

SAS® Visual Data Mining and Machine Learning (VDMML)

Getting Started

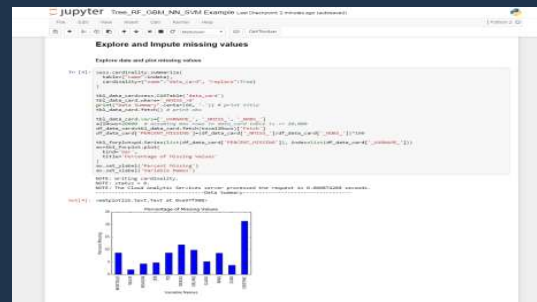
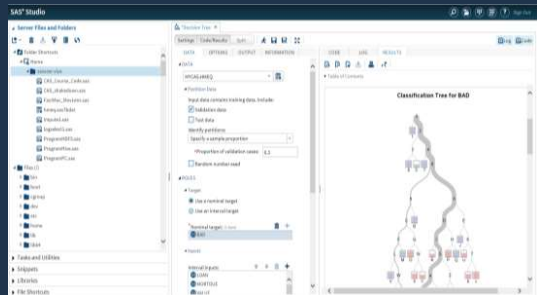
Presenter: Melodie Rush, Principal Data Scientist

Q&A: Twanda Baker, Data Scientist

Host: Lauren Gray, Webinar Host

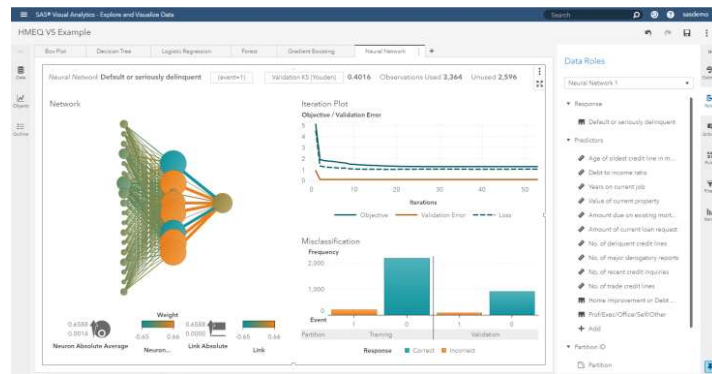
- Increase awareness of and comfort with capabilities in SAS[®] Visual Data Mining and Machine Learning[™] (VDMML)
- Share resources for learning more

SAS® Visual Data Mining and Machine Learning



Presentation Content

- Introduction to SAS® Visual Data Mining and Machine Learning
- Value of SAS® Visual Data Mining and Machine Learning
- Included Algorithms
- Tour of the interfaces
 - Visual
 - Programming
 - Open Source



A photograph of three business professionals in a meeting. A man with glasses and a beard is leaning over a desk, looking at a laptop. A woman with dark hair is looking at the laptop. An older man with white hair is also looking at the laptop. They are in a modern office setting with large windows in the background.

“ The volume of inquiry calls from Gartner clients about AI, advanced ML and related topics increased by 200% between 2015 and 2016. The data science platform segment grew 9.0% from 2015 to 2016

Source: Gartner, July 2017



Pitfalls of Machine Learning



Lack of architecture to support
the analytics life cycle



Need for data science
talent



Missing Data

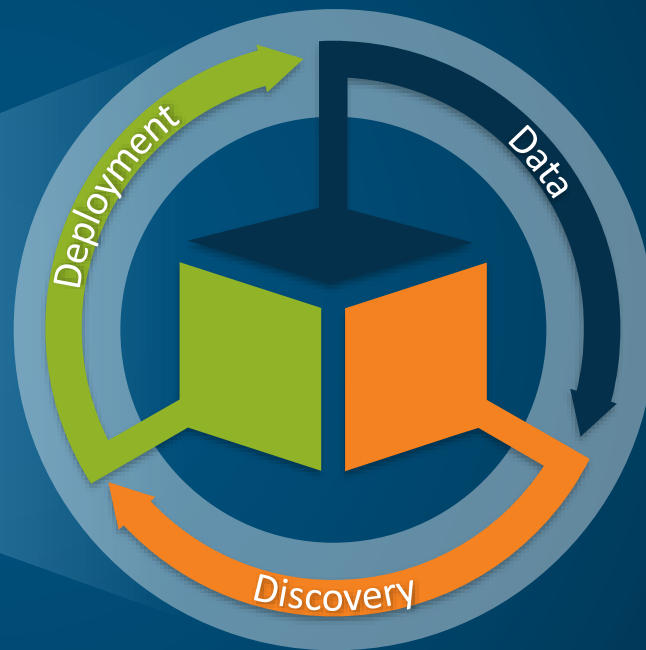


Model performance
deterioration



Long time to interactively
build analytical models

The SAS Platform



Analytics Lifecycle





Integrated
data mining and
machine learning

1

Open platform with
access from SAS and
Open Source

2

Modern machine
learning algorithms

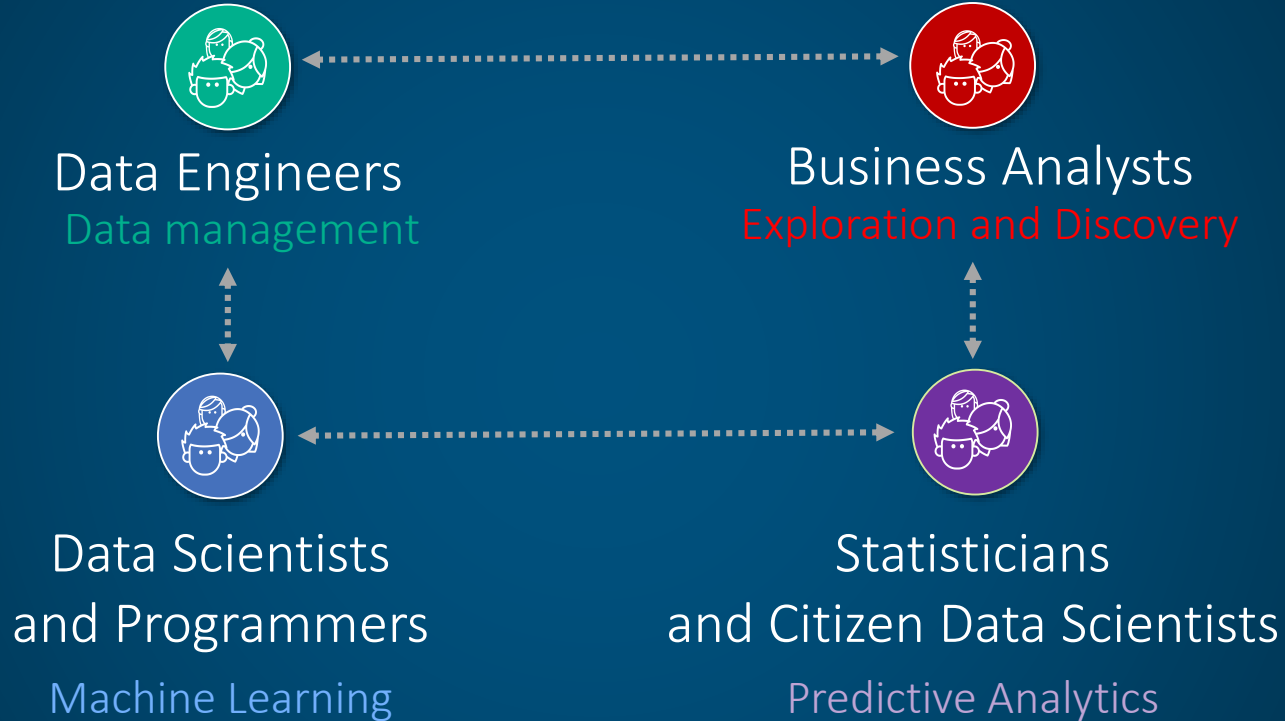
3

Cloud-ready and industrial
scalable analytics

4

SAS® Visual Data Mining and Machine Learning is an end-to-end machine learning solution on the most advanced analytics platform.

Collaboration and Personas



SAS[®] Visual Data Mining and Machine Learning

Key Benefits



SPEED

Get answers fast through integrated advanced analytics workflows in single environment



PRECISION

Get accurate answers to your questions by applying modern machine learning and predictive analytics algorithms



EFFICIENCY

Boost productivity of your data scientists through access from open source



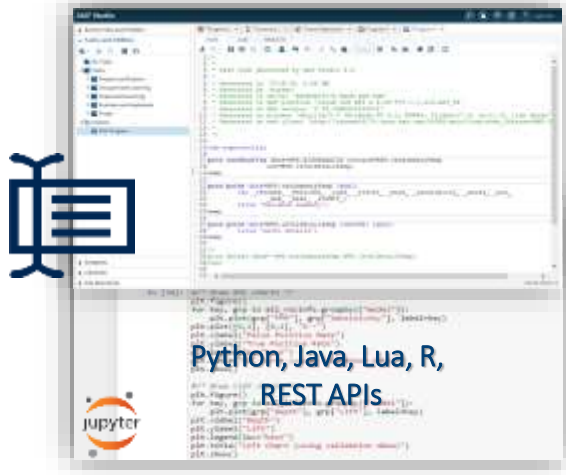
SCALABLE

Scale environment with your growing needs

End to end data preparation, analytics, machine learning and model scoring in a single in-memory, scalable environment

Visual Data Mining and Machine Learning

What does it include?



Requires Visual Analytics

Requires Visual Statistics

Visual Analytics

Visual Statistics

Visual Data Mining
and Machine
Learning

Baseline
Procedures

VS Procedures

VDMML
Procedures



Baseline
Action sets

VS Action
sets

VDMML
Action sets

Visual Data Mining and Machine Learning

What do you get?

Visualizations

- Forest
- Gradient Boosting
- Neural Networks
- Support Vector Machines
- Factorization Machines
- Bayesian Networks

VDMML PROCS

- FOREST
- GRADBOOST
- NNET
- SVMACHINE
- FACTMAC
- TEXTMINE
- TMSCORE
- BOOLRULE
- ASTORE
- CAS
- NETWORK
- BNET
- FASTKNN
- more...

VDMML CAS action sets

- MLEARNING
- TEXTMINE
- DMMLVISSET
- CRSBOOLRULE
- CRSNEURALNET
- CRSSVM
- CRSTKFACTMAC
- TKCAS
- CRSNETSOC
- CSRNETCOMMON
- CRSASTORE
- CRSCMPTRVSN
- CRSDTREEADV
- CRSTXTMINADV
- more...

SAS® Visual Data Mining and Machine Learning 8.3

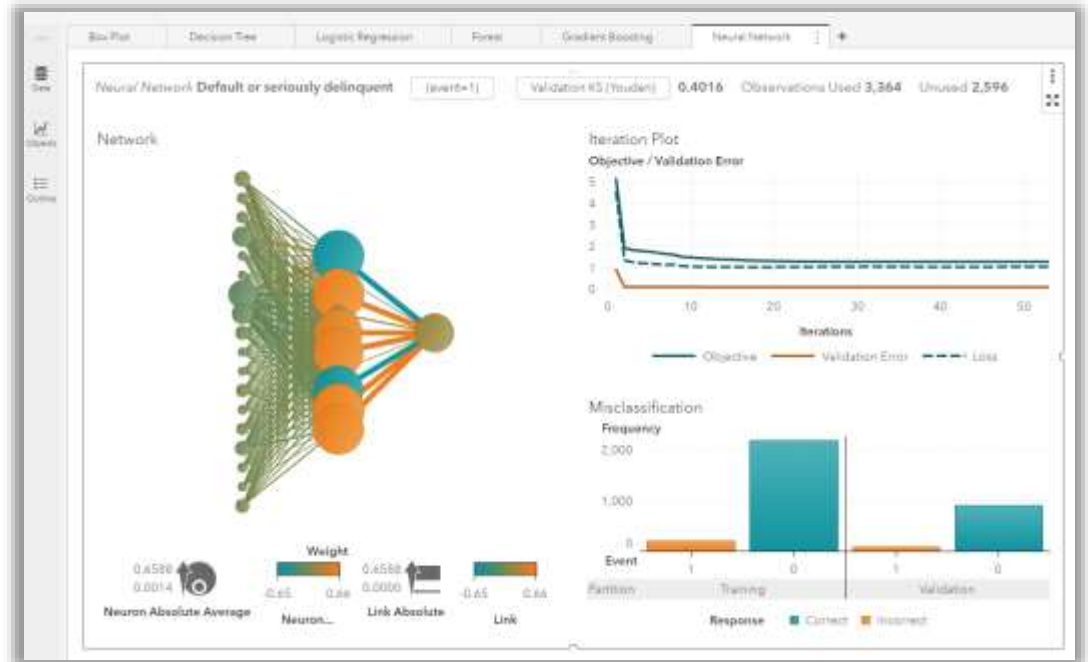
Visual Interface

Machine Learning Techniques

- Forest
- Factorization Machine
- Gradient Boosting
- Neural Network
- Support Vector Machine
- Bayesian Networks

Common Features

- Training-Validation
- Model Assessment
- Model Comparison
- Score Code or Astore Table
- Ability to export model statistics into Excel



SAS Visual Data Mining and Machine Learning 8.3

Programming Tasks in SAS Studio

SAS Viya Prepare and Explore

- Summary
- Transform Data
- Variable Selection
- Sampling
- Partitioning
- Binning
- Imputation

SAS Viya Unsupervised Learning

- Clustering
- Principal Component Analysis
- Moving Window Principal Component Analysis
- Support Vector Data Description
- Robust Principal Component Analysis

SAS Viya Supervised Learning

- Linear Regression
- Logistic Regression
- Generalized Linear Models
- Partial Least Squares Regression
- Quantile Regression
- Decision Tree
- Neural Network
- Forest
- Gradient Boosting
- Factorization Machine
- Support Vector Machine
- Bayesian Network

Includes algorithms in the visual interface plus

- Unsupervised Learning
 - Moving Window PCA
 - Robust PCA
 - Support Vector Data Description
 - Text Parsing and Topic Discovery
- Supervised Learning
 - Quantile Regression
 - Partial Least Squares Regression



