To gain a competitive edge, you need powerful analytic solutions that extract insights from data so you can act on them. More than ever, organizations are turning to predictive analytics and data mining software to uncover patterns in data and discover hidden relationships.

SAS® Enterprise Miner™ for Desktop delivers the power of data mining and machine learning right to your PC. Highly evolved data exploration, variable transformation and analytical modeling techniques make it easy for analytic professionals (e.g., statisticians, data miners and modelers), business analysts and researchers to design and collaborate on predictive analytics projects that are essential to gaining a competitive advantage.

**Benefits**

- **Develop better models faster with an easy-to-use GUI.** An interactive process flow diagram environment eliminates the need for manual coding and dramatically shortens model development time. The SAS SEMMA (sample, explore, model, modify and assess) data mining approach combines a step-by-step process with the logical organization of data mining capabilities so analytic professionals can develop better predictive analytical models and decision makers can begin using insights faster.

- **Give business analysts a fast, easy and self-sufficient way to build their own models.** Business users with limited statistical knowledge can quickly and easily generate predictive models and derive insights in an automated fashion using the SAS Rapid Predictive Modeler task.

- **Spot trends and recognize opportunities with a rich set of data mining tools and machine-learning algorithms on your desktop.** SAS Enterprise Miner for Desktop provides advanced predictive and descriptive modeling capabilities and model assessment features that deliver more precise insights. Dynamically linked graphs and charts make it easier to understand relationships and results.

- **Quickly and easily install the software yourself.** Desktop-based deployment enables users to get up and running quickly. Start using predictive analytics to solve a multitude of business issues quickly and confidently.

**What does SAS® Enterprise Miner™ for Desktop do?**

SAS Enterprise Miner for Desktop is a complete data mining workbench that runs entirely on Microsoft Windows desktops. It uses a wide variety of descriptive and predictive modeling techniques to provide you with the insights needed to make better decisions.

**Why is SAS® Enterprise Miner™ for Desktop important?**

Business leaders and researchers from organizations of all sizes are turning to predictive analytics to gain an unbeatable advantage in today’s dynamic marketplace. SAS Enterprise Miner for Desktop helps you solve a variety of business problems, improve outcomes and consistently make the right decisions.

**For whom is SAS® Enterprise Miner™ for Desktop designed?**

It is designed for analytical professionals (data mining specialists, modelers and quantitative analysts) in small to midsize organizations, or those who work independently in departments.
Product Overview
SAS Enterprise Miner for Desktop offers a wide variety of predictive modeling and machine learning algorithms. A drag-and-drop interface and self-documenting process flow enable users to quickly develop models. With SAS Enterprise Miner for Desktop, it is now possible for individual users to reap the benefits of a complete data mining workbench on their PCs.

Data mining projects are set up and managed within a visual workspace. Users build their own process flow diagrams, add data mining nodes, compare models and generate SAS score code.

Data exploration capabilities help data miners and business analysts spot trends and anomalies quickly, so they can focus their energies on developing better candidate models. The SAS Code node enables users to integrate SAS DATA step procedures into a data mining process flow diagram.

An organized and logical GUI for data mining success
SAS Enterprise Miner for Desktop provides a flexible framework for conducting all phases of data mining using the SEMMA approach. Using drag-and-drop tools, process flow diagrams are created, updated and easily modified. These saved diagrams can be referred to later.

The interface guides users as they:
• Apply statistical and visualization techniques to see and become familiar with relationships and trends.
• Explore and transform the data to identify a candidate set of predictor variables.
• Create models with those variables to predict outcomes. Novice data miners can build initial models quickly with default settings, while more experienced users can tweak settings to specify unique parameters to enhance their models.
• Combine modeling techniques for additional accuracy.
• Compare models and try multiple approaches and options. Easy-to-interpret displays help users communicate why a particular model is the best predictor.
• Validate the accuracy of decision models with holdout data before deploying models with operational systems.
• Apply the champion model against new data using automatically generated, complete score code.

The easy-to-use drag-and-drop process flow diagram approach shortens the model development time for both analytic professionals and business analysts. The process flow diagrams also serve as self-documenting templates that can be updated later or applied to a related set of problems without starting over from scratch.

