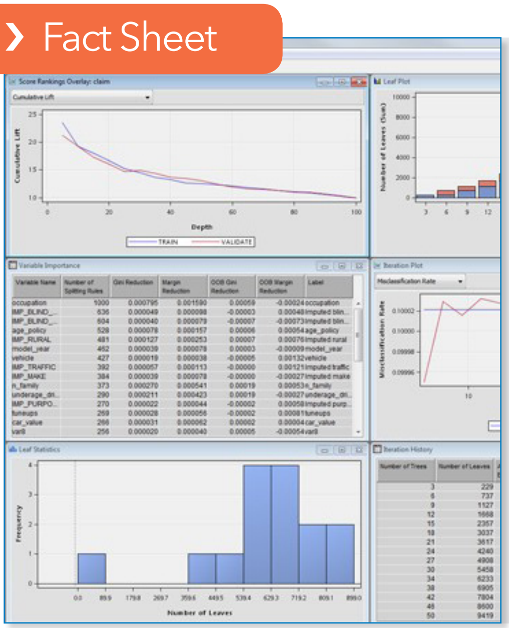
› Fact Sheet



Variable Importance

Variable Name	Number of Splitting Rules	Gini Reduction	Margin Reduction	OOB Gini Reduction	OOB Margin Reduction	Label
occupation	1000	0.000795	0.001590	0.00059	-0.00024	occupation
IMP_BLIND_...	636	0.000049	0.000098	-0.00003	0.00048	imputed blin...
IMP_BLIND_...	604	0.000040	0.000079	-0.00007	-0.00073	imputed blin...
age_policy	528	0.000078	0.000157	0.00006	0.00054	age_policy
IMP_RURAL	481	0.000127	0.000253	0.00007	0.00076	imputed rural
model_year	462	0.000039	0.000078	0.00003	-0.00009	model_year
vehicle	427	0.000019	0.000038	-0.00005	0.00132	vehicle
IMP_TRAFFIC	392	0.000057	0.000113	-0.00000	0.00121	imputed traffic
IMP_MAKE	384	0.000039	0.000078	-0.00000	-0.00027	imputed make
n_family	373	0.000270	0.000541	0.00019	0.00053	n_family
underage_dri...	290	0.000211	0.000423	0.00019	-0.00027	underage_dri...
IMP_PURPO...	270	0.000022	0.000044	-0.00002	0.00058	imputed purp...
tuneups	269	0.000028	0.000056	-0.00002	0.00081	tuneups
car_value	266	0.000031	0.000062	0.00002	0.00004	car_value
var8	256	0.000020	0.000040	0.00005	-0.00054	var8

Iteration History

Number of Trees	Number of Leaves
3	229
6	737
9	1127
12	1668
15	2357
18	3037
21	3617
24	4240
27	4908
30	5458
34	6233
38	6905
42	7804
46	8600
50	9419

What do SAS® High-Performance Analytics products do?

With high-performance analytics products from SAS, you can develop and process models that use huge amounts of diverse data. These products – for statistics, data mining, text mining, econometrics and optimization – are available in a highly scalable, distributed in-memory processing architecture.

Why are SAS® High-Performance Analytics products important?

You can analyze big data to derive more accurate insights and make timely business decisions. The ability to solve difficult problems, test more ideas and evaluate complex scenarios helps you seize new opportunities and reduce uncertainties.

For whom are these products designed?

These products are designed for analytics professionals (e.g., data miners, statisticians, data scientists and business analysts) who need to develop and process models quickly and efficiently. They also provide IT with a highly scalable and reliable infrastructure for managing and processing analytic jobs.

SAS® High-Performance Analytics Products

Produce faster, more accurate insights and solve your most complex problems

Complex business problems require sophisticated, high-end analytics and the ability to integrate big data sources, including vast collections of text-based data.

SAS offers five high-performance analytics products that run analytical computations in a distributed, in-memory environment. This enables you to quickly prepare, explore and model multiple scenarios using data volumes never before possible. Accurate and rapid insights are delivered in near-real time (typically in minutes, rather than hours).

If you can significantly reduce analytic processing from days or hours to minutes or seconds, you can ask more what-if questions. Models can be quickly adjusted and run again.

Combining unstructured and structured data, using more variables and running frequent model iterations – faster than ever before – provides transformative predictive power.

Benefits

- **Quickly and confidently seize new opportunities, detect unknown risks and make the right choices.** SAS High-Performance Analytics products exploit all available computing resources to perform faster statistical modeling and model selection, whether from a single machine or in a distributed computing environment. You get finer, more accurate results to drive new opportunities for your organization.

- **Use all data (including unstructured) with advanced modeling techniques and perform more model iterations to get answers to your difficult questions**. By applying sophisticated analytics against all of your data, you gain improved accuracy for better decision making. Use the best modeling techniques and perform more model iterations. Combining structured data with text data uncovers relationships that were previously undetected and adds more predictive power to your models.

- **Derive insights at breakthrough speeds for high-value and time-sensitive decision making**. Shrink analytical model processing time and derive rapid insights to improve decision making across your enterprise. High-performance analytics products from SAS deliver blazingly fast performance. They can evaluate many alternative scenarios, quickly detect changes in volatile markets and make timely, optimal recommendations.

