

# Viscovery<sup>®</sup> Predictor 5.2 – Data Sheet

---

## Main Functions

### Generation of linear and non-linear predictive models and scorings

#### Glossary

- Prediction: A model for the dependency of a target attribute (that is known in historical data, e.g. churn probability, revenues, loss) on other - descriptive - attributes that is used to predict the target attribute by extrapolating the modeled dependency into the future.
- Scoring: An ordering of the data records by their predicted value in order to
  - assess the quality of the predictive model;
  - to select data records for a campaign that have the highest or lowest predicted values (e.g. select the 10% of customers that have the highest probability to buy).

#### Data Preprocessing

- Definition of nominal variables.
- Removals (including conditional removal using a combination of several attributes).
- Statistical und deterministic sampling and over-sampling.
- Replacements.
- Transformations.
- Sampling to accelerate model generation.
- Definition of new variables using a custom formula language.
- Renaming of attributes, attribute descriptions.
- Treatment of outliers.

#### Technology / Local SOM Prediction (LSP)

- Combination of multi-variate statistical methods with SOM technology.
- Exclusive use of statistically valid models (white-box models).
- Account for non-linear dependencies in the data by restricting linear regressions to local subsets of the data ("local regression") – piece-wise linear approximation.
- Automatic optimization of sizes of these local data subsets.
- Automatic parameterization to avoid overfitting.
- Automatic compensation of correlations in the data.
- Well-defined treatment of missing values.
- Automatic optimization of the priorities of the descriptive attributes for the SOM (for optimal local regression models).
- Support of various regression methods like multi-linear regression, stepwise linear regression.

#### Model Validation

- Automatic sampling to partition the data into training and test data.
- Visual and quantitative validation using score charts, gains charts, and mean square error; graphical comparison of different models.
- Extensive graphical and tabular analytical information about the models.

- Display of statistical parameters of the local models in the SOM representation.

### Application of Predictive Models

- Computation of predicted values by applying the predictive model to new data.
- Result of these computations is a table of the predicted value for each data record.

### Definition of Campaign Groups

- An arbitrary number of campaign groups ("score groups") can be formed by selecting them in the score (e.g. to address the 10% of the customer base that have the highest predicted churn probability).
- Selection of the optimal campaign size via cost-benefit calculation using customizable formulas (e.g. customer value).
- Output of these computations as a flat table that contains for each data record (identified by the key) the predicted value and the score group.

### Evaluation of Models and Applications

- Control groups can be generated automatically.
- Visual and quantitative model and campaign evaluation by score charts and gains charts.
- All model attributes and additional values visually available.

## Features

### Workflow-Oriented

- Optimized workflows leading through the application (Create Data Mart, Create Model, Apply Model, Evaluate).
- Workflows can optionally be processed automatically.
- Integrated project documentation by description of completed workflows.

### Extensive Visual Representation

- Visualization of the Viscovery SOM as well as the local regression coefficients and additional statistical parameters (significance, determination coefficient, etc.) of the local models.
- Histograms, box plots, scatter plots.
- Gains charts, score charts.
- Charts of important parameters (e.g. group profile, etc.)
- Possibility to define labels and paths over the map.

### Statistical Information

- Available for each workflow step; context dependent.
- Data record browser shows the records the SOM is based on.
- Descriptive statistics.
- Correlation matrix.
- Histograms and outlier treatment.
- Principal Components Analysis (attributes can be selected).
- Frequency table.

# Viscovery<sup>®</sup> Predictor 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 statistical algorithms.
- Can be operated by business users.
- Info pop-ups in the workflow and in the SOM visualization, attribute descriptions are included.
- Branching of workflows and copying of workflow steps.
- Project management in project directories and “clean directories” function.
- Supportive context menus at numerous places.
- Tables can be sorted, also by absolute value.

## IT-Technicalities

### Architecture

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

### Interfaces and Formats

#### Data

- Most formats are “read-only”, some formats can also be written (results).
- Uses ODBC/OLEDB to access all common RDBMS available, such as Oracle, SQL Server, DB2, Informix, Sybase, etc.
- Tab separated flat files.
- Space separated flat files (read-only).
- Microsoft Excel files (read-only).
- SPSS files.
- XML files (by default Viscovery XML, can be adjusted to special formats on request – “customized XML”).
- \*.DMS files: Viscovery Data Mart (proprietary file format optimized for the analysis in Viscovery; accompanied by \*.DMD and \*.DMM files). Can be used in both Viscovery Predictor and Viscovery Profiler.
- Data marts can be exported to files (Text, SPSS, XML) and to databases.

### Predictive Models

- \*.SOM files: Viscovery Predictor models (SOM data representation including local regression models).
  - Used by Viscovery Predictor to compute the predicted values.
  - Real-time integration using Viscovery One(2)One Engine into third-party software applications (to compute predicted values and to provide all analytical information in the model).
  - \*.SOM files can also be accessed by Viscovery Profiler and Viscovery Decision Maker (does not impact the local regression models).
- PMML files (\*.VXM): Contain the global regression model in PMML format.
  - Used by Viscovery Predictor to compute the predicted values.
  - Real-time integration using Viscovery One(2)One Engine into third-party software applications to compute the predicted values.
  - These files can also be imported into the engines of other vendors that support PMML.

### Projects

- \*.VPP files: Viscovery Predictor project file. Contains all relevant information of a Viscovery Predictor analysis project in a proprietary format.
- Completed workflows (Create Data Mart, Create Model, Apply Model) can be exported to Viscovery Scheduler via drag & drop and can also be imported back (after Viscovery Scheduler has completed them).

## System Requirements

### Minimal requirements for a workstation configuration (recommended workstation 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).
- The installed Viscovery Predictor software requires approx. 22 MB disk space; to work comfortably on a project we recommend free disk space per project that is 10 times the size of the data file.