SAS® Factory Miner

Run model tournaments at scale with advanced machine learning – quickly, easily and automatically

What does SAS® Factory Miner do?

SAS Factory Miner provides a web-based, automated, yet customizable, environment for building, comparing and retraining predictive models at scale across multiple business segments. Test your favorite machine-learning algorithms in tournaments and find the best performing models with just a few clicks.

Why is SAS® Factory Miner important?

The automation, ease of use and scalability of SAS Factory Miner ramps up your predictive model building power and increases the productivity of your analytics staff. They can quickly share models, templates, results and best practices. Better, more precise models ensure that your organization won't miss profitable opportunities or engage in unknown risks.

For whom is SAS® Factory Miner designed?

It is primarily designed for data miners, statisticians, data scientists, database marketers, business analysts, etc., who need an automated, high-volume mechanism to build and evaluate many models in a short amount of time.

As organizations begin to apply analytics to growing numbers of customer and business segments, predictive models often must be developed at increasingly granular levels. And while this makes it necessary to create more models, most analysts and data scientists don't have the luxury of more time.

SAS Factory Miner provides a web-based interactive environment that makes it extremely easy to create, modify and run hundreds, or even thousands, of models very quickly.

With just a few clicks, you can modify and transform your data, choose which machine-learning techniques you want to use and run the models in an automated model tournament environment to quickly identify the best performer for each segment. If your models are underperforming in a specific segment, it's easy to manually fine-tune these models to improve the accuracy of your predictions.

Benefits

- **Boost discovery productivity.** Reap huge productivity gains by automating time-consuming model development processes – including data preparation, variable transformation, predictor variable and algorithm selection, etc. SAS Factory Miner's easy-to-use web-based interface lets you build multiple models for each segment and automatically identifies the champion.

- **Automate model development**. Choose the best segmentation strategy to solve your business problems. And jump-start your predictive modeling with a set of prebuilt model building templates that you can customize to fit your needs. Share successful templates with your co-workers, and use automated reporting and documentation to implement best practices on model design and results across your organization.

- **Explore new ideas faster.** Apply machine learning and predictive analytic techniques to large, complex data sets, and get fast results. If a model fails, you can try again quickly using different inputs or ideas. As variables change or new variables are found, you can test them without having to rebuild the entire data mining flow or challenge an existing set of algorithms.

- **Put models into operation quickly.** Deploy champion models with just the click of a button. SAS Factory Miner automatically generates complete scoring code – including all necessary data preparation and transformation steps. And retraining models is easy because all assets related to model development and deployment are centrally managed and accessible via REST endpoints.

Product Overview

SAS Factory Miner provides an automated framework for building, comparing and retraining hundreds, even thousands, of models quickly, easily and automatically.

You work through a web-based, drag-and-drop interface to build models for multiple business segments (such as products, regions and channels) or for different customer segments (low- and high-value customers, etc.). The software tests multiple models simultaneously using statistical and machine-learning algorithms, and it identifies the best performing model for each segment based on predefined performance statistics. If you are not satisfied with the automated approach, you can manually fine-tune every single model.