- **Take advantage of a highly scalable and reliable analytics infrastructure to test more ideas and multiple scenarios with all your data**. Analytical professionals can take full advantage of the in-memory infrastructure to solve the most complex questions without architecture constraints. IT can efficiently manage demands for more processing power now and in the future.

Overview

SAS® High-Performance Analytics products enable organizations to analyze big data to produce more accurate insights in minutes. These high-powered products are available for:

- Statistics.
- Data mining.
- Text mining.
- Econometrics.
- Optimization.

In addition to the specialized features in each product, a core set of common procedures is available in all five products to help you prepare and summarize data.

Single machine or distributed mode

SAS High-Performance Analytics products are engineered to run either on a single server or in a distributed mode using a cluster of computers. All high-performance procedures are multithreaded and can exploit all available cores, whether on a single machine or in a distributed computing environment.

In single-machine mode, the high-performance procedures use the numbers of CPUs (cores) on the machines to determine the number of concurrent threads. In a nutshell, single-machine mode means multithreading on the client machine.

When high-performance procedures execute in distributed mode, several nodes in a distributed computing environment are used for calculations. Data is distributed across the machines in a cluster, and the massive computing power of the cluster is used to solve a single large analytic task. Distributed mode enables analytical computations to be performed simultaneously on multiple machines in the cluster and across multiple concurrently scheduled threads on each machine.

On single machines, high-performance modeling procedures achieve scalability by exploiting all cores on a single machine. In distributed computing environments, these procedures exploit parallel access to data, along with all of the cores and huge amounts of memory that are available.

SAS® High-Performance Statistics

With SAS High-Performance Statistics, you can build and run analytical models faster than ever. Modeling methods include regression, logistic regression, generalized linear models, linear mixed models, nonlinear models and decision trees. The procedures provide model selection, dimension reduction and identification of important variables whenever this is appropriate for the analysis.

SAS® High-Performance Data Mining

SAS High-Performance Data Mining lets you analyze large volumes of diverse data using a drag-and-drop interface and powerful descriptive, predictive and machine-learning methods. A variety of modeling techniques, including random forests, support vector machines, neural networks, clustering, etc., are combined with data preparation, data exploration and scoring capabilities. Because you're able to build and run more models faster, you can ask more difficult questions and bring new ideas into your data mining process. (SAS High-Performance Data Mining includes SAS High-Performance Statistics.)

SAS® High-Performance Text Mining

With SAS High-Performance Text Mining, you can gain quick insights from large unstructured data collections involving millions of documents, emails, notes, report snippets, social media sources, etc. Support is included for parsing, entity extraction, automatic stemming and synonym detection, topic discovery and singular value decomposition (SVD). Text mining results can be used as inputs into high-performance data mining to improve your predictive modeling power.

SAS® High-Performance Econometrics

SAS High-Performance Econometrics modeling methods include linear regression, univariate and bivariate logit/probit models, stochastic frontier models, censored/truncated regression, sample selection models, count models and loss distribution models. Some of these methods can also be applied to panel data. There are also tools for simulating distributions from both multivariate copulas as well as compound distribution models.

SAS® High-Performance Optimization

High-performance optimization is useful for certain classes of linear, mixed integer linear and nonlinear problems. Key tasks, including individual optimizations for algorithms such as multistart (nonlinear), decomposition (linear, mixed integer linear), option tuning (mixed integer linear), and global/local search optimization, are executed in parallel, often substantially reducing the time needed to complete the overall optimization effort.

Key Features

Core Capabilities in SAS® High-Performance Analytics Products

High-performance data summarization

- Enables large-scale data exploration and summarization through a series of parallelized procedures.
- Generates descriptive statistics, in the form of a SAS output data set, on a large scale, very quickly.
- Creates mean, min, max, range and measures of spread and centrality along with data for cardinality, summary and levels of variables.

High-performance DS2

- Provides a vehicle for the parallel execution of DS2 code from a Base SAS session in a distributed, in-memory computing environment.
- Enables control of the level of parallelism per execution node and the number of nodes to engage.

High-performance data mining database

- Creates summary statistics of key input data sources using sum, count, min, max, standard deviation and measure of asymmetry.

High-performance correlation

- Compute correlations for big data sets that have large numbers of both rows and columns.

High-performance sampling

- Performs either high-performance simple random sampling or stratified sampling.

High-performance binning

- Bucket (equal-length) binning method.
- Winsorized binning method and Winsorized statistics.
- Pseudo-quantile binning method, which is similar to quantile binning.
- Provides a mapping table for the selected binning method.
- Provides a basic statistical table that contains the minimum, maximum, mean, pseudo-median and so on.
- Histogram table that shows the output mapping statistics.
- Estimation of a pseudo-quantile table.
- Calculates weight of evidence (WOE) and information value (IV) based on binning results.

High-performance imputation

- Executes high-performance numeric variable imputation with a specified value.
- Can also replace numeric missing values with the mean, the pseudo-median, or some random value between the minimum value and the maximum value of the nonmissing values.

SAS® High-Performance Statistics

High-performance logistic regression and model selection

- Predicts binary, binomial and multinomial outcomes.
- Provides model-building syntax with the CLASS and effect-based MODEL statements.
- Provides a variety of link functions for modeling multinomial response variables with ordered or unordered categories.
- Provides predicted values via an OUTPUT data set and generated scoring code.

