

APPENDIX A

Data Mining Products

A.1 BIBLIOGRAPHIC NOTES

The following paragraphs provide a superficial overview of data mining tools on the market. Links are provided to resources where more detailed information can be found.

4Thought

- Product: 4Thought
- Vendor: Cognos Inc.
- URL: www.cognos.com/products/4thought
- Functions: Prediction
- Techniques: Neural networks (MLP)
- Platforms: Windows

4Thought supports decision support applications, including forecasting. A spreadsheet interface is provided. Data can be accessed as text files, Excel, Lotus, dBase, Impromptu, Powerplay, and Scenario files.

AC²

- Product: AC²
- Vendor: ISoft
- URL: www.alice-soft.com/html/prod_ac2.htm
- Functions: Clustering, classification, prediction, segmentation
- Techniques: Decision trees
- Platforms: Unix, Windows

AC² is a complete data mining toolkit designed for knowledgeable users. AC² has a graphical object-oriented modeling language and C/C++ libraries. Interactive editing of trees is supported. It behaves as a multiplatform library of data mining functions.

Affinium Model

- Product: Affinium Model
- Vendor: Unica Corporation
- URL: www.unicacorp.com
- Functions: Classification, segmentation
- Platforms: Unix, Windows

Affinium Model provides CRM applications with segmentation capabilities.

AI Trilogy

- Product: AI Trilogy
- Vendor: Ward Systems Group, Inc.
- URL: www.wardsystems.com
- Functions: Classification, forecasting, prediction
- Techniques: Genetic algorithms, jackknife estimation, neural networks
- Platforms: Windows

AI Trilogy is a suite of three products: *NeuroShell Predictor*, *Neuroshell Classifier* and *GeneHunter*. Ascii, CSV, and Excel files are supported. The *NeuroShell Run-Time Server* is also included as a utility for trained neural networks to run in Excel and user programs.

Alice

- Product: Alice d'ISoft & Alice Server
- Vendor: ISoft
- URL: www.alice-soft.com/html/prod_alice.htm
- Functions: Clustering, correlation, segmentation
- Techniques: Decision trees
- Platforms: Metaframe, TSE, Windows, Unix

Alice d'ISoft is an interactive data mining tool that covers all aspects of the KDD process. Alice can access data in many different formats, including relational databases via both ODBC and OLE DB drivers. Alice integrates decision tree techniques with an OLAP engine. A companion tool, *Amadea*, supports data transformation and cleansing. The *Alice Server* version supports a client/server architecture.

AnswerTree

- Product: AnswerTree 3.0
- Vendor: SPSS Inc.
- URL: www.spss.com/answertree
- Functions: Classification
- Techniques: Decision trees [CHAID, Exhaustive CHAID, C&RT (a variation of CART), QUEST]
- Platforms: Client (Windows), Server (Solaris, Windows)

AnswerTree is an SPSS product used to create decision trees. As a data mining tool it targets profiling groups for marketing and sales. Four basic decision tree algorithms are used. Included are two CHAID algorithms, both of which SPSS has extended to handle nominal categorical, ordinal categorical, and continuous dependent variables.

Braincel

- Product: Braincel
- Vendor: Jurik Research and Consulting
- URL: www.jurikres.com/catalog/ms_bcel.htm
- Functions: Forecasting
- Techniques: Neural networks (back-percolation)

Braincel is an Excel add-in that uses neural networks to perform forecasting. An improved version of backpropagation called back-percolation is used. Promised Land Technologies developed Braincel, and Jurik Research developed the back-percolation approach.

BrainMaker

- Product: BrainMaker
- Vendor: California Scientific Software
- URL: www.calsci.com
- Functions: Forecasting
- Techniques: Neural networks
- Platforms: Macintosh, Windows

BrainMaker is neural network software that can be used with many different data sources, including Lotus, Excel, dBase, ASCII, or binary. It is one of the most popular NN software tools. A companion product, *NetMaker*, is included to assist in the construction of the neural networks. An optional package, *Genetic Training Option (GTO)*, uses a genetic algorithm to create several possible neural nets. Through several iterations of training and genetic evolution, the best network can be chosen. This process also helps to choose the best training data.

