# Announcing CART® Version 5 Decision Tree

CART 5.0 takes the Salford Systems award-winning data mining software to a new level of analytical power, convenience, and ease of use. We have enhanced the industry's best interface with new functions and features and added significant new analytical capability to the core model-building engine. CART 5.0 is even easier to use for data exploration and model building, offering convenient new controls for model evaluation and testing, scoring, deployment of models in stand alone production systems, and integration into existing data analysis systems. CART 5.0 is built on Salford Systems new data mining platform and can be licensed with TreeNet™ in an integrated data mining suite.

## MAJOR ENHANCEMENTS

### Improved data access

-Full support of character variables.
-Categorical predictors are no longer required to have contiguous variables.
-Direct support for plain delimited text ASCII files, CSV,. txt, etc.
-Direct support for SAS® Version 8 files, including v8x, sas7bdat and UNIX SAS files,
-No need to convert data. Read and write directly to hundreds of native Windows and UNIX file formats.

### Long variable names

-Up to 32 character names allowed. Longer names are temporarily shortened for display purposes and a cross-reference table is printed.

### Long text data

-Character data can contain text of any length, permitting you to work with extra long URLs in web mining.

### Variable labels

-Automatically recognize labels from SAS, SPSS, and other data management systems and add your own labels. No length limit imposed.

### Large number of variables

-Read up to 8,000 columns of data in standard CART 5. Our bioinformatics/chemical data mining version handles hundreds of thousands of variables.

## PREPROCESSING DATA REPORTS

### Summary data reports

-High-speed data summary includes means, univariate stats, frequency distributions and percentile tables. All are available in a single compact report that is exportable to Excel.

## INTERFACE, DISPLAY AND REPORTS

### Streamlined model setup

-Quickly and easily specify the target, predictor, categorical, weighting, and auxiliary variables in a single setup tab.

### Auxiliary variables external to model

-Display within-node summary stats and tables for any data variable, even if the variable is not in the model.

### New tree navigator displays

-Navigator display color codes all trees with equivalent accuracy (one standard error rule), helping you to select smaller or larger trees using statistical criteria.
-Bar chart below the tree diagram displays which nodes contain most of the data.

### Separate sub-tree and node displays

-Separate display of child nodes facilitates easy study of a specific split anywhere in the tree.
-Display train versus test for any specific node in the tree or for all nodes in a single display.

### Flexible tree detail display

-Display the tree in full flow-chart detail with either training or test data information within the nodes.
-Scalable tree diagrams allow you to shrink or expand the tree and font from 30% to 300% of normal size.
-Separate controls for the terminal nodes allow display of different details for the terminal versus internal nodes.

### Automated gains chart overlay

- Generate an automated overlaid gains chart to compare models. Manually select the models to be included or automatically overlay all models in a session.

### New scoring features

- Directly score with an on-screen tree using the size of tree displayed.
-Report every tree, or specify specific trees.
-Easily add any on-screen report with a right-mouse click.
- Extract any tree or collection of trees from a committee or ensemble of trees for scoring.
- Select trees on the basis of accuracy or lift/gains/ROC criteria.

### Modeling session summary

-Summarize modeling sessions in a table with one row per model.
-Display growing method, core control settings (priors, costs), and tree performance (relative error, %correct in each class).

### Automated reports

-Define a report by selecting the elements of standard CART output to be included, then obtain this report for specific models on request or automatically generate reports for all models.
-Rich Text Format (.rtf) document is ready to be published to the Web via MS-Word.
-Improved prediction results, gains, lift charts.

### Export tree code

- Export trees as SAS-compatible code, C, or PMML (predictive modeling markup language).

## NEW ANALYTICAL FEATURES

### Entropy splitting rule

-A close cousin of the CART classic gini rule, the entropy rule has a strong following in the computer science world and can sometimes yield quite good results.

### Missing value indicator predictors

-Missing value indicators are added to the list of model predictors.
-These predictors serve to inform the user of missing value patterns and can sometimes substantially improve the model.

### Enhanced penalties

-Penalize variables based on fraction missing, number of levels in a categorical predictor, or to reflect acquisition or other costs.

SALFORD SYSTEMS
8880 Rio San Diego Drive
San Diego, CA 92108
Tel: 619-543-8880
Fax: 619-543-8880
www.salford-systems.com