Open Source and SAS
Extending SAS and R models to Model Manager
Agenda

Open source and SAS

• The Analytics Lifecycle and Open source
• How SAS embraces/extends Open source
• Demo example (Model Manager)
• Open Source Support Summary and Questions
The Analytic Lifecycle

Discovery & Development of Analytics

- Lots of Data
- New Data
- Experimentation
- Fail Fast
- Test & Learn
- Interactive
- Iterative
- Innovation
- Flexibility
- Data Science

Discovery & Development of Analytics

- EXPLORE
- PREPARE
- MODEL

Deployment & Execution of Analytics

- INVENTORY
- EXECUTE
- MONITOR
- ASK

Lots of Data
New Data
Experimentation
Fail Fast
Test & Learn
Interactive
Iterative
Innovation
Flexibility
Data Science

Regulated Automated
Governed
Embed
Reliable
Decisions
Consistent
Documented
Actions
IT

Data

Copyright © 2016, SAS Institute Inc. All rights reserved.
USE SAS to integrate R

Why?
• Model Comparison
• Leverage R for new algorithms
• Ensemble Modelling
• Generate Score Code
• Deploy R models
Productivity

SAS Models (4)
- Neural Network
- Regression LRM
- Decision Tree
- Regression Forest

Gradient Boost

Open Source (2)
- SVM
- Random Forest

What happens if this replicates to many analytical projects?

Compare 7 models
Choose champion
Inventory Model
Generate score code
Deploy in database/Hadoop
You have many models, ...now what?

The ‘Analytics’ folks

I have 200 models that are deployment-ready. Can can you help me manage them?

The ‘IT’ folks

!!!???!!!

How do you deploy models?
THE ANALYTIC LIFECYCLE

**What customers want?**
- Model Management platform
- Inventory ALL models
- Know who’s working on what
- Ability to deploy models
- Auditable models

**How SAS Extends...**
- Central model management platform
- Repository for SAS Models and Open Source (R, Python, PMML)
- Model History
- Version control
- Model and data lineage
- Model Governance
Model Inventory

Model inventory and search

SAS and Open Source models

Model Metadata

Model lineage
THE ANALYTIC LIFECYCLE

Discovery & Development of Analytics

- PREPARE DATA
- EXPLORE
- MODEL
- INVENTORY
- EXECUTE
- MONITOR

What customers want?
- Deployable analytics
- Automation
- Faster time to model execution
- In Hadoop/Database model execution

How SAS Extends...
- Model execution platform
- Execute models as database functions (UDP or via Embedded Process)
- No Language Conversion
- Purpose built model execution engines
Model execution

- Model Publishing and automation
- In Hadoop/database deployment
THE ANALYTIC LIFECYCLE

Discovery & Development of Analytics

PREPARE DATA
EXPLORATION
MODEL
INVENTORY
EXECUTE

What customers want?
• Best possible models
• Model tournaments
• Visibility into performance
• Easy retraining
• Champion/challenger modelling

Deployment & Execution of Analytics

How SAS Extends...
• Model performance platform to keep models “fresh”
• Compare Multiple Models at once
• Assess model accuracy (Lift, ROC, K-S)
• Champion/Challenger modeling
• Model Retraining including open source

Copyright © 2016, SAS Institute Inc. All rights reserved.
Model performance

- Retrain models
- Model performance reports
- Monitor data drift
- Model comparisons
Demonstration

Model Manager
## Support for SAS & Open Source models

- **Supported**
- **Supported via customization**
- **Not supported**

<table>
<thead>
<tr>
<th></th>
<th>Supported PMML</th>
<th>Non-PMML</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inventory</strong></td>
<td><img src="https://example.com/supported" alt="Supported" /></td>
<td><img src="https://example.com/not_supported" alt="Not supported" /></td>
</tr>
<tr>
<td><strong>Publish &amp; Score</strong></td>
<td><img src="https://example.com/supported" alt="Supported" /></td>
<td><img src="https://example.com/not_supported" alt="Not supported" /></td>
</tr>
<tr>
<td>Batch</td>
<td><img src="https://example.com/supported" alt="Supported" /></td>
<td><img src="https://example.com/not_supported" alt="Not supported" /></td>
</tr>
<tr>
<td>In-Database</td>
<td><img src="https://example.com/supported" alt="Supported" /></td>
<td><img src="https://example.com/not_supported" alt="Not supported" /></td>
</tr>
<tr>
<td>Web Service</td>
<td><img src="https://example.com/supported" alt="Supported" /></td>
<td><img src="https://example.com/not_supported" alt="Not supported" /></td>
</tr>
<tr>
<td>Streaming</td>
<td><img src="https://example.com/supported" alt="Supported" /></td>
<td><img src="https://example.com/not_supported" alt="Not supported" /></td>
</tr>
<tr>
<td><strong>Monitor</strong></td>
<td><img src="https://example.com/supported" alt="Supported" /></td>
<td><img src="https://example.com/not_supported" alt="Not supported" /></td>
</tr>
<tr>
<td><strong>Retrain</strong></td>
<td><img src="https://example.com/supported" alt="Supported" /></td>
<td><img src="https://example.com/not_supported" alt="Not supported" /></td>
</tr>
</tbody>
</table>
Thank you
pat.valente@sas.com
Appendix

