

SAS AND OPEN SOURCE

MATT MALCZEWSKI, SAS CANADA



Your Trial, Your Data

Visual Analytics – [Register for Trial](#)

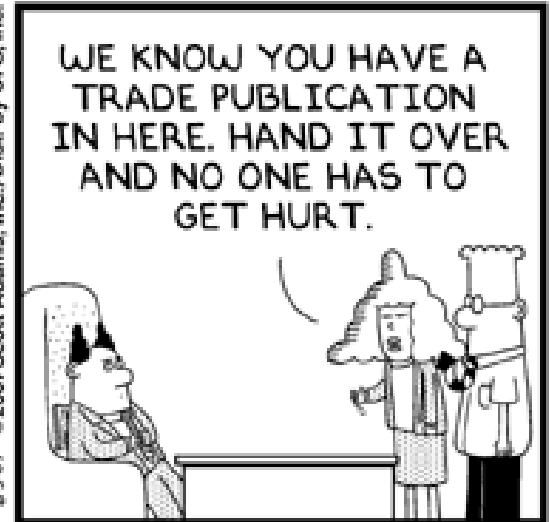
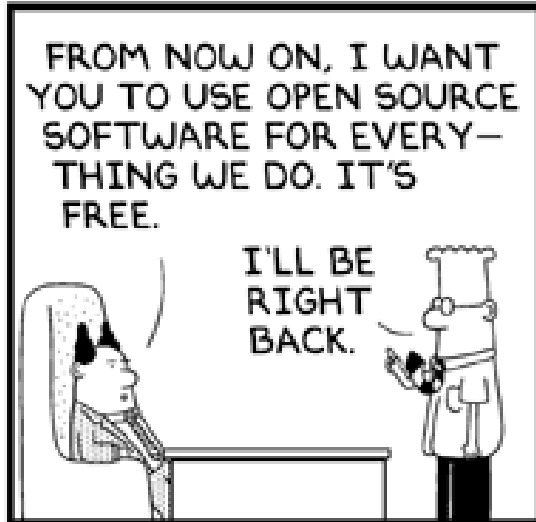
- Smart data exploration with self-services analytics makes this product usable for anyone. Interactive reporting makes it collaborative. Scalability and governance make it fit the needs of your organization, no matter the size.

Visual Statistics – [Register for Trial](#)

- Multiple users can explore and visualize data, then interactively create and refine descriptive and predictive models. Distributed, in-memory processing reduces model development time so you can run complex analytic computations – and get precise results – in minutes.

ACKNOWLEDGEMENTS

TAMARA DULL, SAS BEST PRACTICES
STEVE HOLDER, NATIONAL ANALYTICS LEAD, SAS CANADA



© Scott Adams, Inc./Dist. by UFS, Inc.

WHY OPEN SOURCE?

Why the drive to open source?

- Cost effective –considering total cost of ownership
- Flexible – customers can “build anything”
- Immediate access & easy to get started
- Latest technology and latest algorithms
- Strong community and online support
- Many new data scientists learn in open source



So why use SAS to extend open source?

SAS AS AN ENHANCEMENT



AND



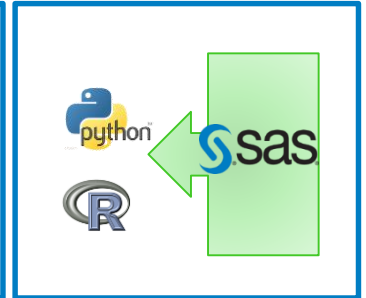
SAS can augment open source

- Increase productivity
- Leverage your assets, people and platforms
- Bring the power of SAS to open source
- Create deployable analytics
- Goal is to 'embrace' and 'extend'

