How does SAS® Support Machine Learning (ML)?

Getting the most out of your data

Melodie Rush
Customer Success Principal Analytical Engineer
https://www.linkedin.com/in/melodierush

THE POWER TO KNOW.

How does SAS support Machine Learning?

Agenda

- What is Machine Learning?
- Terminology and key characteristics
- How you can use machine learning in SAS
- Examples in SAS 9 and SAS Enterprise Miner
- Examples in SAS Viya

Machine Learning & Artificial Intelligence

What is Machine Learning?

Definition

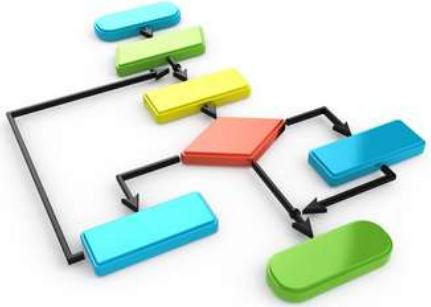
- Automatic
- Adaptive

Using iterative processes, machine learning builds models that ***automatically adapt*** *with little or no human intervention.*

Why is it so important now?

Data

Computing Power



Algorithms

Terminology

Terminology

Machine learning terms versus inferential statistics terms

What are all these archaic, outmoded and confusing terms?

What are all these new fangled and confusing terms?

- Feature
- Input
- Target
- Object

- Variable
- Independent Variable
- Dependent Variable
- Observation

Terminology

What are Machine Learning terminology?

- In statistics we predict a Y or a dependent variable.
- In data mining, Y is called a target.
- In machine learning, a target is called a label.
- In statistics and data mining our inputs are called X's.
- In machine learning our inputs are called features.
- In statistics and data mining we transform our X's.
- In machine learning we do feature creation.

How Does Machine Learning Work?

Distinguish apple from orange

How Does Machine Learning Work?

Distinguish Granny Smith apple from Fuji apple

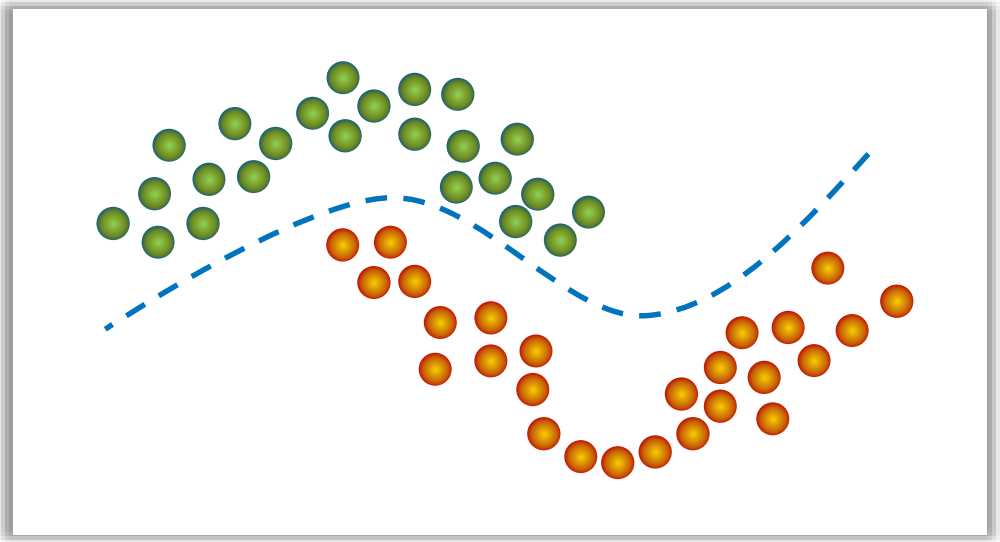
How Does Machine Learning Work?

Finding the rotten apple

How Does Machine Learning Work?

Supervised Learning

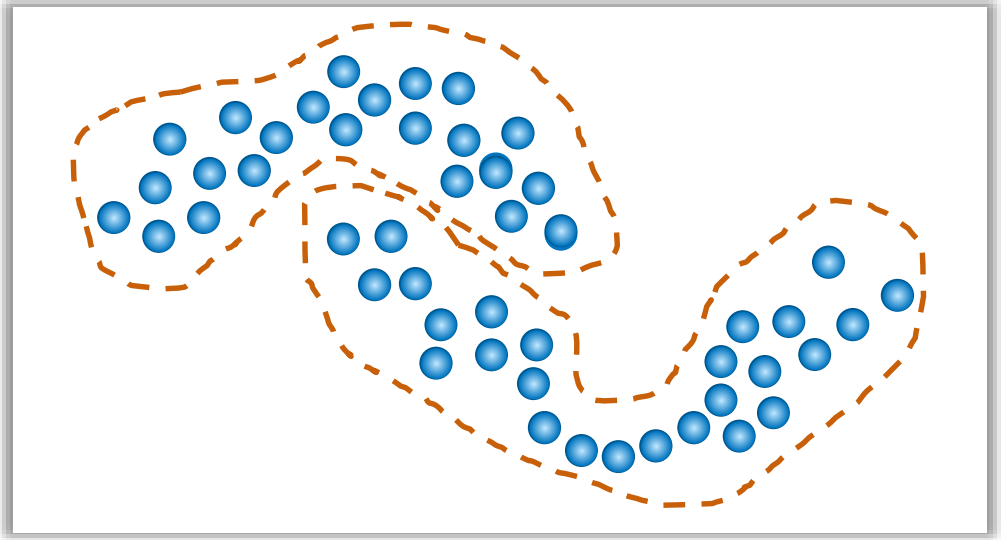
Trained on labeled examples



How Does Machine Learning Work?

Unsupervised Learning

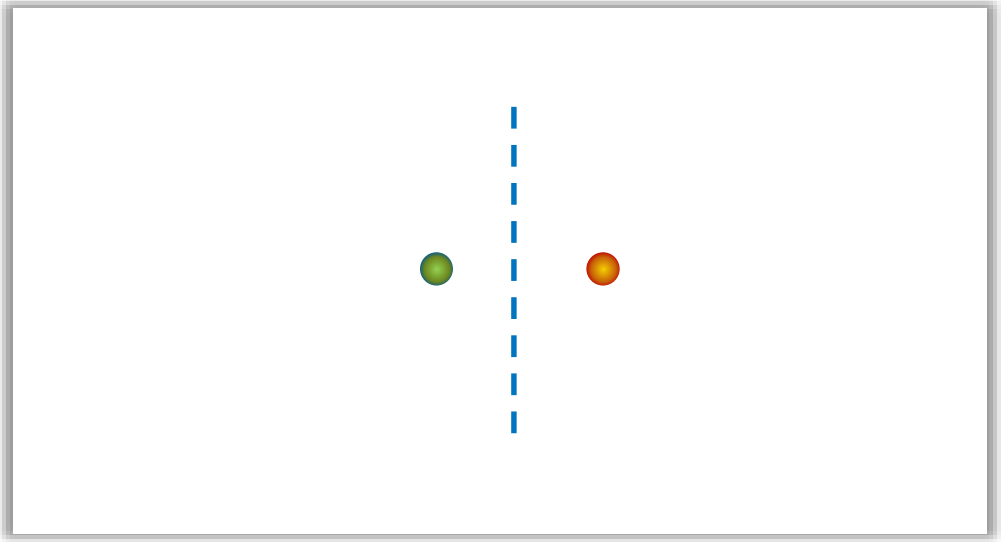
Trained on unlabeled examples



How Does Machine Learning Work?

