	0.15	3	0.45	3	0.45	5	0.75	2	0.3	2	0.3
Data Visualization	0.2	3	0.6	2	0.4	2	0.4	2	0.4	1	0.2
Error Reporting	0.15	3	0.45	3	0.45	2	0.3	1	0.15	2	0.3
Action History	0.15	3	0.45	2	0.3	3	0.45	5	0.75	3	0.45
Domain Variety	0	3	0	2	0	4	0	4	0	4	0
Category Score			**3**		**2.65**		**2.8**		**2.55**		**1.8**
Ancillary Task Support	*0.25*	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**
Data Cleansing	0.2	3	0.6	3	0.6	5	1	4	0.8	4	0.8
Value Substitution	0.1	3	0.3	2	0.2	4	0.4	4	0.4	3	0.3
Data Filtering	0.15	3	0.45	3	0.45	4	0.6	5	0.75	5	0.75
Binning	0.05	3	0.15	4	0.2	3	0.15	2	0.1	2	0.1
Deriving Attributes	0.1	3	0.3	3	0.3	5	0.5	5	0.5	5	0.5
Randomization	0.05	3	0.15	3	0.15	3	0.15	3	0.15	3	0.15
Record Deletion	0.05	3	0.15	3	0.15	4	0.2	4	0.2	4	0.2
Handling Blanks	0.2	3	0.6	3	0.6	3	0.6	2	0.4	3	0.6
Metadata Manipulation	0	3	0	2	0	4	0	3	0	3	0
Result Feedback	0.1	3	0.3	3	0.3	3	0.3	5	0.5	4	0.4
Category Score			**3**		**2.95**		**3.9**		**3.8**		**3.8**
Other Criteria	*0.2*	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**	**Rating**	**Score**
Cost	1	3	3	2	2	2	2	3	3	2	2
insert others	0	3	0	0	0		0		0		0
Category Score			**3**		**2**		**2**		**3**		**2**
Weighted Average			**3**		**2.842**		**3.225**		**3.16**		**2.77**

Table 7 Result of Financial Aid Tool Evaluation

Step 6: The evaluation methodology suggested that Unica's Model 1 tool would be best for the Office of Financial Aid. Based on experience with all of these tools against a sample of Financial

Aid's data, the client agreed that Model 1 was well-suited for their environment and problem domain.

Prior to applying the tool evaluation methodology the client spent several hours on each tool, under the guidance of CDI staff, to develop an understanding of the data mining process and of the tools. The client liked many things about each tool and felt that it was difficult to make an "apples-to-apples" comparison of the tools. The evaluation process itself only required 2.5 hours to complete and eliminated a great deal of confusion as to which tool would best serve the client's needs.

SUMMARY

Experience with a variety of commercial tools and data sets has led to a data mining tool assessment framework and methodology for using the framework. The framework considers performance, functionality, usability, and ancillary task support to evaluate data mining tools. The assessment methodology takes advantage of decision matrix concepts to objectify an inherently subjective process. Furthermore, using a standard spreadsheet application this framework is easily automatable, thus rendering it easy and feasible to employ.

Data mining software is costly and generally accompanied by moderately steep learning curves. Selection of the wrong tool is expensive both in terms of wasted money and lost time. The methodology presented in this paper provides a means of avoiding the selection of an inappropriate tool. This framework should help practitioners avoid spending needless dollars only to discover that a particular tool does not provide the necessary solution. Furthermore, this methodology provides a method for publishing tool comparisons and evaluations in the literature.

Our experience has shown that there is no single best tool for all data mining applications. Furthermore, there are a several data mining software tools that share the market leadership. Although there have been side-by-side tool comparisons published in some of the database journals and publications, we believe that valid side-by-side tool comparisons must be made with respect to the intended usage domain. This methodology provides the flexibility to accommodate a wide-variety of environments and data mining foci.

The Center for Data Insight currently uses this methodology with clients who are seeking guidance in appropriate tool selection. An existing user of a particular data-mining product can use this process to determine if the cost of a software upgrade or a shift to another product is warranted. Preliminary results and feedback from CDI clientele suggest that this framework and methodology is extremely useful.

BIBLIOGRAPHY

Adriaans, P. and Santinge, D., "Data Mining", Addison-Wesley Longman, 1996.

Fayyad, U., Piatetsky-Shapiro, G. and Smyth, S, "Knowledge Discovery and Data Mining: Towards a Unifying Framework", Proceedings KDD-96, August 2-4, 1996, Portland, Oregon.

Morrill, L., "Enterprise Mining: More than a Tool", Database Programming and Design, vol. 11 no.2, February 1998.

Morrill, L., "Intelligence in the Mix", Database Programming and Design, vol. 11 no. 3, March 1998.

Nakhaeizadeh, G. and Schnabl, A., "Development of Multi-Criteria Metrics for Evaluation of Data Mining Algorithms", Proceedings KDD-97, August 14-17, 1997, Newport Beach, California.

Piatetsky-Shapiro, G., Brachman, R., Khabaza, T., Kloesgen, W. and Simoudis, E., "An Overview of Issues in Developing Industrial Data Mining and Knowledge Discovery Applications", Proceedings KDD-96, August 2-4, 1996, Portland, Oregon.

Ulrich, K. and Eppinger, S., "Product Design and Development", McGraw-Hill, 1995.