Open to SAS



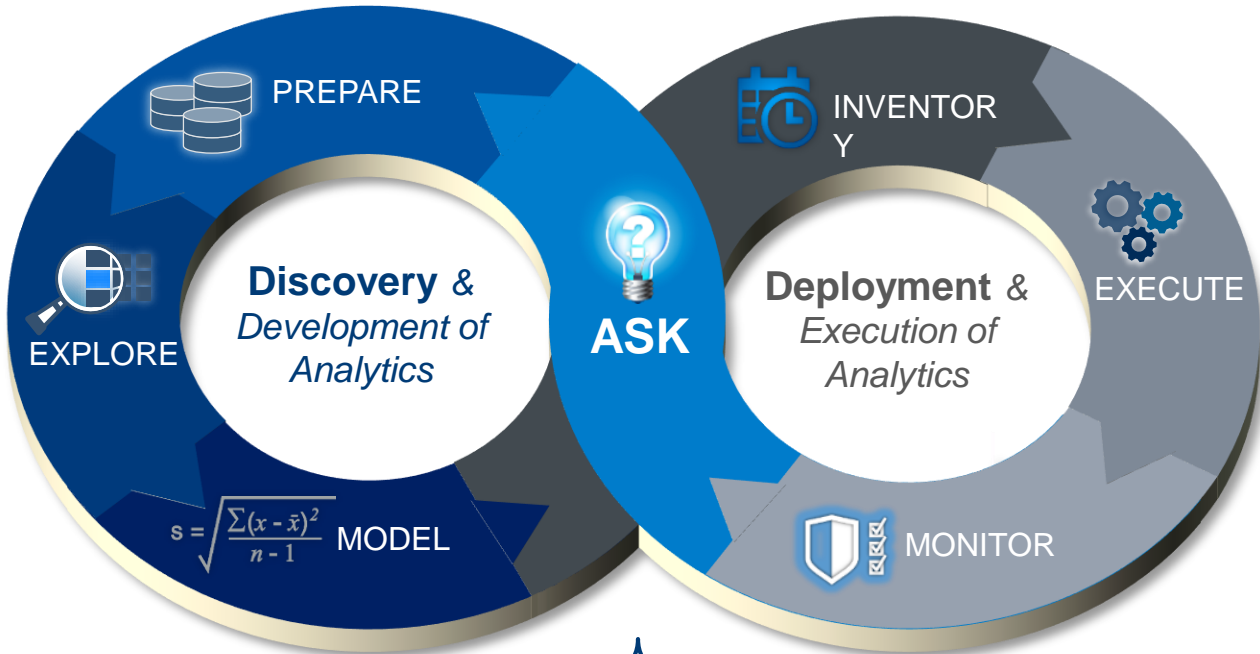
SAS to Open



THE ANALYTIC LIFECYCLE



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



THE ANALYTIC LIFECYCLE: SAS AND OPEN SOURCE

Discovery & Development of Analytics

Deployment & Execution of Analytics



SAS

Open source

- SAS embraces open source for Data Prep
- Open source and SAS work well for Discovery and Development
- SAS can extend open source
 - inventory, register and manage models
 - deploy and execute models in Hadoop and in database
 - enhance models and provide monitoring and reporting

EMBRACE



EXTEND



THE ANALYTIC LIFECYCLE

EMBRACE

Discovery & Development of Analytics

Deployment & Execution of Analytics



Enterprise Wish List

- Ability to connect to Hadoop
- Run natively in Hadoop
- Minimize data movement

How SAS Embraces...

- Optimized engine to access Hadoop
- Embedded engine so Hadoop can run SAS



HADOOP AS PROCESSING ENGINE

EMBRACE



- Use Hadoop as the horsepower for analytics
- Run SAS in Hadoop - no data movement
- Expose Hadoop data to more people through a range of interfaces
- Predictive analytics and machine learning
- SAS for Model Deployment / Scoring



THE ANALYTIC LIFECYCLE

EMBRACE

Discovery & Development of Analytics

Deployment & Execution of Analytics



Enterprise Wish List

- A way for users to interact with Hadoop
- Ability to create analytic views and tables
- Ability to assess data quality

How SAS embraces...

- A business user interface to facilitate:
 - Querying Hadoop
 - Adding data
 - Profiling data
 - Cleansing data
 - Transforming data
- With no data movement



SELF SERVE ACCESS TO HADOOP



What directive do you want to perform?

- Browse Tables**
Browse tables or open a table to see its contents
- Query or Join Data in ...**
Query a table, or join data from multiple tables
- Copy Data from Hadoop**
Copy Data from Hadoop into a database
- Profile Data**
Generate a profile report of the data in a table
- Saved Directives**
Open a previously created directive to run, view or edit
- Sort and De-Duplicate...**
Query, sort, or de-duplicate the data in an existing Hadoop table
- Copy Data to Hadoop**
Copy data from a database into Hadoop
- Saved Profile Reports**
Explore previously generated profile reports
- Run Status**
Show the status of current and previous directive executions
- Run a SAS Program**
Run in-database data quality SAS programs
- Load Data to LASR**
Copy data from a source and load it into LASR. Existing data in the target table...
- Delete Rows**
Delete rows from a selected table. Requires Hive 14 or above.
- Transform Data in Hadoop**
Transform data from a Hadoop table
- Import a File**
Import data from a file into Hadoop
- Run a SAS Program**
Run in-database data quality SAS programs
- Transp**
Transp
- Cleanse**
Cleanse

SAS Data Loader - Profile Reports

df_test4_cust_clean_profile

default#_df_test4_customer_clean

Count: 5000

Data Quality Metrics

Column	#	Unique (%)	Patterns (%)	Null (%)	Blank (%)
customer_state_standardized	2	2344	41	100	0
customer_city	3	2165	97	37	0
customer_id	4	29	1	12	0
customer_name	5	4322	85	10	2
customer_address	6	4949	99	30	0
customer_state	7	500	100	0	0
customer_name	8	4949	100	330	0
customer_lastname	9	2942	95	0	0
customer_address	10	2	0	2	0
customer_address	11	4978	100	919	0
customer_city	12	3919	87	230	0
customer_state	13	471	11	21	0
customer_address	14	3944	79	11	0

* indicates data not available or not applicable for this column.

Descriptive Measures

Metadata Measures

Profile Data

Business user UI

SAS Data Loader

Cleanse Data in Hadoop

SOURCE TABLE: default#_df_b_mca

TRANSFORMATION

Choose the transformation you want to perform on the data

- Change Case**
Change the case of data to comply with expected data type
- Field Extraction**
Extract fields from a column
- Filter Data**
Select rows of data to include
- Gender Analysis**
Identify the gender of the data to be cleaned
- Generate Match Codes**
Create match codes for related rows in the table
- Identification Analysis**
Identify the potential data type of each selected column
- Manage Columns**
Select the columns to include
- Parse Data**
Select the delimiters, quotation, and other options to apply and enter 2
- Pattern Analysis**
Compare the data to an expected pattern
- Standardize Data**
Apply data standards to selected columns
- Summary Rows**
Create a new row with the summary of selected columns

Next Add Another Transform

Create Trusted Data



THE ANALYTIC LIFECYCLE

EXTEND

Discovery & Development of Analytics

Deployment & Execution of Analytics



Enterprise Wish List

- Best possible analytics
- Flexibility of tools
- Productivity
- Greater insights = models
- Trusted models

How SAS Extends...