Semi-Supervised Learning

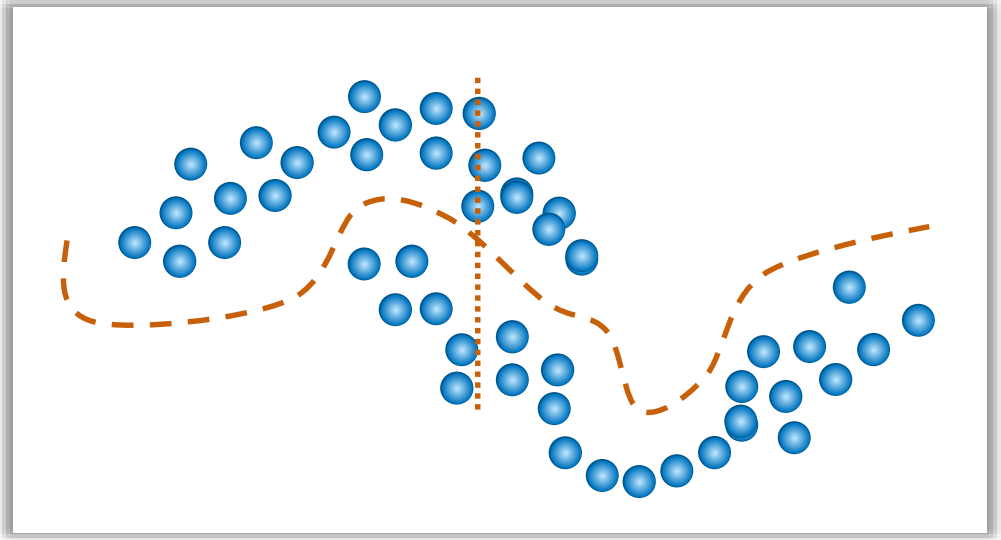
Use labeled and unlabeled observations



How Does Machine Learning Work?

Semi-Supervised Learning

Use labeled and unlabeled observations

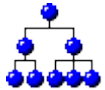


How Does Machine Learning Work?

Not New for SAS

Machine Learning has been available in both SAS/STAT and Enterprise Miner for decades


Neural Networks

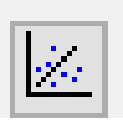

Decision Trees


Random Forests


Clustering

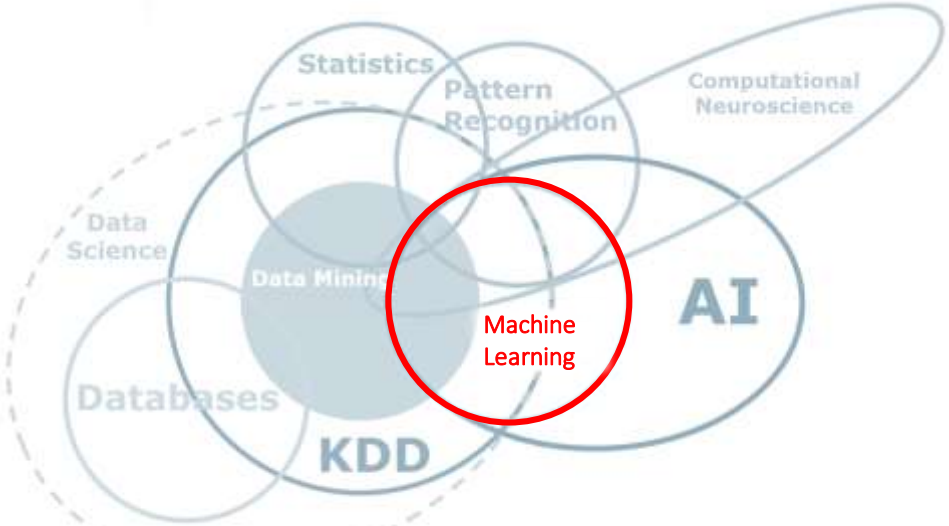

Gradient Boosting


Text Analytics


Regression

PROC DISCRIM (K-nearest-neighbor discriminant analysis) – James Goodnight, SAS founder and CEO, 1979

- 1950's:
  - Samuel's checker program



SAS Data Mining Primer Course
SAS Institute, 1998

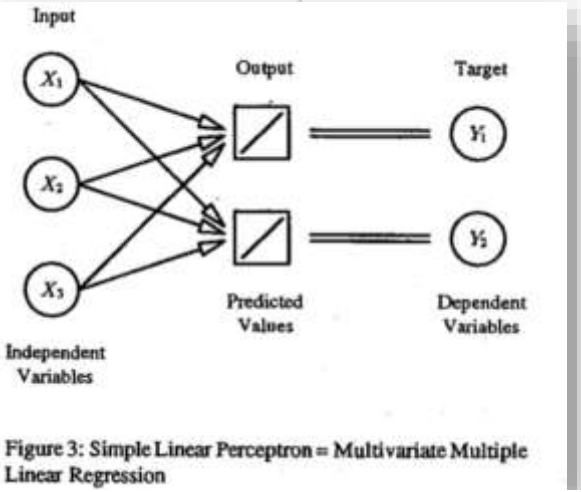


Figure 3: Simple Linear Perceptron = Multivariate Multiple Linear Regression

Neural Networks and Statistical Models,
SAS Institute, 1994

SAS Machine Learning

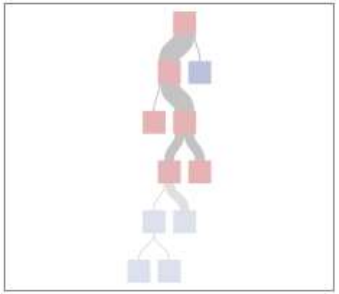
ALGORITHMS

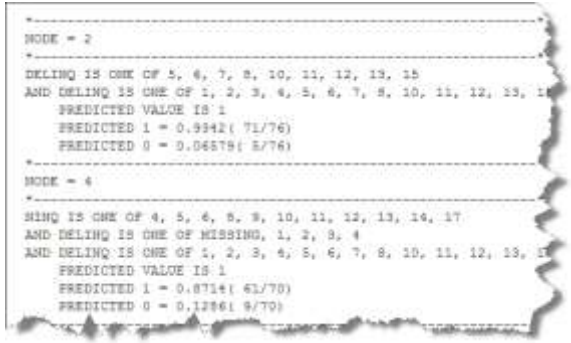
- Neural networks
- Decision trees
- Random forests
- Associations and sequence discovery
- Gradient boosting and bagging
- Support vector machines
- Nearest-neighbor mapping
- k-means clustering
- Self-organizing maps
- Local search optimization techniques such as Genetic algorithms
- Regression
- Expectation maximization
- Multivariate adaptive regression splines
- Bayesian networks
- Factorization Machines
- Kernel density estimation
- Principal components analysis
- Singular value decomposition
- Gaussian mixture models
- Sequential covering rule building
- Model Ensembles
- And More…….

Did you know?