CART

- Product: CART
- Vendor: Salford Systems
- URL: www.salford-systems.com
- Functions: Classification
- Techniques: Decision tree (CART)
- Platforms: CMS, MVS, Unix (Linux), Windows

CART is a decision tree tool that uses the CART algorithm and boosting. Missing data are handled through specialized backup rules that do not always assume that all values for a missing attribute are the same. Seven different splitting criteria (including Gini) are used. Through the use of a data-translation engine, *DBMS/Copy*, data from over 80 different file formats (including Excel, Informix, Lotus, Oracle) may be used. An optional product, *TreeCoder*, allows the use of CART directly in SAS. *TreeViewer* allows trees created on a Unix platform to be viewed on Windows.

Clementine

- Product: Clementine
- Vendor: SPSS Inc. (formerly Integral Solutions, Ltd.)
- URL: www.spss.com/clementine
- Functions: Association rules, classification, clustering, factor analysis, forecasting, prediction, sequence discovery
- Techniques: Apriori, BIRCH, CARMA, Decision trees (C5.0, C&RT a variation of CART), K-means clustering, neural networks (Kohonen, MLP, RBFN), regression (linear, logistic) rule induction (C5.0, GRI)
- Platforms: HP/UX, IBM AIX, Sun Solaris, Windows NT

Perhaps the most unique feature of *Clementine* is its GUI approach to data mining, which Clementine pioneered in 1994. Through the use of descriptive icons, a user creates a data flow description of the functions to be performed. Each icon represents a step in the overall KDD process. Included are icons for such functions as accessing data, preparing data, visualization, and modeling. By dragging and dropping the icons onto the Clementine desktop, a stream of functions is created. Through the use of predefined templates of streams, Clementine provides support for common Web usage mining applications. To assist in the creation of sequences, Clementine uses Capri.

Clementine mines large data sets using a client/server model. When applicable, the server converts data access requests into SQL queries, which can then access a relational database. Clementine supports a wide variety of data formats. Clementine solutions are exported and deployed outside of Clementine using Clementine Solution Publisher.

Cubist

- Product: Cubist 1.10
- Vendor: RuleQuest Research Pty Ltd
- URL: www.rulequest.com/cubist-info.html
- Functions: Numerical modeling
- Techniques: Regression, rules
- Platforms: Windows, Unix (Linux, Irix, Solaris)

As a companion product to See5/C5.0, *Cubist* generates actual prediction values with the rules. Source code in C is provided so that models generated by Cubist can be embedded in applications.

Darwin

- Product: Darwin
- Vendor: Oracle Corporation
- URL: www.oracle.com/ip/analyze/warehouse/datamining/content.html
- Functions: Clustering, prediction, classification, association rules
- Techniques: Decision trees (CART), K-means, K-nearest neighbors, neural networks, regression (linear, logistic)
- Platforms: Windows, Sun Solaris, HP-UX

Darwin, or *Oracle Data Mining Suite*, provides a simple-to-use Windows-based GUI interface and parallel implementations of various data mining algorithms. It implements a complete KDD process model. Through OCI direct data access and ODBC, access to many relational systems is facilitated. Access to text and SAS files is also supported. Oracle Data Mining uses a client–server approach. C, C++, or Java code can be exported.

DataEngine

- Product: DataEngine
- Vendor: Management Intelligenter Technologien GmbH
- URL: www.dataengine.de
- Functions: Classification, clustering, decision trees, time series analysis
- Techniques: Decision trees, fuzzy rules, K-means, neural networks (MLP, Kohonen), regression (linear)
- Platforms: Windows

DataEngine supports many different data mining tasks as well as statistical methods. DataEngine supports many methods for data cleansing, data transformation, and handling missing data. Many types of data are supported via an import–export interface. With DataEngine ADL you can generate C code or produce DLLs, which can be incorporated in application code for subsequent use of a generated model.

