WHY CHOOSE STATISTICA DATA MINER?

A flexible and powerful solution:

- Highly scalable, leading-edge algorithms, trees, boosting, ensembles
- All standard analytic modeling tools, logistic regression, linear models...
- Industry standard interfaces
- Ability to connect to all industry standard database and file formats
- Comprehensive API allows quick and straightforward connection to existing IT assets, including scheduling engines

Preferred tools and solutions recognized by industry experts, experienced users, and novices alike:

- Rated highest in customer satisfaction in largest independent surveys of users for three consecutive years¹
- Double-Victor in Predictive Analytics Industry Quadrant²
- Reviewers’ Choice: STATISTICA software offers superior price to performance ratio³
- Used by hundreds of thousands of users worldwide; supported by 30 offices on all continents

STATISTICA Data Miner is the highest ROI solution platform to support all data mining and predictive modeling analytics applications:

- Data management and preparation
- Modeling and model management, version control
- Deployment to databases or as code, e.g., Java, C#¹
- Efficient, high-performance real-time and batch scoring

Comprehensive BIG DATA support

Enterprise versions of Data Miner offer:

- Access to Hadoop and other distributed file system formats
- Proprietary algorithms to optimize calculation of analytics on extremely large data sets (with map-reduce analytics options)
- Multithreading optimizations to leverage multi-core CPU architecture and provide high performance even on relatively inexpensive hardware
- Managed automation to build, recalibrate, and deploy models, so you can work with thousands of models to extract big insights from big data

At the same time, STATISTICA Data Miner is highly customizable and can be tailored to meet very specific and demanding analysis requirements through its open architecture. It also features a straightforward, wizard-style Data Miner Recipes user interface that guides the user step-by-step with plain English prompts to create a desired model or a work-flow, or alternatively creates a model in one step by bypassing all prompts and using intelligent defaults selected by the program.

STATISTICA has achieved an unprecedented record of recognition and has received the highest rating in EVERY comparative review in which it has been featured since its first release in 1993.
UNIQUE FEATURES

Comprehensiveness

- STATISTICA Data Miner increases your chances of finding crucial insights with more than 14,000 analytic, graphical, and ETL/data-management functions and the most exhaustive collection of data mining and machine learning algorithms on the market at your disposal; it includes among many others:
  - Automatic Feature Selection
  - Automated Neural Networks Selector
  - Association and Sequence Rules
  - Classification and Regression Trees
  - Generalized Linear Models
  - EM, k-Means Clustering
  - C&RT, CHAID Trees
  - Generalized Additive Models
  - Independent Components Analysis
  - Partial Least Squares (PLS, NIPALS, PCA)
  - Multivariate Adaptive Regression Splines
  - Nearest Neighbor Methods
  - Random Tree Forests (Bagging)
  - Stochastic Gradient Boosted Trees

Data Miner Recipes

- This wizard-like user interface option is perfect for novices and for saving reusable workflows.

Repeatable Data Mining Templates

Through the Enterprise platform, Data Miner workflows and approaches can be quickly turned into highly customizable analytic templates that can be shared through the enterprise to create single-click predictive modeling, automated model calibration, or libraries of templates.

Model Management

- The Enterprise platform provides all tools for integrated model management, version control, and rapid deployment from development to production, without any programming.

Integrated Rules

- The Decisioning Platform™ (an optional extension for enterprise installations) is fully integrated with the Model Management tools, to support segmentation (pre-scoring) and policy rules, conditional scoring rules, or rules-based-only batch and real-time scoring.

Real-time and Batch Scoring

- Models and model flows generated by STATISTICA Data Miner can be used for batch scoring through Enterprise or real-time scoring through the Live Score® platform.

Customizability

- STATISTICA Data Miner integrates seamlessly into your existing architecture; simple customization of all analyses and interfaces is achieved via conventional drag-and-drop methods or the industry-standard Visual Basic scripting language; custom tools can be quickly and easily built for individual users or applications.

Comprehensive Data Management Tools

- The STATISTICA platform provides a large number of procedures for data cleaning and preparation, such as missing data removal, substitution, and imputation; data filtering: outlier and invalid value detection and removal; optimal binning of class variables; automatic feature selection and extraction methods; simple and comprehensive batch transformations using date functions, lag functions, logical operators, and statistical aggregation functions; and a wide variety of other features to support the critical data preparation step; these tools are optimized for processing extremely large data sets and are among the fastest in the industry.

Visual Data Mining

- STATISTICA Data Miner is recognized for its superior data visualization (e.g., Computerwoche, 2011); it is fully integrated into the STATISTICA platform, the most comprehensive solution for visual data exploration and visual data mining available, to accelerate the critical data preparation step (identification of outliers, valid data ranges, basic patterns).

Performance and Technology

- STATISTICA Data Miner features highly optimized and refined algorithms, highly-optimized multi-threading and multitasking (Client-Server), native support for 64-bit platforms in-place processing of remote databases, complete Web enablement, and unrivaled scalability to databases and other file systems.

Flexible Deployment Options, STATISTICA Live Score

- Deploy predictive models via XML/PMMML, C/C++, code, JAVA, SAS data step, and as-stored procedures. STATISTICA Data Miner is also seamlessly integrated with the STATISTICA Enterprise platforms for efficient deployment and updating of prediction models throughout the enterprise; models can be pushed to STATISTICA Live Score, a server-based real-time automatic scoring solution that has been successfully deployed at leading banks worldwide.

Compatibility and Open Architecture

- STATISTICA Data Miner supports all industry-standard data file formats as well as ODBC, OLE DB, and SQL; STATISTICA is SAP certified and also provides connectors to special-purpose databases (OSI PI); there is no proprietary database/data warehouse format or programming environment/language to learn; all functionality is exposed via COM (and SOAP in Client-Server version), and third-party procedures (e.g., R, SAS) are easily incorporated.

Unmatched User Friendliness

- STATISTICA Data Miner (rated highest in satisfaction in largest independent surveys of users for three consecutive years) offers the most advanced methods with simple and elegant user interfaces; the Data Miner Recipe interface provides a simple wizard-style interface to completely automate the creation of effective workflows; high-quality graphical output is easily created with a few mouse clicks; all aspects of the interface can be customized with industry-standard tools; extremely large data files can be efficiently analyzed even on a laptop computer.

Also Available:

- STATISTICA Decisioning Platform (an optional extension for enterprise installations) is fully integrated with the Model Management tools to support segmentation (pre-scoring) and policy rules, conditional scoring rules, or rules-based-only batch and real-time scoring. Designed for large-scale, validated deployment of predictive models in banking, insurance, and other industries.

- STATISTICA Text Miner (an optional extension of STATISTICA Data Miner,) features a large selection of text retrieval, pre-processing, and analytic & interpretive mining procedures for unstructured text data (including Web pages), with numerous options for converting text into numeric information (for mapping, clustering, predictive data mining, etc.), language-specific stemming algorithms, and options to support other unstructured input (e.g., sound or image files).

- STATISTICA Process Optimization is a powerful software solution that integrates all quality control charts, process capability analyses, experimental design procedures, and Six Sigma methods with a comprehensive library of cutting-edge techniques for exploratory and predictive data mining and optimization (of data mining models). Includes tools for general optimization of (prediction) models (e.g., campaign optimization) using Simplex, GA, and/or Grid optimization.

- STATISTICA Sequence / Link Analysis is a state-of-the-art, comprehensive application for extracting rules from complex data sets.

2 2011 Hurwitz Victory Index for Predictive Analytics
3 Computerwoche, March, 2011