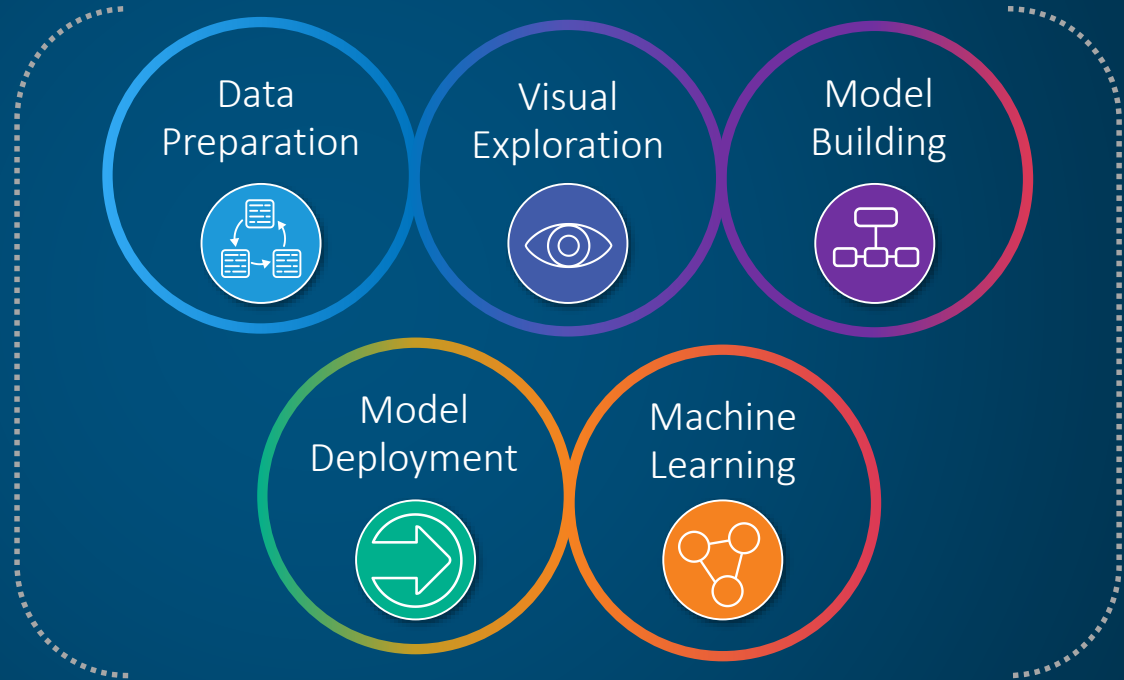
What's Included

SAS® Visual Data Mining and Machine Learning

Visual "drag & drop"
Interface



Programming
Interface





Data Preparation

| Name | Age | Sex | Height | Weight | Cholesterol | Glucose | Insulin | Hemoglobin | Hematocrit | Hemoglobin A1c | Diabetes | Cholesterol | Glucose | Insulin | Hemoglobin | Hematocrit | Hemoglobin A1c | Diabetes |
|---------------|-----|--------|--------|--------|-------------|---------|---------|------------|------------|----------------|----------|-------------|---------|---------|------------|------------|----------------|----------|
| John Doe | 35 | Male | 175 | 70 | 200 | 100 | 10 | 15 | 45 | 5.5 | No | 150 | 100 | 10 | 15 | 45 | 5.5 | No |
| Jane Smith | 42 | Female | 160 | 60 | 180 | 90 | 8 | 12 | 40 | 5.2 | No | 140 | 90 | 8 | 12 | 40 | 5.2 | No |
| Bob Johnson | 28 | Male | 180 | 80 | 220 | 110 | 12 | 18 | 50 | 5.8 | No | 160 | 110 | 12 | 18 | 50 | 5.8 | No |
| Alice Brown | 55 | Female | 155 | 55 | 160 | 85 | 6 | 10 | 35 | 5.1 | No | 130 | 85 | 6 | 10 | 35 | 5.1 | No |
| Charlie Davis | 30 | Male | 170 | 75 | 190 | 105 | 10 | 15 | 45 | 5.6 | No | 155 | 105 | 10 | 15 | 45 | 5.6 | No |
| Diana Evans | 48 | Female | 165 | 65 | 175 | 95 | 9 | 13 | 42 | 5.3 | No | 145 | 95 | 9 | 13 | 42 | 5.3 | No |
| Frank Green | 25 | Male | 185 | 85 | 210 | 115 | 13 | 19 | 52 | 5.9 | No | 165 | 115 | 13 | 19 | 52 | 5.9 | No |
| Grace Hill | 60 | Female | 150 | 50 | 150 | 80 | 5 | 8 | 30 | 5.0 | No | 125 | 80 | 5 | 8 | 30 | 5.0 | No |
| Henry King | 38 | Male | 178 | 78 | 205 | 108 | 11 | 16 | 48 | 5.7 | No | 158 | 108 | 11 | 16 | 48 | 5.7 | No |
| Ivy Lee | 45 | Female | 162 | 62 | 170 | 92 | 8 | 12 | 40 | 5.2 | No | 142 | 92 | 8 | 12 | 40 | 5.2 | No |
| Jack Miller | 22 | Male | 182 | 82 | 215 | 118 | 14 | 20 | 55 | 6.0 | No | 170 | 118 | 14 | 20 | 55 | 6.0 | No |
| Karen Wilson | 52 | Female | 158 | 58 | 165 | 88 | 7 | 11 | 38 | 5.1 | No | 135 | 88 | 7 | 11 | 38 | 5.1 | No |
| Leo Young | 33 | Male | 175 | 75 | 195 | 105 | 10 | 15 | 45 | 5.6 | No | 155 | 105 | 10 | 15 | 45 | 5.6 | No |
| Mia Zane | 40 | Female | 160 | 60 | 170 | 90 | 9 | 13 | 40 | 5.3 | No | 140 | 90 | 9 | 13 | 40 | 5.3 | No |

- Access to different data sources
- Training-Validation Data Partitioning
- Feature Engineering (e.g. parameters, interactions)
- Variable selection and missing values



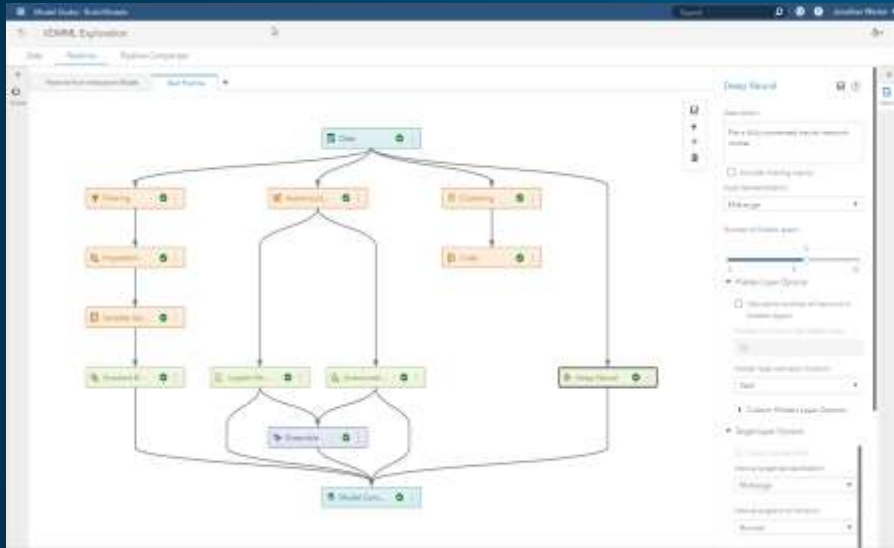
Visual Exploration



- Interactively discover relationships, trends, outliers
- Smart autocharting
- Analytics driven visualizations
- Explore predicted outputs
- Variable transformation



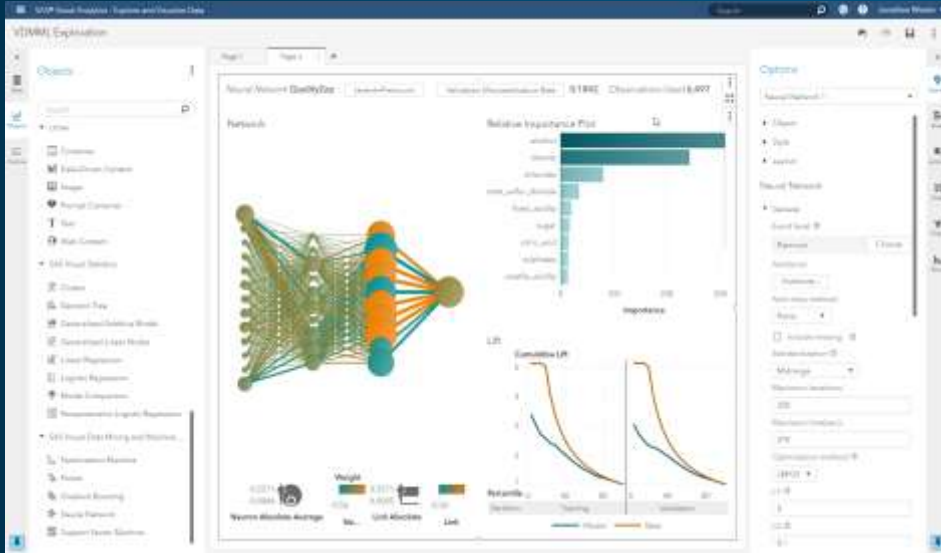
Model Building



- Pipeline of activities
- Drag and drop and access to code
- Nodes are run asynchronously
- Reproducibility
- SAS best practice toolkit



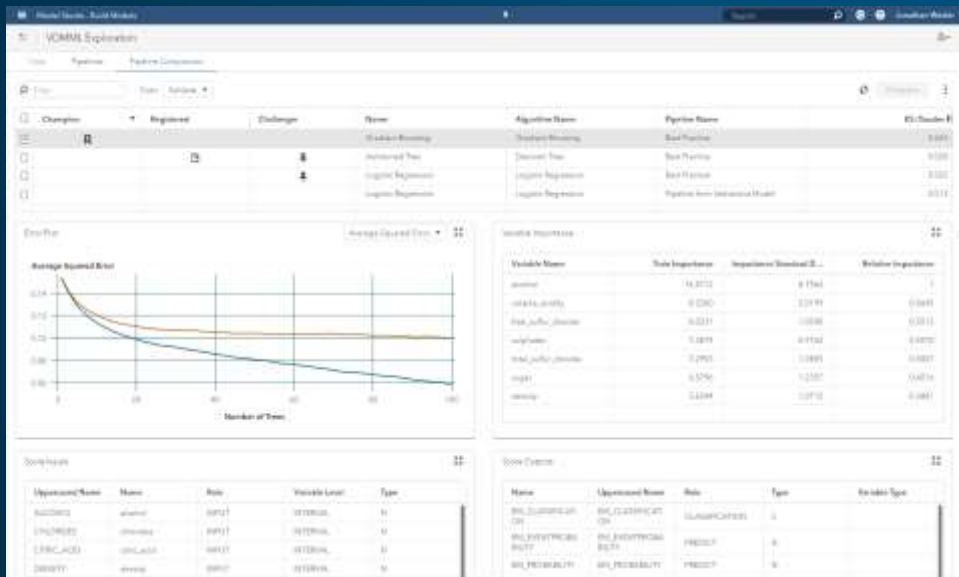
Modern Machine Learning



- Forest
- Neural Network (including Deep Learning)
- Gradient Boosting
- Support Vector Machines
- Factorization Machines
- Bayesian Networks
- Autotuning



Comparison and Deploy



- Model comparison summaries
- Interactively assess models
- Assessment charts for partitioned data
- Publish score code; batch, API call, in-database

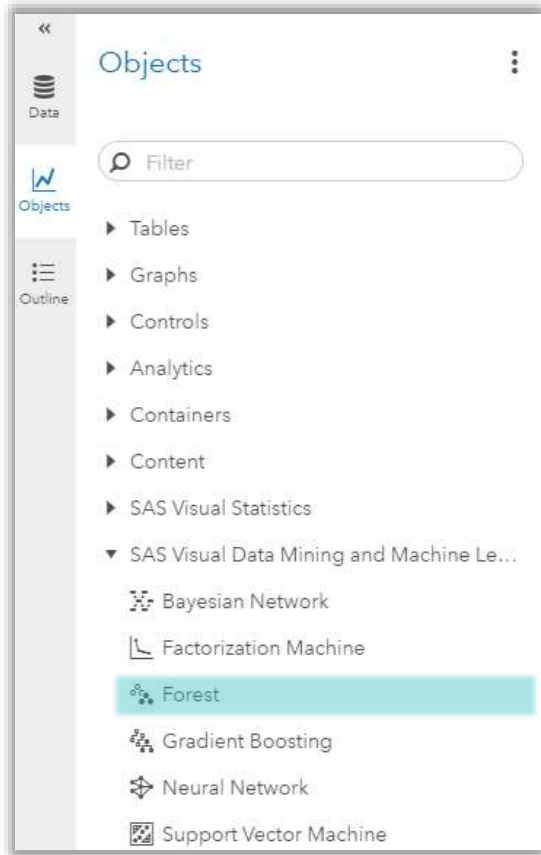


Visual Interface

SAS Visual Analytics

Supervised Learning: VDMML

Forest: PROC FOREST



FOREST

A random forest is used primarily when building classification models on large datasets.

It is an ensemble model of many decision trees from slightly different samples of the training data. Specifically:

- Each tree is built on subset of observations (rows)
- The features (variables) available to each splitting node are subset

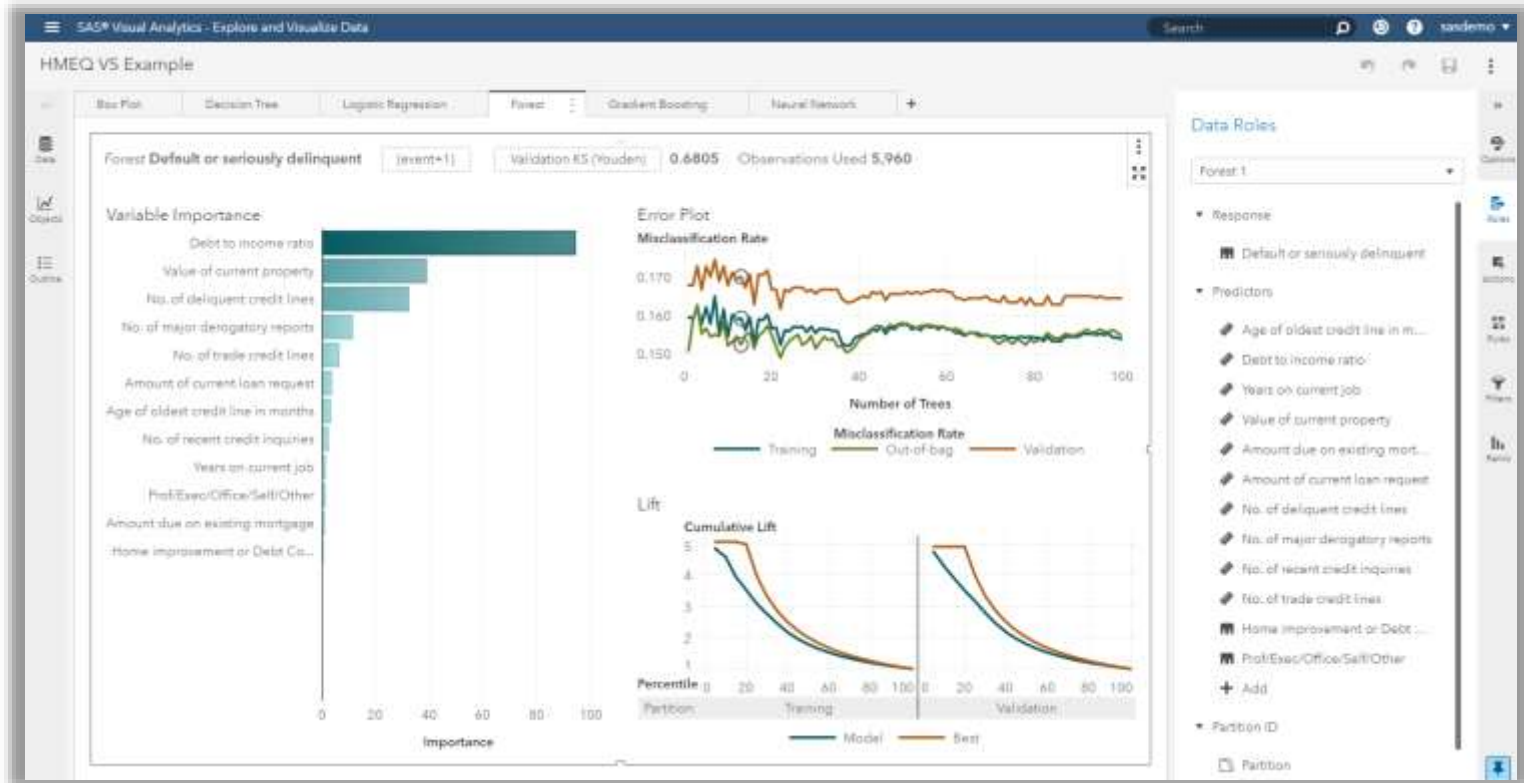
Forest Business Use Cases:

- Customer Churn. Determining the factors that can cause churn.
- Loan Default. Predicting who will default on their loan using customer behavior patterns.