A fast, easy and self-sufficient way for business analysts to generate models
SAS Rapid Predictive Modeler automatically steps through a workflow of data mining tasks (e.g., transforming data, selecting variables, fitting a variety of algorithms and assessing models) to quickly generate predictive models for a wide range of business problems. SAS Rapid Predictive Modeler is a SAS® Enterprise Guide® or SAS Add-In for Microsoft Office (Microsoft Excel only) task. Models generated using the SAS Rapid Predictive Modeler task can be modified further in SAS Enterprise Miner for Desktop.

Data preparation and exploration
Preparing data is the most time-consuming aspect of data mining endeavors. SAS Enterprise Miner for Desktop combines powerful data mining capabilities with data exploration and data preparation features, making it easy to read in data from files other than SAS as a fully integrated part of the data mining process.

Extensive data summarization and data visualization features enable users to examine large amounts of data in dynamically linked, multidimensional plots that support varied exploration tasks.

Critical preprocessing tasks include merging files, appending data, sampling, choosing appropriate methods for handling incomplete entries and missing values, binning variables, clustering observations, dropping variables and filtering outliers. Bad data is bad business, and only by starting with quality inputs (careful treatment of variables) can you expect to get quality results.

An integrated suite of unmatched modeling techniques
SAS Enterprise Miner for Desktop provides sophisticated analytical depth for the desktop user with a suite of advanced predictive and descriptive modeling algorithms, including decision trees, gradient boosting, neural networks, clustering, linear and logistic regression, associations and
Figure 1: Automatically generate predictive models for a variety of business problems using the SAS Rapid Predictive Modeler task in SAS Enterprise Guide or SAS Add-In for Microsoft Office (Microsoft Excel only).

Figure 2: Use interactive binning to create bins or buckets or classes of all input variables. You can create bins in order to reduce the number of unique levels as well as attempt to improve the predictive power of each input.

more. You can incorporate previously created SAS/STAT® software models into the SAS Enterprise Miner for Desktop environment for even further fine-tuning and integrated model comparisons.

Model comparisons, reporting and management
SAS Enterprise Miner for Desktop offers numerous assessment tools for comparing results from different models. Results are presented in both statistical and business terms within a single, easy-to-interpret chart.

Scoring with unprecedented ease
The final and most important phase in data mining projects occurs when new data is “scored.” When the model scoring code is applied to new data, a predicted outcome is produced and appropriate decisions are identified for action.

Once the data mining models are developed, SAS Enterprise Miner for Desktop allows you to export SAS score code for rapid deployment into your operational environment. Manual conversion of scoring code not only causes delays for model implementation, it also can introduce potentially costly mistakes.

Unless the entire process that led to the final model is mirrored in the score code (including all data preprocessing steps), the real-world application will miss the mark. SAS Enterprise Miner for Desktop automatically generates score code for the entire process flow and supplies this scoring code in SAS.
Key Features

Intuitive interfaces
• Easy-to-use GUI for building process flow diagrams:
  • Build more and better models faster.
  • Access the SAS programming environment.
  • Provides XML diagram exchange.
  • Reuse diagrams as templates for other projects or users.
  • Directly load a specific data mining project or diagram, or choose from a Project Navigator tree that contains the most recent projects or diagrams.
• Batch processing (program development interface):
  • Encapsulates all features of the GUI.
  • SAS macro-based.
  • Embed training and scoring processes into customized applications.

Accessing and managing data
• Access and integrate structured and unstructured data sources, including time series data, market baskets, web paths and survey data as candidate predictors.
• File Import node for easy access to Microsoft Excel, comma-delimited files, SAS and other common file formats.
• Support for variables with special characters.
• Enhanced Explorer window to quickly locate and view table listings or to develop a plot using interactive graph components.
• SAS Library Explorer and Library Assignment wizard.
• Drop Variables node.
• Merge Data node.
• Append node.
• Filter outliers:
  • Apply various distributional thresholds to eliminate extreme interval values.

Intuitive interfaces
• Combine class values with fewer than n occurrences.
• Interactively filter class and numeric values.
• Metadata node for modifying columns of metadata such as role, measurement level and order.