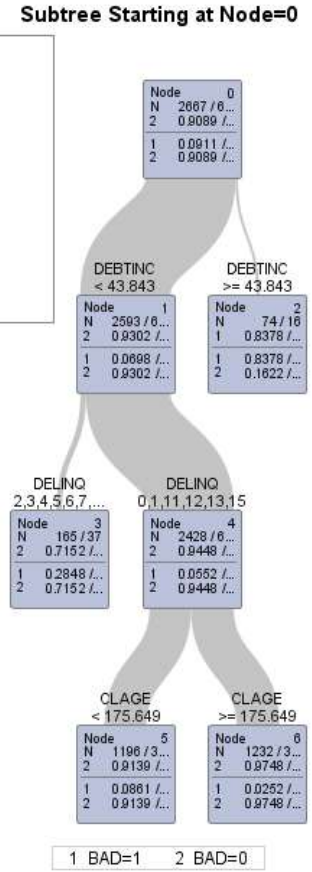
HPSplit Procedure

```
proc hpsplit data=sashelp.hmeq maxdepth=7 maxbranch=2;
    target BAD;
    input DELINQ DEROG JOB NINQ REASON / level=nom;
    input CLAGE CLNO DEBTINC LOAN MORTDUE VALUE YOJ / level=int;
    criterion entropy;
    prune misc / N <= 6;
    partition fraction(validate=0.2);
    rules file='hpsplhme2-rules.txt';
    score out=scored2;
run;
```





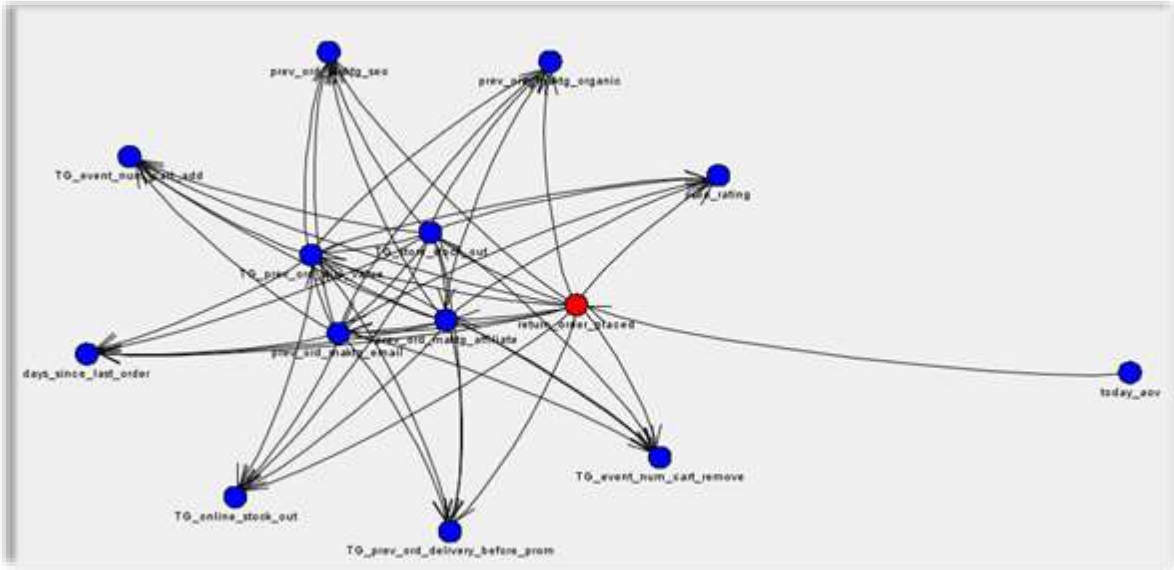
```
*-----------------------------------------------------------------*
NODE = 2
*-----------------------------------------------------------------*
DELINQ IS ONE OF 5, 6, 7, 8, 10, 11, 12, 13, 15
AND DELINQ IS ONE OF 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 1
    PREDICTED VALUE IS 1
    PREDICTED 1 = 0.9342( 71/76)
    PREDICTED 0 = 0.0658( 5/76)
*-----------------------------------------------------------------*
NODE = 4
*-----------------------------------------------------------------*
NINQ IS ONE OF 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 17
AND DELINQ IS ONE OF MISSING, 1, 2, 3, 4
AND DELINQ IS ONE OF 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 1
    PREDICTED VALUE IS 1
    PREDICTED 1 = 0.8714( 61/70)
    PREDICTED 0 = 0.1286( 9/70)
```



[HPSPLIT Procedure Documentation]

SAS Enterprise Miner

Algorithms – basic and advanced



SAS Enterprise Miner

- Linear & Logistic Regression
- Decision Trees
- Random Forest
- Gradient Boosting
- Support Vector Machines
- Neural Networks
- Clustering
- Bayesian Networks
- Principal Components
- Open Source Models

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

Alphabetic List of Variables and Attributes				
#	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 deliquent 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

SAS Enterprise Miner

Machine Learning and Artificial Intelligence Demo

SAS Viya

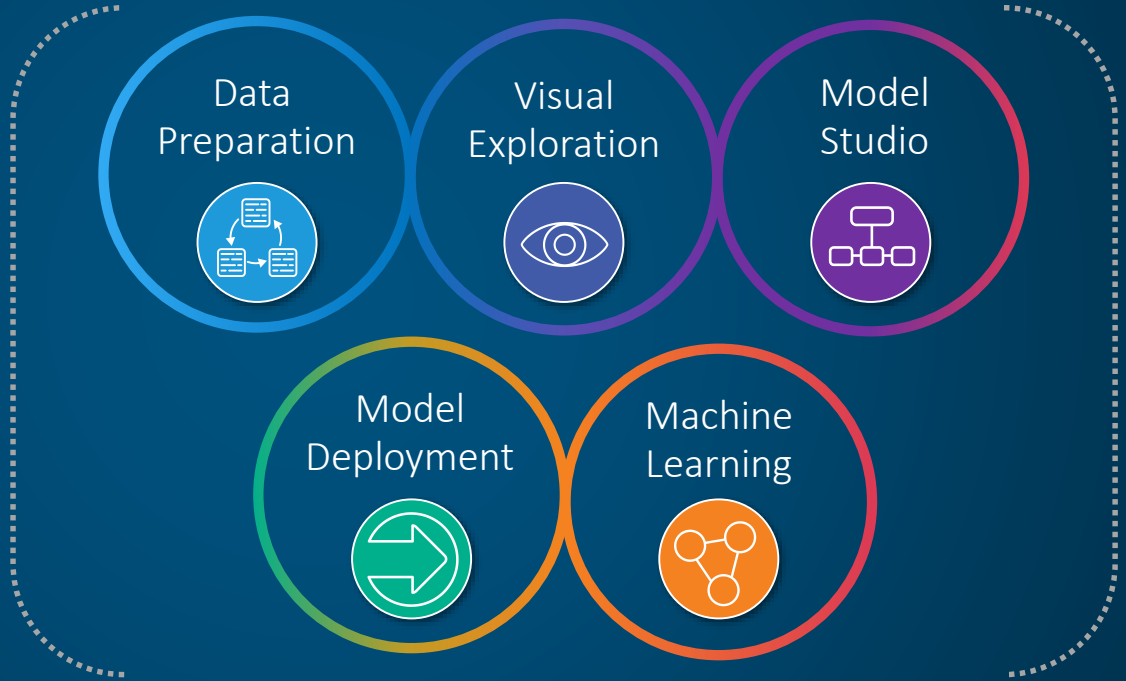
SAS Viya Products

- SAS Viya is an underlying foundation for additional products that will take advantage of a cloud-enabled, open platform. Most offerings include both a coding interface as well a visual interface.
  - SAS Visual Analytics
  - SAS Visual Statistics
  - SAS Visual Data Mining and Machine Learning
  - SAS Visual Forecasting
  - SAS Visual Text Mining
  - SAS Optimization
  - SAS Econometrics
  - SAS Visual Investigator