Data Mining Suite

- Product: Data Mining Suite
- Vendor: Information Discovery, Inc.
- URL: www.datamining.com/dmsuite.htm
- Functions: Association rules, forecasting, prediction, rules
- Techniques: Rule induction

The Data Mining Suite is designed to mine large relational databases. The suite of tools actually consists of six different modules, each targeted to different data mining functions. Incremental data mining is supported, as are multidimensional mining and ROLAP. The results of all data mining are viewed as patterns or rules. The most unique feature of the Data Mining Suite is the fact that generated patterns are stored and are accessible using the *Pattern Query Language (PQL)*.

DataMite

- Product: DataMite
- Vendor: Logic Programming Associates Ltd.
- URL: www.lpa.co.uk/ind_pro.htm
- Functions: Rules
- Techniques: Decision trees, rule induction
- Platforms: ODBC, Windows

DataMite performs mining against relational databases, which can be accessed via ODBC. If-then rules are generated based on outcomes desired by the user.

DBMiner

- Product: DBMiner Analytical System
- Vendor: DBMiner Technologies Inc.
- URL: www.dbminer.com
- Functions: Association rules, classification, clustering
- Techniques: Decision trees, K-means
- Platforms: Windows

DBMiner accesses data from many sources, including Microsoft SQL Server, Excel, OLEDB, and other relational databases through ODBC. Users can use either a DMSQL interface or a GUI interface. Interface through MS SQL server's OLAP allows a cube view of data. There actually are three different products: *DBMiner AX 2000* targets association rules, *DBMiner SX 2000* performs sequential mining, and *DBMiner GX 2000* targets OLAP mining.

Decider

- Product: Decider and Decider-Online
- Vendor: Neural Technologies
- URL: www.neuralt.com
- Functions: Classification, prediction, rules
- Techniques: Neural networks

Decider and *Decider-Online* are DSS tools that incorporate data mining functionality to identify credit risks and fraud.

DecisionTime

- Product: DecisionTime and WhatIf
- Vendor: SPSS Inc.
- URL: www.spss.com/decisiontime
- Functions: Time series forecasting and analysis
- Techniques: ARIMA, exponential smoothing
- Platforms: Windows

DecisionTime is a time series forecasting tool. Through user interaction, DecisionTime chooses the best of the included models via a wizard to perform the forecasting or allows the more sophisticated user to build his or her own models. Input is accepted from many different types of files, including Excel, SPSS, ASCII, and an ODBC source. The *WhatIf?* part of the solution allows the user to change ARIMA model predictor variables to create various scenarios of the forecasts generated in DecisionTime.

Enterprise Miner

- Product: Enterprise Miner
- Vendor: SAS Institute Inc.
- URL: www.sas.com/products/miner
- Functions: Association rules, classification, clustering, prediction, time series
- Techniques: Decision trees (CART, CHAID), K nearest neighbors, regression (linear, logistic), memory-based reasoning, neural networks (Kohonen, MLP, RBF, SOM)
- Platforms: Client(Windows), Server (Unix, Windows)

Enterprise Miner from SAS implements the complete KDD process using their own process model called *SEMMA* (*Sample, Explore, Modify, Model, Assess*). An icon-based, point and click GUI (not unlike Clementine) creates a process flow to be performed by the data mining task. In addition, Enterprise Miner contains many tools for bagging and boosting, sampling, visualization, imputation, filtering, transformations, and model assessment. An experiment text mining feature is also available. Enterprise Miner generates the complete scoring formula for all stages of model development in the form of SAS, C, and Java for subsequent model deployment and scoring.