Sampling
• Simple random.
• Stratified.
• Weighted.
• Cluster.
• Systematic.
• First N.
• Rare event sampling.

Transformations
• Simple: log, log 10, square root, inverse, square, exponential, standardized.
• Binning: bucketed, quantile, optimal binning for relationship to target.
• Best power: maximize normality, maximize correlation with target, equalize spread with target levels.
• Interactions editor: define polynomial and nth degree interaction effects.
• Interactively define transformations:
  • Define customized transformations using the Expression Builder or SAS Code editor.
  • Compare the distribution of the new variable with the original variable.
  • Predefine global transformation code for reuse.

Data partitioning
• Create training, validation and test data sets.
• Ensure good generalization of your models through use of holdout data.
• Default stratification by the class target.
• Balanced partitioning by any class variable.
• Output SAS tables or views.

Interactive variable binning
• Quantile or bucket.
• Gini variable selection.
• Handle missing values as separate group.
Develop customized transformations using the interactive Transform Variables node Expression Builder.

- Fine and coarse classing detail.
- Profile bins by target.
- Modify groups interactively.
- Save binning definitions.

Rules Builder node
- Create ad hoc, data-driven rules and policies.
- Interactively define the value of the outcome variable and paths to the outcome.

Data replacement
- Measures of centrality.
- Distribution-based.
- Tree imputation with surrogates.
- Mid-medium spacing.
- Robust M-estimators.
- Default constant.

Replacement Editor
- Specify new values for class variables.
- Assign replacement values for unknown values.
- Interactively cap extreme interval values to a replacement threshold.

Descriptive statistics
- Univariate statistics and plots:
  - Interval variables: \( n \), mean, median, min, max, standard deviation, scaled deviation and percent missing.
  - Class variables: number of categories, counts, mode, percent mode and percent missing.
- Distribution plots.
- Statistics breakdown for each level of the class target.
- Bivariate statistics and plots:
  - Ordered Pearson and Spearman correlation plot.
  - Ordered chi-square plot with option for binning continuous inputs into bins.
  - Coefficient of variation plot.
- Variable selection by logworth.
- Other interactive plots:
- Variable worth plot ranking inputs based on their worth with the target.
- Class variable distributions across the target and/or the segment variable.
- Scaled mean deviation plots.

Graphs/visualization
- Batch and interactive plots: scatter, matrix, box, constellation, contour, needle, lattice, density and multidimensional plots; 3-D, pie and area bar charts; and histograms.
- Segment profile plots:
  - Interactively profile segments of data created by clustering and modeling tools.
  - Easily identify variables that determine the profiles and the differences between groups.
- Easy-to-use Graphics Explorer wizard and Graph Explore node:
  - Create titles and footnotes.
  - Apply a WHERE clause.
  - Choose from several color schemes.
  - Easily rescale axes.
  - Surface underlying data from standard SAS Enterprise Miner for Desktop results to develop customized graphics.
  - Plots and tables are interactively linked, supporting tasks such as brushing and banding.
  - Data and plots can be easily copied and pasted into other applications or saved as BMP files.
  - Interactive graphs are automatically saved in the Results window of the node.
Clustering and self-organizing maps
- Clustering:
  - User defined or automatically chooses the best clusters.
  - Several strategies for encoding class variables into the analysis.
  - Handles missing values.
  - Variable segment profile plots show the distribution of the inputs and other factors within each cluster.
  - Decision tree profile uses the inputs to predict cluster membership.
- Self-organizing maps (SOM):
  - Batch SOMs with Nadaraya-Watson or local-linear smoothing.
  - Kohonen networks.
  - Overlay the distribution of other variables onto the map.
  - Handles missing variables.

Market basket analysis
- Associations and sequence discovery:
  - Grid plot of the rules ordered by confidence.
  - Expected confidence versus confidence scatter plot.
  - Statistics line plot of the lift, confidence, expected confidence and support for the rules.
  - Statistics histogram of the frequency counts for given ranges of support and confidence.
  - Rules description table.
  - Network plot of the rules.
  - Interactively subset rules based on lift, confidence, support, chain length, etc.
  - Seamless integration of rules with other inputs for enriched predictive modeling.
  - Hierarchical associations:
    - Derive rules at multiple levels.
    - Specify parent and child mappings for the dimensional input table.