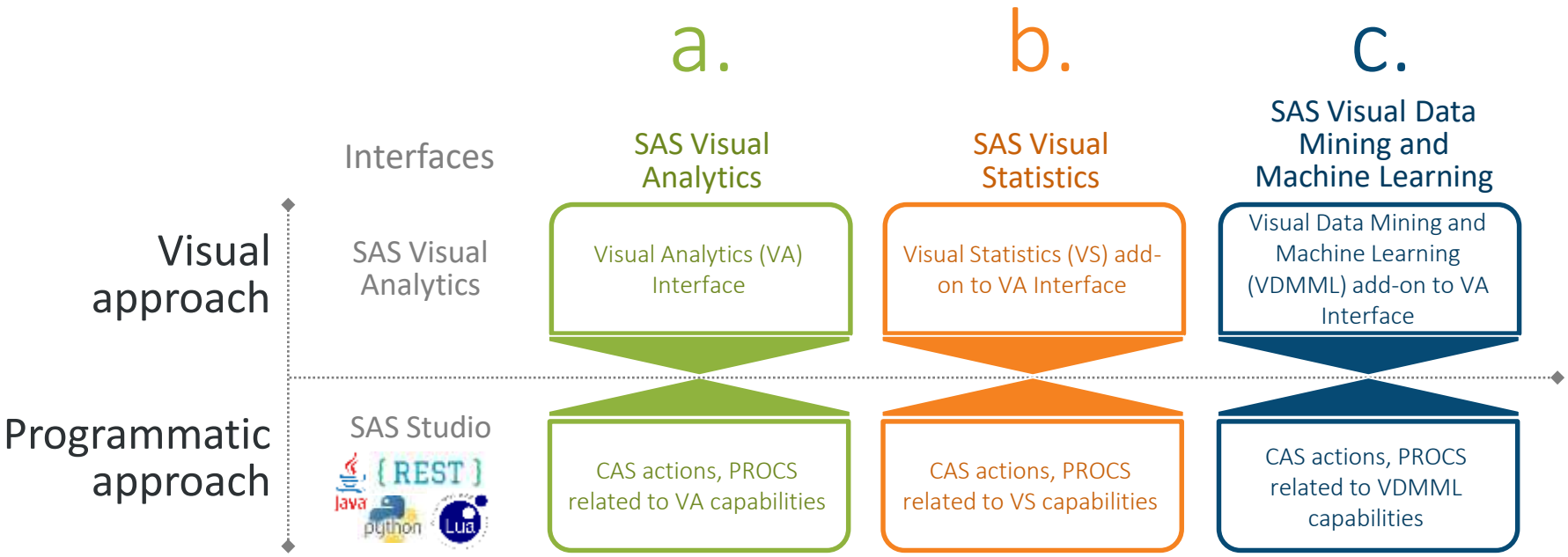
SAS® Visual Data Mining and Machine Learning

Visual "drag & drop" Interface

Programming Interface



SAS® Viya™ Based Product offerings



How Does Machine Learning Work?

Example Algorithms in SAS VS/VDMML

SUPERVISED LEARNING

- Tree-based models (Decision Trees, Random Forests, Gradient Boosting)
- Regressions
- Neural Networks
- Support Vector Machines
- Recommender Systems

UNSUPERVISED LEARNING

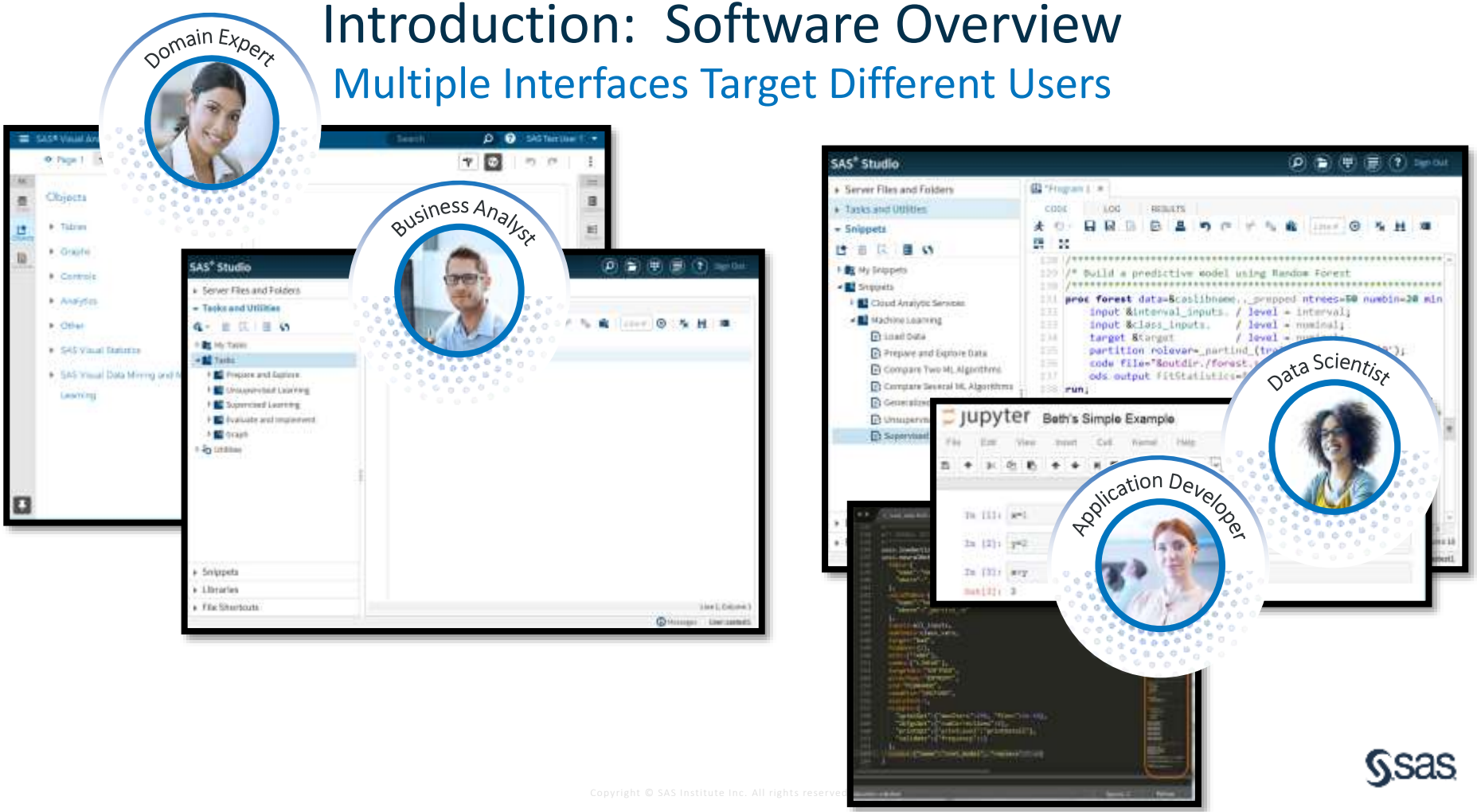
- Clustering
- Principal Component Analysis
- Text Topic Detection