GainSmarts

- Product: GainSmarts
- Vendor: Urban Science
- URL: www.urbanscience.com/main/gainpage.htm
- Functions: Classification, clustering, prediction
- Techniques: Bayesian classification, decision trees (AID, CHAID), genetic algorithms, K-means, neural networks, regression (linear, logistic)
- Platforms: SAS, Windows

GainSmarts is a comprehensive product that supports all steps in the KDD process. It is flexible in the data sources that are supported, including any supported by SAS or ACCESS. Collaborative filtering and survival analysis are also supported.

HNC Risk Suite

- Product: HNC Risk Suite
- Vendor: HNC Software
- URL: www.hnc.com
- Functions: Prediction
- Techniques: Neural networks, rules

HNC's *Risk Suite* actually is a set of tools targeting risk analysis and prediction. The *Falcon Fraud Manager* is an integral part of this suite.

Intelligent Miner

- Product: Intelligent Miner
- Vendor: IBM Corporation
- URL: www.software.ibm.com/data/iminer
- Functions: Association rules, clustering, classification, prediction, sequential patterns, time series
- Techniques: Decision trees (modified CART), K-means, neural networks (MLP, back-propagation, RBF), regression (linear)
- Platforms: Windows, Solaris, AIX, OS/390, OS/400

DB2 Intelligent Miner for Data performs mining functions against traditional DB2 databases or flat files. It also has capabilities to access data in other relational DBMSs using ODBC. However to do this IBM's DataJoiner must be used. It is implemented using a client-server approach with a straightforward GUI interface provided to assist the user in choosing data mining functions. Several visualization techniques are used.

There are two other products in the IBM Intelligent Miner family. *Intelligent Miner for Text* performs mining activities against textual data, including e-mail and Web pages. It consists of text analysis tools, a search engine, *NetQuestion Solution*, and a Web crawler package. The text analysis tools include the ability to cluster, classify, summarize, and extract important features from a document. *NetQuestion Solution* is a set of tools to facilitate indexing and searching Web documents. *DB2 Intelligent Miner Scoring* allows SQL applications the ability to request data mining applications against a DB2 or Oracle database. It is a user-defined extension to DB2. It can be used to determine the actual score that a record has with respect to user-defined ranking criteria.

JDA Intellect

- Product: JDA Intellect (formerly Decision Series)
- Vendor: JDA Software Group, Inc.
- URL: www.jda.com/jdad/ProductsServices/Intellect_facts.htm
- Functions: Association rules, classification, clustering, prediction
- Techniques: Naive Bayes, decision trees, K-means, K nearest neighbors, neural networks (back-propagation, RBF)
- Platforms: Solaris, Windows

JDA purchased NeoVista in June 2001. The new *JDA Intellect* product is based on Neovista's *Decision Series* data mining suite. JDA Intellect actually is a set of KDD and data mining tools. *Seasonal Profiling Intellect* can be used to extract and describe profiles reflecting seasonal sales trends. *Channel Clustering Intellect* is targeted to the clustering of retail stores based on products sold. The Decision Series consists of many different data mining engines, including DecisionNet, DecisionCubist, DecisionTree, DecisionBayes, DecisionCluster, DecisionKmean, and DecisionAR. Each of these engines implements one or more of the data mining techniques. A GUI interface facilitates creation of data flow designs by naive users. More experienced users can use a command line and programming interface.

JDBCMine

- Product: JDBCMine
- Vendor: Intelligent Systems Research
- URL: www.intsysr.com/JDBCMine.htm
- Functions: Classification
- Techniques: Decision trees (C4.5)
- Platforms: Windows, Java

JDBCMine performs classification against JDBC and ODBC databases using C4.5. The decision trees created can be visually browsed, printed, or saved as JPEG files.

Kaidara Advisor

- Product: Kaidara Advisor (formerly KATE)
- Vendor: Kaidara
- URL: www.kaidara.com
- Functions: Classification, clustering, prediction
- Techniques: Case-based reasoning, decision trees

Kaidara Advisor (formerly *KATE*) is a data mining tool targeted to support CRM applications. Kaidara Advisor supports sales automation and intelligent catalogs.