High-performance linear regression and model selection

- Supports general linear models and reference parameterization for classification effects.
- Provides multiple methods for model effect selection.
- Provides model-specification syntax with CLASS and effect-based MODEL statements.
- Supports partitioning of data into training, validation and testing roles.
- Enables selection from a large number of effects (tens of thousands).

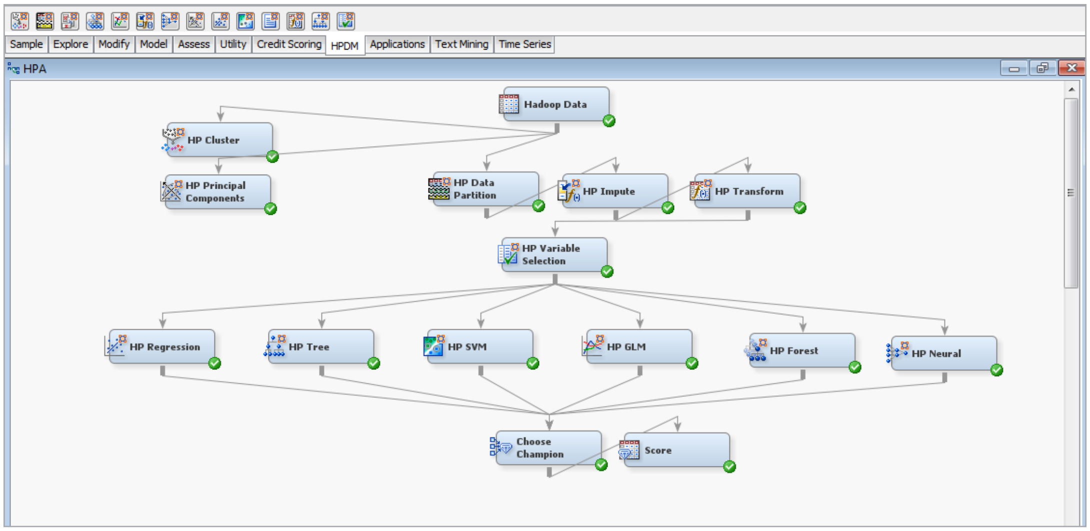


Figure 1: SAS High-Performance Data Mining process flow diagram using high-performance nodes.

Key Features (continued)

- Provides stopping rules based on a variety of model evaluation criteria.
- Supports stopping and selection rules based on external validation and leave-one-out cross-validation.
- Provides predicted values via an OUTPUT data set and generated scoring code.

High-performance nonlinear regression

- Estimates parameters using least squares and maximum likelihood estimation.
- Provides a variety of optimization techniques for computing parameter estimates.
- Computes confidence limits for user-provided functions of parameters.

High-performance mixed linear models

- Supports multiple covariance structures, including variance components, compound symmetry, unstructured, AR(1), Toeplitz and factor analytic.
- Implements REML and maximum likelihood estimation with a variety of optimization techniques.
- Supports data with many subjects.

High-performance partial least squares

- Supports GLM and reference parameterization for classification effects.
- Permits any degree of interaction effects that involve classification and continuous variables.
- Supports partitioning of data into training and testing roles.

High-performance quantile regression analysis

- Supports quantile regression for single or multiple quantile levels.
- Supports GLM and reference cell parameterization for classification effects.
- Supports any degree of interaction (crossed effects) and nested effects.

High-performance generalized linear modeling and model selection

- Estimates parameters of a generalized linear model by maximum likelihood.
- Provides model training, validation and testing.
- Provides model-building syntax with the CLASS and effect-based MODEL statements.
- Provides multiple link functions and distributions, including the Tweedie family of distributions.

High-performance decision trees

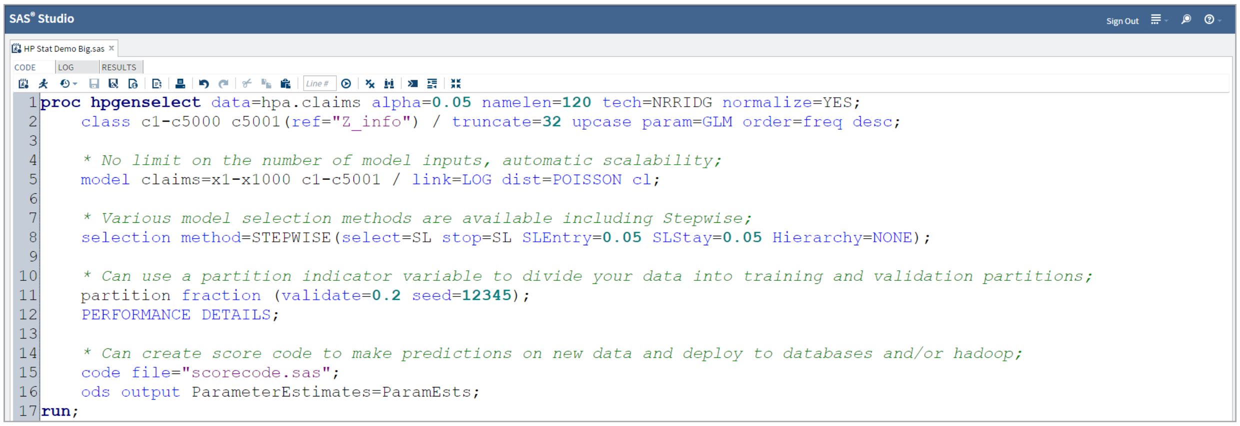
- Creates decision tree models.
- Supports interval and nominal inputs and target variables.
- Provides the entropy, Gini and FastCHAID, CHAID, information gain ratio (IGR) and chi-square methods for decision tree growth (for nominal targets).
- Provides the variance, CHAID and F test methods for regression tree growth (for interval targets).
- Supports growing and pruning decision tree models.
- Provides C4.5-style pruning.
- Provides English rules that describe the leaves of the tree.