SEMI-SUPERVISED LEARNING

- Clustering
- Factorization Machines

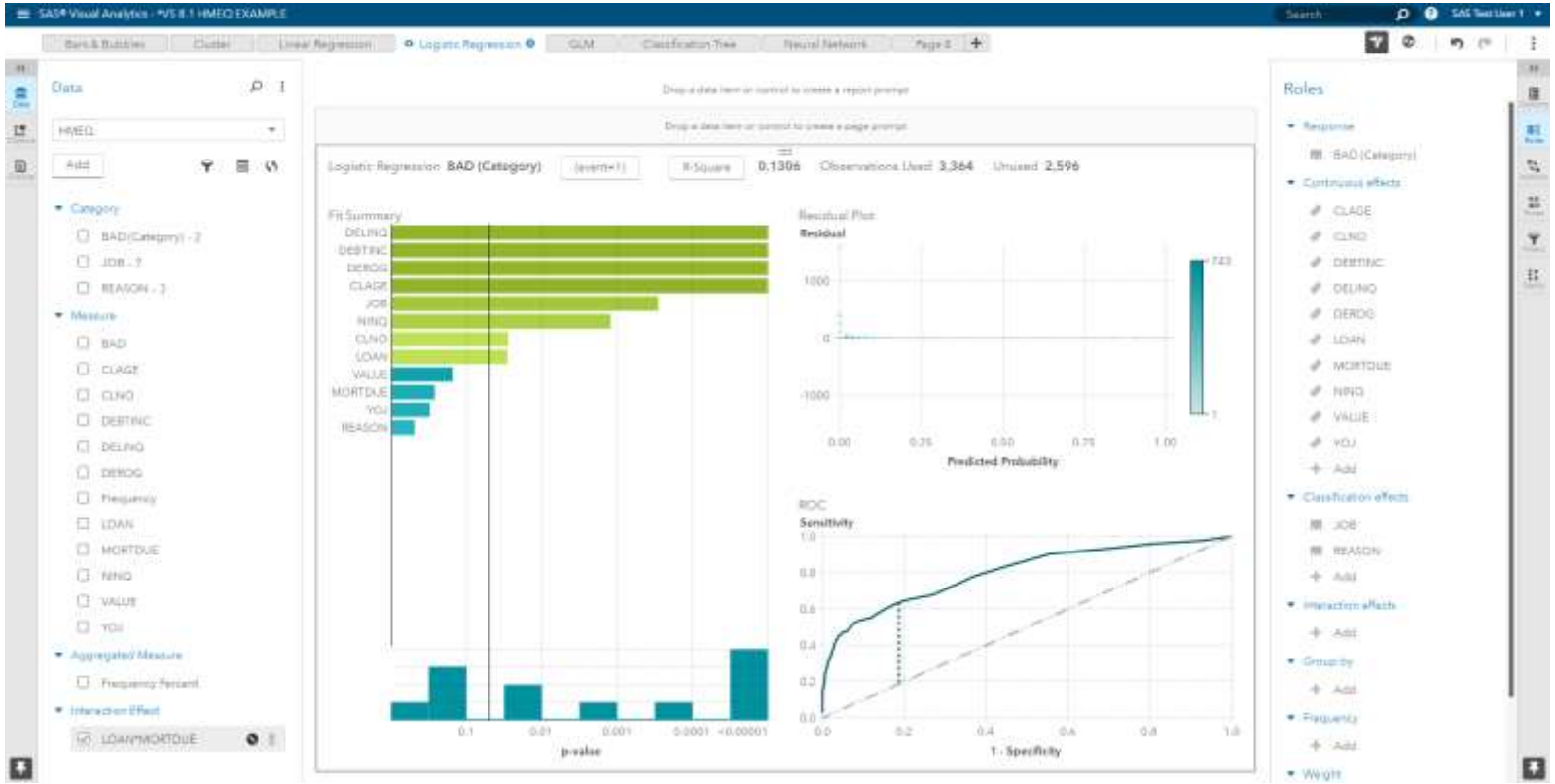
Introduction: Software Overview

Multiple Interfaces Target Different Users



Interfaces

Building a Model from Scratch in the Visual Reporting Interface



Interfaces

Building a Model Using SAS Studio Tasks

SAS® Studio

Server Files and Folders

Tasks and Utilities

My Tasks

Tasks

Prepare and Explore

Summary

Transform Data

Variable Selection

Sampling

Partitioning

Binning

Imputation

Unsupervised Learning

Clustering

Principal Component Analysis

Moving Window Principal Component Analysis

Support Vector Data Description

Robust Principal Component Analysis

Text Parsing and Topic Discovery

Supervised Learning

Linear Regression

Logistic Regression

Generalized Linear Models

Partial Least Squares Regression

Quantile Regression

Snippets

Libraries

File Shortcuts

*Logistic Regression 1

Settings | Code/Results | Split

DATA | MODEL | SELECTION

*Event of interest:

1

Link function:

Logit

Inputs

Interval inputs:

LOAN

MORTDUE

VALUE

DEROG

DELINQ

NINQ

DEBTINC

Nominal inputs:

REASON

JOB

CODE | LOG | RESULTS

```
/*
 *
 * Task code generated by SAS Studio 4.2
 *
 * Generated on '4/12/17, 6:15 PM'
 * Generated by 'sastest1'
 * Generated on server 'SASVIYA01.RACE.SAS.COM'
 * Generated on SAS platform 'Linux LIN X64 3.10.0-514.6.1.el7,x86_64'
 * Generated on SAS version 'V.03.02M0P03082017'
 * Generated on browser 'Mozilla/5.0 (Windows NT 6.3; WOW64) AppleWebKit/537.36 (KHTML,
 * Generated on web client 'http://sasviya01.race.sas.com/SASStudio/main?locale=en_US&
 *
 */

ods noproctitle;

proc logselect data=CASUSER.CASHMEQ;
    class REASON JOB;
    model BAD(event='1')=REASON JOB LOAN MORTDUE VALUE DEROG DELINQ NINQ DEBTINC /
        link=logit;
run;
```

Interfaces

Building a Model Using SAS Studio Snippets

SAS® Studio

Sign Out

- Server Files and Folders
- Tasks and Utilities
- Snippets
  - My Snippets
    - My First CAS
  - Snippets
    - Cloud Analytic Services
    - Machine Learning
      - Load Data
      - Prepare and Explore Data
      - Compare Two ML Algorithms
      - Compare Several ML Algorithms
      - Generalized Linear Models
      - Unsupervised Learning
      - Supervised Learning
- Libraries
- File Shortcuts

Supervised Learning 1

CODE | LOG | RESULTS

```
    by _NAME_;
run;

/* Variance explained by Iteration plot */
proc sgplot data=out_iter_trans;
    title "Variance Explained by Iteration";
    yaxis label="Variance Explained";
    vbar Iteration / response=COL1 group=_NAME_;
run;

/*********************************************************************/
/* Build a predictive model using Random Forest                      */
/*********************************************************************/
proc forest data=&caslibname.._prepped ntrees=50 numbin=20 minleafsize=5;
    input &interval_inputs. / level = interval;
    input &class_inputs.    / level = nominal;
    target &target          / level = nominal;
    partition rolevar=_partind_(train='1' validate='0');
    code file="&outdir./forest.sas";
    ods output FitStatistics=fitstats;
```

Messages: 1 User: sastest1

Interfaces

Building a Model Using Python

jupyter

Control Panel | Logout

GELANALYTIC_03ModelDev_04PythonSupervised (unsaved changes)

File | Edit | View | Insert | Cell | Kernel | Help | Python [Root]

Markdown | CellToolbar

BAD. BAD indicates whether or not a home equity loan recipient defaulted (BAD =1) or not (BAD =0).