- A variety of options to develop models
- Allows data scientist to code in language of choice
- Ability to scale to any data volume
- Handle complex graphics

Discovery



SAS FROM R



Discovery



The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for connecting to a SAS server and retrieving data.


```

1 library("RCurl")
2 tempDir <- tempfile()
3 dir.create(tempDir)
4
5 myhtml <- getURL(url="http://joeloonix-sasbiws2.na.sas.com:7980/SASBIWS/rest/storedProcesses/CABdemo/RandomForest/d:
6             httpheader=c(Accept="application/xml,text/html",
7             "Content-Type" = "application/xml", Authorization="Basic c2FzZG9tbzpwYXNzd29yZA=="),
8             postFields="<RandomForest><parameters><dataset>shoptrain</dataset><numtrees>100</numtrees><varstot:
9             verbose = TRUE)
10
11 write(myhtml, file.path(tempDir, "sasout.html"))
12 rstudio::viewer(file.path(tempDir, "sasout.html"))
13
14
15
            
```
- Console:** Shows the output of the R script, including network logs and SAS response headers.

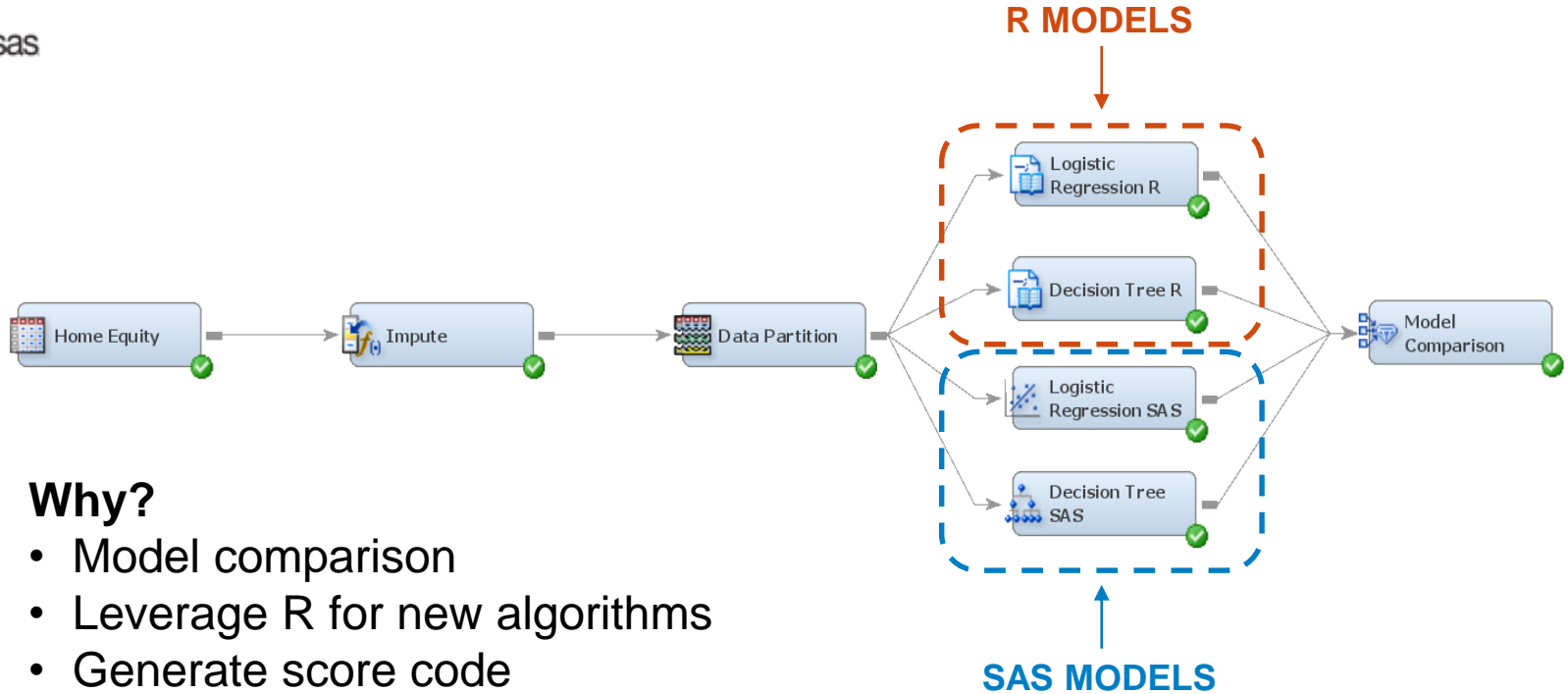

```

* upload completely sent off: 131 out of 131 bytes
< HTTP/1.1 200 OK
< Date: Tue, 26 May 2015 13:05:29 GMT
< Server: Apache-Coyote/1.1
< Content-type: text/html;charset=utf-8
< Transfer-Encoding: chunked
<
* connection #0 to host joeloonix-sasbiws2.na.sas.com left intact
> source("~/active-rstudio-document")
* Hostname was NOT found in DNS cache
* Trying 10.12.38.157...
* connected to joeloonix-sasbiws2.na.sas.com (10.12.38.157) port 7980 (#0)
> POST /SASBIWS/rest/storedProcesses/CABdemo/RandomForest/datatargets/_WEBOUT HTTP/1.1
Host: joeloonix-sasbiws2.na.sas.com:7980
Accept: application/xml,text/html
Content-type: application/xml
Authorization: Basic c2FzZG9tbzpwYXNzd29yZA==
Content-Length: 131
* upload completely sent off: 131 out of 131 bytes
< HTTP/1.1 200 OK
< Date: Tue, 26 May 2015 14:29:25 GMT
< Server: Apache-Coyote/1.1
< Content-type: text/html;charset=utf-8
< Transfer-Encoding: chunked
<
* connection #0 to host joeloonix-sasbiws2.na.sas.com left intact
            
```
- Environment:** Displays a table of loaded packages.