High-performance finite mixture models

- Provides maximum likelihood estimation for univariate finite mixture models.
- Provides Markov chain Monte Carlo estimation for some models.
- Provides many built-in link and distribution functions.
- Models classification and regression effects in the mixing probabilities.

High-performance principal components analysis

- Provides a multivariate technique for examining relationships among quantitative variables.
- Computes eigenvalues, eigenvectors and principal component scores.



```
proc hpgenselect data=hpa.claims alpha=0.05 namelen=120 tech=NRRIDG normalize=YES;
    class c1-c5000 c5001(ref="Z_info") / truncate=32 upcase param=GLM order=freq desc;

    * No limit on the number of model inputs, automatic scalability;
    model claims=x1-x1000 c1-c5001 / link=LOG dist=POISSON cl;

    * Various model selection methods are available including Stepwise;
    selection method=STEPWISE(select=SL stop=SL SLEntry=0.05 SLStay=0.05 Hierarchy=NONE);

    * Can use a partition indicator variable to divide your data into training and validation partitions;
    partition fraction (validate=0.2 seed=12345);
    PERFORMANCE DETAILS;

    * Can create score code to make predictions on new data and deploy to databases and/or hadoop;
    code file="scorecode.sas";
    ods output ParameterEstimates=ParamEsts;
run;
```

Figure 2: With SAS High-Performance Statistics, you can build a generalized linear model using thousands of inputs, using variable selection, while taking advantage of in-memory technology.

Key Features (continued)

High-performance canonical discriminant analysis

- Provides dimension reduction.
- Computes squared Mahalanobis distances between class means.
- Produces canonical coefficients and scored canonical variables.

SAS® High-Performance Data Mining

High-performance variable reduction

- Reduces dimensionality for structured inputs and to select a subset of the original variables.
- Performs unsupervised variable selection by identifying a set of variables that jointly explains the maximum amount of data variance (covariance analysis).
- Provides distributed computation and output of the CORR, COV or SSCP matrix.
- Uses the CLASS statement to support categorical inputs.
- Outputs statistics and matrix information that can be used for statistical procedures.

High-performance time series dimensional reduction

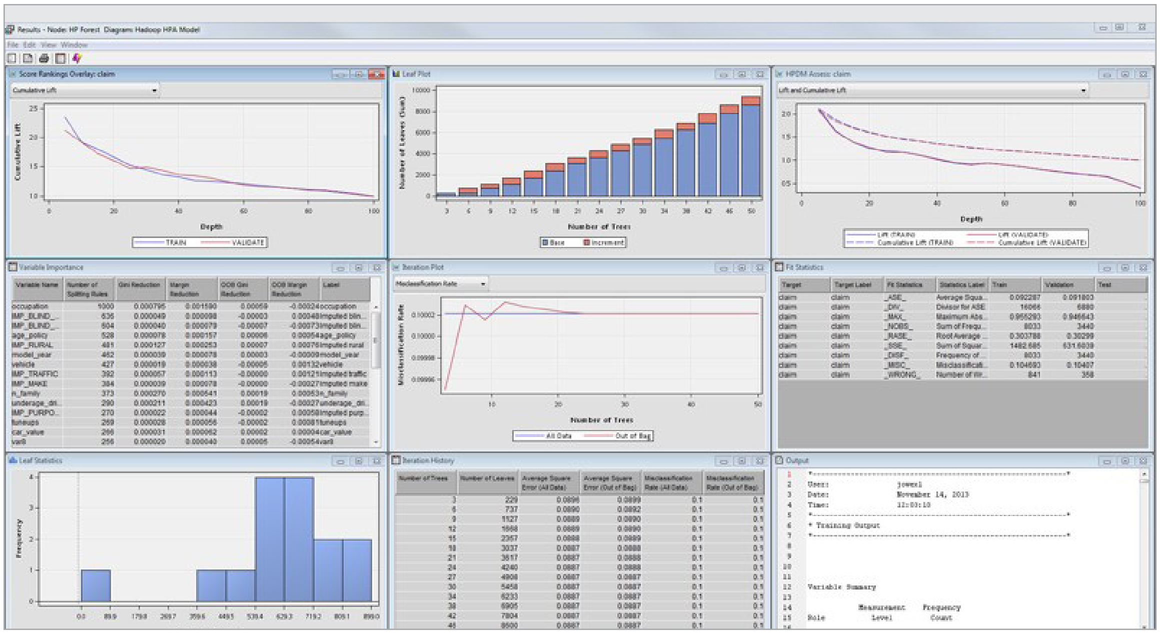
- Reduces dimensionality to perform tasks such as similarity, clustering, etc.
- Accepts three time series formats for the input data: transactional, transposed and columnwise.
- Outputs reduced-dimensional time series in three formats: transactional, transposed and columnwise.
- Handles multiple time series variables in transactional format for the input data.

