

# SAS® Enterprise Miner™ 13.1

*Helping to Solve Your Big Data Needs*

---

## Overview

SAS Enterprise Miner 13.1 introduces ten new nodes and several enhancements to previously existing nodes.

---

## SAS® Enterprise Miner™ Core

The Open Source node enables integration with the R language inside a SAS Enterprise Miner flow. You can perform data transformation and exploration in addition to training and scoring supervised and unsupervised models in R. You can then seamlessly integrate the results, assess the model, and compare it to models generated by SAS Enterprise Miner. In several cases, corresponding SAS DATA step scoring code can also be generated to enable full integration.

The Register Model node enables you to register models directly in the SAS® Metadata Server. Information about variables, target levels, scoring code and the mining function is accumulated and registered in the metadata. This information can be used as input in SAS® Model Manager or SAS® Enterprise Guide® or used to score data in SAS Enterprise Miner.

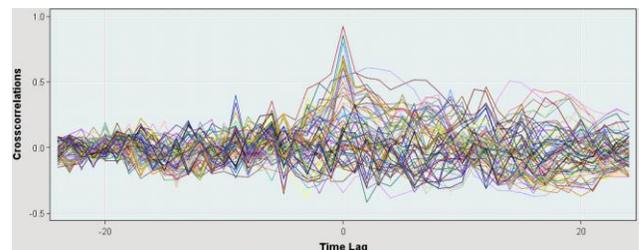
The Save Data node provides you with a simple way to save training, validation, test, score, or transaction data from a SAS Enterprise Miner path to a previously defined SAS library or specified path. You can export a SAS data set, JMP table, Excel 2010 spreadsheet, comma-separated values (CSV) or tab delimited file.

---

## SAS® Enterprise Miner™ Applications

The Time Series Dimension Reduction node is available with several dimension reduction techniques, including discrete wavelet and Fourier transforms, singular value decomposition, and line segment methods. It extracts features from each time series and reduces the dimension of time. It is useful for clustering extremely long time series as a preprocessing step.

The Time Series Correlation node helps you perform correlation and cross-correlation analysis. It calculates numerous autocorrelation and cross-correlation statistics on time series data. It provides time covariate selection functionality through the cross-correlation analysis among multiple time series when target time series are specified. The autocorrelation statistics can also be used for clustering tasks.



The Time Series Decomposition node enables you to perform seasonal decomposition of time series. It provides classical seasonal decomposition components and uses that information in subsequent nodes. The functionality includes the transpose option of the exported data for the performing time series clustering tasks.

---

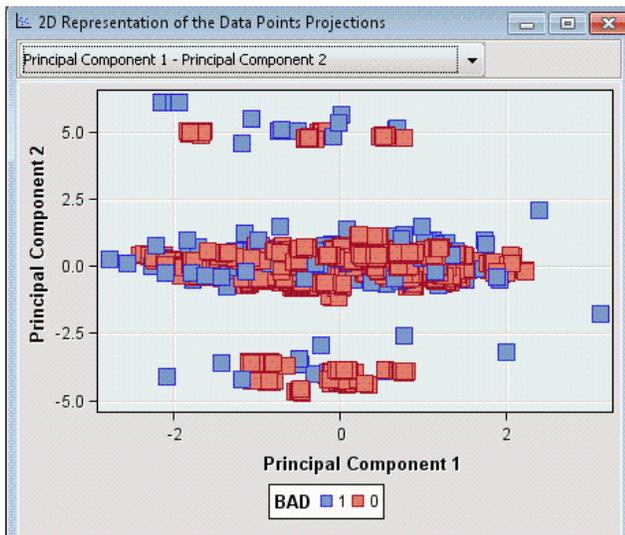
## SAS® Enterprise Miner™ High Performance Data Mining

The HP Cluster node uses the high-performance HPCLUSTER procedure and performs  $k$ -means clustering analysis in distributed computing environments. It uses the method of least squares estimation in  $k$ -means clustering to compute the cluster centroids, but it accelerates the convergence with a large number of observations by performing clustering and scoring in parallel. This node currently takes only numeric interval variables as inputs.

The HP GLM node uses the high-performance HPGENSELECT procedure to fit a generalized linear model in a distributed computing environment. Several different response probability distributions and link functions are available. You can also select and modify variable-based reference levels.

The HP Neural node now allows a user-defined architecture which gives you more control over the building of the network itself. You can specify the number of hidden layers, the number of hidden neurons, and the associated activation function for each layer. Input and target standardizations as well as target error and activation functions are now available through node properties.

The HP Principal Components node uses the high-performance HPPRINCOMP procedure to perform principal component analysis. The node calculates eigenvalues and eigenvectors from the covariance matrix or the correlation matrix of input variables and produces principal components. The principal components are useful for data dimension reduction, which is frequently an intermediate step in the data mining process.



The HP Support Vector Machine node uses the newly developed high-performance HPSVM procedure for binary classification problems. The node constructs separating hyper-planes that maximize the margin between two classes by using kernels: linear, polynomial, radial basis function, and sigmoid function. Interior point and active set optimization methods are available.

---

## Other Enhancements

---

The following are highlights of other enhancements in SAS Enterprise Miner 13.1:

- The Decision Tree node allows users to import a previously created model and apply this model to new data. This is also useful when importing diagrams from XML, recreating diagrams from a SPK file or when migrating from version to version.
- The HP Forest node now enables you to optionally perform variable selection based on either OOB average error for interval targets or OOB marginal reduction for class targets.
- The HP Neural node now allows a User Defined Architecture which gives users more control over the building of the network itself. Users are able to specify the number of hidden layers, the number of hidden neurons and associated activation function for each layer. Input and Target Standardizations as well as Target Error and Activation Functions are now surfaced through node properties.
- The HP Tree node now supports the generation of models for interval targets. The criterion for determining the associated splitting rules can be based on a reduction in variance or determined by using the F test. More control over missing value handling has also been provided with the addition of mapping these values to the largest branch, the most correlated branch, or a separate branch. Surrogate rules can now also be populated for assigning missing values.

---

## For More Information

---

For more information, ask your organization's SAS representative to contact the SAS Customer Interaction Center at 1.800.727.0025.