Package	Version	Source
USER_FACTOR2	1271	0.000548 -0.00055 0.001095 -0.00001
USER_FACTOR1	1425	0.000563 -0.00056 0.001305 0.00008
USER_FACTOR16	1398	0.000584 -0.00057 0.001168 0.00003
USER_FACTOR10	1492	0.000644 -0.00057 0.001288 0.00007
USER_FACTOR20	1547	0.000594 -0.00057 0.001187 0.00002
USER_FACTOR11	1447	0.000623 -0.00058 0.001245 0.00005
USER_FACTOR9	1412	0.000615 -0.00059 0.001230 0.00006
USER_FACTOR17	1557	0.000604 -0.00060 0.001208 0.00001
USER_FACTOR19	1610	0.000658 -0.00062 0.001316 0.00005
USER_FACTOR18	1793	0.000783 -0.00068 0.001527 0.00011
- Procedure Task Timing:**

Task	Seconds	Percent
Reading Data	2.18	22.14%
Training Forest	7.64	77.79%
Saving Model	0.01	0.07%
- OOB vs Training Plot:** A line graph showing Misclassification Rate on the y-axis (ranging from 0.220 to 0.230). The plot compares training performance (solid blue line) and out-of-bag (OOB) performance (dashed red line). Both rates decrease sharply from approximately 0.230 to 0.220 within the first few iterations and then gradually increase, with the OOB rate consistently higher than the training rate.

USE SAS TO INTEGRATE R



Why?

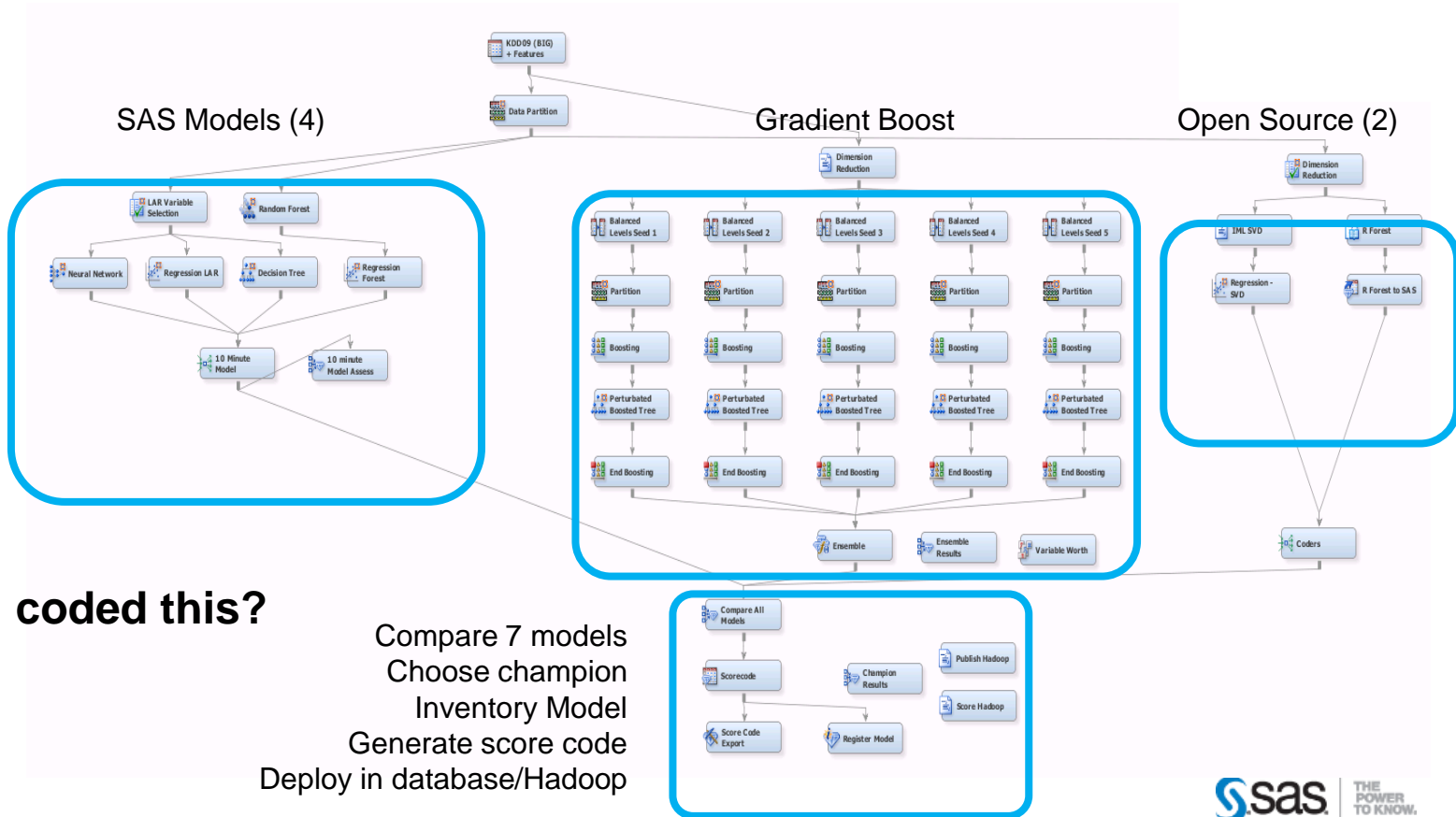
- Model comparison
- Leverage R for new algorithms
- Generate score code
- Deploy R models

Discovery



PRODUCTIVITY

EXTEND



What if you coded this?