High-performance neural networks

- Provides automatic standardization of input and target variables.
- Provides intelligent defaults for most neural network parameters (e.g., activation and error functions).
- Provides automatic selection and use of a validation data subset.
- Provides automatic termination of training when the validation error stops improving.
- Provides the ability to weight individual observations.
- Accepts inputs to enhance predictive power from unstructured text.
- Lets you use an arbitrary number of hidden layers to support deep learning.
- Lets you specify the Poisson and gamma error function and the exponential output layer activation function to support modeling of count data.
- Lets you specify an activation function (identity, tanh or sin) for hidden layers and for the output layer.

High-performance random forests

- Creates an ensemble of hundreds of decision trees to predict a single target.
- Trains hundreds of decision trees in parallel independently on different grid nodes.
- Randomly selects the input variables considered for splitting a node from all available inputs.
- Considers only a single variable that is most associated with the target for splitting.
- Accepts inputs to enhance predictive power from unstructured text.



Variable Name	Number of Splitting Rules	Gini Reduction	Margin Reduction	OOB Gini Reduction	OOB Margin Reduction	Label
occupation	1000	0.000795	0.001590	0.00059	-0.00024	occupation
IMP_BLIND_...	636	0.000049	0.000098	-0.00003	0.00048	imputed blin...
IMP_BLIND_...	604	0.000040	0.000079	-0.00007	-0.00073	imputed blin...
age_policy	528	0.000078	0.000157	0.00006	0.00054	age_policy
IMP_RURAL	481	0.000127	0.000253	0.00007	0.00076	imputed rural
model_year	462	0.000039	0.000078	0.00003	-0.00009	model_year
vehicle	427	0.000019	0.000038	-0.00005	0.00132	vehicle
IMP_TRAFFIC	392	0.000057	0.000113	-0.00000	0.00121	imputed traffic
IMP_MAKE	384	0.000039	0.000078	-0.00000	-0.00027	imputed make
n_family	373	0.000270	0.000541	0.00019	0.00053	n_family
underage_dri...	290	0.000211	0.000423	0.00019	-0.00027	underage_dri...
IMP_PURPO...	270	0.000022	0.000044	-0.00002	0.00058	imputed purp...
tuneups	269	0.000028	0.000056	-0.00002	0.00081	tuneups
car_value	266	0.000031	0.000062	0.00002	0.00004	car_value
var8	256	0.000020	0.000040	0.00005	-0.00054	var8

Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
claim	claim	_ASE_	Average Squa...	0.092287	0.091803	
claim	claim	_DIV_	Divisor for ASE	16066	6880	
claim	claim	_MAX_	Maximum Abs...	0.955293	0.946643	
claim	claim	_NOBS_	Sum of Frequ...	8033	3440	
claim	claim	_RASE_	Root Average ...	0.303788	0.30299	
claim	claim	_SSE_	Sum of Squar...	1482.685	631.6039	
claim	claim	_DISF_	Frequency of ...	8033	3440	
claim	claim	_MISC_	Misclassificati...	0.104693	0.10407	
claim	claim	_WRONG_	Number of Wr...	841	358	

Number of Trees	Number of Leaves	Average Square Error (All Data)	Average Square Error (Out of Bag)	Misclassification Rate (All Data)	Misclassification Rate (Out of Bag)
3	229	0.0896	0.0899	0.1	0.1
6	737	0.0890	0.0892	0.1	0.1
9	1127	0.0889	0.0890	0.1	0.1
12	1668	0.0889	0.0890	0.1	0.1
15	2357	0.0888	0.0889	0.1	0.1
18	3037	0.0887	0.0888	0.1	0.1
21	3617	0.0887	0.0888	0.1	0.1
24	4240	0.0887	0.0888	0.1	0.1
27	4906	0.0887	0.0887	0.1	0.1
30	5458	0.0887	0.0887	0.1	0.1
34	6233	0.0887	0.0887	0.1	0.1
38	6905	0.0887	0.0887	0.1	0.1
42	7804	0.0887	0.0887	0.1	0.1
46	8600	0.0887	0.0887	0.1	0.1

```
*------------------------------------------------------------*
User:                   jowexl
Date:                   November 14, 2013
Time:                   12:03:10
*------------------------------------------------------------*
* Training Output
*------------------------------------------------------------*




Variable Summary

             Measurement     Frequency
Role            Level          Count
```

Figure 3: Use sophisticated techniques like random forests to get fast answers to complex problems with SAS High-Performance Data Mining.

Key Features (continued)

High-performance random forest scoring

- Scores a previously trained forest model produced by the HPFOREST procedure.

High-performance decisions

- Creates optimal decisions that are based on a user-specified decision matrix, on prior probabilities and on output from a modeling procedure (which can be predicted values for an interval target variable).
- The decision matrix contains columns (decision variables) that correspond to each decision and rows (observations) that correspond to target values. The values of the decision variables represent target-specific consequences that could be profit, loss or revenue.