1. Set up (same as previous exercise)

In [1]:

```
import swat
from swat import *
from pprint import pprint
from matplotlib import pyplot as plt
import pandas as pd
import sys
%matplotlib inline

cashost='gatekrbhdp01.gatehadoop.com'
casport=5570
casauth='/home/sasdemo/.authinfo'
indata_dir='/opt/sas/data'
indata="hmeq"

sess = swat.CAS(cashost, casport, authinfo=casauth, caslib="mycaslib")

caslib=sess.table.addcaslib(name="mycaslib", path="/opt/sas/data/",datasource={
results=sess.table.caslibInfo(caslib="mycaslib",verbose="True")
info=sess.table.fileinfo(caslib="mycaslib")
res=sess.table.loadtable(path="hmeq.sas7bdat",caslib="mycaslib",casout={"name":
res=sess.table.tableinfo()
```

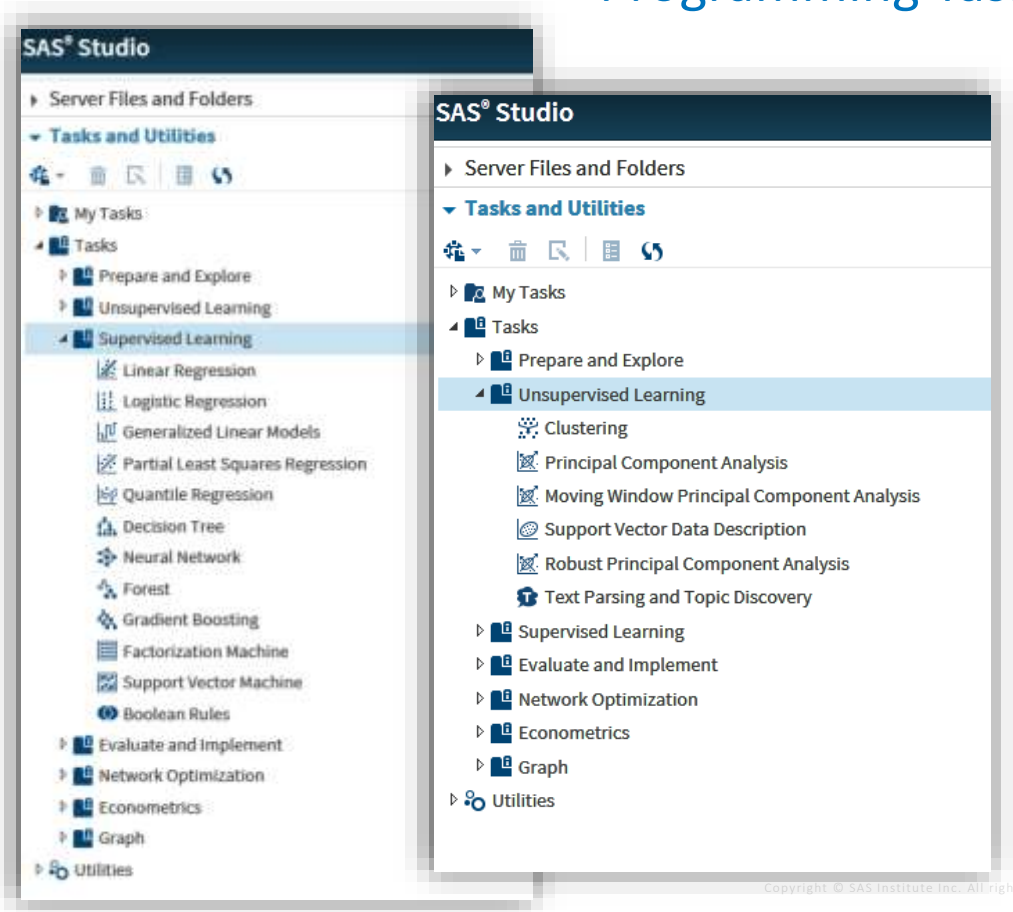
SAS Viya Key Features

Visual Interface

- Modeling Techniques
  - Clustering (k-means)
  - Linear Regression
  - Logistic Regression
  - GLM Regression
  - Decision Trees
  - Forest
  - Factorization Machine
  - Gradient Boosting
  - Neural Network
  - Support Vector Machine
- Common Features
  - Training & Validation partitioning
  - Variable Importance / Profile
  - Model Assessment
  - Model comparison
  - Derivation of predictive outputs
  - Ability to export model statistics into Excel
  - Score Code or Astore Table