- Compare 7 models
- Choose champion
- Inventory Model
- Generate score code
- Deploy in database/Hadoop



SAS FROM JUPYTER



jupyter CAB Demo Last Checkpoint: 4 hours ago (autosaved)

File Edit View Insert Cell Kernel Help Python 2

Code Cell Toolbar: None

```

TRAIN a random forest model on customer transaction data to predict which ones can be expected to be rep...

Args:
data: Name of the data set to train the model
numTrees: Number of trees to train in the model
numVarsToTry: Number of variables to consider for the splitting rule at each node
...

print "Training random forest from " + data + "..."
sys.stdout.flush()
sascode_uri = 'http://joeloonix-sasbiws2.na.sas.com:7980/SASBIWS/rest/storedProcesses/CARDemo/RandomFore...
headers = {"Accept": "application/xml,text/html", "Content-Type": "application/xml", "A...
xml_payload = '<RandomForest><parameters><dataset>' + data + '</dataset><numtrees>' + $
resp = requests.post(sascode_uri, headers=headers, data=xml_payload)
display(HTML(resp.text))
return

trainForest(data = "shoptrain", numTrees = 250, numVarsToTry = 20)
    
```

249	1120	0.158	0.163	0.218
250	139950	0.158	0.163	0.218

Loss Reduction Variable Importance

Variable	Number of Rules	Gini	OOB Gini
ITEM_FACTOR20	257	0.006625	0.00...
ITEM_FACTOR18	208	0.004304	0.00...

Using the ABC method search for the best number of clusters

cm = stat.cl

cm.DIAGNOST

The REG Procedure
Model MODEL1
Dependent Variable:



THE ANALYTIC LIFECYCLE

EXTEND

Discovery & Development of Analytics

Deployment & Execution of Analytics



Enterprise Wish List

- 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

EXTEND

SAS® Decision Manager

File Help

My Tasks

Data

Business Rules

Models

Projects

Portfolios

Workflows

HMEQ

Properties | Versions | Models | Variables | Scoring | Performance | Retrain | Reports | History | Attachments | Comments

Search

Name	Role	Version	Description	Model Type	Date Published	Date Modified
EM Regression		1.0		Classification	Oct 17, 2014 12:20	
HPForest		1.0		Classification	Oct 17, 2014 01:52	
Neural Net		1.0		Classification	Oct 17, 2014 12:37	
R Decision Tree	Champion	1.0		Classification	Mar 2, 2015 05:07 PM	sasdemo
Regression	Challenger	1.0		Classification	Oct 17, 2014 12:38	sas
STAT		1.0		Classification	Oct 17, 2014 12:40	sas

Alerts: 0 Total, 0 New

User: SAS Installer ID

Model inventory and search

SAS and Open Source models

Search Inventory

Save Search

Filter

Last 30 days

Last 100 days

Start date: 01/01/2016

End date: 04/05/2016

Properties

Model function:

Algorithm:

Modeler:

Input variable:

Target variable:

Keywords:

User-Defined Text Properties

User-Defined Numeric Properties

Model Properties | Versions | Attachments | Comments

Version: Current

Variables

Input

Output

Output Mapping

Properties

General

Specific

System

User-Defined

Scoring

Advanced

Score Code

Model Files

History

Log

Published

Specific Properties

Default scoring input table: HMEQ Tutorial HMEQ_SCORE_INPU X Browse

Default scoring output table: HMEQ Tutorial HMEQ_SCORE_OUTPU X Browse

Default performance table: HMEQ Tutorial HMEQ_SEQ_Q1 X Browse

Default train table: Demo Data BANK X Browse

Expiration date:

Model label:

Subject:

Status:

Algorithm: DecisionTree

Function: Classification

Modeler: sasdemo

Tool: Enterprise Miner

Tool version:

Score code type: DATA step

Template: EMSPClassification

Score code filename: score.sas

Suggested champion: No

Retrainable: Yes

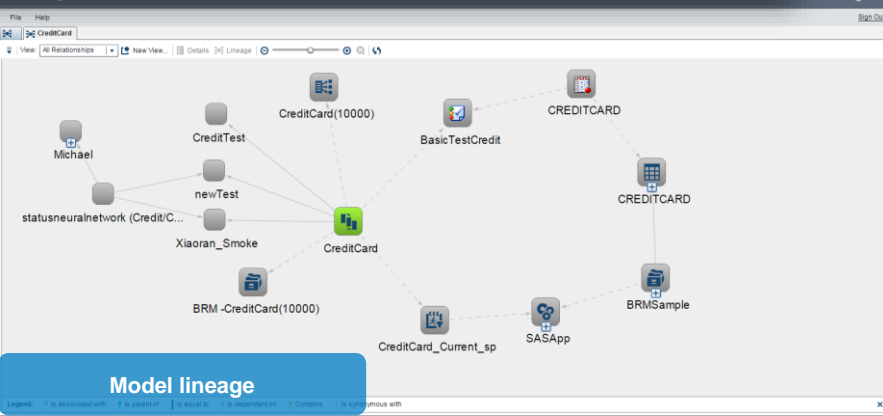
Train code filename: batch.sas

Copied from:

Target variable:

Target event value:

Model Metadata



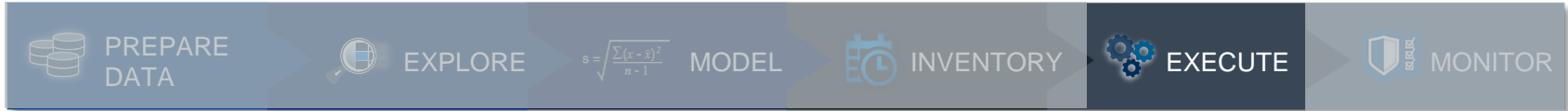
Model lineage

THE ANALYTIC LIFECYCLE

EXTEND

Discovery & Development of Analytics

Deployment & Execution of Analytics



Enterprise Wish List

- Deployable analytics
- Automation
- Faster time to model execution
- In Hadoop/database model execution

How SAS Extends...