High-performance Bayesian network

- Learns a Bayesian network.
- Learns different types of Bayesian network structures, including naive, tree augmented naive (TAN), Bayesian network-augmented naive (BAN), parent-child Bayesian network and Markov blanket.
- Performs efficient variable selection through independence tests and selects the best model automatically from the specified parameters using a validation data subset.
- Generates SAS DATA step code to score a data set.

High-performance clustering

- Performs a cluster analysis on the basis of distances that are computed from one or more quantitative variables. The observations are divided into clusters such that every observation belongs to only one cluster.
- Performs k-means for clustering and takes only numeric interval variables as input.
- Provides a new technique called the aligned box criterion (ABC) for estimating the number of clusters in the data set.

High-performance support vector machines

- Uses both linear and nonlinear kernels to conduct training.
- Provides two optimization techniques: the interior point method and the active-set method.
- Supports both continuous and categorical inputs in the model training of a binary target.
- The interior-point method can run in either single-machine mode or distributed mode, whereas active-set method runs only in single-machine mode.

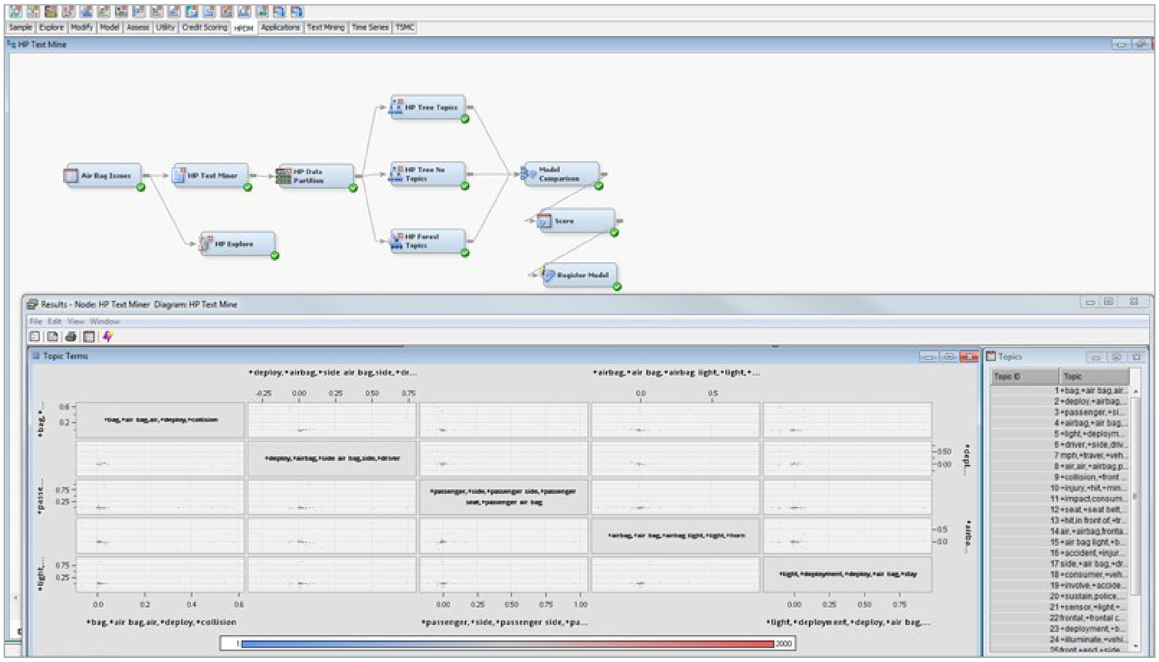
High-performance enabled SAS® Enterprise Miner™ nodes

- HP Data Partition.
- HP Explore.
- HP Transform.
- HP Variable Selection.
- HP Regression.
- HP Neural.
- HP Forest.
- HP Impute.
- HP Tree.

SAS® High-Performance Text Mining

Natural language processing (NLP)

- Identify term part of speech automatically (more than 15 different definitions are system defined).
- Choose to extract standard entities such as location, time, date and address from 17 predefined options.
- Detect noun groups and multi-term lists and treat as single terms in machine-learned processing.
- Detect different term stems without manual intervention.
- Find term variants automatically with synonym detection.


Sample
Explore
Modify
Model
Assess
Utility
Credit Scoring
HPDM
Applications
Text Mining
Time Series
TSMC
HP Text Mine
Air Bag Issues
HP Text Miner
HP Data Partition
HP Explore
HP Tree Topics
HP Tree No Topics
HP Forest Topics
Model Comparison
Score
Register Model
Results - Node: HP Text Miner Diagram: HP Text Mine
File Edit View Window
Topic Terms
+deploy,+airbag,+side air bag,side,+dr...
+airbag,+air bag,+airbag light,+light,+...
+bag,+air bag,air,+deploy,+collision
+deploy,+airbag,+side air bag,side,+driver
+passenger,+side,+passenger side,+passenger seat,+passenger air bag
+airbag,+air bag,+airbag light,+light,+horn
+light,+deployment,+deploy,+air bag,+stay
+passenger,+side,+passenger side,+pa...
+light,+deployment,+deploy,+air bag...
+bag,+air bag,air,+deploy,+collision
Topics

