What has changed?

- New High-Performance Bayesian Network node
- HP Variable Selection node adds a new tree-based selection method.
- HP Clustering supports automatic selection of the number of clusters, using the ABC criterion
- HP SVM and HP Forest nodes support a portable format of the model that can be used to score observations within a database.
- Score code supports PMML 4.2.

Why it matters

- Users can test more analytical algorithms to find optimal model
- Model scoring via PMML requires less manual customization
- Score code can be more easily ported to production environment
# Complete List of SAS® Enterprise Miner™ Nodes

<table>
<thead>
<tr>
<th>Sample</th>
<th>Append</th>
<th>Data Partition</th>
<th>File Import</th>
<th>Filter</th>
<th>Merge</th>
<th>Sample</th>
<th>Input Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPLORE</td>
<td>Association Cluster</td>
<td>Graph Explore</td>
<td>Variable Clustering</td>
<td>DMDB MultiPlot</td>
<td>Market Basket StatExplore</td>
<td>Link Analysis Path Analysis</td>
<td>Variable Selection SOM/Kohonen</td>
</tr>
<tr>
<td>MODIFY</td>
<td>Drop</td>
<td>Impute</td>
<td>Interactive Binning</td>
<td>Principal Components</td>
<td>Replacement</td>
<td>Rules Builder</td>
<td>Transform Variables</td>
</tr>
<tr>
<td>MODEL</td>
<td>Decision Tree AutoNeural Regression Neural Network Partial Least Squares Dmine Regression DM Neural Ensemble Rule Induction Gradient Boosting LARS MBR Two Stage Model Import</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incremental Response Survival Analysis Credit Scoring*</td>
<td>TS Correlation</td>
<td>TS Data Prep</td>
<td>TS Dimension Reduction</td>
<td>TS Decomp.</td>
<td>TS Similarity</td>
<td>TS Exponential Smoothing</td>
<td></td>
</tr>
<tr>
<td>HP Explore HP Impute HP BNET Transform HP Variable Selection HP Neural HP Forest HP Decision Tree HP Data Partition HP GLM HP SVM HP Cluster HP Principal Components</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASSESS</td>
<td>Cutoff</td>
<td>Decisions</td>
<td>Model Comparison</td>
<td>Score</td>
<td>Segment Profile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UTILITY</td>
<td>Control Point</td>
<td>End Groups Start Groups</td>
<td>Open Source Integration Reporter Score Code Export Metadata SAS Code Ext Demo Save Data Register Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Requires Credit Scoring for SAS Enterprise Miner Add-on License.
# SAS HIGH-PERFORMANCE PROCEDURES

<table>
<thead>
<tr>
<th>High Performance Statistics</th>
<th>High Performance Data Mining</th>
<th>High Performance Econometrics</th>
<th>High Performance Optimization</th>
<th>High Performance Text Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>HPLOGISTIC</td>
<td>HPREDUCE</td>
<td>HPCOUNTREG</td>
<td>OPTLSO</td>
<td>HPTMINE</td>
</tr>
<tr>
<td>HPREG</td>
<td>HPREDUCE</td>
<td>HPSEVERITY</td>
<td>Select features in</td>
<td>HPTMSCORE</td>
</tr>
<tr>
<td>HPL MIXED</td>
<td>HPREDUCE</td>
<td>HPQLIM</td>
<td>OPTMILP</td>
<td>HPBOOLRULE</td>
</tr>
<tr>
<td>HPSPLIT</td>
<td>HPREDUCE</td>
<td>HP sleeve</td>
<td>OPTLP</td>
<td></td>
</tr>
<tr>
<td>HPGEN SELECT</td>
<td>HPREDUCE</td>
<td>HPPANEL</td>
<td>OPTMODEL</td>
<td></td>
</tr>
<tr>
<td>HPFMM</td>
<td>HPREDUCE</td>
<td>HPCOPULA</td>
<td>OPTGRAPH</td>
<td></td>
</tr>
<tr>
<td>HPCANDISC</td>
<td>HPREDUCE</td>
<td>HPCDM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPRINCOMP</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPLS</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP QUANT SELECT GAMPL</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPFOREST</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPFOREST</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPFOREST</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP FOREST</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPSCORE</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HP DECIDE</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPLUS</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPSVM</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPBNET</td>
<td>HPREDUCE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common Set: HPDS2, HPDMDB, HPSAMPLE, HPSUMMARY, HPIMPUTE, HPBIN, HPCorr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Test multiple algorithms for each segment and then choose the “best” model for each segment.
ORGANIZATIONAL CHALLENGE