- Model execution platform
- Execute models as database functions
- No language conversion
- Purpose built model execution engines

Deployment



MODEL EXECUTION

EXTEND

Publish Models

Publish destination: SAS Metadata Repository

Select the models to publish, and specify a publish name for each model

Select	Model Name	Role	Version	Model Type	Publish Name	Date Published
<input checked="" type="checkbox"/>	badcar-bin-reg	Champion	1.0	Classification	badcar-bin-reg_f	

Model Publishing and automation

Create a Scoring Output Table

Specify the name, library, and the variables to include in the scoring output table.

Name: *

Library: MMLib

Input variables:

<input type="checkbox"/> All	Name	Description	Type	Length
<input type="checkbox"/>	customer_id		C	20
<input type="checkbox"/>	BAD		N	8
<input type="checkbox"/>	LOAN		N	8
<input type="checkbox"/>	MORTDUE		N	8
<input type="checkbox"/>	VALUE		N	8

Output variables:

<input type="checkbox"/> All	Name	Description	Type	Length
<input type="checkbox"/>	P_BAD0	Predicted: BA...	N	8
<input type="checkbox"/>	EM_PROBABI...	Probability of ...	N	8
<input type="checkbox"/>	P_BAD1	Predicted: BA...	N	8
<input type="checkbox"/>	_WARN_	Warnings	C	4
<input type="checkbox"/>	EM_EVENTP...	Probability for ...	N	8

Add model ID

Use project mappings

Model Score Code Creation

Add Variables

Publish Models

Publish destination: Teradata

Publish method: SAS Embedded Process

Select the models to publish, and specify a publish name for each model

Select	Model Name	Role	Version	Date Published
<input type="checkbox"/>	badcar-bin-reg	Champion	1.0	

Replace scoring files that have the same publish name

Specify an identifier to add to the database target table for each model

BadCar

Validate scoring results

Train table: auto.BADCAR_TRAIN

Teradata Settings

Database server:

Database:

User ID: Password:

In Hadoop/database deployment

Publish Cancel

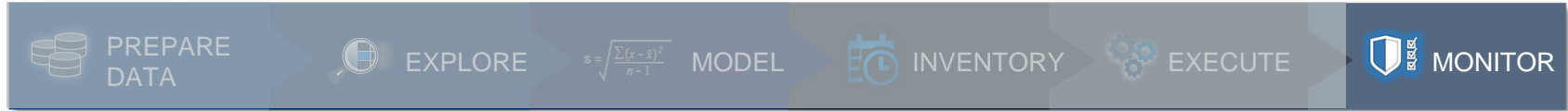


THE ANALYTIC LIFECYCLE

EXTEND

Discovery & Development of Analytics

Deployment & Execution of Analytics



Enterprise Wish List

- Best possible models
- Model tournaments
- Visibility into performance
- Easy retraining
- Champion/challenger modelling

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

Deployment



MODEL PERFORMANCE

EXTEND

Retrain Settings

Models

Name	Version	Model Type	Role
Tree1	1.0	Classification	

Data processing method: Standard configuration
 Destination version for new models: New version
 Training data source: Tutorials:HMEQ_TRAIN
 SAS Application Server: SASApp
 Report folder: D:\SMM131Tutorials\Reports
 Retrain results folder: D:\SMM131Tutorials\Retrain

Register new trained model
 Trace on

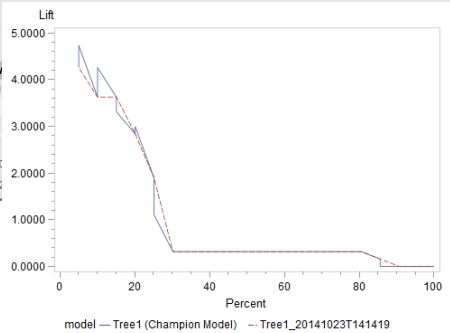
Comparison Settings

Models

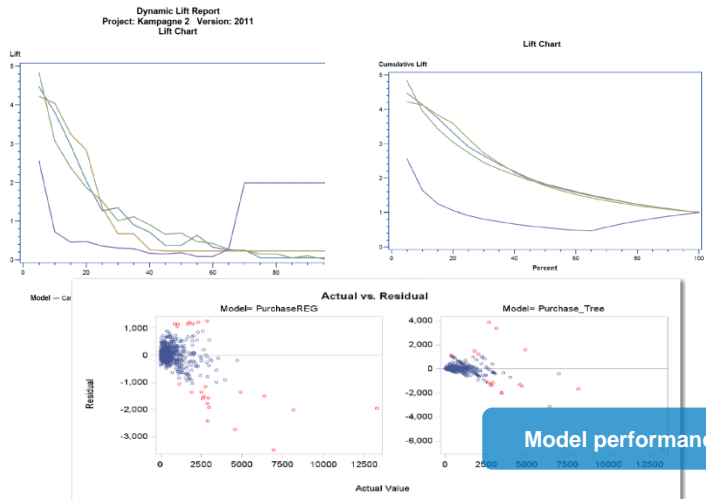
Name	V
Tree1	1

Partition percent: 3C
 Random seed: 12

Retrain models



Model comparisons



Model performance reports

BadCar

Properties | Versions | Models | Variables | Scoring | Performance | Retrain | Reports | History | Attachments | Comments