KnowledgeSTUDIO

- Product: KnowledgeSTUDIO
- Vendor: ANGROSS Software Corporation
- URL: www.angoss.com/ProdServ/AnalyticalTools/index.html
- Functions: Classification, clustering, prediction, rules
- Techniques: Decision trees (CHAID), expectation-maximization, K-means, neural networks (MLP, RBF), regression (linear, logistic)
- Platforms: Windows, Server (Solaris, Windows)

KnowledgeSTUDIO is a complete data mining workbench that performs many different tasks. It reads data from all major statistical packages and can import data from relational databases using ODBC. The optional *KnowledgeSTUDIO Software Development Kit* can be used to create in-house data mining applications. Application code can be generated for Visual Basic, PowerBuilder, Delphi, C++, and Java and then embedded into user code. *KnowledgeSERVER* can be used to create a client–server architecture. It also can be used to access data directly in relational databases.

KnowledgeSEEKER

- Product: KnowledgeSEEKER
- Vendor: ANGROSS Software Corporation
- URL: www.angoss.com/ProdServ/AnalyticalTools/index.html
- Functions: Classification
- Techniques: Decision trees (CHAID, XAID, entropy-based algorithms)
- Platforms: Windows, Unix (AIX, HP-UX, IRIX, Digital Alpha, Sinux, Solaris, SCO, LINUX)

KnowledgeSEEKER has a graphically based GUI and supports decision tree classification algorithms.

LOGIT

- Product: LOGIT
- Vendor: Salford Systems
- URL: www.salford-systems.com
- Functions: Forecasting, hypothesis testing
- Techniques: Regression (logistic)
- Platforms: DOS, MacOS, Unix

LOGIT is a logistic regression tool. As with MARS, DBMS/Copy allows the use of many different data formats.

Magnify

- Product: Magnify
- Vendor: Magnify
- URL: www.magnify.com
- Functions: Prediction
- Platforms: Unix

Magnify offers complete solutions (technology suite, consulting) for customer-related prediction problems. Magnify targets the insurance, direct marketing, and government industries. The technology used is designed to be scalable to any size dataset.

Magnum Opus

- Product: Magnum Opus 1.3
- Vendor: RuleQuest Research Pty Ltd
- URL: www.rulequest.com/MagnumOpus-info.html
- Functions: Association Rules
- Techniques: Opus
- Platforms: Windows, Unix (Solaris, Linux)

Magnum Opus generates association rules using the measures of leverage, lift, strength, coverage, and support. Only rules that satisfy the desired measurement constraints are generated. Filtering of rules is also performed. Association rules can be generated for market basket-type data as well as other attribute values. Numeric attributes are automatically partitioned into subranges.

Mantas

- Product: Mantas
- Vendor: Mantas, Inc.
- URL: www.mantas.com
- Functions: Association rules, classification, clustering, link analysis, prediction, sequence analysis, time series
- Techniques: Decision trees, neural networks
- Platforms: Unix, Windows

Mantas, Inc., a spin-off from SRA International in May 2001, provides knowledge discovery solutions for the global financial services industry. The *Mantas Knowledge Discovery Platform* provides an extensible foundation that proactively collects, analyzes, highlights, and disseminates relevant, actionable information. The Mantas product suite includes five products. *Mantas Best Execution* is used to analyze brokerage trading and identify opportunities to improve execution quality. *Mantas Equities Trading Compliance* monitors trading behaviors to ensure adherence to industry rules and regulations. *Mantas Brokerage Fraud and Anti-Money Laundering* monitors security and monetary transaction for potential fraud and money laundering activities. *Mantas Brokerage and Investor Protection* monitors broker and investor behavior and trading activity to identify potential risks to both the investor and the firm. *Mantas Enterprise Anti-Money Laundering* is used to identify potential money laundering risks relating to enterprise products and transaction services.