TIME IS INCREASINGLY COSTLY

- Producing a new score or adjusting an existing model for the business takes too long with respect to the fast changing market
- Many stakeholders in many steps of the Predictive Analytical process are involved adding more complexity
- Near real-time Predictive Analytics support needed in certain domains like fraud and e-commerce
WHAT ARE WE LOOKING FOR

- Automation
- Scalability
- Integration
- Sophistication

Introducing SAS® Factory Miner !!
1. Automated, white box modeling
2. Sharable best practice process templates
3. Modern machine learning algorithms
4. Industrial scalable analytics
SAS® FACTORY MINER

BOOST MODEL PERFORMANCE AND DATA SCIENTIST PRODUCTIVITY

Segments

Alternatives

Train

Assess

Champion

Test Data

New Data

Score
KEY FEATURES: MODEL TEMPLATES

- Define properties and select from provided templates.
- Modern machine learning algorithms
  - Decision trees, Random forests, Gradient boosting, Neural networks, Bayesian networks,
  - Support vector machines, Regression, Generalized linear models

- Automated data transformation
- Principal component analysis
- Unsupervised and supervised variable selection
KEY FEATURES:
MODELING BY SEGMENT

Check overall project or drill down into individual segments.
Do more models guarantee greater accuracy?  **NO!**

- But more models let us:
  - Determine the correct level for the analysis
  - Go to levels of granularity not possible before
  - Try different combinations of techniques
  - Apply automation to the entire modeling process
<table>
<thead>
<tr>
<th></th>
<th>Manual Approaches</th>
<th>Automated approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Productivity</td>
<td>1-10 models per week</td>
<td>100s to 100s models per week allow testing of more what-if scenarios and higher confidence in answers</td>
</tr>
<tr>
<td>Modeling at scale</td>
<td>Need to sample</td>
<td>Use all records and all features to identify optimal driving factors for models</td>
</tr>
<tr>
<td>Accurate</td>
<td>Small number of iterations limit confidence in accuracy</td>
<td>Scalable testing capabilities allow for highest accuracy in acceptable time frame</td>
</tr>
<tr>
<td>Self-Service</td>
<td>Manual steps through process required deep analytical skills</td>
<td>Guided self-service approach extends analytical talent pool</td>
</tr>
<tr>
<td>Collaboration</td>
<td>Analyst often working in silos</td>
<td>Analysts sharing insights, process and best practices</td>
</tr>
<tr>
<td>Easy access</td>
<td>Install and maintain software across large number of desktops</td>
<td>Easy access through web based interface</td>
</tr>
</tbody>
</table>
• Key goals
  • Identify customer with high likelihood to respond or attrite
  • Segmented models across plans, channels, and campaigns
  • Automated machine learning, model selection and deployment

• Solutions
  • Development Partner: SAS Factory Miner
  • Currently Using SAS Enterprise Miner and SAS Rapid Predictive Modeler

• Key Benefits and Results (ongoing)
  • Increased Model Lift vs Control (Significant Lift in top demi-decile)
  • Collaborative Modeling Methods and Best Practices
  • Faster deployment (automated selection within minutes across all segments)

“SAS Factory Miner will allow us to quickly automate our modeling approach across our customer segments”.

“You can experiment with multiple machine learning methods while deploying these models quickly to production”.

“This solution offers collaboration across your department, enabling best practice modeling at scale”.

