

CT Global Solutions & SAS®

# EasyAI™ Visual Data Mining & Machine Learning

Taking Demand Planning to the Next Level with  
More Accurate Predictive Models &  
Machine-Driven Automation

## SAS® Visual Data Mining & Machine Learning: What is it?

**SAS® Visual Data Mining & Machine Learning (VDMML)** is one of the main solutions in CT Global's **EasyAI™**. It provides a comprehensive, visual interface for accomplishing all steps related to the analytical life cycle. In addition to innovative machine learning and deep learning techniques for analyzing structured and unstructured data, it integrates all other tasks in your analytical processes. From data preparation and exploration to model development and deployment, multiple personas work in the same, integrated environment. Scalable and elastic processing provides flexibility and speed for faster answers to complex questions.

## When is SAS Visual Data Mining & Machine Learning the Best Option for Your Analytics Needs?

SAS® VDMML is the next step beyond time-series forecasting and initial predictive models. Building on this solid analytic foundation SAS VDMML uses Machine Learning and Artificial Intelligence to take your results to a higher level of statistical accuracy and business value. SAS VDMML also adds causal factors to create more complete and accurate models of your demand and supply environment.

SAS® VDMML is an easy step up to a higher level of capability because it is so easy to use. Quickly explore, mine, develop, and compare predictive models within a single interactive environment. State of the art Machine Learning and AI automatically deliver optimal insights that empower users to make actionable *and profitable* business decisions.

### Examples of Data Mining & Machine Learning

- > Decision Trees
- > Neural Networks
- > Logistic Regression
- > Linear Regression
- > Gradient Boosting
- > Random Forests
- > And so much more....



# What You Can Do with SAS® VDMML?

## Interactive programming in a web-based development environment

- Visual interface includes interactive drag-and-drop for the entire analytical life cycle
- Create code automatically at each node in the pipeline.
- Explore and view data within each node and run data mining using Model Studio.
- Collaborate with other users to share data and best practices.

## Model development with modern machine learning algorithms

- Automate model development using decision forests, ensemble of decision trees, distribution of independent training runs, autotuning of model parameters, generation of SAS code for production scoring, search for optimal partitions, and much more.
- Customize tree training based on splitting criteria, depth of subtrees and compute resources.

## Support vector machines

- Use linear and polynomial kernels for model training.
- Include continuous and categorical in/out features.
- Automated scaling of input features.
- Apply the interior-point method and the active-set method.
- Use data partition for model validation.
- Use cross-validation for penalty selection.

## Bayesian networks

- Use different Bayesian network structures, including naive, tree-augmented naive (TAN), Bayesian network-augmented naive (BAN), parent-child Bayesian networks and Markov blanket.
- Use independence tests to efficiently select variables.
- Select the best model automatically from specified parameters.

## Analytical data preparation

- Leverage distributed data management, and large-scale data exploration and summarization.
- Profile large-scale data and cardinality for input data sources.
- Impute missing values.
- Use random and stratified sampling, oversampling for rare events and indicator variables for sampled records.

## Scalable, distributed in-memory analytical processing

- Distributed, in-memory processing of complex analytical calculations on large data sets provide low-latency answers.
- Chain analytical tasks as a single, in-memory job without reloading data or writing to disk.
- Concurrent access to the same data in memory by many users.
- Built-in workload management and parallel processing for efficient use of resources.
- Built-in failover management guarantees for job completion.
- Automated I/O disk spillover for improved memory management.

## Neural networks

- Automatically tune parameters to identify optimal model.
- Set intelligent defaults for most neural network parameters.
- Customize neural networks architecture and weights.
- Use deep forward neural network (DNN), convolutional neural networks (CNNs), recurrent neural networks (RNNs) and autoencoders.
- Use hidden layers to support deep learning.

## Factorization machines

- Create systems based on sparse matrices of user IDs and item ratings.
- Apply full pairwise-interaction tensor factorization.
- Use categorical and numerical input features for more accurate models.
- Supercharge models with timestamps, demographic data and context information.
- Restart (update models with new transactions) without full retraining.

## Semi-supervised learning algorithm

- Create predicted labels for unlabeled and labeled data tables.
- Use distributed and multithreaded stochastic modeling.
- Use parallel implementation to guide low-dimensional embeddings.