Forest Competitive Differentiators

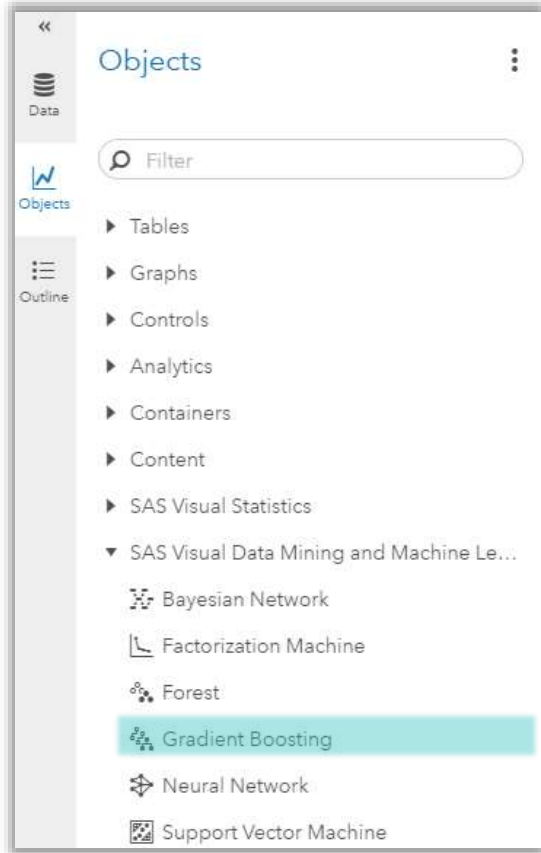
- Distributed and massively parallel
- Faster, more memory-efficient, and more scalable algorithm
- Truly deployable

Supervised Learning: VDMML Forest



Supervised Learning: VDMML

Gradient Boosting: PROC GRADBOOST



Gradient Boosting

Gradient boosting is used primarily when building classification models on large datasets.

It generates many decision trees sequentially from slightly different subsamples of the training data. In the end, all models are given a weight depending on their accuracy, and the model results are combined into one consolidated result.

GB Business Use Cases:

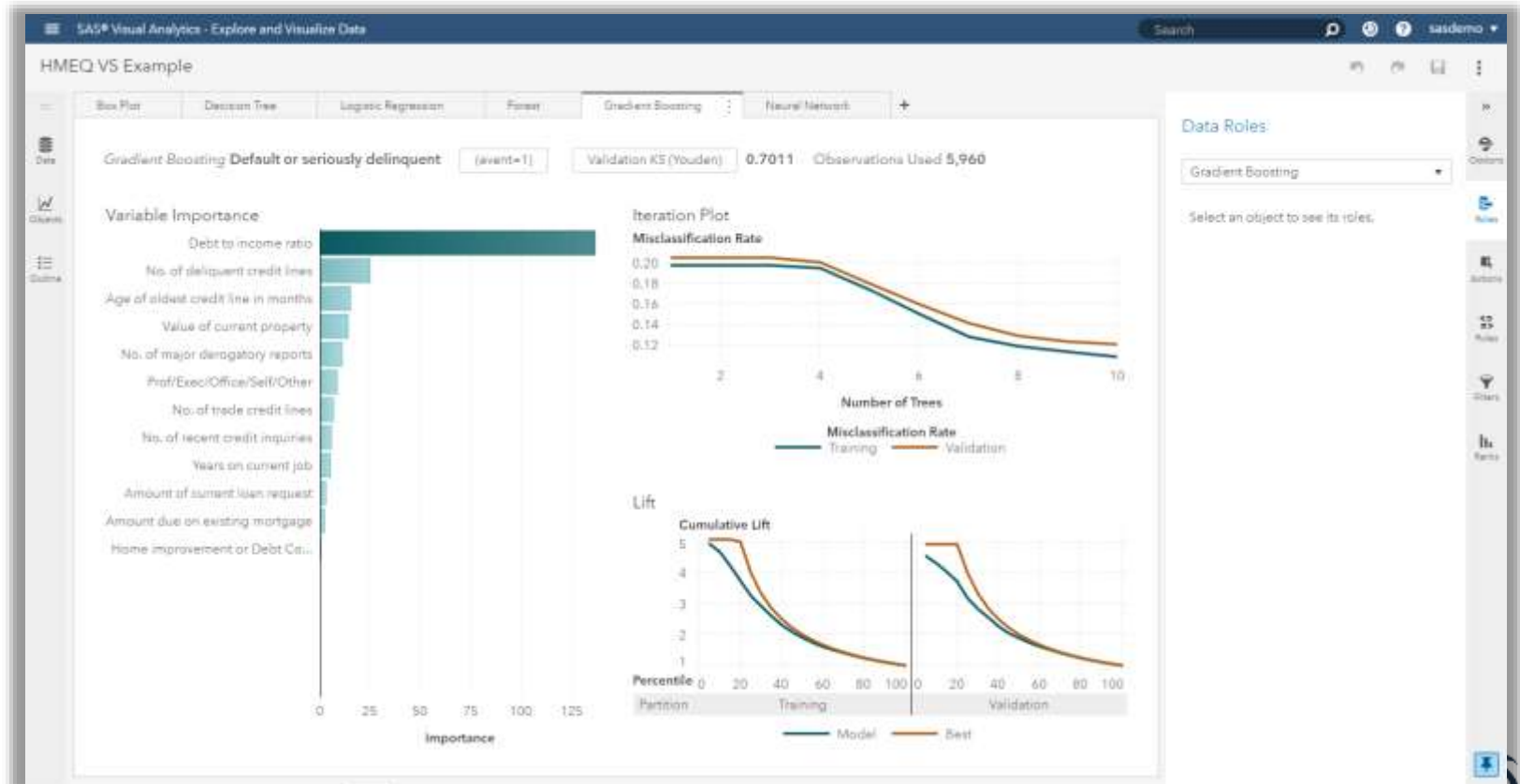
- Customer Churn. Determining the factors that can cause churn.
- Loan Default. Predicting who will default on their loan using customer behavior patterns.

GB Competitive Differentiators

- Distributed and massively parallel
- Faster, more memory-efficient, and more scalable algorithm
- Truly deployable

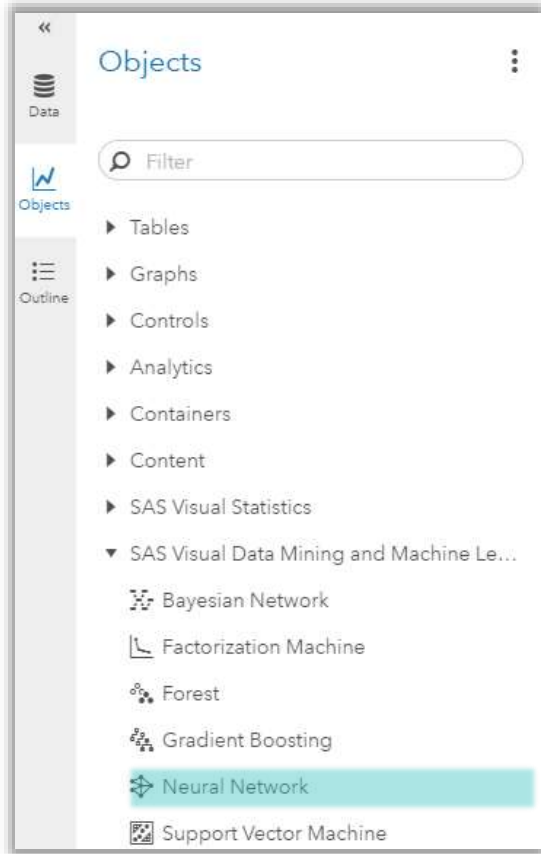
Supervised Learning: VDMML

Gradient Boosting



Supervised Learning: VDMML

Neural Network: PROC NNET



Neural Network

Neural networks are used to solve a wide variety of tasks that are hard to solve using other methods.

The goal of the neural network is to solve problems in the same way that the human brain would. They are comprised of processing elements called units or neurons.

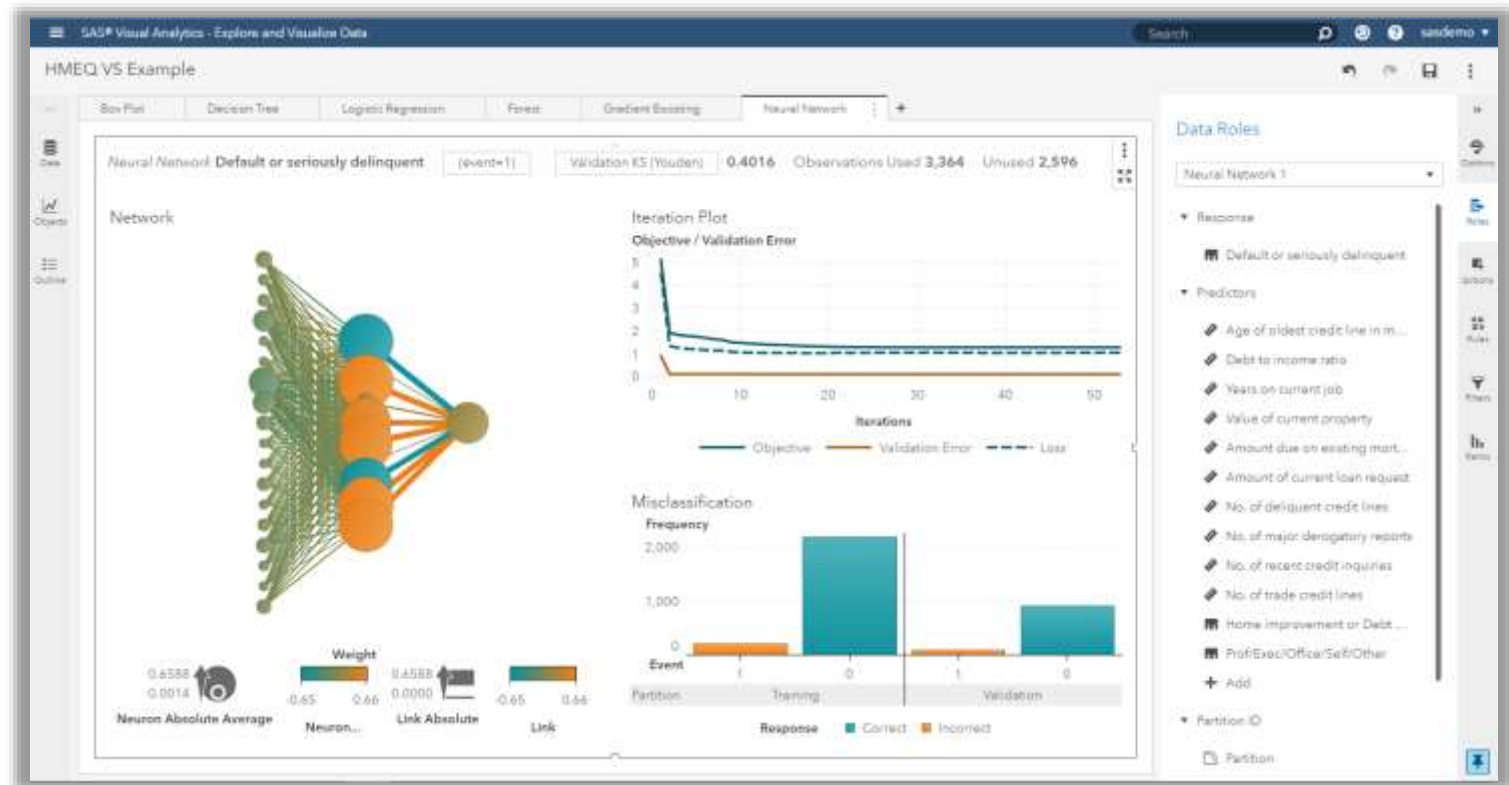
NNET Business Use Cases:

- Computer vision.
- Speech recognition.

NNET Competitive Differentiators

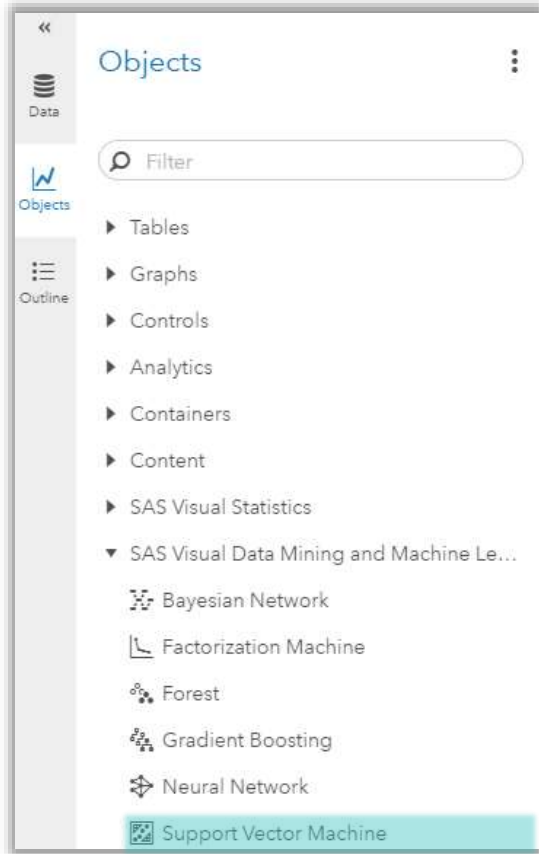
- Distributed and massively parallel
- Faster, more memory-efficient, and more scalable algorithm
- Can be auto-tuned

Supervised Learning: VDMML Neural Networks



Supervised Learning: VDMML

Support Vector Machine: PROC SVMACHINE



Support Vector Machine

Support vector machine is used to separate or classify data into groups.