Web path analysis
- Scalable and efficient mining of the most frequently navigated paths from clickstream data.
- Mine frequent consecutive subsequences from any type of sequence data.

Link analysis
- Converts data into a set of interconnected linked objects (or entities) that can be visualized as a network of effects.
- Provides a visual model of how two variables’ levels in relational data or between two items’ conoccurrence in transactional data are linked.
- Provides centrality measures and community information to understand linkage graphs.
- Provides weighted confidence statistics for next-best offer information.
- Generates cluster scores for data reduction and segmentation.

Dimension reduction
- Variable selection:
  - Remove variables unrelated to target, based on a chi-square or R2 selection criterion.
  - Remove variables in hierarchies.
  - Remove variables with many missing values.
  - Reduce class variables with a large number of levels.
  - Bin continuous inputs to identify nonlinear relationships.
  - Detect interactions.
- Least Angle Regression (LARS) variable selection:
  - AIC, SBC, Mallows C(p), cross-validation and other selection criteria.
  - Plots include: parameter estimates, coefficient paths, iteration plot, score rankings and more.
  - Generalizes to support LASSO (least absolute shrinkage and selection operator).
  - Supports class inputs and targets as well as continuous variables.
  - Score code generation.
- Principal components:
  - Calculate Eigenvalues and Eigenvectors from correlation and covariance matrices.
  - Plots include: principal components coefficients, principal components matrix, Eigenvalue, Log Eigenvalue, Cumulative Proportional Eigenvalue.
• Interactively choose the number of components to be retained.
• Mine the selected principal components using predictive modeling techniques.
• Variable clustering:
  • Divide variables into disjoint or hierarchical clusters.
  • Eigenvalue or principal components learning.
  • Includes class variable support.
  • Dendrogram tree of the clusters.
  • Selected variables table with cluster and correlation statistics.
  • Cluster network and R-square plot.
  • Interactive user override of selected variables.
• Time series mining:
  • Reduce transactional data into a time series using several accumulation methods and transformations.
  • Analysis methods include seasonal, trend, time domain and seasonal decomposition.
  • Mine the reduced time series using clustering and predictive modeling techniques.

Regression
• Linear and logistic.
• Stepwise, forward and backward selection.
• Equation terms builder: polynomials, general interactions, and effect hierarchy support.
• Cross-validation.
• Effect hierarchy rules.
• Optimization techniques include: Conjugate Gradient, Double Dogleg, Newton-Raphson with Line Search or Ridging, Quasi-Newton, Trust Region.
• Dmine Regression node:
  • Fast forward stepwise least squares regression.
  • Optional variable binning to detect nonlinear relationships.
  • Optional class variable reduction.
  • Include interaction terms.

SAS Code node
• Write SAS code for easy-to-complex data preparation and transformation tasks.
• Incorporate procedures from other SAS products.
• Develop custom models.
• Create SAS Enterprise Miner for Desktop extension nodes.
• Augment score code logic.
• Support for SAS procedures.
• Batch code uses input tables of different names and locations.
• Batch code integrates project-start code that you can use to define libraries and options.
• Easy-to-use program development interface:
  • Macro variables to reference data sources, variables, etc.

Decision trees
• Methodologies:
  • CHAID, classification and regression trees, bagging and boosting, gradient boosting and bootstrap forest.
  • Select tree based on profit or lift objectives and prune accordingly.
  • K-fold cross validation.
• Splitting criterion: Prob Chi-square test, Prob F-test, Gini, Entropy, and variance reduction.
• Switch targets for designing multi-objective segmentation strategies.
• Automatically output leaf IDs as inputs for modeling and group processing.
• Displays English rules.
• Calculates variable importance for preliminary variable selection and model interpretation.
• Display variable precision values in split branches and nodes.
• Unique consolidated tree map representation of the tree diagram.
• Interactive tree capabilities:
  • Interactive growing/pruning of trees; expand/collapse tree nodes.
  • Incorporates validation data to evaluate tree stability.
- Define customized split points, including binary or multiway splits.
- Split on any candidate variable.
- Copy split.
- Tables and plots are dynamically linked to better evaluate the tree performance.
- Print tree diagrams easily on a single page or across multiple pages.
- Interactive subtree selection.
- User-specified display of text and statistics in the Tree node.
- User-controlled sample size within interactive trees.
- Based on the fast ARBORETUM procedure.