With automation for the predictive model discovery pipeline (data definitions, data cleaning, feature engineering and selection, modeling and model assessment), SAS Factory Miner shortens model development time.
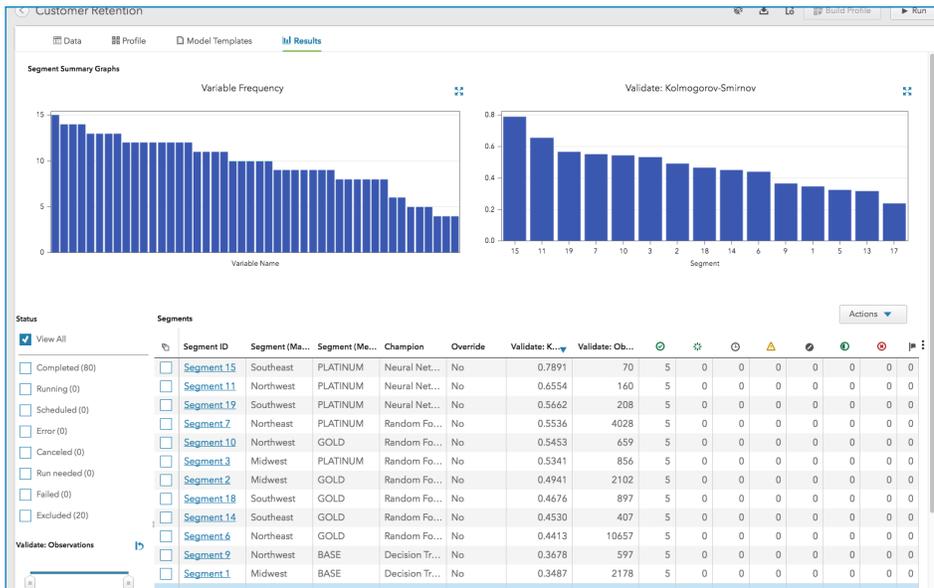
Both experienced and novice analysts can become much more productive, efficient and focused on deriving valuable outcomes from granular segments for more accurate insights and better decisions.

Data access and preparation

Through the interface users can easily access SAS data sets, database tables and Hadoop files.

The next step is preparing data, which is usually the most time-consuming aspect of analytical modeling. With SAS Factory Miner, users can assign variable roles and levels once and then share the rules with others. This means every user can rely on consistent definitions for essential rules, such as which attribute is the target, which variables define the segmentation strategy and which variables are input features.

Automated data profiling provides a quick overview of the analytical base table with the ability to identify issues with the data, including missing values, outliers or skewed variable distributions. Interactive data preparation tools make it easy to apply required data transformations, derive new variables and run intelligent feature selection methods, such as variable selection based on trees and random forests.

Customizable model templates

Easy-to-use model building templates that take advantage of cutting-edge machine-learning techniques such as random forests, gradient boosting, support vector machines and Bayesian networks, as well as decision trees and regression models, are provided out of the box. All of these templates include a list of building blocks for the model development pipeline and default settings for the required parameters based on SAS best practices.

You can customize, edit and share your model templates across projects and with others to build your own set of best-practice modeling pipelines for the organization. Drill into any model and customize it to improve model lift. Decide whether and how to apply redundancy reduction techniques such as

Customizable assessment techniques allow you to generate champion models for every segment in your data.

You can access model fit and make any necessary changes for all of your predictive models.

variable selection or principal components, filtering for outliers, missing imputation and variable transformation – by simply dragging and dropping building blocks in your model template.

For every factory project, you can choose which modeling templates you want to compete in the tournament – selecting from default templates, your own templates or those shared by your colleagues.

Self-service machine-learning

Self-service machine learning and predictive analytics enable you to efficiently test new ideas within an intuitive interface. Without writing a single line of code or having a deep understanding of how the machine-learning techniques work, you can create a set of baseline models very quickly.

For models that underperform and do not meet organizational standards, your analytical experts can easily drill deep into the engine, customize the building blocks of the modeling pipeline and even fine-tune parameter settings of each building block. The fine-tuned settings can be saved and shared in customized templates.

SAS Factory Miner automatically selects the champion model for each segment using customizable assessment techniques. You also have the ability to override system-selected models and manually identify the champion model using your favorite criteria.

Models can also be compared across segments to identify relevant differences in the performance as well as the main effects (drivers) for each model.

Model exception identification

To help improve model performance, SAS Factory Miner provides standardized, easy-to-understand reports that pinpoint issues with models and identify the best models with high confidence, even for nontechnical users. You can then easily recognize under-performing models and drill into details for models that need fine-tuning. Automated documentation makes it easy to share best practices on model design, as well as the

Key Features

Data access and preparation
- Access data sources registered in SAS Metadata Server, including SAS data sets, database tables and Hadoop files.
- Interactively assign data source metadata, such as variable roles, levels and order, or use automated settings to share variable settings across projects.
- Define segments in your data for stratified modeling.
- Assess data issues with automated data profiling and interactive variable distribution graphs.
- Filter your analytical input data with these techniques: Winsorized, standard deviations, trimmed, and rare values.
- Apply transformations to your data for better models: log, log 10, square root, inverse, square, exponential, centering, standardized, range, bucketed, pseudo-quantile, optimal binning and principal component analysis.
- Clean your data with statistical and machine-learning imputation methods: mean, min, max, median, midrange, constant, count-based and distribution-based.