The standard SVM model solves binary classification problems that produce non-probability output (only sign +1/-1) by constructing a set of hyperplanes that maximize the margin between two classes.

SVM Business Use Cases:

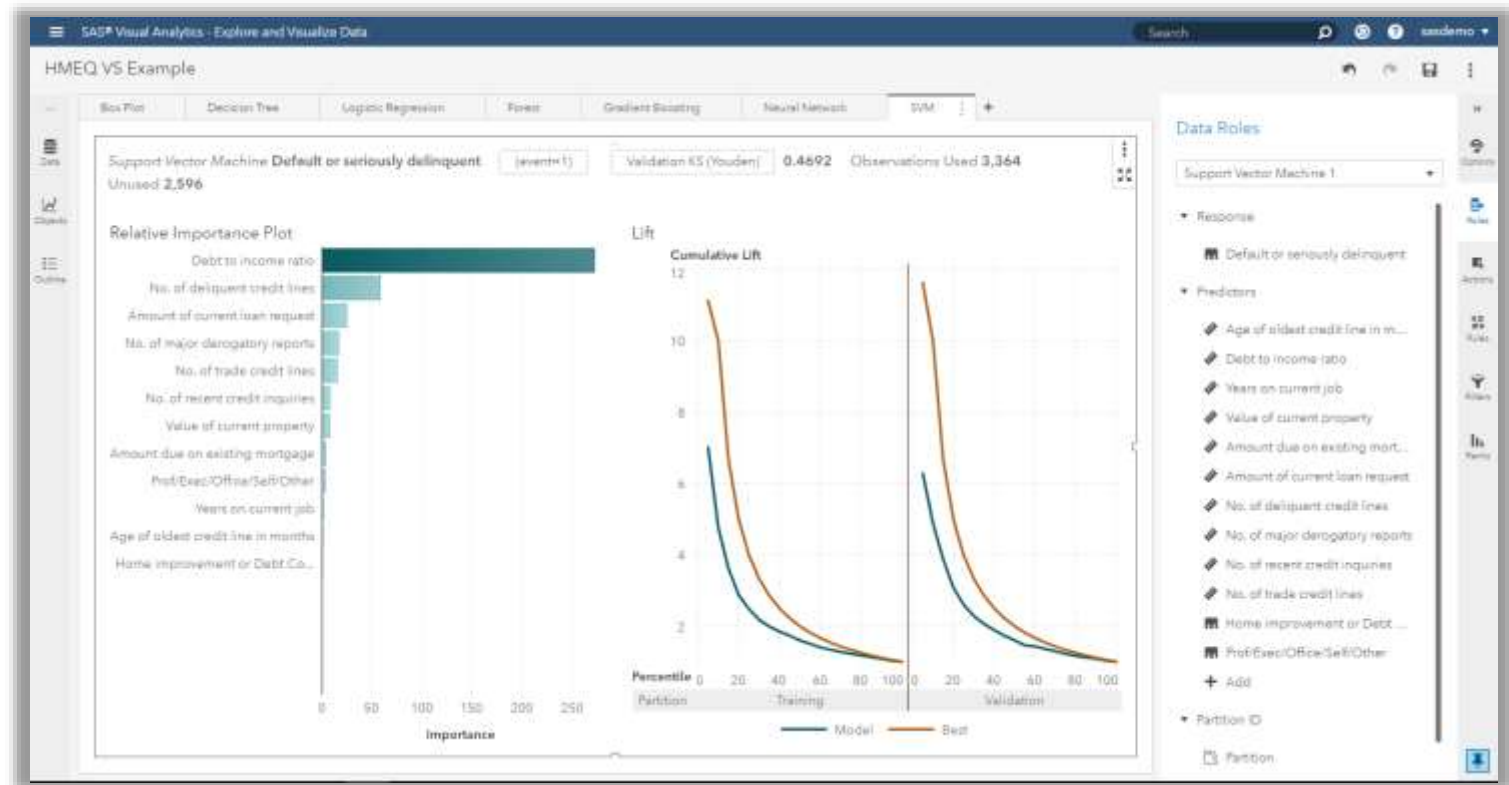
- Customer survey. Predicting which product a customer might buy next based on previous behavior and survey results.
- Predicting bad debt or loan default.

SVM Competitive Differentiators

- Distributed and massively parallel
- Faster, more memory-efficient, and more scalable algorithm
- Truly deployable

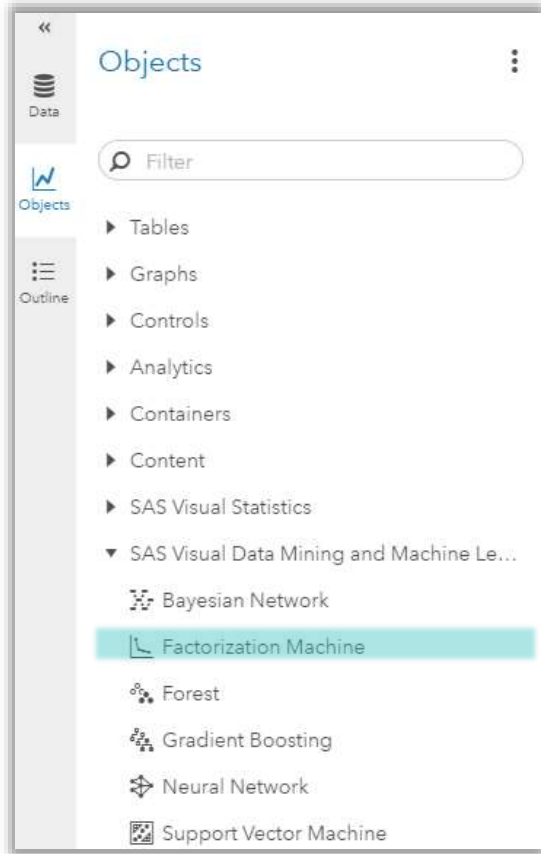
Supervised Learning: VDMML

Support Vector Machines



Supervised Learning: VDMML

Factorization Machine: PROC FACTMAC



Factorization Machine

Factorization machines are used when data has high dimensionality and sparse data.

The most common use case is in recommender engines.

“If you like item A, you may also like item B”.

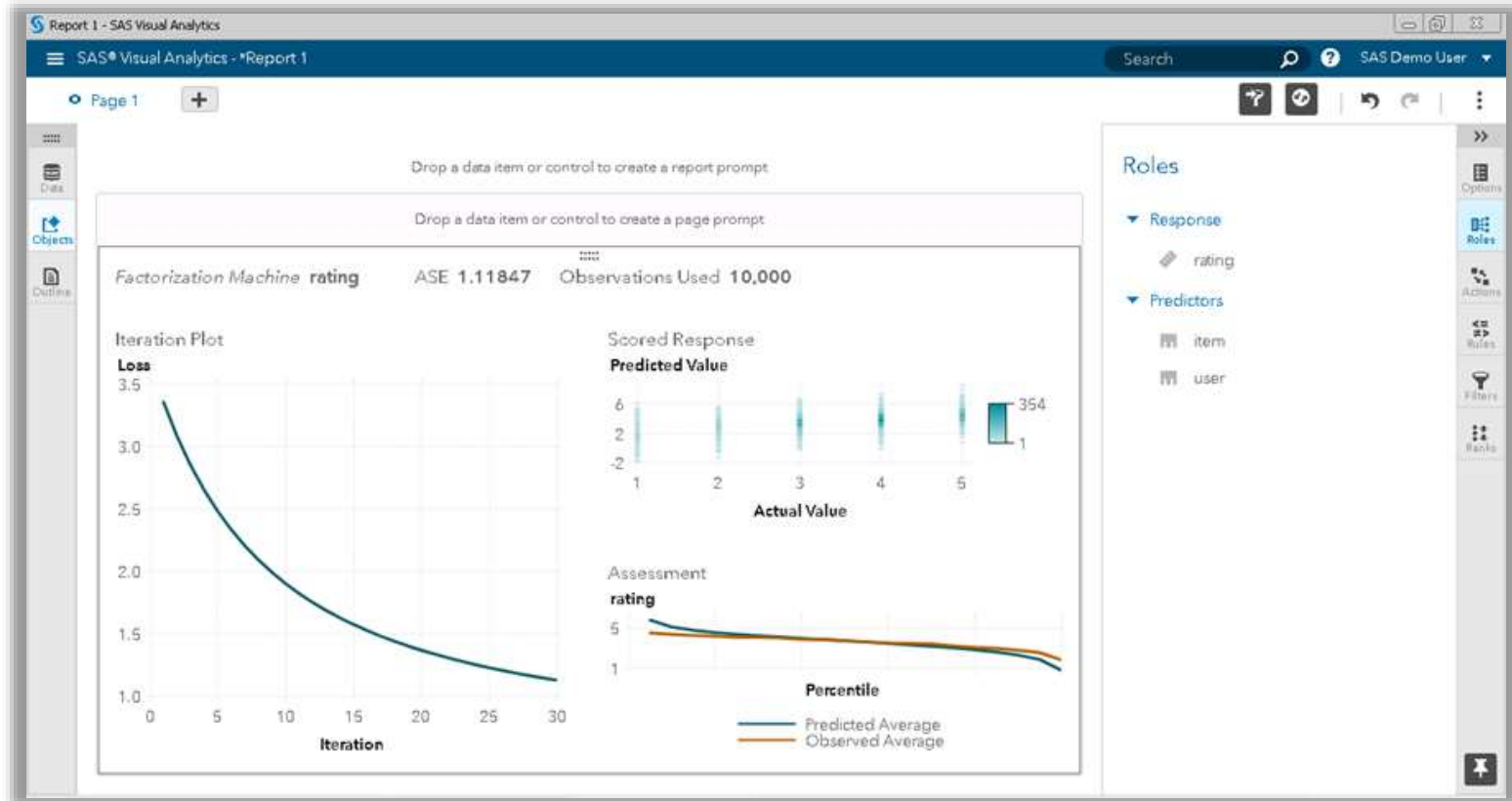
FM Business Use Cases:

- Customer survey. Predicting which product a customer might buy next based on previous behavior and survey results.
- Online real-time recommendations. Surfacing relevant recommendations based on products that the customer views.

FM Competitive Differentiators

- Distributed and massively parallel
- Faster, more memory-efficient, and more scalable algorithm
- Truly deployable

Supervised Learning: VDMML Factorization Machine



Supervised Learning: VDMML

Bayesian Network: PROC BNETH

Bayesian Network

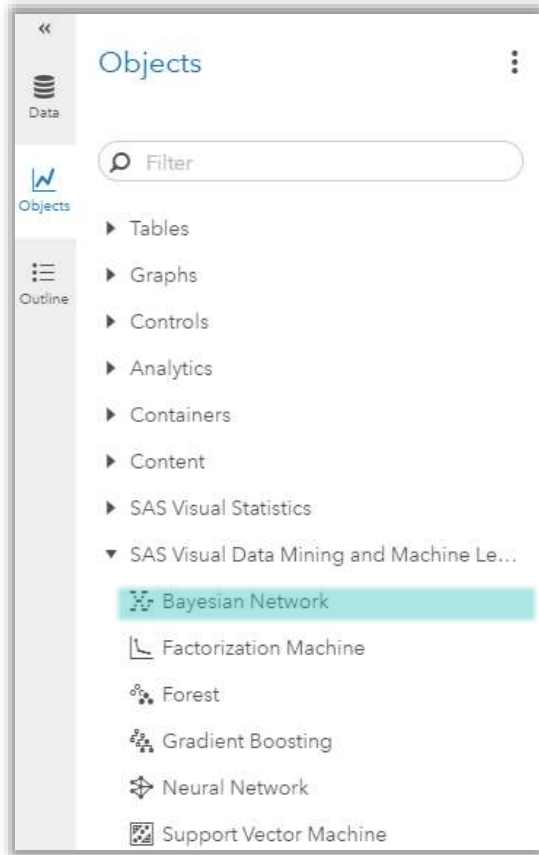
A Bayesian network is a directed, acyclic graphical model in which the nodes represent random variables and the links between the nodes represent conditional dependency between two random variables.

BNETH Business Use Cases:

- They can be used for a wide range of tasks including prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction and decision making under uncertainty.