Neural networks
- Neural Network node:
  - Flexible network architectures with combination and activation functions.
  - 10 training techniques.
  - Preliminary optimization.
  - Automatic standardization of inputs.
  - Supports direction connections.
- Autoneural Neural node:
  - Automated multilayer perceptron building searches for optimal configuration.
  - Type and activation function selected from four different types of architectures.
- DM Neural node:
  - Model building with dimension reduction and function selection.
  - Fast training; linear and nonlinear estimation.

Partial Least Squares node
- Especially useful for extracting factors from a large number of potential correlated variables.
- Also performs principal components regression and reduced rank regression.

Model ensembles
- Combine model predictions to form a potentially stronger solution.
- Methods include: Averaging, Voting and Maximum.

Rule induction
- Recursive predictive modeling technique.
- Especially useful for modeling rare events.

Two-stage modeling
- Sequential and concurrent modeling for both the class and interval target.
- Choose a decision tree, regression or neural network model for each stage.
- Control how the class prediction is applied to the interval prediction.
- Accurately estimate customer value.

User or automatic selection of the number of the factors.
- Choose from five cross-validation strategies.
- Supports variable selection.

Incremental response/net lift models
- Net treatment versus control models.
- Binary and interval targets.
- Stepwise selection.
- Fixed or variable revenue calculations.
- Net information value variable selection.
- Users can specify the treatment level of the treatment variable.
- User can specify a cost variable in addition to a constant cost.
- Penalized Net Information Value (PNIV) available for variable selection.
- Separate model selection options available for an incremental sales model.

Memory-based reasoning
- k-nearest neighbor technique to categorize or predict observations.
- Patented Reduced Dimensionality Tree and Scan.

Time series data mining
- Time series data preparation:
  - Aggregate, transform and summarize transactional and sequence data.
• Automatically transpose the time series to support similarity analysis, clustering and predictive modeling.
• Process data with or without TimeID variable.
• Similarity analysis:
  • Useful for new product forecasting, pattern recognition and short lifecycle forecasting.
  • Computes similarity measures between the target and input series, or among input time series.
  • Similarity matrix for all combinations of the series.
  • Hierarchical clustering using the similarity matrix with dendrogram results.
  • Constellation plot for evaluating the clusters.
• Exponential smoothing:
  • Control weights decay using one or more smoothing parameters.
  • Best-fitting smoothing method (simple, double, linear, damped trend, seasonal or Winters') is selected automatically.
• Dimension reduction:
  • Supports five time series dimensions techniques: Discrete Wavelet Transform, Discrete Fourier Transform, Singular Value Decomposition, Line Segment Approximation with the Mean, and Line Segment Approximation with the Sum.
• Cross-correlation:
  • Provides autocorrelation and cross-correlation analysis for time-stamped data.
  • The Time Series Correlation node outputs time-domain statistics based on whether autocorrelation or cross-correlation is performed.
• Seasonal decomposition.
Figure 10: Build a high-performance random forest model, which consists of ensembling several decision trees. Through multiple iterations, randomly select variables for splitting while reducing the dependence on a sample selection. Use out-of-bag samples to form predictions.

Open Source Integration node
- Write code in the R language inside of SAS Enterprise Miner for Desktop.
- SAS Enterprise Miner for Desktop data and metadata are available to your R code with R results returned to SAS Enterprise Miner for Desktop.
- Train and score supervised and unsupervised R models. The node allows for data transformation and exploration.
- Generate model comparisons and SAS score code for supported models.

