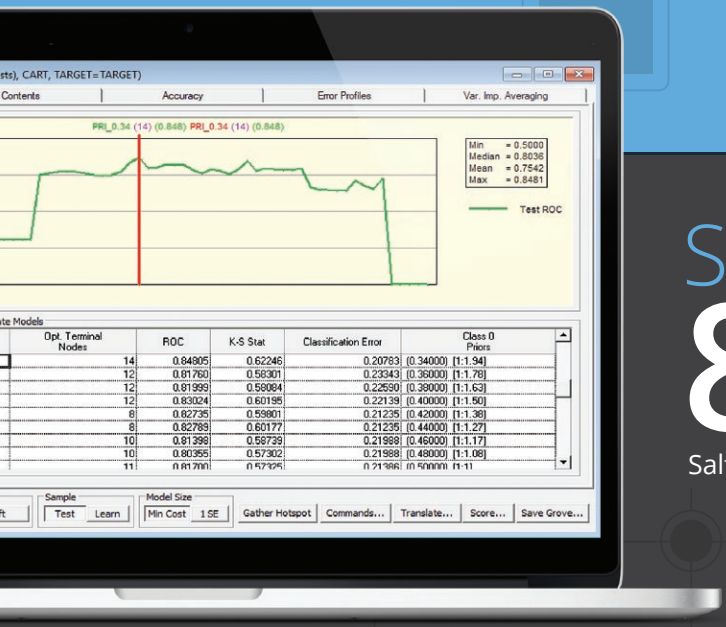
SALFORD SYSTEMS
a Minitab company



SPM® 8.2

Salford Predictive Modeler®

SPM SALFORD PREDICTIVE MODELER® | CART® | Random Forests® | MARS® | TREENET®

SPM® 8.2

The SPM Salford Predictive Modeler® software suite is a highly accurate and ultra-fast platform for developing predictive, descriptive, and analytical models from databases and datasets of any size, complexity, or organization.

The SPM software suite's data mining technologies span classification, regression, survival analysis, missing value analysis, data binning and clustering/segmentation to cover all aspects of your data science projects. SPM algorithms are considered to be essential in sophisticated data science circles.

Features

- 70+ pre-packaged automation scenarios inspired by the way leading model analysts structure their work.
- Tools to relieve gruntwork, allowing the analyst to focus on the creative aspects of model development.
- Regression, Classification, and Logistic Regression enhanced to support massive datasets.

The SPM software suite's automation accelerates the process of model building by conducting substantial portions of the model exploration and refinement process for the analyst. While the analyst is always in full control we optionally anticipate the analysts' next best steps and package a complete set of results from alternative modeling strategies for easy review.

"Do in one day what normally requires a week or more using other systems!"

SPM® 8.2 Product Versions

ULTRA: The best of the best. For the modeler who must have access to leading edge technology and fastest run times including major advances in ensemble modeling, interaction detection and automation.

PROEX: For the modeler who needs cutting-edge data mining technology, including extensive automation of modeling experiments typical for experienced data analysts and dozens of extensions to the Salford data mining engines.

PRO: A true predictive modeling workbench designed for the professional data miner. Variety of supporting conventional statistical modeling tools, programming language, reporting services, and a modest selection of workflow automation options.

BASIC: Literally the basics. Salford Systems award winning data mining engines without extensions or automation or surrounding statistical services, programming language, and sophisticated reporting. Designed for small budgets while still delivering our world famous engines.

Features
SPM® 8.2

- Performance improvements
- Descriptive statistics controls
- Model-based imputation or imputation using summary statistics
- Correlation analysis and Multi-dimensional Scaling
- Exact test partitioning fractions
- KEEP and EXCLUDE controls
- Expanded Set of Performance Statistics: Variance of ROC and Kolmogorov-Smirnov (K-5) Statistic
- New Time Series support mechanisms
- New automated model experiments
- New scoring distribution controls and visualization
- Recursive feature elimination

CART®

Classification and Regression Trees

CART® software is the ultimate classification tree that has revolutionized the field of advanced analytics, and inaugurated the current era of data science. CART is one of the most important tools in modern data mining.

Others have tried to copy CART, but no one has succeeded, as evidenced by accuracy, performance, feature set, built-in automation and ease of use.

Designed for both non-technical and technical users, CART can quickly reveal important data relationships that could remain hidden using other analytical tools.