Topic ID	Topic
1	+bag,+air bag,air...
2	+deploy,+airbag...
3	+passenger,+si...
4	+airbag,+air bag...
5	+light,+deploym...
6	+driver,+side,driv...
7	mph,+travel,+veh...
8	+air,air,+airbag,p...
9	+collision,+front ...
10	+injury,+hit,+min...
11	+impact,consum...
12	+seat,+seat belt,...
13	+hit,in front of,+tr...
14	air,+airbag,fronta...
15	+air bag light,+b...
16	+accident,+injur...
17	side,+air bag,+dr...
18	+consumer,+veh...
19	+involve,+accide...
20	+sustain,police,...
21	+sensor,+light,+...
22	frontal,+frontal c...
23	+deployment,+b...
24	+illuminate,+vehi...
25	front,+seat,+side

Figure 4: Using in-memory processing, easily build a predictive model that incorporates text -mined topics.

Key Features (continued)

- Natively examine text in English or German, retaining the meaning intended by the author.

Text processing options

- Choose desired parsing options to customize machine-learned text models.
- Vary term importance based on frequency weights for both individual documents, and across the corpus.
- Select frequency-term weighting to dampen term occurrence impacts.
- Distinguish more important terms from others using term weights.
- Use duplicate document identifiers and control processing with keyword options in both text preprocessing and scoring.

Text filtering

- Specify a start list as a SAS data set to include specific terms in parsing and downstream processing.
- Include a stop list as a SAS data set to exclude terms from parsing and further analysis.
- Refine start and stop lists by adding, deleting and editing terms, including multi-word terms.
- Define the minimum number of documents a term must appear in to restrict candidate terms in further analysis.

Topic generation

- Machine-learned topics represent the generated term-by-document matrix as a structured numeric representation of the document collection.
- Extend data mining analysis using results of semantically related topics as input into high-performance structured data mining nodes or procedures.
- Use labeled, machine-generated topics in other SAS applications.

Graphs and tabular output

- Examine text model terms by reviewing detailed term characteristics table.
- Review graphs describing term frequency in the entire collection relative to term weight and number of documents by term frequency to evaluate or refine the text model.
- Evaluate the document collection from graphs depicting the frequency of terms by role or attribute.
- Create your own results plot, or modify an existing results plot with the graph wizard.

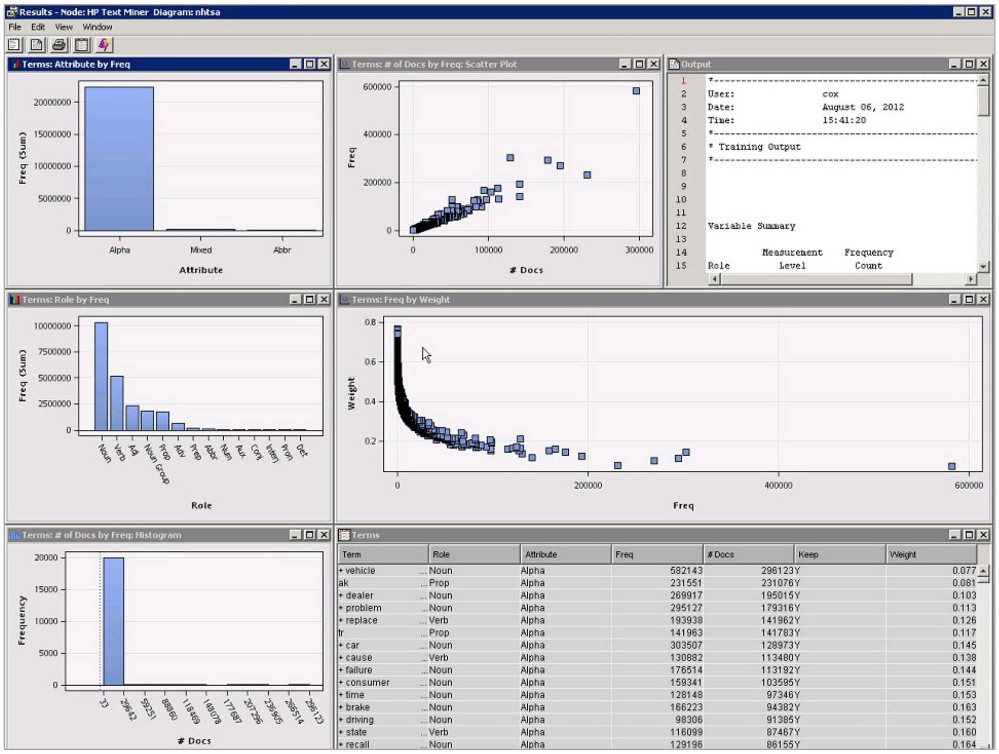
SAS® High-Performance Econometrics

High-performance count regression

- Fits regression models where the dependent variable represents counts.
- Supports Poisson and negative binomial models, zero-inflated Poisson and negative binomial models, and can fit separate regressors for the zero-inflated distribution.