Screen Shots
<table>
<thead>
<tr>
<th>Name</th>
<th>Role</th>
<th>Version</th>
<th>Description</th>
<th>Model Function</th>
<th>Date Published</th>
<th>Date Modified</th>
<th>Created By</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensemble R Combo_20160811TD</td>
<td></td>
<td>2.0</td>
<td></td>
<td>Classification</td>
<td>Aug 11, 2016 09:23 AM</td>
<td></td>
<td>sasdemo</td>
</tr>
<tr>
<td>Ensemble SAS Combo</td>
<td>Champion</td>
<td>2.0</td>
<td></td>
<td>Classification</td>
<td>Nov 28, 2016 12:34 PM</td>
<td></td>
<td>sasdemo</td>
</tr>
<tr>
<td>Ensemble SAS and R</td>
<td>Challenger</td>
<td>2.0</td>
<td></td>
<td>Classification</td>
<td>Nov 28, 2016 12:24 PM</td>
<td></td>
<td>sasdemo</td>
</tr>
</tbody>
</table>
### Project 2

<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>from the SAS Metadata Repository</td>
<td></td>
<td></td>
</tr>
<tr>
<td>from a SAS package file</td>
<td></td>
<td></td>
</tr>
<tr>
<td>from a PMML file</td>
<td></td>
<td></td>
</tr>
<tr>
<td>from local files</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Name</td>
<td>Role</td>
<td>Version</td>
</tr>
<tr>
<td>----------</td>
<td>-------</td>
<td>---------</td>
</tr>
<tr>
<td>EMRPML</td>
<td>1.0</td>
<td></td>
</tr>
</tbody>
</table>

**Variables**

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Measurement</th>
<th>Length</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEPALLEN</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEMALE</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEPALWIDTH</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEMALE</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Properties**

