HOW TO TURBOCHARGE YOUR MODELING, GO FROM ZERO TO HERO

DR MARK CHIA
PRACTICE LEAD, ADVANCED ANALYTICS
• Introduction
• Factory Miner
• Summary
• Q & A
ARE YOUR DECISIONS KEEPING UP WITH YOUR DATA GROWTH?
EXTRACTING BUSINESS VALUE FROM THE DATA

Decisions with Confidence

- Reduce Uncertainty
- More Accurate
- More Timely
- Optimize
DECISIONS AT SCALE

- More Models
- More Granular Segments
- More Predictive Machine Learning Algorithms
- Model Tournaments
- Enable Non-Technical Users

- Integration
- Embedded Analytics
- Move Insights Closer to the Decision Maker
- Combine Models with Business Rules

- Using ALL the Relevant Data
- Creating More Attributes
CREATING EVEN MORE GRANULAR MODELS

Creating a **hierarchy** of models, choosing the “best” model at each level

DATA

- Summed customer transactions
- Customer type
- Community type
- Distance
- Spend frequency

Increasing refinement / granularity
Creating Even More Granular Models

All 17 potential segments are modeled and “the best" algorithm and prediction level is chosen for each branch.

New

Fico >= 620
Fico < 620

Stable

Metro

Lives <= 5 mi. retail store

Lives > 5 mi. retail store

Rural

@Risk

Spent >= $300 last 12 mo.
Spent >= $100 & < $300 last 12 mo.
Spent < $100 last 12 mo.
Spent > $200 last 12 mo.
Spent <= $200 last 12 mo.
Spent >= $50 last 6 mo.
Spent < $50 last 6 mo.

Now, what if we want to do this analysis by state?
850 Potential Models to Evaluate

Introducing SAS® Factory Miner!!
SAS® FACTORY MINER

1. Automated, white box modeling
2. Sharable best practice process templates
3. Modern machine learning algorithms
4. Scalable Analytics
- Run model tournaments on segments in your data
- Define as many model process flows as required
- Test ideas with statistical and machine learning algorithms in unified environment
- Identify champion model by segment using customizable assessment statistics
- Identify and fine-tune underperforming models fast
- Edit any model pipeline in easy to use interface
• Sharable projects
• Global data source metadata
• Model templates with building blocks
  • Include variable engineering and reduction
• Customizable and shareable
• Generate best-practice building blocks for organizational model process
MODERN MACHINE LEARNING

- 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
• Train models using multithreaded procedures on SAS servers
• Train models using asynchronous processes via SAS Grid Manager for workload balancing and scheduling
• Train models in memory using SAS High-Performance Data Mining on database appliance or on Hadoop
• Same user experience from single machine to SAS grids and In-memory computing
## MANUAL VERSUS AUTOMATED MACHINE LEARNING

<table>
<thead>
<tr>
<th></th>
<th>Manual Approaches</th>
<th>Automated approach</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Productivity</strong></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><strong>Modeling at scale</strong></td>
<td>Need to sample</td>
<td>Use all records and all features to identify optimal driving factors for models</td>
</tr>
<tr>
<td><strong>Accurate</strong></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><strong>Self-Service</strong></td>
<td>Manual steps through process required deep analytical skills</td>
<td>Guided self-service approach extends analytical talent pool</td>
</tr>
<tr>
<td><strong>Collaboration</strong></td>
<td>Analyst often working in silos</td>
<td>Analysts sharing insights, process and best practices</td>
</tr>
<tr>
<td><strong>Easy access</strong></td>
<td>Install and maintain software across large number of desktops</td>
<td>Easy access through web based interface</td>
</tr>
</tbody>
</table>
EXTEND ORGANIZATIONAL TALENT POOL

Data Scientist Superhero

Analytics Collaboration

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

Data Miner
- Adapts model workflow templates

Data Scientist
- Builds model workflow templates
NOW YOU HAVE LOTS OF MODELS. WHAT NOW?

The ‘IT’ folks

!!!????!!!

The ‘Analytics’ folks

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