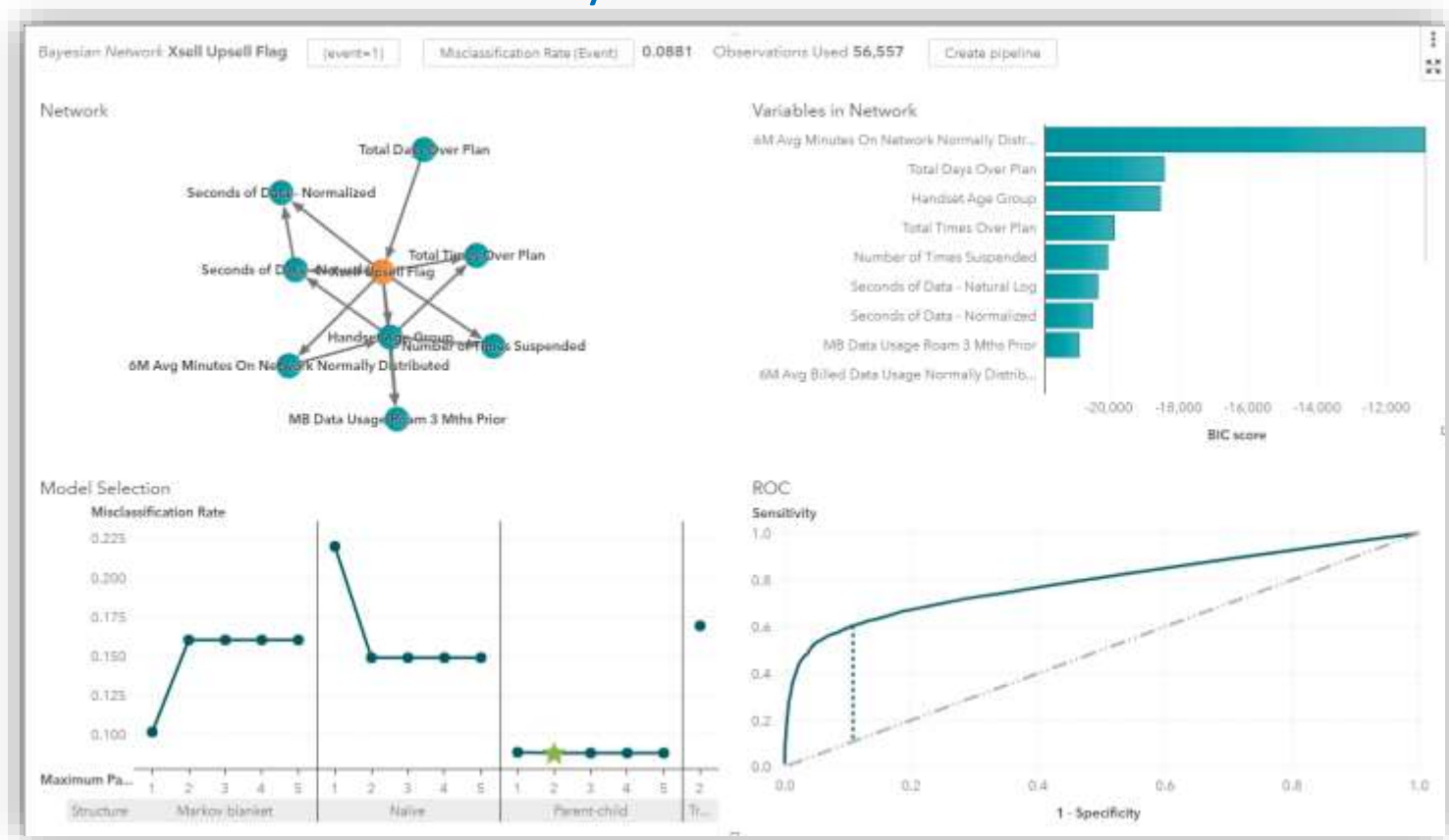
BNETH Competitive Differentiators

- When you have a lot of missing data, e.g. in medicine, BN's can be very effective since modeling the joint reduces your dependency in having a fully observed dataset.
- When you want to model a domain in a way that is visually transparent, and also aims to capture cause→effect relationships, BN's can be very powerful. Note that the causality assumption in BN's is open to debate though.



Supervised Learning: VDMML

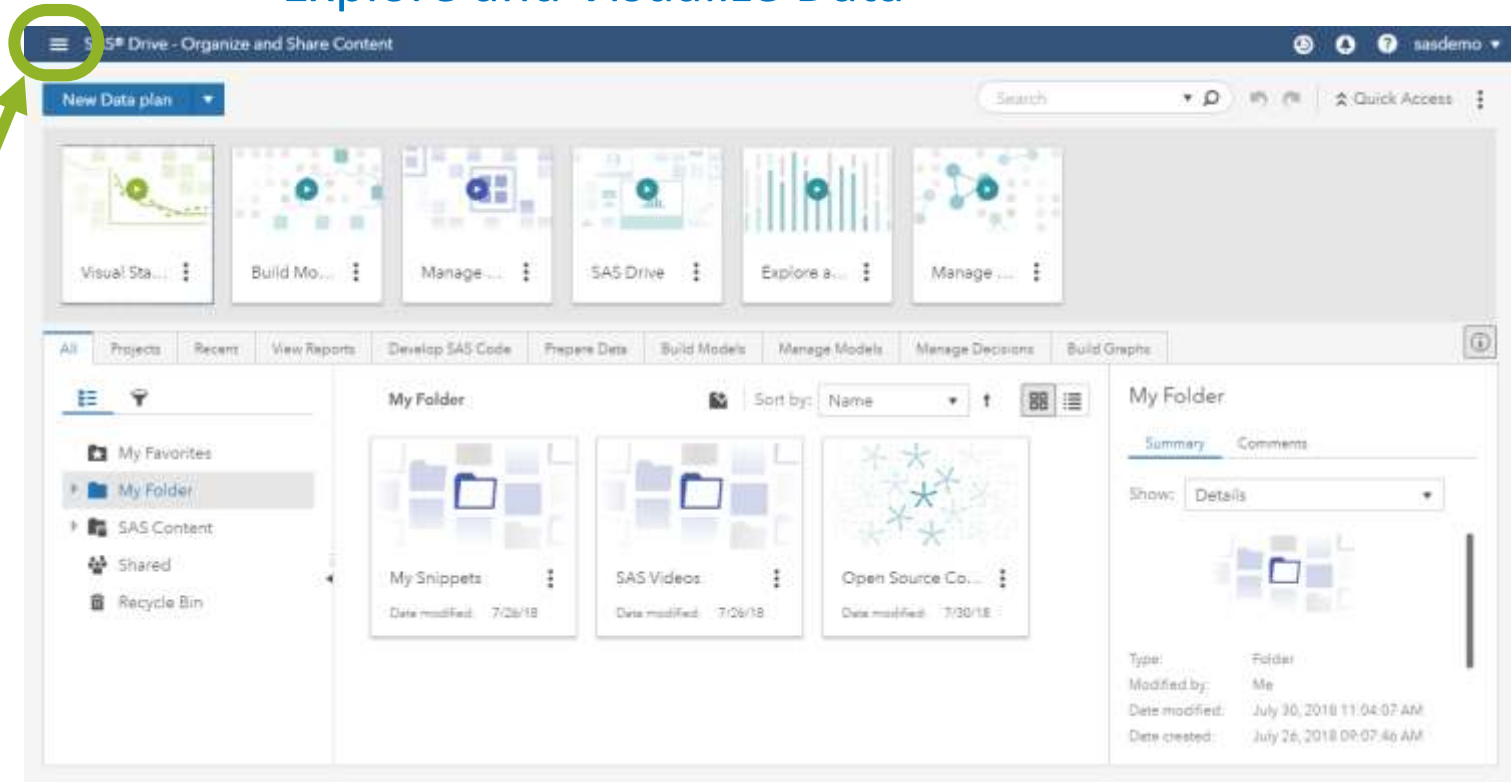
Bayesian Network



SAS® Drive

Explore and Visualize Data

Click on applications menu and select Explore and Visualize Data



SAS® Drive

Explore and Visualize Data

Click on Blue
Box and
Select New
Report

The screenshot displays the SAS Drive web interface. At the top, a dark blue header contains the text 'SAS® Drive - Organize and Share Content' and user information 'sasdemo'. Below the header is a search bar and a 'Quick Access' menu. The main content area features a row of six tiles: 'Visual Sta...', 'Build Mo...', 'Manage...', 'SAS Drive', 'Explore a...', and 'Manage...'. A green circle highlights the 'New Data plan' button in the top left corner, with a green arrow pointing to it from the text 'Click on Blue Box and Select New Report'. Below the tiles is a navigation bar with tabs: 'All', 'Projects', 'Recent', 'View Reports', 'Develop SAS Code', 'Prepare Data', 'Build Models', 'Manage Models', 'Manage Decisions', and 'Build Graphs'. The main workspace shows a 'My Folder' view with three sub-items: 'My Snippets', 'SAS Videos', and 'Open Source Co...'. A right-hand pane displays details for 'My Folder', including 'Summary', 'Comments', 'Show: Details', and metadata such as 'Type: Folder', 'Modified by: Me', 'Date modified: July 30, 2018 11:04:07 AM', and 'Date created: July 26, 2018 09:07:46 AM'.

Classification

Our example today

- The dataset is from a financial institution with customer demographics and loan/credit behavior.
- The goal of this modeling exercise is to **predict which people are likely to default on a home equity loan.**
- The data are at the customer-level (subject-level).
- n=5960
- columns = 13

| # | Variable | Type | Len | Label |
|----|----------|------|-----|--|
| 1 | BAD | Num | 8 | Default or seriously delinquent |
| 10 | CLAGE | Num | 8 | Age of oldest credit line in months |
| 12 | CLNO | Num | 8 | No. of trade credit lines |
| 13 | DEBTINC | Num | 8 | Debt to income ratio |
| 9 | DELINQ | Num | 8 | No. of delinquent credit lines |
| 8 | DEROG | Num | 8 | No. of major derogatory reports |
| 6 | JOB | Char | 7 | Prof/Exec/Office/Self/Other |
| 2 | LOAN | Num | 8 | Amount of current loan request |
| 3 | MORTDUE | Num | 8 | Amount due on existing mortgage |
| 11 | NINQ | Num | 8 | No. of recent credit inquiries |
| 5 | REASON | Char | 7 | Home improvement or Debt Consolidation |
| 4 | VALUE | Num | 8 | Value of current property |
| 7 | YOJ | Num | 8 | Years on current job |



Visual Interface Demo

SAS Visual Analytics



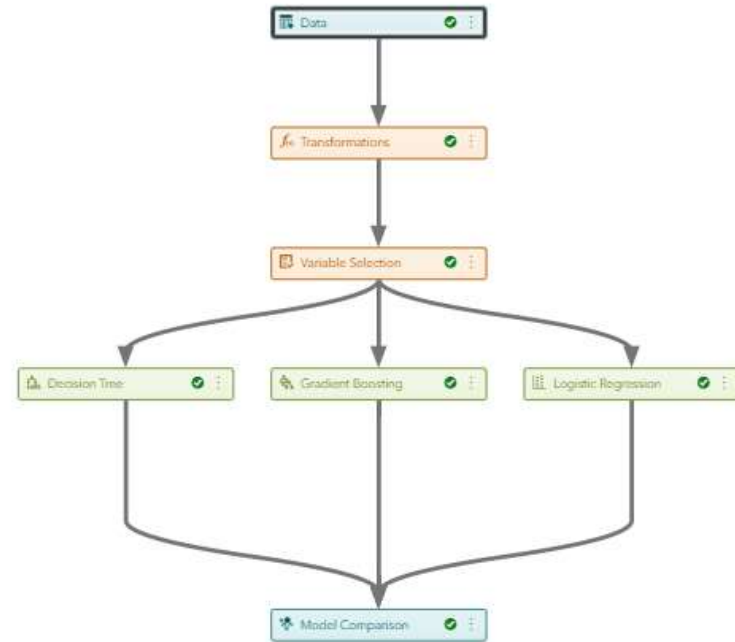
Visual Interface

Pipelines

SAS® Visual Data Mining and Machine Learning 8.3

Pipelines

- Drag-and-drop pipelines including preprocessing and machine learning techniques
- Customizable and portable nodes and SAS best practice pipelines (Toolbox)
- Support for SAS coding (macro, data step, procs, batch Enterprise Miner) within pipelines
- Collaboration through the use of the “Toolbox” – a collection of SAS Best Practice Pipelines, in addition to user-generated templates








[Example Code for Pipeline](#)













SAS[®] Visual Data Mining and Machine Learning 8.3

Pipelines

▼ Data Mining Preprocessing

-  Anomaly Detection
-  Clustering
-  Feature Extraction
-  Filtering
-  Imputation
-  Manage Variables
-  Replacement
-  Text Mining
-  Transformations
-  Variable Clustering
-  Variable Selection

▼ Supervised Learning

-  Batch Code
-  Bayesian Network
-  Decision Tree
-  Forest
-  GLM
-  Gradient Boosting
-  Linear Regression
-  Logistic Regression
-  Neural Network
-  Quantile Regression
-  Score Code Import
-  SVM

▼ Postprocessing

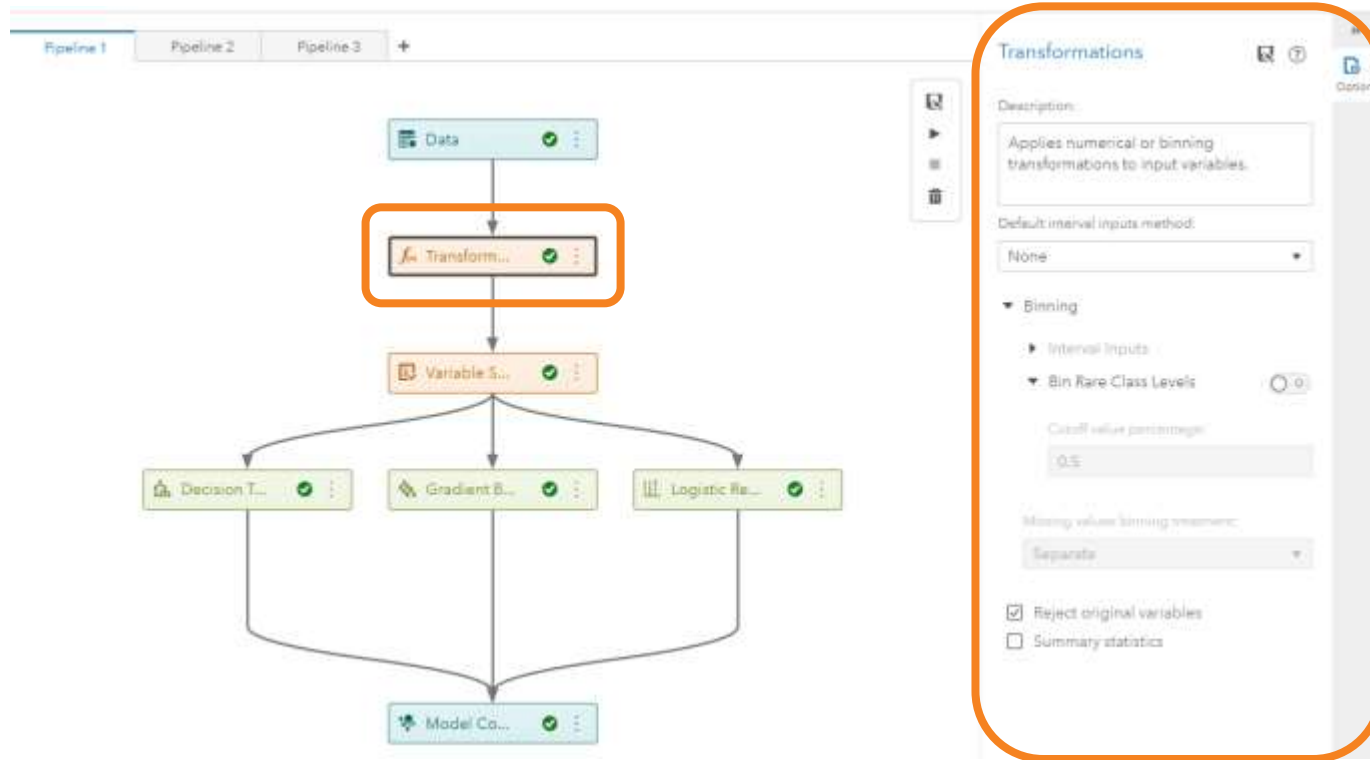
-  Ensemble

▼ Miscellaneous

-  Data Exploration
-  Open Source Code
-  SAS Code
-  Save Data

SAS® Visual Data Mining and Machine Learning 8.3

Pipelines



SAS® Visual Data Mining and Machine Learning 8.3

Pipelines

- Automated API generation for retraining and scoring
- Ability to deploy models in to databases directly
- Assessment against imported Test datasets
- Integration with Model Manager for versioning, tracking and deployment
- Integration with SAS 9.4 Enterprise Miner score code & Batch Code
- Integration with Open Source (R and Python)

The image displays two screenshots of the SAS Visual Data Mining and Machine Learning 8.3 interface. The top screenshot shows a 'Claims' pipeline with a table of models and a context menu for 'Pipeline 2'. The bottom screenshot shows a dashboard with two pipeline cards and a 'Toolbox' section.

| Change | Name | Algorithm Name | Pipeline Name |
|--------------------------|-------------------|-------------------|---------------|
| <input type="checkbox"/> | Example | Example | Pipeline 2 |
| <input type="checkbox"/> | Logit Regression | Logit Regression | Pipeline 1 |
| <input type="checkbox"/> | Gradient Boosting | Gradient Boosting | Pipeline 1 |

The dashboard below shows two pipeline cards:

- Claims**: Date modified: Feb 07, 2018, 10:47:20...; Created by: sasadm; Modified by: sasadm; Description: Main Claims; Model with: Example; Deployment: Example.
- SAS Visual Analytic**: Date modified: Sep 26, 2018, 9:46:49 AM; Created by: sasadm; Modified by: sasadm; Description: SAS Visual Analytic; Model with: Example; Deployment: Forward Logit Regression.