- General
- System
- User-Defined

**Scoring**

- Status
- Advanced
- Score Code
- Model Files

**History**

- Log
- Published
Lift Chart

Lift Table

<table>
<thead>
<tr>
<th>#</th>
<th>model</th>
<th>Event Count</th>
<th>Count</th>
<th>Percent</th>
<th>Cumulative Event Count</th>
<th>Cumulative Count</th>
<th>Lift</th>
<th>Cumulative Lift</th>
<th>Percent Responses</th>
<th>Cumulative Percent Responses</th>
<th>Captured Responses</th>
<th>Cumulative Captured Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Baseline R Combo_2</td>
<td>0.0000</td>
<td>0</td>
<td>0</td>
<td>0.0000</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>Baseline R Combo_2</td>
<td>23.0000</td>
<td>30</td>
<td>0.5</td>
<td>23.0000</td>
<td>30</td>
<td>2.3750</td>
<td>3.3750</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1610</td>
<td>0.1610</td>
</tr>
<tr>
<td>3</td>
<td>Baseline R Combo_2</td>
<td>21.0000</td>
<td>30</td>
<td>1</td>
<td>21.0000</td>
<td>59</td>
<td>2.5250</td>
<td>3.9000</td>
<td>0.7000</td>
<td>0.7000</td>
<td>0.3200</td>
<td>0.3200</td>
</tr>
<tr>
<td>4</td>
<td>Baseline R Combo_2</td>
<td>16.0000</td>
<td>30</td>
<td>1.5</td>
<td>16.0000</td>
<td>89</td>
<td>3.8750</td>
<td>7.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.4888</td>
<td>0.4888</td>
</tr>
<tr>
<td>5</td>
<td>Baseline R Combo_2</td>
<td>14.0000</td>
<td>30</td>
<td>2</td>
<td>14.0000</td>
<td>129</td>
<td>5.0000</td>
<td>12.7750</td>
<td>3.0000</td>
<td>0.9000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>6</td>
<td>Baseline R Combo_2</td>
<td>10.0000</td>
<td>30</td>
<td>5</td>
<td>10.0000</td>
<td>158</td>
<td>8.0000</td>
<td>17.7750</td>
<td>5.0000</td>
<td>1.0000</td>
<td>0.7875</td>
<td>0.7875</td>
</tr>
<tr>
<td>7</td>
<td>Baseline R Combo_2</td>
<td>12.0000</td>
<td>30</td>
<td>20</td>
<td>12.0000</td>
<td>218</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>8</td>
<td>Baseline R Combo_2</td>
<td>0.0000</td>
<td>30</td>
<td>0.5</td>
<td>0.0000</td>
<td>218</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>9</td>
<td>Baseline R Combo_2</td>
<td>14.0000</td>
<td>30</td>
<td>40</td>
<td>14.0000</td>
<td>187</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>10</td>
<td>Baseline R Combo_2</td>
<td>14.0000</td>
<td>30</td>
<td>45</td>
<td>14.0000</td>
<td>187</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>11</td>
<td>Baseline R Combo_2</td>
<td>16.0000</td>
<td>30</td>
<td>50</td>
<td>16.0000</td>
<td>218</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>12</td>
<td>Baseline R Combo_2</td>
<td>14.0000</td>
<td>30</td>
<td>55</td>
<td>14.0000</td>
<td>218</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>13</td>
<td>Baseline R Combo_2</td>
<td>10.0000</td>
<td>30</td>
<td>60</td>
<td>10.0000</td>
<td>218</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>14</td>
<td>Baseline R Combo_2</td>
<td>12.0000</td>
<td>30</td>
<td>65</td>
<td>12.0000</td>
<td>218</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>15</td>
<td>Baseline R Combo_2</td>
<td>4.0000</td>
<td>30</td>
<td>70</td>
<td>4.0000</td>
<td>218</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>16</td>
<td>Baseline R Combo_2</td>
<td>0.0000</td>
<td>30</td>
<td>75</td>
<td>0.0000</td>
<td>218</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>17</td>
<td>Baseline R Combo_2</td>
<td>2.0000</td>
<td>30</td>
<td>80</td>
<td>2.0000</td>
<td>218</td>
<td>10.0000</td>
<td>27.7750</td>
<td>1.5533</td>
<td>0.8000</td>
<td>0.5825</td>
<td>0.5825</td>
</tr>
<tr>
<td>Name</td>
<td>Model Function</td>
<td>Location</td>
<td>Algorithm</td>
<td>Date Modified</td>
<td>Created By</td>
<td>Keywords</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------</td>
<td>---------------------------------</td>
<td>-----------</td>
<td>------------------</td>
<td>------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensemble SAS and R</td>
<td>Classification</td>
<td>SAS_R_Integration/SAS...</td>
<td>Ensemble</td>
<td>Aug 2, 2016 04:47 AM</td>
<td>sandemon</td>
<td>sandemon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensemble SAS Combo</td>
<td>Classification</td>
<td>SAS_R_Integration/SAS...</td>
<td>Ensemble</td>
<td>Aug 2, 2016 04:47 AM</td>
<td>sandemon</td>
<td>sandemon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensemble SAS and R</td>
<td>Classification</td>
<td>SAS_R_Integration/SAS...</td>
<td>Ensemble</td>
<td>Nov 29, 2016 12:24 PM</td>
<td>sandemon</td>
<td>sandemon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ENHMM</td>
<td>Classification</td>
<td>R_Models_PMLMixProject...</td>
<td>rpart</td>
<td>Dec 3, 2016 12:16 AM</td>
<td>sandemon</td>
<td>sandemon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensemble R Combo</td>
<td>Classification</td>
<td>SAS_R_Integration/SAS...</td>
<td>Ensemble</td>
<td>Aug 2, 2016 04:47 AM</td>
<td>sandemon</td>
<td>sandemon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensemble R Combo 20160811T</td>
<td>Classification</td>
<td>SAS_R_Integration/SAS...</td>
<td>Ensemble</td>
<td>Aug 11, 2016 08:23 AM</td>
<td>sandemon</td>
<td>sandemon</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>