SAS Visual Data Mining and Machine Learning 8.3

Programming Tasks in SAS Studio



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
  - Boolean Rules

Visual Interface Demo

SAS Visual Analytics

Visual Interface

Pipelines

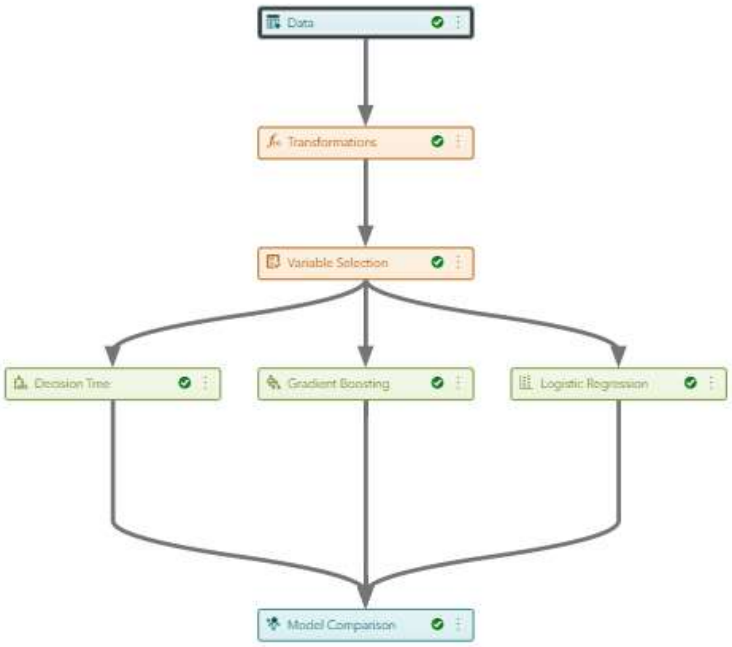
SAS® Viya
Pipelines

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

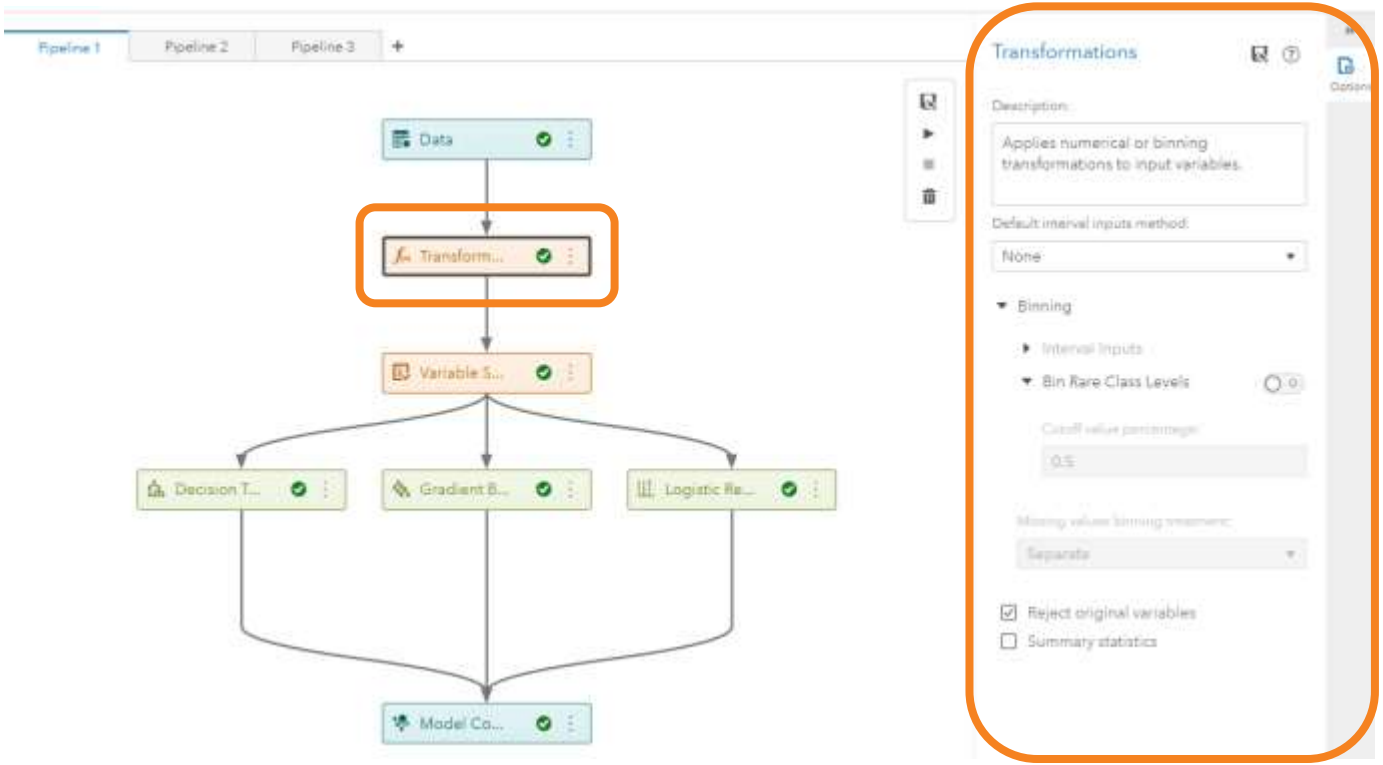
SAS® Viya

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® Viya

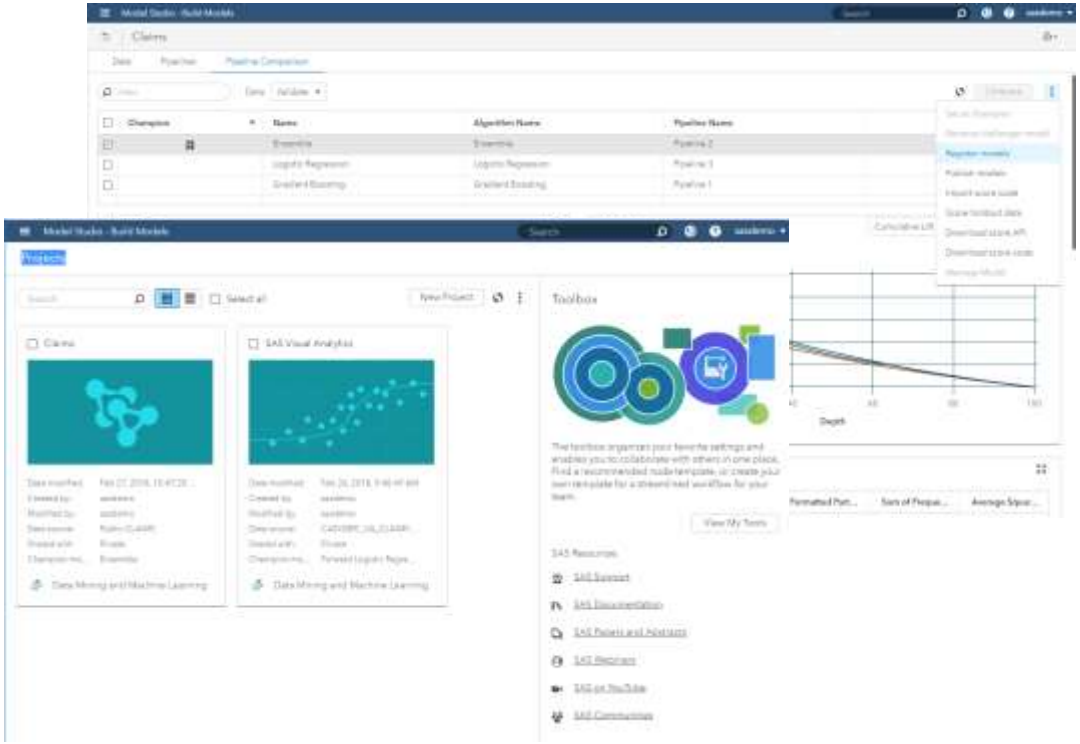
Pipelines



SAS® Viya

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



Visual Interface Demo

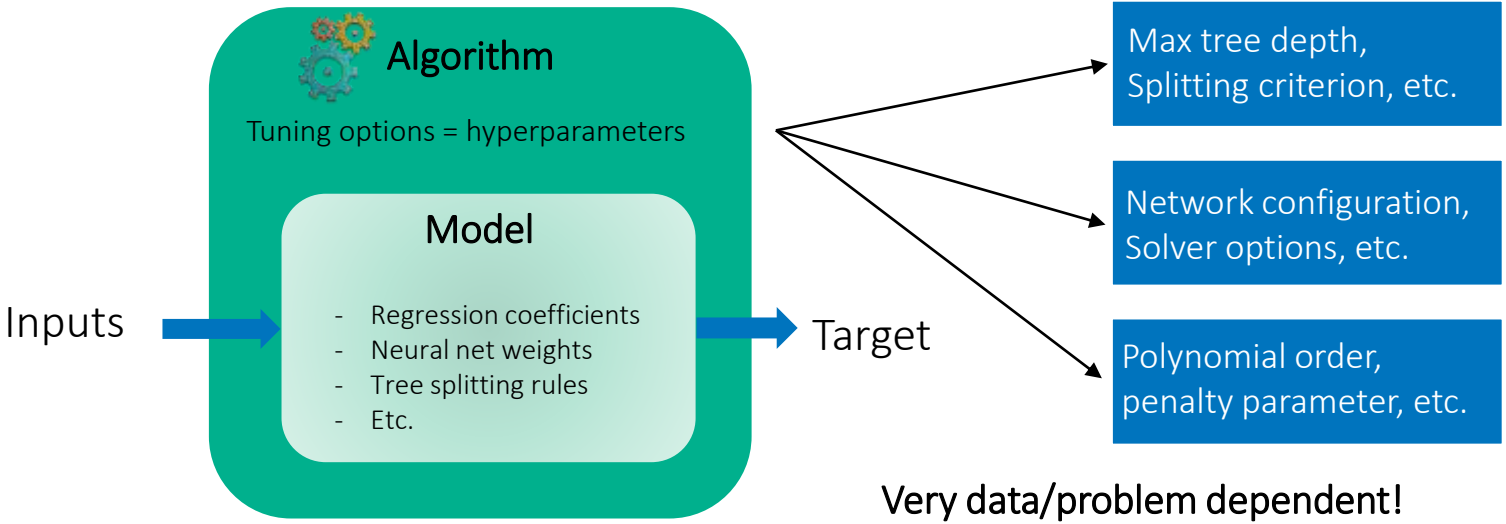
Pipelines

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



Very data/problem dependent!