The 'Toolbox' section includes a graph with 'Depth' on the x-axis and a 'View My Tools' button. Below the toolbox are links for SAS Resources, SAS Support, SAS Documentation, SAS Products and Systems, SAS Products, SAS on YouTube, and SAS Community.



Visual Interface Demo

Pipelines



SAS Programming Interface

SAS Studio

SAS® Visual Data Mining and Machine Learning

Openness for a Non-SAS Programmer



SAS Language



Python, R, Lua & Java



REST APIs

SAS® Visual Data Mining and Machine Learning

Analytical Capabilities

Data Manipulation

- In-Memory Data Step
- Frequency / Crosstab
- Data Transpose
- Variable Binning
- Variable Cardinality Analysis
- Variable Summary
- Sampling and Partitioning
- Missing Value Imputation
- Variable Selection
- Model Assessment
- DS2
- FedSQL

Statistics

- Cox Proportional Hazards
- Decision Trees
- Design Matrix
- General Additive Models
- Generalized Linear Models
- K-means and K-modes Clustering
- Linear Regression
- Logistic Regression
- Nonlinear Regression
- Ordinary Least Squares Regression
- Partial Least Squares Regression
- Pearson Correlation
- Principal Component Analysis
- Quantile Regression
- Shewhart Control Chart Analysis

Machine Learning

- Bayesian Networks
- Boolean Rules
- Factorization Machines
- Frequent Item Set Mining
- Gradient Boosting
- K Nearest Neighbor
- Image Processing
- Market Basket Analysis
- Moving Windows PCA
- Network Analytics/Community Detection
- Neural Networks / Deep Learning
- Random Forest
- Robust PCA
- Support Vector Data Description
- Support Vector Machines
- Text Mining
- Variable Clustering



SAS Programming Interface Demo

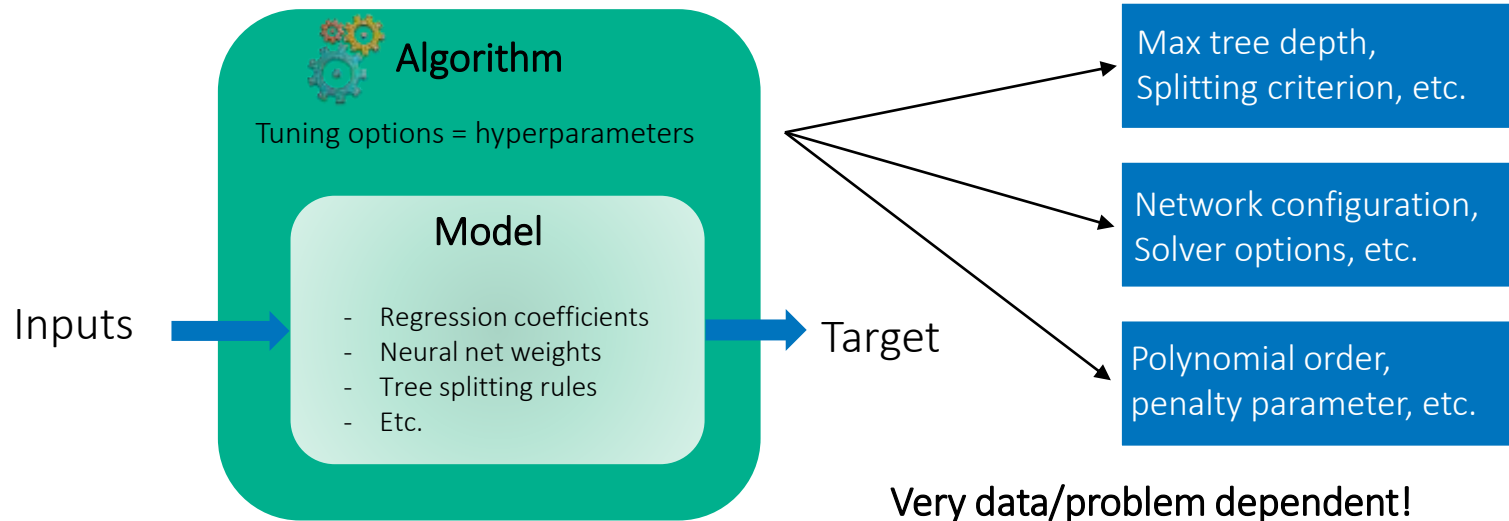
SAS Studio



Autotuning

Automating Autotuning: Hyperparameters

- Training a model involves using an algorithm to determine **model parameters** or other logic to map inputs to a target
- Tuning a model involves determining the **algorithm hyperparameters** (tuning options) that result in the model which maximizes predictability on an independent data set



Autotuning Methods



- SAS Visual Data Mining and Machine Learning offers:
 - Random search (highly parallelizable)
 - Latin Hypercube (highly parallelizable)
 - LH + proprietary SAS/OR algorithm (sequential in nature – *named Genetic in Options*)
- All 5 VDMML procs as well as Decision Trees can be autotuned via SAS Visual interface and SAS Studio interface

Autotune Statement

How SAS proprietary tuning is done

Decision tree: PROC TREESPLIT

- Depth of tree
- Splitting criterion
- Number of bins for interval variables

Forest: PROC FOREST

- Number of trees
- Number of levels in each tree
- Bootstrap sampling rate
- Number of inputs used for splitting a node

Gradient Boosting: PROC GRADBOOST

- Number of iterations (trees)
- Sampling proportion
- LASSO (L1) regularization
- Ridge (L2) regularization
- Number of inputs used for splitting a node
- Learning Rate

Neural Networks: PROC NNET

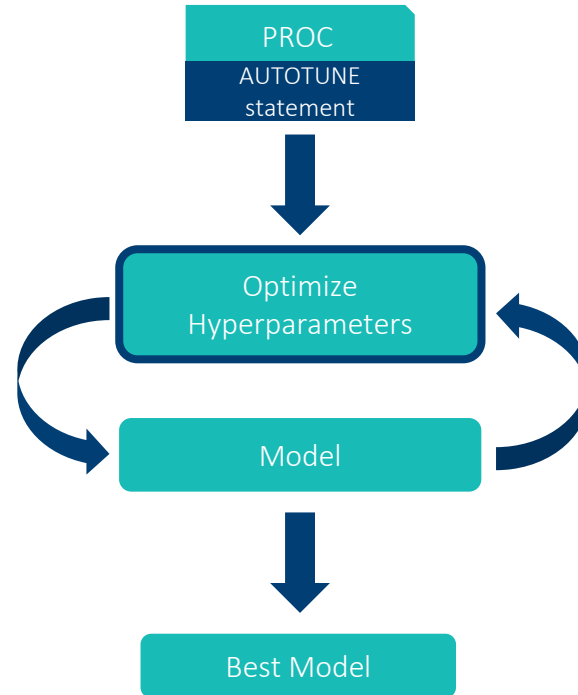
- Number of hidden layers
- Number of neurons in each hidden layer
- L1 regularization
- L2 regularization
- SGD options (annealing rate, learning rate)

Support Vector Machines: PROC SVMACHINE

- Polynomial degree
- Penalty value

Factorization Machine: PROC FACTMAC

- Number of factors
- Step size (learning rate)
- Number of iterations



Uses Standard Grid, Random Search or Latin Hypercube to seed the Genetic algorithm

LOOP until stop criterion (e.g max time, max models, max iterations, population size etc.)



SAS Programming Interface - AutoTuning Demo

SAS Visual Analytics & SAS Studio



Open Source Interface

Jupyter Notebooks

Open Access



SAS language



APIs



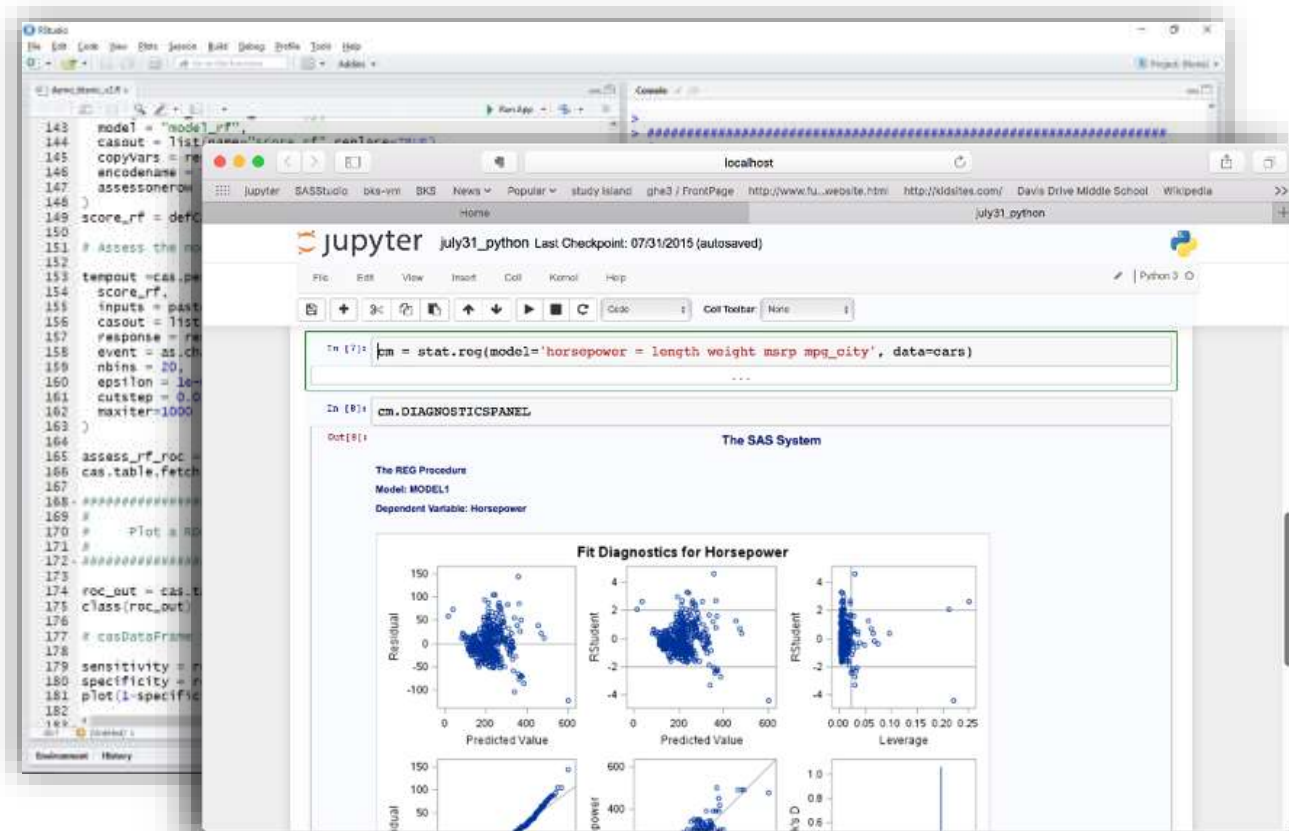
Other programming languages – R, Lua Python and Java



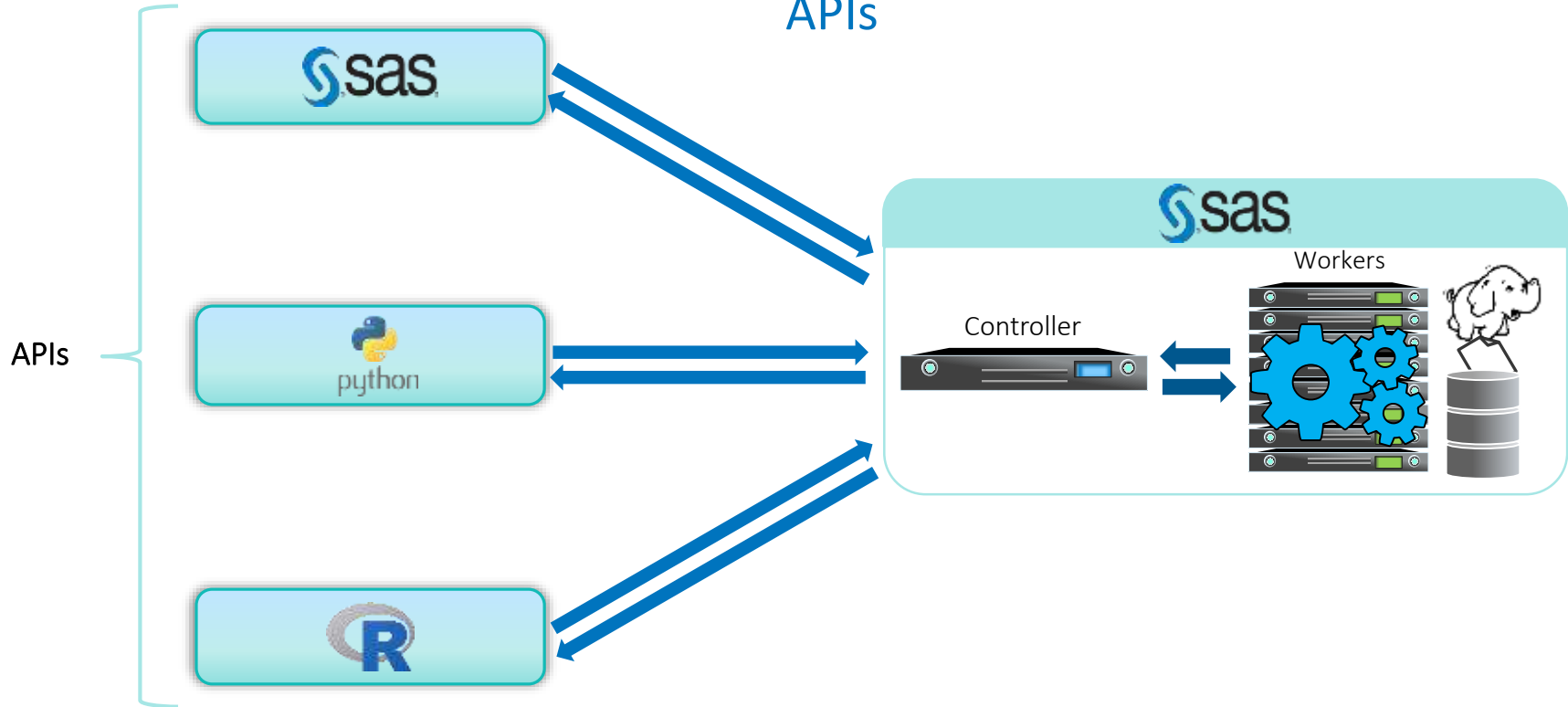
[Developer & user communities](#)

