

Viscovery[®] SOMine 8 – Data Sheet

Extension Module | **Predict and Score**

Non-linear predictive models based on self-organizing maps

Main Functions and Features

General characteristics

- **Predict and Score** is an extension module of the Viscovery SOMine Suite. It requires the **Visual Explorer** core module and can optionally be combined with other modules of the suite.
- The module provides two additional workflows: the **Create Predictor** workflow guides the user through the creation and validation of linear, logistic and non-linear prediction models, the **Apply Predictor** workflow provides routines for the application of prediction models to new data, scorings and evaluation of prediction and scoring results for application and control group.

Prediction models

- Creation of prediction models with the dedicated **Create Predictor** workflow
- Deterministic and random partitioning of data into a training set and up to two test data sets
- Global prediction models (linear regression, logistic regression)
- Local prediction models combining SOM data representation and piecewise linear regression

Global linear regression

- Computation of global models as multivariate linear or, optionally, stepwise linear regressions
- Option for logistic approximation of binary target variables
- Model statistics for resulting regression and common estimates for determination coefficient, standard error, and prediction interval for selected confidence level
- Beta coefficients for independent variables, t-statistics, tolerance, and “linear influence” on the target variable
- Storage of linear regression models as PMML code in *.vxml auxiliary files, which can be directly imported into PMML-supporting engines from other vendors

Logistic regression

- Logistic regression and regularized logistic regression (lasso, ridge, elastic net) using established R routines
- Model statistics for resulting regression and common estimates for determination coefficient (McFadden), AIC, BIC and Log-loss
- Beta coefficients for independent variables, Z-statistics, and odds ratios

Local regressions and non-linearity analysis

- Based on a patented procedure that combines SOM technology with statistical “white-box” models, designed for explaining residual variance in the global linear model
 - Approximation of non-linear dependence of the target variable from the independent variables with piecewise linear local regressions, which are computed for local subsets of data across the SOM
 - Automatic optimization of sizes of these local data subsets, with parameterization to avoid overfitting
 - Automatic iteration of attribute priorities, to optimally represent non-linear dependences in the SOM
 - Visualization of the resulting SOM and associated regression coefficients for the local models, as well as of additional statistical estimates, such as significance, explained variance, and residuals
 - Charts with mean descriptives of local models, as well as variance estimates and gain factors for the overall model
-

Viscovery[®] SOMine 8 – Data Sheet

- Non-linearity diagnostics providing estimates for the overall model, such as error reduction due to the resolution of non-linearities and the “non-linear influences” of independent variables on the target variable

Graphical validation and comparison of models

- Visual and quantitative validation using score charts, gains charts, and prediction error
- Subdivision of the charts into up to 1000 groups
- Comparison of predicted values with actual values for each model
- Display of all models generated from the same data in the same chart
- Performance comparison of different models, using the model data or the test data sets

Definition and application of scoring models

- Guidance for definition and application of scoring models through the dedicated **Score** workflow
- Interactive definition of “score groups” in the model chart for assigning different measures to subsets of data
- Definition of objective functions using the built-in formula language (e.g., by computing a “customer value” from the prediction value, to select the optimal campaign size)
- Application of scoring models to new data marts by batch computation of prediction values
- Automatic random generation and administration of control groups
- Export of prediction results, score group and, optionally, objective function computation to tab-delimited flat-text files (*.txt) and Microsoft Excel files (*.xlsx)

Evaluation of application results

- Specification of additional attributes in the evaluation data mart, containing the actual value (for which the prediction was computed) and, optionally, control group information, for test or evaluation purposes
- Score, gains, and scenario charts for comparison of results from the application with the control group
- Comparison of actual values and predicted values for the application, control group, or entire data
- Simultaneous display in the charts of further attributes, which are available in the evaluation data mart