

Viscovery[®] SOMine 5.2 – Data Sheet

Main Functions

Explorative data mining based on Self-Organizing Maps (SOM)

Data preprocessing

- Derivation of new variables using a built-in formula language.
- Definition and automatic management of nominal variables.
- Transformation of variables.
- Definable replacements in value ranges.
- Well-defined treatment of missing values.
- Outlier treatment.
- Statistical and deterministic sampling and over-sampling as well as sampling to accelerate model creation.
- Removals (including conditional removals using combinations of several attributes).

Data representation through Self-Organizing Maps

- High-performance computation of classical batch SOM with predefined training schedules and selectable granularity.
- Two-dimensional SOM data representation with interactive visualization.
- Ability to define the influences of individual attributes on the data ordering by setting attribute priorities.
- Automatic compensation of correlations in the data.

Data visualization and exploration

- SOM visualization with automatic color-coding and numerous graphical tools (attributes, clusters, segments, U-matrix, frequency, quantization error, etc.).
- Ability to define selections, labels and trajectories in the map.
- Ad-hoc data queries through selections on the model.
- Statistical group profiles for any selection – down to the microcluster level.
- Charts of important parameters (histograms, segment comparisons, etc.).

Statistical analysis and access to original data (only in Expert Edition and Enterprise Edition)

- Available for each workflow step and for each selection in the map (context dependent).
- Display of original data records corresponding to selected areas in the model.
- Descriptive statistics and histograms.
- Correlation and principal component analysis.
- Frequency table.
- Box plots and scatter plots.

Clustering, segmentation and classification

Visual cluster analysis

- Automatic cluster methods (SOM-Ward, Ward, SOM-Single-Linkage) with integrated cluster visualization.
- Separate display of cluster boundaries, cluster centers and inner structure.
- Interactively available statistical description and cluster profiles.

Definition of segments

- Definition of segments/classes based on statistical clustering.
- Manual adaptation of segment boundaries through mouse click selection.
- Any number of segmentations possible in a single model.
- Assignment of operational actions to segments as codes or formulas (“business rules”).
- Management of actions (per project, can be imported and exported).
- Integrated documentation of segmentations and segments.

Application of models to new data sets

- Assignment of each data record to segments or classes as well as evaluation of the defined formulas (business rules) and actions.
- Output of these computations as a flat table that contains for each data record the segment, the assigned actions, and optionally the results of formulas, segment statistics (including profile), and map node values.

Evaluation of applications

- Import of the application results and join with the model data.
- Visualization of the classification error over the map if actual classification is known.
- Charts for visual and quantitative evaluation of the model application.
- All model attributes and values of the evaluated application are available for visual inspection.

Features

Workflow orientation

- Optimized analysis workflows with proven default settings, providing goal-oriented navigation through the application (Create Data Mart, Create Model, Apply Model and Evaluate)
- Ability to generate model variations through workflow branching.
- Ability to copy parameterizations of workflow steps.
- Dedicated decision-making workflow for segment definition (including inline documentation).

Viscovery[®] SOMine 5.2 – Data Sheet

Reports

- Instant reports for each workflow step as well as the entire workflow available anytime.
- Automatic reporting of completed workflow steps.
- Integrated project documentation.

Usability

- Simple operation because the user is shielded from the technology core and the statistical algorithms.
- Info pop-ups in the map, the workflows and the tables as well as supportive context menus in numerous places.
- Ability to save current combinations of visible windows with selected properties as "Arrangements".
- Tables can be sorted, also by absolute value.
- Attributes can be renamed and described.
- Project administration in project directories and "clean directories" function.

Technical Information

Architecture

- File-oriented, stand-alone software suite implemented in Visual C++ using Microsoft Foundation Classes (MFC).

Interfaces and formats

Data

- Tab-delimited flat-files.
- Space-separated flat-files (read-only).
- Microsoft Excel files (read-only).
- SPSS files (*.sav).
- XML files (by default Viscovery XML, as "Customized XML" also adaptable to special formats).
- SOM_PAK model files (*.cod).
- *.DMS files: Viscovery Data Mart (proprietary file format optimized for analysis in Viscovery, accompanied by *.DMD and *.DMM files).

Viscovery models

- *.SOM files: Viscovery SOMine Model (SOM data representation including segmentations and business rules).

- Used by Viscovery SOMine to assign segments and actions, compute business rules and evaluate campaigns.

Viscovery Projects

- *.VSP files: Viscovery SOMine Project files. Contain all relevant information of a Viscovery SOMine analysis project in a proprietary format.

System requirements

Minimal requirements for a work station configuration (suggested configuration)

- 1 GHz CPU (3 GHz or higher).
- 512 MB RAM (2 GB or more).
- Windows XP SP2 (or higher).
- 24 bit color graphics, 1280x800 (or more).
- Viscovery SOMine requires approximately 22 MB disk space for the installation; to work comfortably on a project we recommend free disk space per project that is 10 times the size of the data file.

Basic Edition

- Ability to process up to 100,000 data records with up to 100 variables.

Expert Edition

- All features and functions of the Basic Edition.
- Context-sensitive display of the original data records corresponding to selected areas in the model or to the workflow steps.
- Extended statistical and graphical analysis of the data records corresponding to the selected areas or workflow steps.

Enterprise Edition

- All features and functions of the Expert Edition.
- Optimized performance for extremely high-dimensional data (N >2000).
- Ability to handle an unlimited number of data records and variables.