-Tamer Cagatay, Business Analytics Lead at Turkcell
SAS® FACTORY MINER™

KEY BENEFITS

SPEED
Get answers fast through automated model development using customizable templates

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

EFFICIENCY
Boost productivity of your data scientists through automation and collaboration

SCALABLE
Scale environment with your growing needs
The ‘IT’ folks

I just built 850 new models. When can you put them into production?

The ‘Analytics’ folks

!!!????!!!
SAS DECISION MANAGER
WHERE ANALYTICS MEETS THE REAL WORLD
BRIDGING THE GAP

AUTOMATED MODELS & BUSINESS RULES…TOGETHER

Data Prep
• Analytical data preparation
• Batch Deployment

Model Building
• Segmented Models
• Fullscale process flow
• Open Source Models

Model Deployment
• Model Monitoring
• In-Database deployment
• Data driven Rules

Deployment Targets
- Batch
- In Database
- **Hadoop
- ***In Stream

DB, Hadoop

Factory Miner
Ent Miner
Open Source
Rules

Data Integration

Decision Manager

Decision Service

** available Q2 2015
*** available late 2015
SINGLE DECISION MANAGEMENT PLATFORM: BACK-END

- Business Rule Manager
- Data Sources
- Model Management
DEPLOYMENT OPTIONS: FRONT-END

Website

External Application

Mobile application
CLOSING THE LOOP: MONITOR AND REPORT
CREDIT APPLICATION CUSTOMER CASE

CREDIT PROCESS

Origination
- Risk rating
- Risk assessment & collateral evaluation
- Application scoring

Servicing
- Annual rating review
- Credit limit setting
- Ongoing risk assessment
- Behavioural scoring

Collections
- Alerts for non-performing loans
- Collections strategy

BUSINESS BENEFITS

Product Management
- Improve margins
- Speed to market for new products
- Better customer experience
- Reduced attrition

Risk Management
- Fewer overrides
- Single platform (integrate with multiple bank systems)
- Consistent credit decisioning
SO WHAT ARE YOU GOING TO SEE TODAY?
Bringing the analytical lifecycle to life: who is involved??

Data Scientist Superhero

Analytics Collaboration

Business Analyst (Citizen Data Scientist)
- Builds model workflow templates
- Adapts model workflow templates
- Uses model workflow templates

Data Miner
- Adapts model workflow templates

Data Scientist
- Builds model workflow templates
WHAT DO YOU NEED TO MODEL?

HOW HAPPY ARE YOUR CUSTOMERS?

Can you easily identify those unhappy customers before you interact with them?

Can you do something about it?

Are your promoter scores right where you want them?

How can I assess the business impact?
PROMOTER SCORE ANALYTICS

FIND THE DRIVERS AND ACT UPON THEM

DETRACTOR

PASSIVE

PROMOTER

OFFERS

Copyright © 2014, SAS Institute Inc. All rights reserved.
SCENARIO  

CUSTOMER CALL CENTER SCENARIO....

5 different product lines
Handled by 5 different departments
Customers are categorized into four levels

Can we create a model for each account level for each department?
**SCENARIO**

**WHO IS THE DETRACTO???

Current State:
5 models
One for each Department

Cost Per Model:
$50K in house OR $150K outsourced
2-3 months to develop
SCENARIO

REACT TO THE DETRACTORS!!! MAKE SURE YOUR CALL CENTER OPERATORS KNOWS WHAT TO DO...

Customer reaches the call center...

Is this a detractor?

Models and expert based rules are triggered in real time

<table>
<thead>
<tr>
<th>Customer Name</th>
<th>Cristina Conti</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment</td>
<td>Gold</td>
</tr>
<tr>
<td>Products Owned</td>
<td>One</td>
</tr>
<tr>
<td>Location</td>
<td>Northern Italy</td>
</tr>
<tr>
<td>Calling from</td>
<td>Home</td>
</tr>
<tr>
<td>Recent contacts</td>
<td>Five</td>
</tr>
<tr>
<td>Detractor likelihood</td>
<td>3%</td>
</tr>
</tbody>
</table>

Operator knows how he should treat the customer!