## Data exploration, feature engineering and dimension reduction

- T-distributed stochastic neighbor embedding (t-SNE).
- Feature binning & dimension reduction.
- Large-scale principal components analysis (PCA).
- Unsupervised learning with cluster analysis and mixed variable clustering.

### Integrated text analytics

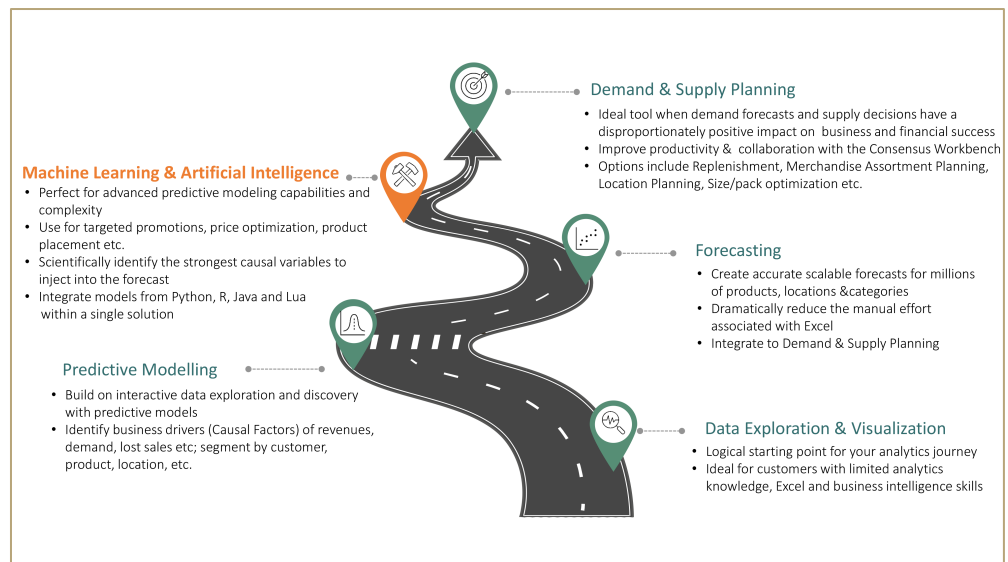
- Supports 32 native languages out of the box
- Automated parsing, tokenization, part-of-speech tagging and lemmatization.
- Predefined extract of common entities such as names, dates, currency values etc.
- Automated feature extraction with machine-generated topics (singular value decomposition and latent Dirichlet allocation).
- Machine Learning and automatic rules generation within a single project.
- Documents classified more accurately with deep learning (recurrent neural networks).

### Model assessment & scoring

- Automatically calculate supervised learning model performance statistics.
- Create output statistics for interval and categorical targets.
- Create lift table for interval and categorical target, and ROC table for categorical target.
- Automatically generate SAS DATA step code for model scoring.
- Apply scoring logic to training, holdout data and new data.


## When is SAS Visual Data Mining & Machine Learning the Best Option for Your Analytics Needs?

SAS® VDMML is the next step beyond time-series forecasting and initial predictive models. Building on this solid analytic foundation SAS VDMML uses Machine Learning and Artificial Intelligence to take your results to a higher level of statistical accuracy and business value. SAS VDMML also adds causal factors to create more complete and accurate models of your demand and supply environment.



SAS® VDMML is an easy step up to a higher level of capability because it is so easy to use. Quickly explore, mine, develop, and compare predictive models within a single interactive environment. State of the art Machine Learning and AI automatically deliver optimal insights that empower users to make actionable *and profitable* business decisions.

## The Power of the PARTNER

 **CT Global Solutions** is a strategic SAS® partner that helps turn your data into profits. A Gold SAS® alliance partner for 18 years, CT Global has demonstrated domain expertise in analytics, cost and profit management, financial management, demand planning and decision optimization. CT Global puts SAS® to work and amplifies its value to MAKE EVERY DECISION COUNT.

**SAS®** is the leader in analytics. Through innovative software and services, SAS empowers and inspires customers around the world to transform data into intelligence. SAS gives you THE POWER TO KNOW®. 92 of the top 100 companies on the 2018 Fortune Global 1000® use SAS customers, [see their stories here](#).



To talk with a CT Global about Demand Planning, please contact us at [info@ctglobalsolutions.com](mailto:info@ctglobalsolutions.com)