Technically, CART is based on landmark mathematical theory introduced in 1984 by four world-renowned statisticians at Stanford University and the University of California, Berkeley. Salford Systems' implementation of CART is the only decision tree software embodying the original proprietary code. The CART creators continue to collaborate with Salford Systems to continually enhance CART with proprietary advances.

Patented extensions to CART are specifically designed to enhance results for market research and web analytics.

Newest**Features**

SPM® 8.2

- Unsupervised learning for advanced anomaly detection
- Association analysis using hotspots
- Forced splits
- Scalable limits on bottom nodes
- Automated prior search
- Automated bias penalty

CART supports high-speed deployment, allowing Salford Systems models to predict and score in real time on a massive scale.

TreeNet®
Gradient Boosting

TreeNet® software is Salford Systems' most flexible and powerful data mining tool, responsible for at least a dozen prizes in major data mining competitions since its introduction in 2002.

SUPREME ACCURACY

TreeNet adds the advantage of a degree of accuracy usually not attainable by a single model or by ensembles such as bagging or conventional boosting. As opposed to neural networks, TreeNet is not sensitive to data errors and needs no time-consuming data preparation, preprocessing or imputation of missing values.

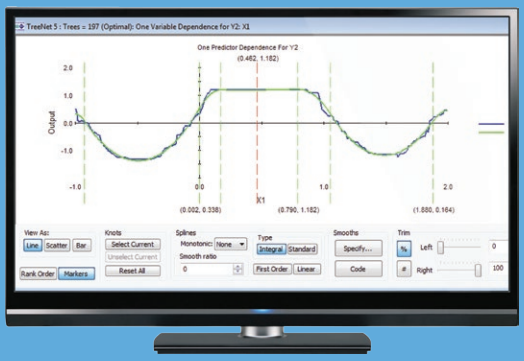
TreeNet's robustness extends to data contaminated with erroneous target labels. This type of data error can be very challenging for conventional data mining methods and will be catastrophic for conventional boosting. In contrast, TreeNet is generally immune to such errors as it dynamically rejects training data points too much at variance with the existing model.

ADVANCED FEATURES

Interaction detection statistics establish whether interactions of any kind are needed in a predictive model, and searches to discover specifically which interactions are important. The interaction detection system not only helps improve model performance (sometimes dramatically) but also assists in the discovery of valuable new segments and previously unrecognized patterns.

NextGen
SPM® 8.2

- Regularized Gradient Boosting™ (RGBOOST)
- Building Random Forests with TreeNet
- RuleLearner®: TreeNet's accuracy plus the interpretability and transparency of regression and CART
- New loss functions including differential lift
- Interaction Strength Reporting: discover, measure, and plot the most important interactions
- Advanced Controls: subsampling and influence trimming
- Monotone constraints
- Vary the number of predictors
- Customized starting values
- Model building with TreeNet splines



MARS®

Automatic Non-linear Regression

MARS® software is ideal for users who prefer results in a form similar to traditional regression while capturing essential nonlinearities and interactions. The MARS approach to regression modeling effectively uncovers important data patterns and relationships that are difficult, if not impossible, for other regression methods to reveal.

REGRESSION & CLASSIFICATION

The MARS model is designed to predict numeric outcomes such as the average monthly bill of a mobile phone customer or the amount that a shopper is expected to spend in a web site visit. MARS is also capable of producing high quality classification models for a yes/no outcome. MARS performs variable selection, variable transformation, interaction detection, and self-testing, all automatically and at high speed.

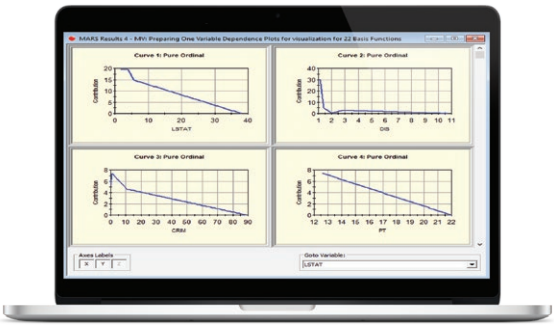
OUTSTANDING PERFORMANCE

Areas where MARS has exhibited excellent performance include forecasting electricity demand for power generating companies, relating customer satisfaction scores to the engineering specifications of products, and presence/absence modeling in geographical information systems (GIS).