MarketMiner

- Product: MarketMiner
- Vendor: MarketMiner Inc.
- URL: www.marketminer.com
- Functions: Classification, pattern recognition, prediction, segmentation
- Techniques: Decision trees (C4.5), KNN, regression (linear, logistic), statistical techniques
- Platforms: Windows

MarketMiner is a set of marketing analysis tools designed to automate the data mining and analysis process for a business user without advanced statistics knowledge.

MARS

- Product: MARS
- Vendor: Salford Systems
- URL: www.salford-systems.com
- Functions: Forecasting
- Techniques: Regression
- Platforms: Unix (Linux, Solaris), Windows

MARS performs forecasting via the *Multivariate Adaptive Regression Splines* technique. It fits separate splines to different intervals for the predictor variables. Automatic selection of predictor variables is performed. As with, CART, DBMS/COPY allows access to many different formats of data. A graphical GUI is provided.

Minotaur

- Product: Minotaur and Minotaur Transcure
- Vendor: Neural Technologies
- URL: www.neuralt.com
- Functions: Classification, prediction, rules
- Techniques: Neural networks

Minotaur and *Minotaur Transcure* are neural network products targeted to support financial applications. *Minotaur* is aimed at fraud detection analysis in the telecommunications industry. *Minotaur Transcure* targets improving the efficiency of credit card transaction processing.

Net Perceptions Retail Analyst

- Product: Net Perceptions Retail Analyst (formerly KD1)
- Vendor: Net Perceptions, Inc.
- URL: www.netperceptions.com
- Functions: Association rules, clustering

Net Perceptions Retail Analyst is a retail analysis and report tool targeted toward retail sales applications. It supports OLAP functions.

Oracle9i Database

- Product: Oracle9i Database
- Vendor: Oracle Corporation
- URL: www.oracle.com/ip/dep/otn/database/oracle9i/bi_dm.html
- Functions: Association rules, classification, prediction
- Techniques: Naïve Bayes
- Platforms: All platforms on which Oracle9i runs

Oracle9i Database has a data mining component embedded within it. The API used is based on a Java data mining approach called *Java Data Mining (JDM)*.

Partek

- Product: Partek
- Vendor: Partek Incorporated
- URL: www.partek.com
- Functions: Clustering, prediction
- Techniques: Genetic algorithms, neural networks, regression
- Platforms: Unix, Windows

Partek actually consists of several companion products designed to perform pattern recognition, exploratory data analysis, statistical inference, and predictive modeling. The pattern recognition product, *Partek Pro*, contains over 20 predefined similarity measures. Partek can access data imported from flat files, ODBC databases, and Web servers. Many normalization and scaling transformation techniques are supported. The *Pattern Visualization System* contains many tools to graphically view the data mining results.

PolyAnalyst

- Product: PolyAnalyst 4.4
- Vendor: Megaputer Intelligence Inc.
- URL: www.megaputer.com/products/pa/index.php3
- Functions: Association rules, classification, clustering, prediction
- Techniques: Decision trees, neural networks
- Platforms: Windows, Unix

PolyAnalyst can access data stored in relational databases using the ODBC interface. In addition, it can process flat files, MS Excel, and DBF files. A point and click object-oriented GUI is provided to facilitate use. Many different visualization tools are included.

Quadstone

- Product: Quadstone System
- Vendor: Quadstone
- URL: www.quadstone.com
- Functions: Prediction, profiling, segmentation
- Techniques: Decision trees, regression
- Platforms: Unix (Solaris, HP, IBM AIX) and Windows NT

Quadstone System is a comprehensive analytical CRM software tool that includes tools to create customer-oriented datasets; segment, profile, and model customer data with advanced data mining algorithms; and deploy customer selections, models and scores in real-time or batch modes. A sophisticated GUI interface and graphical visualization tools are provided.