Definition | Results | Schedule | Job History

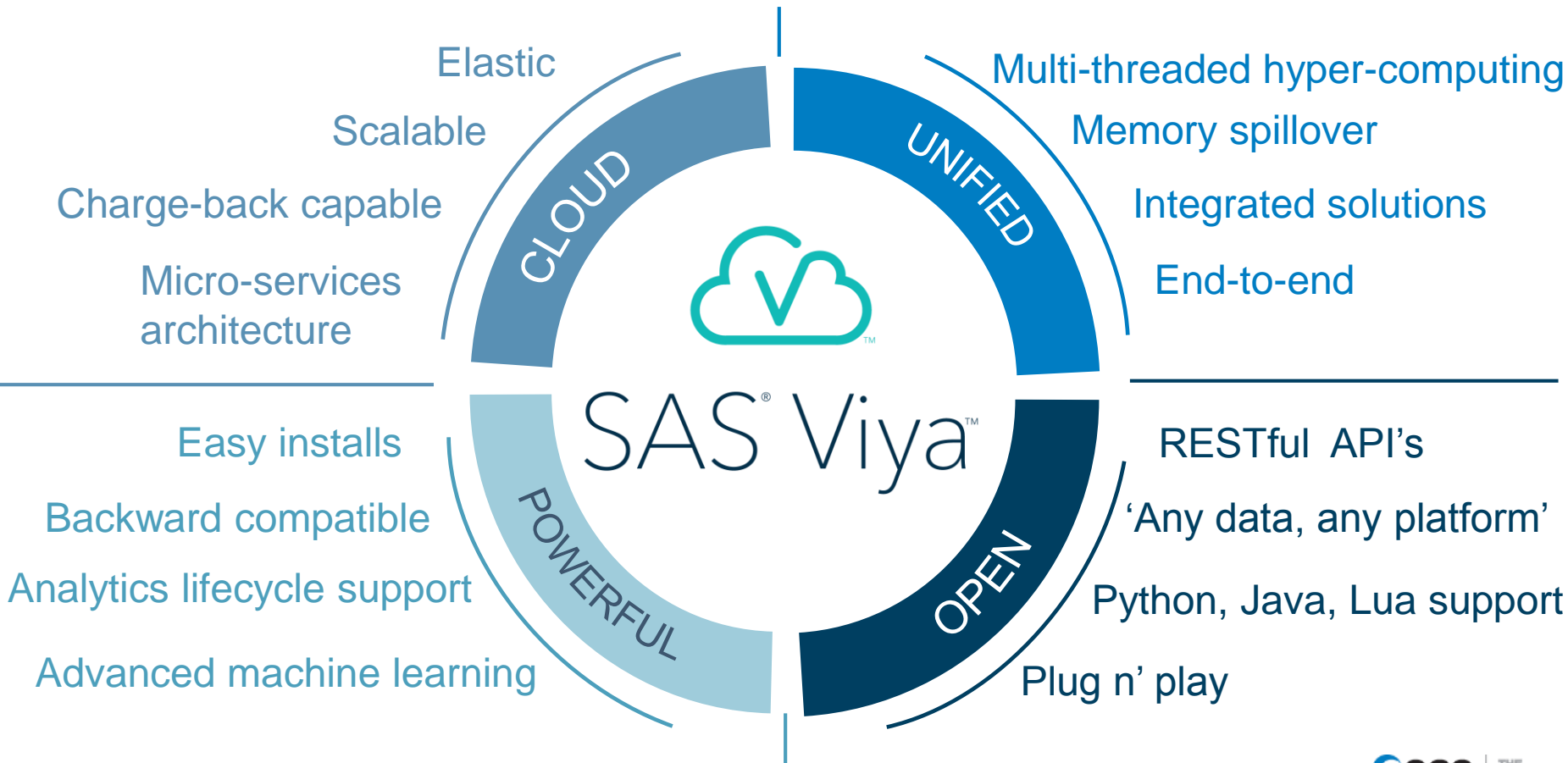
Stability	Lift	KS	Bas.	Even.	Count	Perc.	Com.	Com.	Lift	Com.	Perc.	Com.	Capt.	Com.	Mod.	Cha.	Task	
1	Y	0	0	0.00	0	0	NaN	NaN	NaN	NaN	0	0	0.26	0.177	0	1	Feb 1.	
2	Y	1367	2085	0.00	1367	2085	5.29	0.65	0.65	0.26	0.26	0.26	0.26	0.177	0	1	Feb 1.	
3	Y	471	2085	18.00	1836	4170	1.82	3.55	0.22	0.44	0.09	0.25	0.25	0.25	0.177	0	1	Feb 1.
4	Y	362	2085	15.00	2300	6255	1.40	2.83	0.17	0.35	0.07	0.42	0.26	0.177	0	1	Feb 1.	
5	Y	308	2085	20.00	2508	8340	1.39	2.42	0.14	0.30	0.05	0.48	0.26	0.177	0	1	Feb 1.	
6	Y	286	2085	25.00	2794	10425	1.10	2.16	0.13	0.26	0.05	0.54	0.26	0.177	0	1	Feb 1.	
7	Y	289	2085	30.00	3083	12510	1.11	1.98	0.13	0.24	0.05	0.59	0.26	0.177	0	1	Feb 1.	
8	Y	239	2085	35.00	3322	14595	0.92	1.83	0.11	0.22	0.04	0.64	0.26	0.177	0	1	Feb 1.	
9	Y	235	2085	40.00	3557	16660	0.90	1.72	0.11	0.21	0.04	0.68	0.26	0.177	0	1	Feb 1.	
10	Y	189	2085	45.00	3746	18765	0.73	1.61	0.09	0.19	0.03	0.72	0.26	0.177	0	1	Feb 1.	
11	Y	184	2085	50.00	3930	20950	0.71	1.52	0.08	0.19	0.03	0.76	0.26	0.177	0	1	Feb 1.	
12	Y	171	2085	54.99	4161	22935	0.66	1.44	0.08	0.17	0.03	0.79	0.26	0.177	0	1	Feb 1.	
13	Y	164	2085	59.99	4265	25020	0.63	1.37	0.07	0.17	0.03	0.82	0.26	0.177	0	1	Feb 1.	
14	Y	157	2085	64.99	4422	27105	0.60	1.31	0.07	0.16	0.03	0.85	0.26	0.177	0	1	Feb 1.	
15	Y	138	2085	69.99	4500	29190	0.53	1.26	0.06	0.15	0.02	0.88	0.26	0.177	0	1	Feb 1.	
16	Y	125	2085	74.99	4585	31275	0.48	1.20	0.05	0.14	0.02	0.90	0.26	0.177	0	1	Feb 1.	
17	V	197	2085	78.08	4810	33360	0.49	1.16	0.06	0.14	0.02	0.91	0.26	0.177	0	1	Feb 1.	