Features

SPM® 8.2

- Build a series of models varying the maximum number of basis functions (Automate BASIS)
- Build a series of models varying the smoothness parameter (Automate MINSPAN)
- Build a series of models varying the order of interactions (Automate INTERACTIONS)
- Build a series of models varying the modeling speed (Automate SPEED)
- Build a series of models using varying degree of penalty on added variables (Automate PENALTY MARS)



Random Forest®

Breiman and Cutler's Random Forests

Random Forests® software is a bagging tool that leverages the power of multiple alternative analyses, randomization strategies, and ensemble learning. Its strengths are spotting outliers and anomalies in data, displaying clusters, predicting future outcomes, identifying important predictors, replacing missing values with imputations, and providing insightful graphics.

CLUSTER AND SEGMENT

Much of the insight provided by Random Forests is generated by methods applied after the trees are grown and include new technology for identifying clusters or segments in data as well as new methods for ranking the importance of variables. The method was developed by Leo Breiman and Adele Cutler of the University of California, Berkeley, and is licensed exclusively to Salford Systems. Ongoing research is being undertaken by Salford Systems in collaboration with Professor Adele Cutler, the surviving co-author of Random Forests.

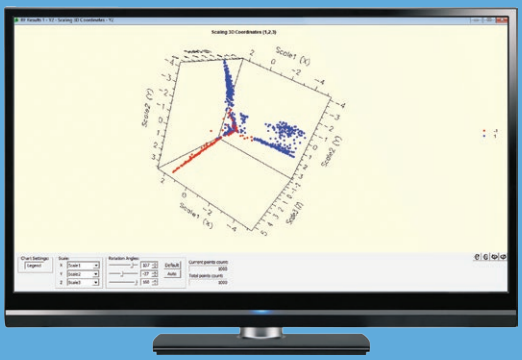
SUITED FOR WIDE DATASETS

Random Forests is a collection of many CART trees that are not influenced by each other when constructed. The sum of the predictions made from decision trees determines the overall prediction of the forest. Random Forests is best suited for the analysis of complex data structures embedded in small to moderate data sets containing less than 10,000 rows but potentially millions of columns.

Advantages

SPM® 8.2

- Easy parameter optimization with automates
- Special RF clustering enhancements
- Extreme Trees: Special controls for random splitting
- Advanced missing value imputation options
- Penalty Configuration: allows a user to penalize the inclusion of particular variables



SPM
SALFORD PREDICTIVE MODELER®

THE COMPANY

Salford Systems specializes in state-of-the-art machine learning technology designed to assist data scientists in all aspects of predictive model development.

PEDIGREE

Salford Systems' technology was developed and enhanced in direct collaboration with original creators of modern machine learning and data mining. We maintain an active R&D program, leveraging our ties to leading universities.

THE SOFTWARE

Salford's tools are known for their ease of use, capability of working with large volumes of data, high-speed model development, robustness and reliability and consistent delivery of ultra-accurate models. Salford's modeling automation tools guide novice data scientists through the complex process of model development and help expert data scientists develop world-class predictive models.

It's not just Salford Systems that's winning. Our customers using Salford Systems' software and algorithms have been recognized for their winning approach to challenges in their industries.

CUSTOMER AWARDS AND ACHIEVEMENTS

PAKDD Data Mining Competition, the annual Pacific-Asia Knowledge Discovery and Data Mining data mining competition

- Cross-selling task, financial dataset

DMA Direct Marketing Association Analytics Challenge

- Predicting customer lifetime value to drive personalized customer interactions
- Healthcare response task
- Make-A-Wish Foundation targeting solution lapsed donor segments
- Targeted marketing task

SALFORD AWARDS AND ACHIEVEMENTS

KDD Cup, the annual Data Mining and Knowledge Discovery competition

- Web analytics task

SIG KDD Innovation Award to Drs. Leo Breiman and Jerome Friedman, creators of algorithms in SPM

- Outstanding technical contributions to the field of knowledge discovery in data and data mining that have had lasting impact in furthering the theory and/or development of commercial systems

Teradata Center for Customer Relationship Management (CRM) at Duke University

- Churn Modeling, CRM

CONTACT

9685 Via Excelencia, Suite 208, San Diego, CA 92126
Telephone: (619)543-8880 | Fax: (619)543-8888
salford-systems.com

SALFORD SYSTEMS
a Minitab company