Re:order

- Product: Re:order (formerly Capri)
- Vendor: Lumio Limited
- URL: www.lumio.com/products.reorder
- Functions: Sequential patterns

The *Re:order* software product targets the efficient discovery of sequences. Sequences may or may not be required to be contiguous in time. Through the use of templates, users indicate the type of sequences to be detected. Templates may include temporal constraints as well as information about required or optional items to appear in the sequence. Capri uses the *Predictive Modeling Mark-Up Language (PMML)* to represent sequences.

Scenario

- Product: Scenario
- Vendor: Cognos Inc.
- URL: www.cognos.com/products/scenario
- Functions: Classification, clustering, outlier detection
- Techniques: Decision trees
- Platforms: Windows

Scenario provides support for clustering and classification applications. Data may be input from various sources, including Excel, Lotus 1-2-3, and from relational databases using ODBC.

See5

- Product: See5/C5.0 1.15
- Vendor: RuleQuest Research Pty Ltd
- URL: www.rulequest.com/see5-info.html
- Functions: Classification
- Techniques: Decision trees, rules
- Platforms: Windows, Unix(Linux, Solaris, Irix)

The classification algorithm C5.0 is implemented in these two products: *See5* runs on Windows machines and *C5.0* runs on Unix. Source code in C is provided so that classifiers generated by See5/C5.0 can be embedded in applications.

S-Plus

- Product: S-Plus
- Vendor: Insightful Corporation
- URL: www.insightful.com
- Functions: Classification, clustering, hypothesis testing, prediction, time series analysis
- Techniques: ARIMA, correlation (Pearson), decision trees, hierarchical clustering (agglomerative, divisive), K-means, regression (linear, logistic, nonlinear, polynomial), statistical techniques (Jackknife, Monte Carlo),
- Platforms: Unix, Windows

The *S-Plus* data mining tool has versions for Unix (including Linux) and Windows, as well as client–server versions. S-PLUS can import or export data from ASCII, SAS, SPSS, Matlab, Excel and Lotus spreadsheets, and other formats. S-PLUS can also import data from databases via ODBC or directly from Oracle on Solaris. S-PLUS is extensible: analytics are created using the object-oriented “S” language, and user-defined C, C++, Fortran, or Java code can be incorporated into S routines. A GUI is provided on both Windows and Unix, and S-PLUS supports over 80 different charting types. Techniques are provided to handle missing data and for analysis of data with outliers.

STATISTICA Data Miner

- Product: STATISTICA Data Miner
- Vendor: StatSoft, Inc.
- URL: www.statsoft.com/dataminer.html
- Functions: Classification, clustering, prediction
- Techniques: ARIMA, decision trees (CART, CHAID), exponential smoothing, neural networks (Back-propagation, MLP, RBF, SOM), regression
- Platforms: Windows

STATISTICA Data Miner provides a comprehensive set of statistical methods to solve data mining problems. It offers a point and click iconic-based GUI to create a workflow description of the KDD and data mining tasks to be performed. Created Visual Basic code can be used to update or modify the tasks at a later date. Extended versions of traditional neural network techniques, association rule algorithms, CART and CHAID, and a wide variety of other techniques are included. The application offers options to process remote databases “in place” (without creating local copies), which greatly increases the performance when data repositories are very large.

SuperQuery

- Product: SuperQuery
- Vendor: AZMY Thinkware, Inc.
- URL: www.azmy.com
- Functions: Rules
- Techniques: Rule induction
- Platforms: Windows

There are two main versions of *SuperQuery*. The Office edition is designed to work with Excel and Access data files, whereas the Discovery edition can be used to access many different types of data sources, including Access, xBASE, Borland Paradox, Excel, text files, and ODBC databases.