High-performance severity models

- Fits parametric probability distributions for the severity (magnitude) of random events from an empirical portfolio of losses.



```
*-----------------------------------------------------------
User:                   cox
Date:                   August 06, 2012
Time:                   15:41:20
*-----------------------------------------------------------
* Training Output
*-----------------------------------------------------------




Variable Summary

            Measurement    Frequency
Role           Level         Count
```

Term	Role	Attribute	Freq	# Docs	Keep	Weight
+ vehicle ...	Noun	Alpha	582143	296123	Y	0.077
ak ...	Prop	Alpha	231551	231076	Y	0.081
+ dealer ...	Noun	Alpha	269917	195015	Y	0.103
+ problem ...	Noun	Alpha	295127	179316	Y	0.113
+ replace ...	Verb	Alpha	193938	141962	Y	0.126
tr ...	Prop	Alpha	141963	141783	Y	0.117
+ car ...	Noun	Alpha	303507	128973	Y	0.145
+ cause ...	Verb	Alpha	130882	113480	Y	0.138
+ failure ...	Noun	Alpha	176514	113192	Y	0.144
+ consumer ...	Noun	Alpha	159341	103595	Y	0.151
+ time ...	Noun	Alpha	128148	97346	Y	0.153
+ brake ...	Noun	Alpha	166223	94382	Y	0.163
+ driving ...	Noun	Alpha	98306	91385	Y	0.152
+ state ...	Verb	Alpha	116099	87467	Y	0.160
+ recall ...	Noun	Alpha	129196	86155	Y	0.164

Figure 5: View graphical results of term attributes analysis in SAS High-Performance Text Mining.

Key Features (continued)

- Fits regression models for the scale of the severity distribution.
- Automatically selects the best-fitting distribution from among nine probability distributions or lets users to choose their fit statistic of choice.
- Enables users to add additional probability distributions.
- Users can model truncation (deductibles) and censoring (policy limits) in loss values.

High-performance qualitative and limited independent variable models

- Fits linear, logit/probit, censored and truncated regression models with heteroscedasticity and stochastic frontier production and cost models.
- Allows for estimation of both univariate and multivariate response models.
- Bayesian tools allow users to find posterior probability distributions for parameters.

High-performance panel data models

- Estimates linear panel models with either one-way or two-way fixed or random effects.

High-performance copula simulation

- Uses information regarding a given correlation structure to simulate data from a specified multivariate copula.

High-performance compound distribution model simulation

- Uses count data models in conjunction with severity models to build aggregate loss distribution models for insurance and banking applications.
- Useful in conducting what-if and other scenario analysis where certain assumptions are varied. It can also be used in bank-reported VaR (Value at Risk) for certain loss types.
- Fully flexible syntax allows simulation to consider certain business rules so many different insurance schemes (deductibles, policy limits, etc.) can be layered.

SAS® High-Performance Optimization

- Global/local search optimizes general user-defined functions (nonlinear, nondifferentiable, etc.) with continuous and integer decision variables and linear and nonlinear constraints.
- Decomposition algorithm implements a more efficient approach for solving block-angular linear and mixed integer linear optimization problems. Single- and multi-objective optimization is supported.
- Multistart optimization increases the likelihood of finding global solutions to nonlinear optimization problems with many locally optimal solutions.
- Option tuning identifies the most effective optimization solver option settings for a specified mixed integer linear optimization problem (or set of problems).

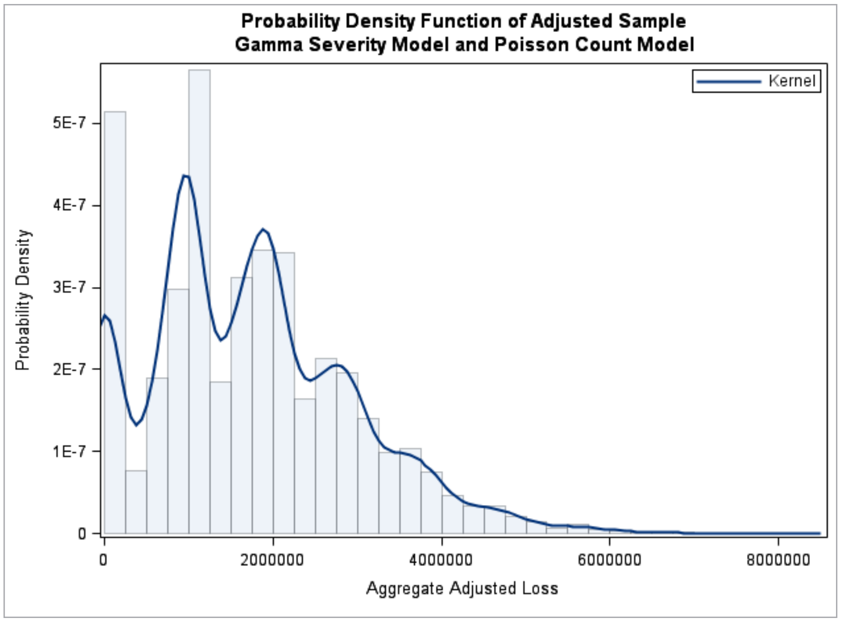


Figure 6: SAS High-Performance Econometrics lets you apply various insurance layering schemes to the simulation of aggregate losses.

To learn more about SAS® High-Performance Analytics products' features and technical requirements, please visit sas.com/hpanalytics.

THE POWER TO KNOW.