Customizable supervised learning templates
- Interactively build custom templates that include models and the following processing steps: filtering, principal components, imputation, transformations, and supervised and unsupervised variable selection.
- Create your own model templates.
- Edit any data preparation or model parameters and save as customized templates.
- Share model templates across projects and users.

Self-service machine learning techniques
- Build models using the following techniques:
  - Bayesian networks.
  - Decision trees.
  - Gradient boosting.
  - Neural networks.
  - Random forests.
  - Support vector machines.
  - Generalized linear models.
  - Linear regression.
  - Logistic regression.
- Interactively view model-specific results.

Champion model identification
- Champion models are automatically selected for each segment using selectable criteria:
  - Kolmogorov-Smirnnov.
  - Lift and cumulative lift.
  - Gain and cumulative gain.
  - Misclassification rate.
  - Percent captured event.
  - Average percent captured event.
  - Average square error.
- Override system-selected models and manually identify your champion model.
- Interactively compare and assess models within a segment and across multiple segments.

Model exceptions identification
- View reports that highlight model performance exceptions.
- Easily identify and drill into details for underperforming models.
- Modify default settings for each model template.

Model tracking and reporting
- Generate summary reports that contain model results, significant variables and model settings.
- Share reports via PDFs and RTFs.

results, among stakeholders across your organization. This can also help you fulfill compliance requirements.

Flexible model deployment and management

Once the model tournaments have identified the champions for each segment, models need to be deployed in operational environments for regular scoring. This can be tedious, especially when it entails manually rewriting or converting code, which delays model implementation and can introduce potentially costly errors.

SAS Factory Miner automatically creates SAS score code for each model, including all necessary steps for data preparation and transformations. Combined with a SAS Scoring Accelerator, SAS Factory Miner models can be published directly into a variety of databases or into Hadoop. In addition, all model development and scoring assets can be registered to SAS Model Manager, a centralized web-based environment for managing the life cycle and governance of your modeling assets from SAS or third-party providers.

Model retraining

Predictive models have a limited lifespan. Relationships identified by the machine-learning algorithms in the data during development often change over time, and it becomes necessary to retrain the models. With SAS Factory Miner, just point your saved model tournament to the new data and rerun. You can also automatically retrain models with new data via a REST endpoint.

The built-in history of model assessment results enables you to easily see if the new models outperform the previous batch. New

Key Features (continued)

Model retraining
- Retrain models in batch through REST endpoints.
- Retrain existing model templates on new data sets.
- Track model-build assessment statistics across retraining iterations.
- Longitudinal model performance degradation reports.

Flexible model management and deployment
- Automatically generate SAS score code for all model templates.
- Register models to SAS Model Manager for centralized model deployment and management. (Requires SAS Model Manager.)
- Deploy models in database and in Hadoop. (Requires SAS Scoring

champion models can then be promoted to production to replace the existing underperforming champion models.

Scalable processing

SAS Factory Miner can run analytical models in a single-machine environment, across grids using SAS Grid Manager to take advantage of workload balancing and scheduling, or in memory using SAS High-Performance Data Mining. This means you can run in the environment that best fits the processing to your problem. And it's easy to scale out when your data grows, your problem becomes more complex or your analytical maturity increases.

Accelerator for Hadoop.)

Scalable processing
- Train models using multithreaded procedures on SAS servers to take advantage of multicore servers.
- Train models using asynchronous processes via SAS Grid Manager for workload balancing and scheduling. (Requires SAS Grid Manager.)
- Train models in memory using SAS High-Performance Data Mining on database appliances (Oracle, Teradata, Greenplum and SAP HANA) or on Hadoop. (Requires SAS High-Performance Data Mining.)

TO LEARN MORE »

To learn more about SAS Factory Miner system requirements, view screenshots and see other related materials please visit sas.com/factory-miner.