Ask the Expert

[The New Languages of SAS 9.4 and SAS Viya: A SAS Programmer's Primer](#)

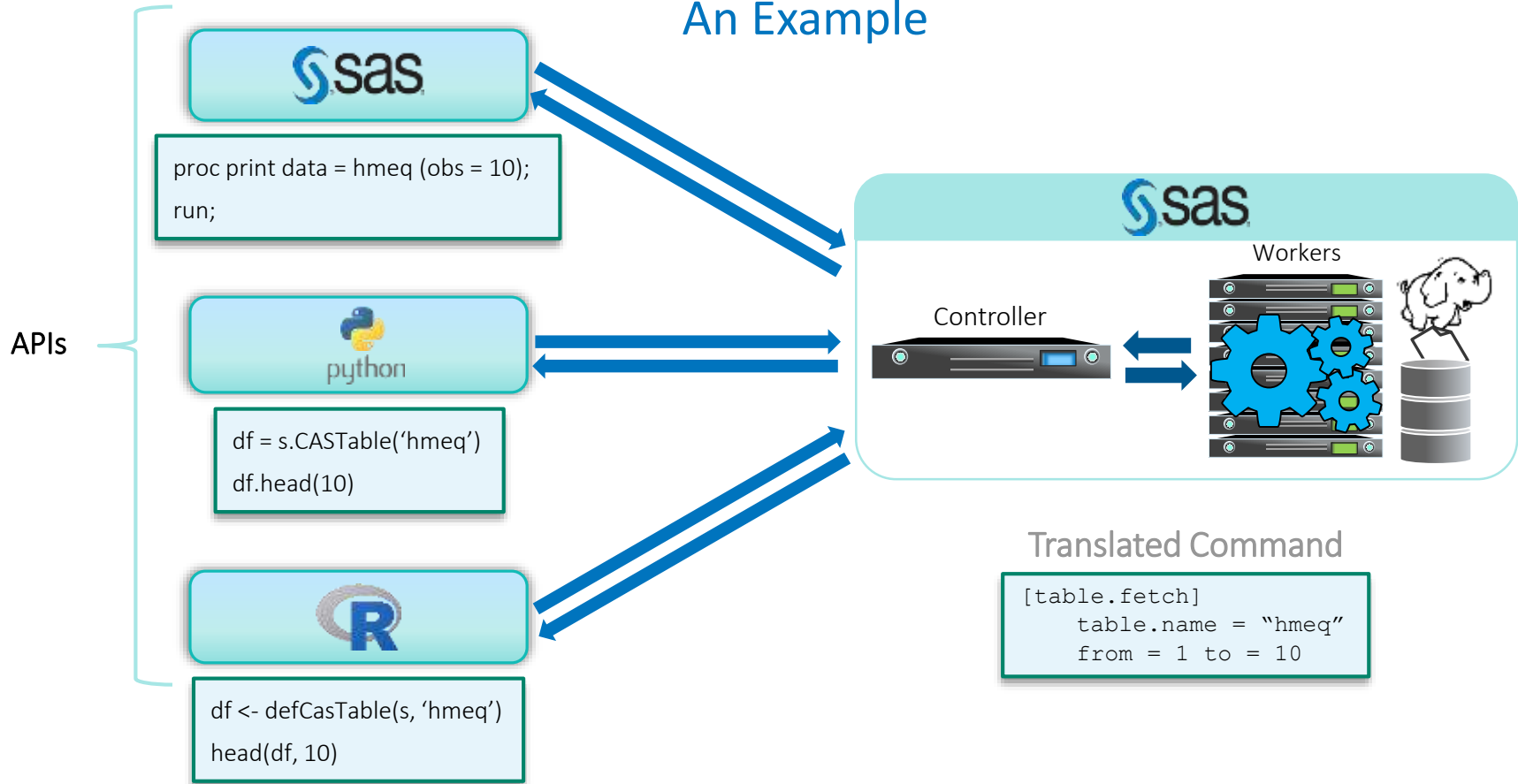


SAS Viya APIs

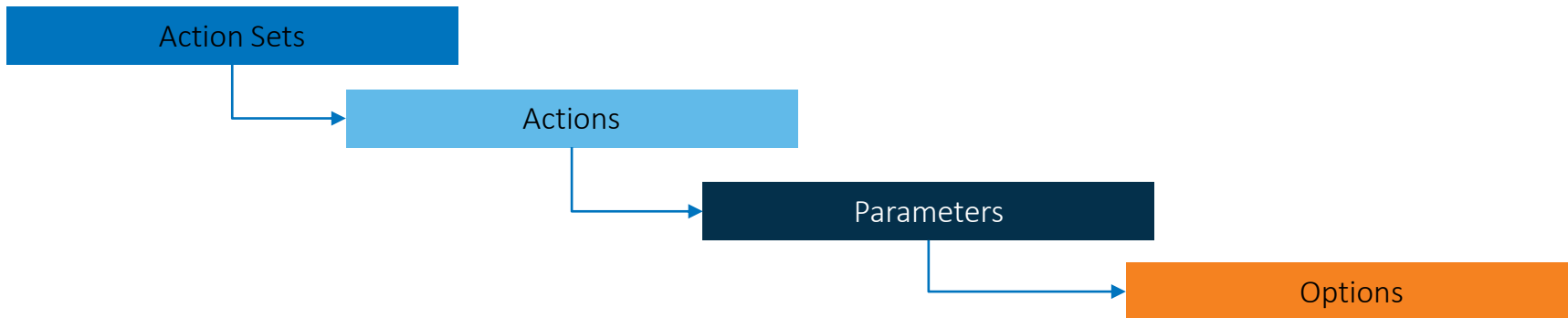


SAS Viya

An Example



CAS Actions Hierarchies



```
table.attribute <result =results> <status=rc> /  
  attributes={{  
    column="string",  
    *key="string",  
    value="string" | 64-bit-integer | integer | double | binary-large-object  
  }, {...}}
```

PROC versus CAS Action

```
proc factmac data=mycas.movlens nfactors=10 learnstep=0.15
            maxiter=20 outmodel=mycas.factors;
  input userid itemid /level=nominal;
  target rating /level=interval;
  output out=mycas.out1 copyvars=(userid itemid rating);
run;
```

```
proc cas;
  action factmac result=R / table={name="movlens"},
  outModel={name="factors_out", replace=true},
  inputs={"userid", "itemid"},
  nominals={"userid", "itemid"},
  target="rating",
  maxIter=20, nFactors=10, learnStep=0.15,
  output={casout={name="score_out", replace="TRUE"},
  copyvars={"userid","itemid","rating"}};
run;
```



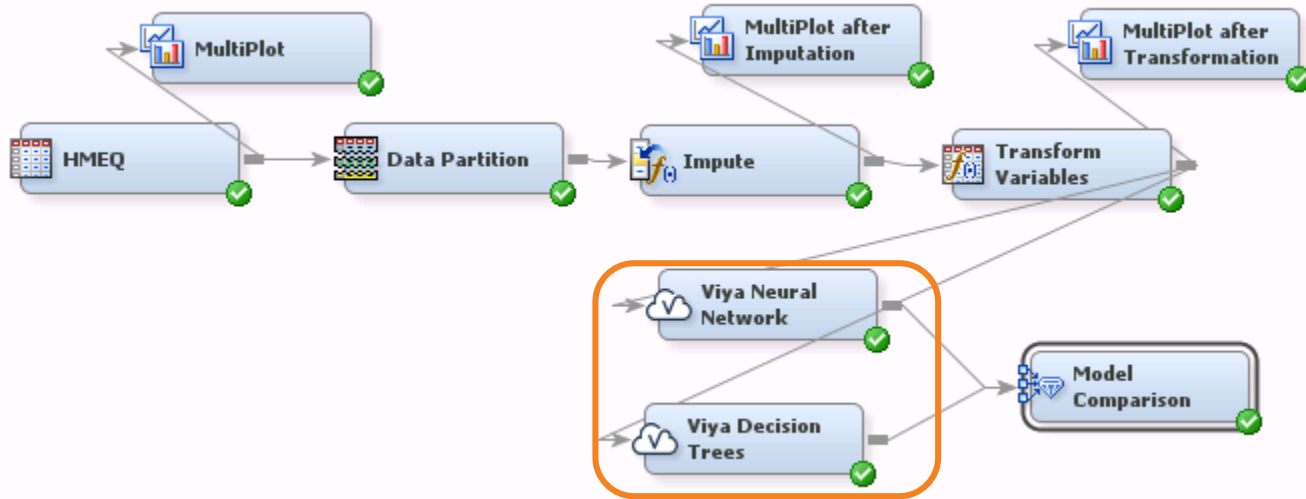
Open Source Interface Demo

Jupyter Notebooks



Additional Capabilities

Integrating Viya and SAS Enterprise Miner Viya Code Node



Other Features in SAS VDMML 8.3

Additional Analytical Algorithms and Options

- Tensor Factorization
- Neural Network Autoencoders
- Clustering mixed variables
- Deep forward neural networks (DNNs), convolutional neural networks (CNNs) and recurrent neural networks (RNNs)
- Bayesian Network
- Market Basket Analysis
- Image Processing (only CAS Actions currently)
 - Load images recursively & at random
 - Retrieve Image labels across all folders when importing
 - Convert image table action (wide format)
 - Support image processing with Deep Learning

What's New in VDMML 8.3

In case you have an earlier version

- SAS Drive replaces SAS Home
- Visual Interface
 - Bayesian Network in Visual Interface
 - Model Interpretability Charts
 - Variable Importance
 - Partial Dependence
 - ICE
 - LIME
 - Create Pipelines or Add Pipelines to existing projects
 - Registering models from Visual Interface
- Model Studio
 - Remembers where you were
 - Added to Pipelines
 - Feature Engineering Template
 - Supervised Learning
 - Batch Code
 - Quantile Regression
 - Score Code Import
 - Miscellaneous – Open Source Code
 - Transformation Node – Best Transformation



Resources

Where can I learn more?

SAS® Visual Data Mining and Machine Learning

[Try it before you buy!](#)

SAS® Visual Data Mining and Machine Learning

[Overview](#)

[Support](#)

[Free Trial](#)

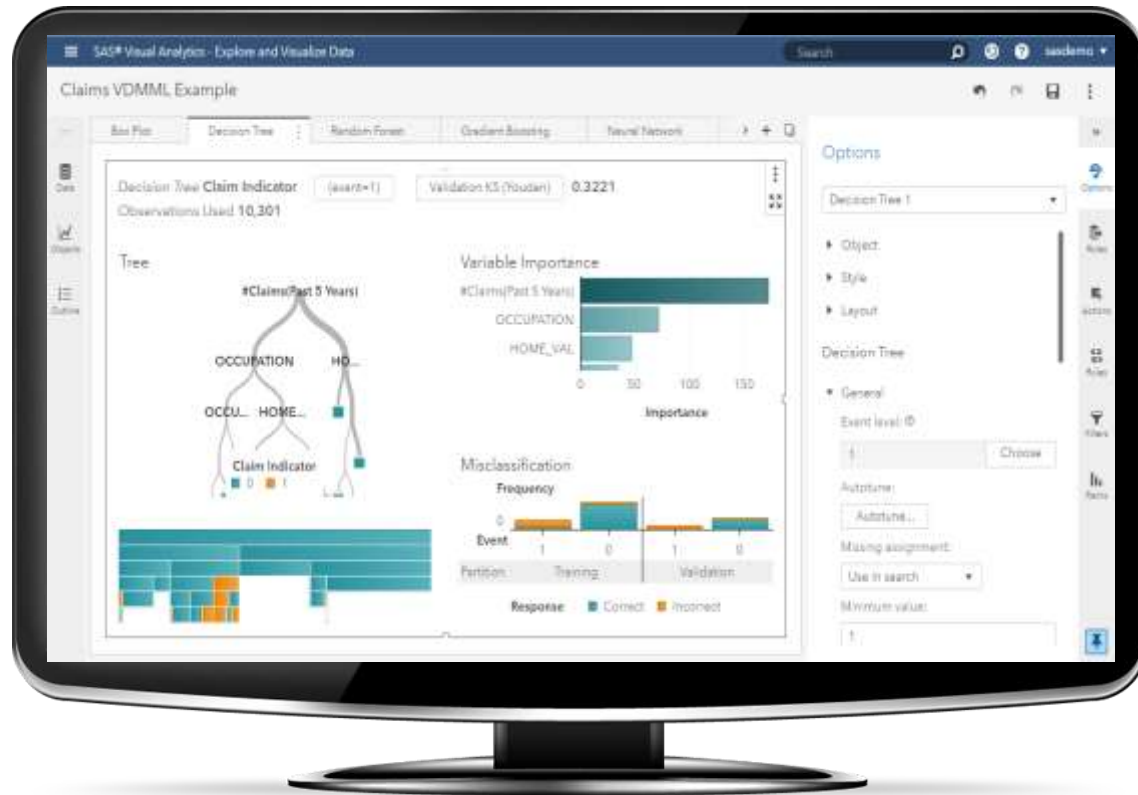
SAS® VISUAL DATA MINING AND MACHINE LEARNING

Everything you need to solve the most complex analytical problems – in a single, integrated, collaborative solution.