SurfAid Analytics

- Product: SurfAid Analytics
- Vendor: IBM Corporation
- URL: www.ibm.com/surfaid
- Functions: Web mining

SurfAid Analytics is an IBM e-business service offering that provides Business Intelligence for Web sites. SurfAid identifies visitor attributes that directly relate to customer retention, navigation patterns, and buying habits. SurfAid is compatible across all servers and platforms, requires no software installation or prerequisite hardware, and is serviced entirely on SurfAid machines by SurfAid consultants. Customers send their Web logs to the IBM SurfAid facility for processing.

Visualizer Workstation

- Product: Visualizer Workstation
- Vendor: Computer Science Innovations, Inc.
- URL: www.csi-inc.com
- Functions: Classification, clustering, correlations, prediction, rules
- Techniques: Neural networks(RBF)
- Platforms: Windows

Visualizer Workstation facilitates the process of identifying trends and correlations in data, as well as outliers and exceptions thereto. It is a module of CSI's *Advisor Toolkit*, which includes software to build predictive models and perform additional data mining functions. CSI supports its own KDD development process called the *Cognitive Engineering Methodology*, which closely matches the CRISP-DM, with an extension to address ROI analysis.

WebAnalyst

- Product: WebAnalyst
- Vendor: Megaputer Intelligence Inc.
- URL: www.megaputer.com/products/wa/index.php3
- Functions: Web mining (prediction, patterns)
- Platforms: Windows

WebAnalyst provides Web usage mining functions in a client-server architecture. WebAnalyst not only analyzes Web usage logs, but also can be used to make real-time predictions as to the future behavior of a Web site visitor. These can then be used to dynamically personalize pages for the user.

WebMiner ASP

- Product: WebMiner ASP
- Vendor: Webminer
- URL: www.webminer.com
- Functions: Prediction, web mining

WebMiner ASP is a subscriber-based data mining tool for e-commerce Web sites. By evaluating both customer databases and clickstream data, rules are generated that are used to segment users allowing customization for future Web site visitors.

WizWhy

- Product: WizWhy
- Vendor: WizSoft
- URL: www.wizsoft.com/Why.html
- Functions: Prediction, rules (if-then and if-and-only-if)
- Techniques: Rule induction
- Platforms: Windows

WizWhy is a rule induction data mining tool that can be used for (1) analyzing the data, (2) issuing predictions, and (3) revealing unexpected cases that deviate from the rules. *WizWhy* accesses many different types of data, including dBase, MS Access, MS SQL, Oracle, OLE databases, ODBC, and ASCII.

XpertRule Miner

- Product: XpertRule Miner
- Vendor: Attar Software Ltd.
- URL: www.attar.com
- Functions: Association rules, classification, clustering
- Techniques: Decision trees, rules
- Platforms: ODBC, Windows

XpertRule Miner supports the complete KDD process through the use of an iconic click and drag GUI interface. It supports the CRISP DM KDD development process. Visualization of results includes various 2D and 3D graphs and figures. *XpertRule Miners* supports a multitier client-server architecture. Access to any database supported by an ODBC connection is allowed.

Attar provides a companion product to *XpertRule Miner*. *XpertRule Knowledge Builder* provides a data flow development environment for KDD applications. The output from *XpertRule Miner* can be exported to a Knowledge Builder application.

A.1 BIBLIOGRAPHIC NOTES

Several surveys of data model software tools have been published, including [GG99]. An extensive list of data mining software products is found at the *StatServ* home page: <http://www.statserv.com/datamsoft.html>. The *DBMS Online* page also provides a list of data mining vendors at: <http://www.dbmsmag.com/9807mbg.html> as does the *KDnuggets* homepage: <http://www.kdnuggets.com/software/index.html>. A recent tutorial provided an overview of data mining tools [IA98]. This tutorial contained a complete comparison of the 17 covered products, including a discussion of algorithm approaches used, input and output techniques, usability, and automation and bundling techniques included. A tutorial comparing leading data mining tools was presented at KDD in 1998 [IA98]. This appendix is kept up-to-date at www.engr.smu.edu/~mhd/book.