

## Insightful Miner Feature List

### Visual Workflow Environment

- Create self-documenting visual programs
- Intuitive drag-and-drop interface
- Link nodes together to describe analytic process
- On-screen annotations
- Node-level change-tracking for multi-user collaboration
- Visual confirmation of validity and caching
- Save and share worksheets as templates for best practices
- Export worksheet image to a file

### Data Access (Input and Output)

- Delimited ASCII files
- Fixed format ASCII
- Data dictionary support
- SAS®, SPSS®, Excel® & many other flat file formats
- ODBC access to compliant databases (Windows®)
- Native access Oracle®, DB2, Microsoft® SQL Server, Sybase

### Data Manipulation

- Powerful sampling, including stratified methods
- Row: Aggregate, Append, Filter, Partition, Sample, Shuffle, Sort, Stack, and Unstack
- Column: Bin, Create, Filter, Join, Modify, Reorder, Transpose and Normalize
- Automatically bin continuous variables
- Continuous, date, categorical and string data types
- Create or modify columns and filter rows using powerful expression language

### Data Cleaning

- Detect and repair missing values with variance-preserving methods
- Detect duplicates
- Missing value handling: drop, replace, impute and last observation carried forward
- Detect multi-dimensional outliers with leading-edge robust methods

### Exploratory Data Analysis and Visualization

- Trellis graphics quickly show structure of high-dimension data
- Univariate descriptive statistics, plus Correlation and Covariance calculations
- Table views and Visual Crosstabs rapidly slice and dice data
- Compare datasets for validation purposes
- 1-D Charts: Pie, Bar, Column, Dot, Histogram, Boxplot
- 2-D Charts: Scatterplot, Boxplot, Strip plot, Quantile-Quantile, Density
- Hexagonal Binning chart to view relationships between variables of very large data sets
- 3-D Charts: Contour, Level plot, Surface plot, Cloud plot
- Multivariate charts: Multiple 2-D plot, Scatterplot matrix, Hexbin Matrix, Parallel plot
- Time series charts: Line plot, High-Low plot, Stacked Bar plot

### Model Types, Algorithms and Visualizers

- Prediction and classification outcome models with basic and advanced model options
- Highly scalable algorithms: train models on very large data sets without the need for sampling or aggregation
- Decision trees for classification and regression with single-tree or ensemble techniques using Block Model Averaging™; K-Fold cross-validation, plus Gini and Entropy splitting rules
- Linear and logistic regression implemented as QR decomposition with Householder transformations
- Neural Networks with Multi-layer perceptrons
- Neural Network training methods: Resilient Propagation, Quick Propagation, Delta-Bar-Delta, Conjugate Gradient, and Online methods.
- Neural Networks: up to three hidden layers with user-specified number of nodes per layer
- Interactive Neural Network visualizer allows real-time control over learning process
- Naïve Bayes Classifier
- Principal components analysis
- Cox Proportional Hazard models for censored data with time-varying covariates

- Customer segmentation models with K-Means Clustering
- Collapsible tree viewer with interactive dendrogram
- Assess models with gain charts, lift charts, ROC charts and agreement matrices
- Variable importance tool for selection of the most significant variables
- Automatic calculation of dummy variable and interaction columns

### Scalability

- All components operate out-of memory and in-memory
- Unique "Pipeline Architecture" moves data in blocks through processing components
- Classical incremental techniques
- Block Model Averaging™ techniques
- Tailor size of blocks to optimize use of computing resources
- Automatic and manual control of caching to balance quick response with massive scalability

### Extensibility

- Compound nodes: create an entire process within a single node
- Create new nodes using S programming language
- Complete access to all S-PLUS 7 Enterprise Developer functions and libraries through S programming language
- Create custom predictive models, charts and reports
- Create and share user libraries of custom nodes
- Manage multiple custom libraries

### Deployment and Scoring

- Web-ready graphical reports
- HTML, PDF, PostScript and RTF model summary exports
- Non-interactive batch execution of all components\*
- Model ports support automatically-updating scoring components
- Score custom predictive models created using S-PLUS on very large databases
- Predictive Model Markup Language (PMML) model import and export

- Generate C code for run-time model scoring\*

Note: \* Requires Insightful Miner Server

### Systems Requirements

- Windows 2000, Windows XP Professional, Windows Server 2003 on 32-bit x86 processors. (Minimum system configuration: Pentium III with 512MB of RAM, 350Mb disk space on C or D drive.)
- Solaris 8 or Solaris 9 on 32-bit SPARC processors (Minimum system configuration: 350Mb disk space)
- Microsoft Terminal Services support
- 5X free disk to data (minimum), 10X recommended