Survival analysis
- Discrete time to event regression with additive logistic regression.
- Event probability for time effect is modeled using cubic splines.
- Users can now enter the cubic spline basis functions as part of the stepwise variable selection procedure in addition to the main effects.
- User-defined time intervals for specifying how to analyze the data and handle censoring.
- Automatically expands the data with optional sampling.
- Supports non-time varying covariates.
- Computes survival function with holdout validation.
- Generates competing risks or sub-hazards.
- Score code generation with mean residual life calculation.
- Incorporate time varying covariates into the analysis with user-specified data formats, including standard, change-time and fully expanded.
- Users can specify left-truncation and censor dates.

High-performance data mining procedures and nodes
- These multithreaded procedures will execute concurrently and take advantage of all available cores on your desktop:
  - HPBIN (high-performance binning).
  - HPBNET (high-performance Bayesian networks).
  - HPCLUS (high-performance clustering).
  - HPCORR (high-performance correlation).
  - HPDECIDE (high-performance decide).
  - HPDMDB (high-performance data mining database).
  - HPDS2 (high-performance DS2).
  - HPFOREST (high-performance forests).
  - HPIMPUTE (high-performance imputation).
  - HPREDUCE (high-performance variable reduction).

Use cases: stratified modeling, bagging and boosting, multiple targets and cross-validation.

Group processing with the Start and End Groups nodes
- Repeat processing over a segment of the process flow diagram.

SAS® Rapid Predictive Modeler customized task in SAS® Enterprise Guide® or SAS® Add-In for Microsoft Office (Excel only)
- Automatically generates predictive models for a variety of business problems.
• Models can be opened, augmented and modified in SAS Enterprise Miner for Desktop.
• Produces concise reports, including variable importance charts, lift charts, ROC charts and model scorecards for easy consumption and review.
• Ability to score the training data with option to save the scored data set.

Consistent modeling features
• Select models based on either the training, validation (default) or test data using several criteria such as profit or loss, AIC, SBC, average square error, misclassification rate, ROC, Gini and KS (Kolmogorov-Smirnov).
• Incorporate prior probabilities into the model development process.
• Supports binary, nominal, ordinal and interval inputs and targets.
• Easy access to score code and all partitioned data sources.
• Display multiple results in one window to help better evaluate model performance.
• Decisions node for setting target event and defining priors and profit/loss matrices.

Model Import node
• Register SAS Enterprise Miner for Desktop models for reuse in other diagrams and projects.
• Import and evaluate external models.

Model evaluation
• Compare multiple models in a single framework for all holdout data sources.
• Automatically selects the best model based on the user-defined model criterion.
• Supports user override.
• Extensive fit and diagnostics statistics.
• Lift charts; ROC curves.

Figure 11: Evaluate multiple models together in one easy-to-interpret framework using the Model Comparison node.

• Interval target score rankings and distributions.
• Profit and loss charts with decision selection; confusion (classification) matrix.
• Class probability score distribution plot; score ranking matrix plots.
• Cutoff node to determine probability cutoff point(s) for binary targets.
• User override for default selection.
• Max KS Statistic.
• Min Misclassification Cost.
• Maximum Cumulative Profile.
• Max True Positive Rate.
• Max Event Precision from Training Prior.
• Event Precision Equal Recall.

Reporter node
• Uses SAS Output Delivery System to create a PDF or RTF document of a process flow.
• Helps document the analysis process and facilitate results sharing.
• Document can be saved and is included in SAS Enterprise Miner for Desktop results packages.
• Includes image of the process flow diagram.

Save Data node
• Save training, validation, test, score or transaction data from a node to either a previously defined SAS library or a specified file path.
• Export Excel 2010, CSV and tab delimited files. Default options are designed so that the node can be deployed in SAS Enterprise Miner for Desktop batch programs without user input.
• Can be connected to any node in a SAS Enterprise Miner for Desktop process flow diagram that exports training, validation, test, score or transaction data.

• User-defined notes entry.

Scoring
• Score node for interactive scoring in the SAS Enterprise Miner for Desktop GUI.
• Optimized score code is created by default, eliminating unused variables.
• Automated score code generation in SAS.
Figure 12: Fit highly complex nonlinear relationships using the Neural Network node.

To learn more about SAS Enterprise Miner for Desktop system requirements, download white papers, view screenshots and see other related material, please visit sas.com/enterpriseminer.