Monitor data drift



THE FUTURE IS NOW...





SAS AND OPEN SOURCE

SAS 9.4



EMBRACE

open source by including it
and leveraging it where we
can



EXTEND

open source by improving
its interoperability and
utility for the enterprise

THANK YOU
MATT MALCZEWSKI
MATT.MALCZEWSKI@SAS.COM



FOR MORE INFORMATION

Empowering the SAS/IML user with the functionality of R

Documentation: *IML User's Guide - Calling Functions in the R Language*

http://support.sas.com/documentation/cdl/en/imlug/66845/HTML/default/viewer.htm#imlug_r_toc.htm

Video: *Calling R Procedures from SAS/IML® Software*

<https://www.youtube.com/watch?v=rUaTTre24kl>

Video: *SAS/IML and R: Using Them Together*

<https://www.youtube.com/watch?v=nmRQ3MtkG6A>

Blogs: *The DO Loop – R tags*

<http://blogs.sas.com/content/iml/tag/r/>

Paper (p 14-17): *Rediscovering SAS/IML® Software: Modern Data Analysis for the Practicing Statistician*

<http://support.sas.com/resources/papers/proceedings10/329-2010.pdf>

Article: *Versions of R that are supported by SAS/IML*

<http://blogs.sas.com/content/iml/2013/09/16/what-versions-of-r-are-supported-by-sas.html>

FOR MORE INFORMATION - EXTENDING R

Video: *Using R in SAS Enterprise Miner*

<https://www.youtube.com/watch?v=TbXo0xQCqDw>

Blogs: *Spectral Clustering in SAS® Enterprise Miner™ Using Open Source Integration Node*

<https://communities.sas.com/docs/DOC-8011>

Blogs: *How to execute a Python script in SAS® Enterprise Miner™*

<https://communities.sas.com/docs/DOC-10832>

Blogs: *Open Source Integration Using the Base SAS Java Object*

<https://communities.sas.com/docs/DOC-10746>

Article: *The Open Source Integration node installation cheat sheet*

<https://communities.sas.com/docs/DOC-9988>

Usage Notes:

<http://support.sas.com/dsearch?Find=Search&ct=&qt=open+source&col=supprdr&nh=25&qp=&qc=suppsas&ws=1&qm=1&st=1&lk=1&rf=0&oq=&rq=0>

FOR MORE INFORMATION MATERIALS ON GITHUB

Sas integration and sample code Integration with R, Python

<https://github.com/sassoftware/enlighten-integration>

Integration with Jupyter Notebook and Python

https://github.com/sassoftware/sas_kernel

<https://github.com/sassoftware/saspy>


Sample codes of SAS Machine Learning methods

<https://github.com/sassoftware/enlighten-apply>

SAS Enterprise Miner process flow diagrams

<https://github.com/sassoftware/dm-flow>

This organization Search

 **sassoftware** ⓘ

Java ★ 23 🍴 20

enlighten-integration

Example code and materials that illustrate techniques for integrating SAS with popular open source analytics technologies like Python and R.

Updated a day ago

Jupyter Notebook ★ 18 🍴 6

sas_kernel

A Jupyter kernel for SAS. This opens 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 and view results inline.

Updated 2 days ago

Python ★ 8 🍴 5

saspy

An interface module to the SAS System. It works with Linux SAS, and is currently intended as a support module for the sas_kernel project which provides a Jupyter Notebook kernel which surfaces the SAS Language and SAS ODS Output to Jupyter Notebooks. Additionally, provides magics which allow SAS code to be submitted for notebooks with other kern...

Updated 4 days ago

SAS ★ 40 🍴 31

enlighten-apply

Example code and materials that illustrate applications of SAS machine learning techniques.

Updated 8 days ago

★ 9 🍴 6

dm-flow

Library of SAS Enterprise Miner process flow diagrams to help you learn by example about specific data mining topics.

Updated 21 days ago