Try it for free

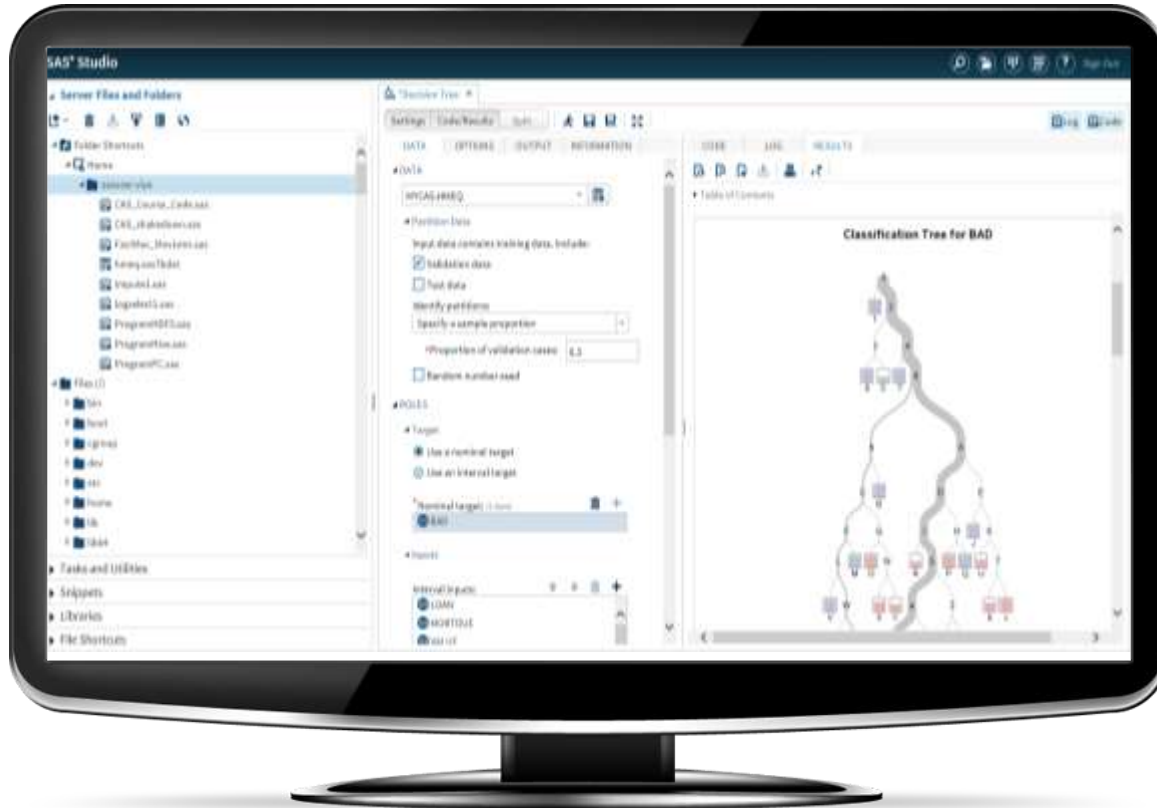


SAS[®] Visual Data Mining and Machine Learning Visual Interface – SAS Visual Analytics



**SAS Visual Data Mining
and Machine Learning on
Visual Interface**
<https://goo.gl/qDNdCS>

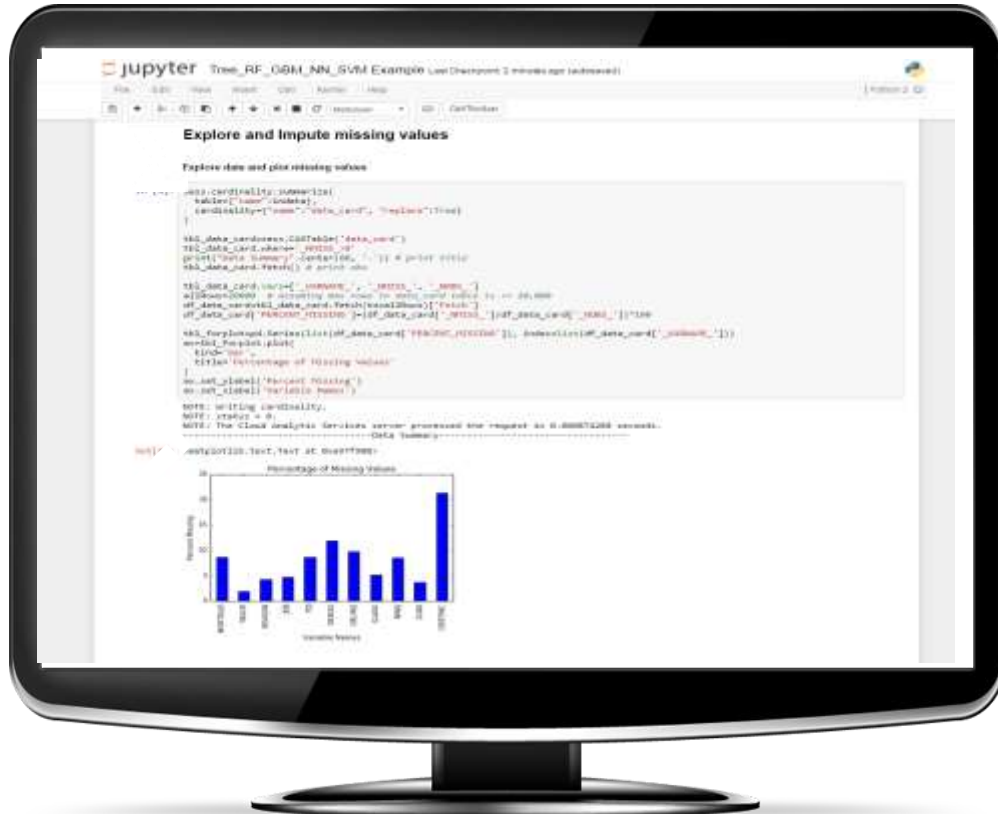
SAS® Visual Data Mining and Machine Learning Programmatic Interface - SAS Studio



**SAS Visual Data Mining
and Machine Learning on
SAS Studio**

https://youtu.be/X0AU4gDUc_Y

SAS® Visual Data Mining and Machine Learning Programming with Open Source



**SAS Visual Data Mining and
Machine Learning with Python
Demo**

<https://youtu.be/LXoikPWQJ3o>

SAS® Visual Data Mining and Machine Learning

Where to learn more?

Video Resources

- [Video Tutorials](#)

SAS Visual Data Mining and Machine Learning

https://youtu.be/X0AU4gDUc_Y

SAS Visual Data Mining and Machine Learning
with Python

<https://youtu.be/LXoikPWQJ3o>

‘How do I’ videos

<http://support.sas.com/training/tutorial/viya/index.html>

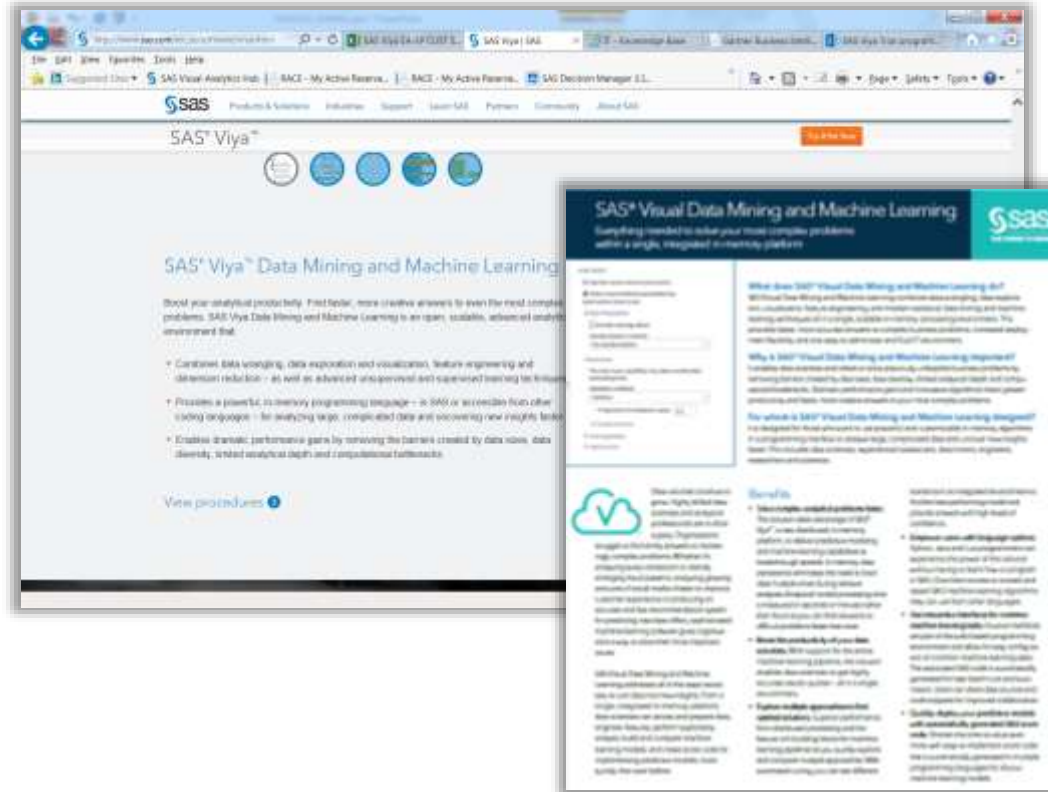
The screenshot shows the SAS Visual Data Mining and Machine Learning tutorial page. On the left is a navigation menu with categories like 'Products & Solutions', 'Analytics in Action', and 'How To Tutorials'. The 'How To Tutorials' section is expanded, showing a list of topics such as 'Administration & Architecture', 'Advanced Analytics', 'SAS Analytics U', 'Customer Intelligence', 'Data Management', 'SAS Enterprise Guide', 'SAS Enterprise Miner', 'Programming', 'Risk Management', 'SAS Office Analytics', 'SAS Studio', 'SAS Visual Analytics', 'SAS Visual Data Mining and Machine Learning', 'SAS Visual Statistics', and 'SAS Viya'. The main content area displays a grid of video thumbnails. The first row includes 'Getting Started with Data Mining and Machine Learning Pipelines on SAS Viya' (1:29), 'Building an Ensemble Model Using a Forest in SAS Visual Data Mining and Machine Learning 8.1 on SAS Viya' (2:31), and 'Building a Factorization Machine on Model Sparse Data in SAS Visual Data Mining and Machine Learning 8.1 on SAS Viya' (6:49). The second row includes 'Getting Started with SAS Visual Data Mining and Machine Learning 8.1 on SAS Viya' (8:11), 'Building a Classifier Model Using Support Vector Machines in SAS Visual Data Mining and Machine Learning 8.1 on SAS Viya' (9:46), and 'Building a Model Comparison in SAS Visual Data Mining and Machine Learning 8.1 on SAS Viya' (7:38). The third row includes 'Building a Neural Network Model in SAS Visual Data Mining and Machine Learning 8.1 on SAS Viya' (13:21) and 'Building a Gradient Boosting Model in SAS Visual Data Mining and Machine Learning 8.1 on SAS Viya' (15:14). The page also features a search bar, a '14 of 9' indicator, and a 'Sort: Most Recent' dropdown.

SAS® Visual Data Mining and Machine Learning

Where to learn more?

Key Resources

- [SAS VDMML Product Web Page](#)
- [Factsheet](#)
- [SAS Viya Brochure](#)
- [Documentation](#)
- [VDMML SAS Community](#)



Communities

The screenshot shows the SAS Data Mining and Machine Learning community page. At the top, there is a navigation bar with links for Products & Solutions, Industries, Support, Learn, Partners, Connect, and About SAS. Below this is a header for the community, including a search bar and a 'Find a Community' button. The main content area features a video titled 'Open source integration with SAS Enterprise Miner' by Tilda Milan, Senior Business Solutions Manager for SAS UK. The video description states: 'You or your colleagues may use other tools, like R or Python, along with SAS to perform production analytics. SAS Enterprise Miner on SAS 9.4 integrates with open source technologies like these so you can bring those results together in one stable, scalable enterprise environment. Tilda Milan, Senior Business Solutions Manager for SAS UK, demonstrates how in this video. Watch the video and check out the related repository on GitHub.' Below the video is a table of messages:

| Subject | Replies | Views | Author | Libs | Views | Latest Post |
|--|---------|-------|------------|------|-------|--|
| What is PROC Arboreman Internal Error | 1 | 2 | EsmondJ_JT | 0 | 22 | 48m ago by PatrickOlivier |
| Scoring data and fit statistics in SAS E Miner | 1 | 2 | gwen_123 | 0 | 125 | 2 hours ago by WendyCrika |

On the right side, there is a 'Resources' section with links to 'Product Information', 'Learn SAS9 Enterprise Miner on SAS 9.4', 'SAS Enterprise Miner on SAS Enterprise Base', 'SAS9.4', 'SAS9.4TSB', and 'SAS9.4TSB Data'. Below that is a 'Videos' section with a link to 'How to Use SAS 9.4 Data Mining & SAS Getting Started with Miner'. At the bottom, there is a 'Technical Papers' section with a link to 'Proceedings from SAS 2017'.

The screenshot shows the GitHub repository page for SAS Software. The page header includes navigation links for Features, Business, Explore, Marketplace, and Pricing, along with a search bar and 'Sign in or Sign up' buttons. The main content area displays the repository name 'sas' and the organization 'SAS Software'. Below this, there are tabs for 'Repositories' and 'People'. The 'Pinned repositories' section lists several repositories:

- sas_kernel**: A Jupyter kernel for SAS. It spins up all the data manipulation and analytics capabilities of your SAS system within a notebook interface. Use the Jupyter Notebook interface to execute SAS code in...
● Jupyter Notebook ★ 71 ▼ 29
- dm-flow**: Library of SAS Enterprise Miner process flow diagrams to help you learn by example about specific data mining topics.
★ 27 ▼ 24
- sas-viya-programming**: Code samples and materials to help you learn to access SAS Viya services by writing programs in Python and other open source languages.
● Jupyter Notebook ★ 31 ▼ 24
- sas-prog-for-r-users**: Teaching and lab material for the "SAS Programming for R Users" course, including course notes, data, and code.
● SAS ★ 20 ▼ 17
- saspy**: A Python interface module to the SAS System. It works with Linux, Windows, and mainframe SAS. It supports the sas_kernel project (a Jupyter notebook kernel for SAS) or can be used on its own.
● Jupyter Notebook ★ 48 ▼ 28
- python-swaf**: The SAS Scripting Wrapper for Analytics Transfer (SWAT) package is the Python client to SAS Cloud Analytic Services (CAS). It allows users to execute CAS actions and process the results all from Py...
● Python ★ 24 ▼ 11

At the bottom, there is a search bar for repositories and filters for 'Type: All' and 'Language: All'.



Questions?

Thank you for your time and attention!

Connect with me:

LinkedIn: <https://www.linkedin.com/in/melodierush>

Twitter: @Melodie_Rush

sas.com