SAS Programming Interface - AutoTuning Demo

SAS Visual Analytics & SAS Studio

Resources

Where can I learn more?

SAS Viya - Ask the Expert

http://support.sas.com/training/askexpert.html

SAS Viya

Accessing and Manipulating Data in SAS Viya

You may have heard of SAS Viya, a modernization of the SAS platform, but you might not be familiar with how to access and manage your data in SAS Viya.

Register for live or on-demand event

Leveraging SAS Viya to Improve Base SAS Processes New Topic!

The SAS Platform is greatly enhanced by SAS Viya. In this session, the presenter will share his experiences in leveraging SAS Viya to improve Base SAS processes. Explore technical insights and implementation details to get your code CAS-enabled, achieving faster results. These are illustrated by real-life case studies.

Register for live or on-demand event

SAS Optimization on SAS Viya: Getting Started New Topic!

Every time we use our GPS system, receive items purchased online or see food, medicine and clothing that are in stock, we benefit from optimization algorithms. In this presentation, you will get coding examples for linear programming, mixed integer linear programming, nonlinear programming and network optimization.

Register for live or on-demand event

SAS Visual Analytics 8.2 on SAS Viya: Getting Started

Learn about the new SAS Visual Analytics architecture on SAS Viya. In Visual Analytics on SAS Viya, data discovery and reporting can be accomplished through a single easy-to-use interface. Join us as we explore in-memory data and create interactive reports and dashboards.

Register for live or on-demand event

SAS Visual Data Mining and Machine Learning on SAS Viya: Getting Started

SAS Visual Data Mining and Machine Learning is designed for the data scientist, statistician and advanced business analyst. Whether you want to program or point and click it provides innovative algorithms and fast, in-memory processing. This session is for the new user. And covers what it is, the value provided and an overview of the capabilities.

Register for live or on-demand event

SAS Visual Forecasting on SAS Viya: Getting Started New Topic!

Join this webinar for an introduction to SAS Visual Forecasting on SAS Viya. Learn what it is, how it supports high-performance automated forecasting and how to start using it. This presentation offers learning opportunities for all.

Register for live or on-demand event

SAS Visual Statistics on SAS Viya: Getting Started New Topic!

SAS Visual Statistics is designed for statisticians and business analysts. You can quickly create predictive models through the visual point-and-click interface or through programming using SAS Studio or other programming interfaces. This session covers the what, why, and how for using SAS Visual Statistics 8.2.

Register for live or on-demand event

The New Languages of SAS 9.4 and SAS Viya: A SAS Programmer's Primer

Did you know there are many more ways to program against SAS 9.4 and SAS Viya products than just PROCs? This session will describe new language interfaces to SAS in SAS programmer terms and how these compare to traditional SAS 9.4 counterparts.

Register for live or on-demand event

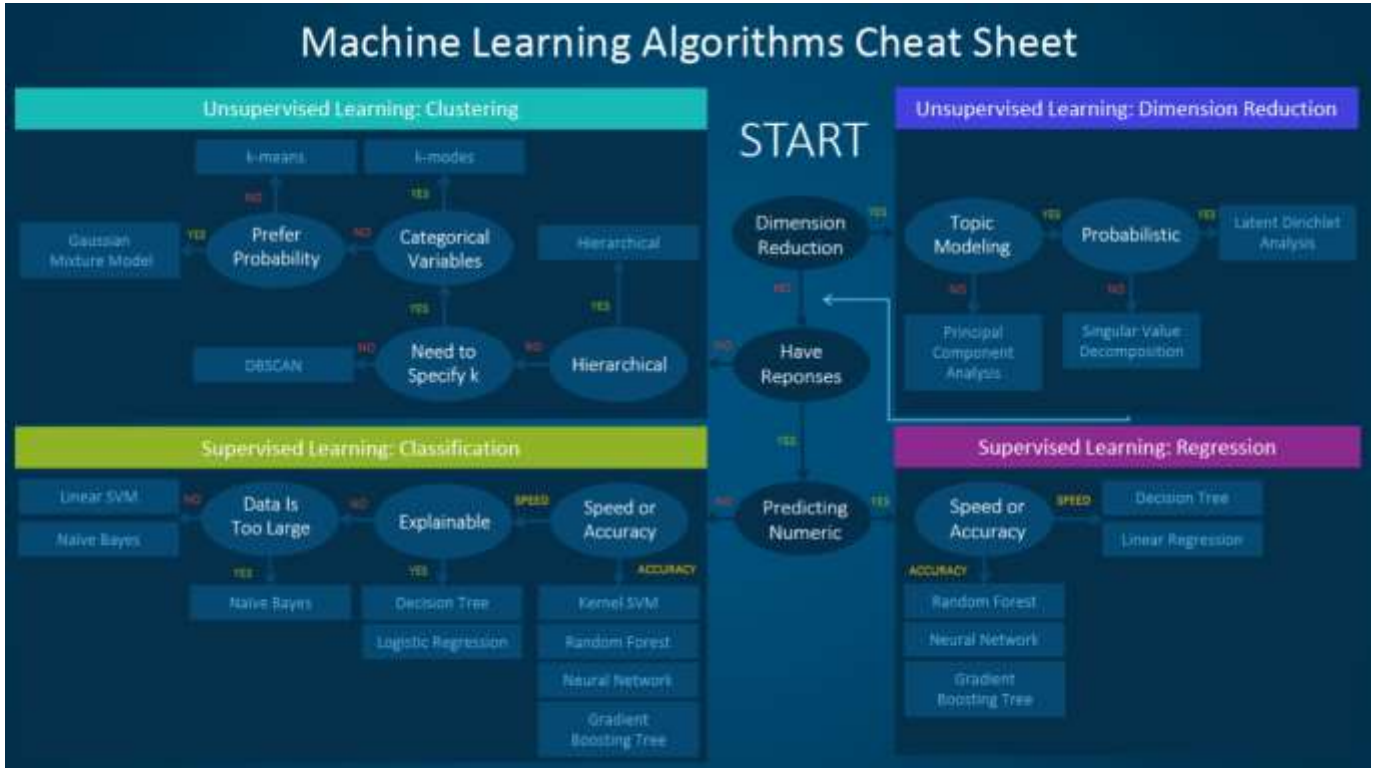
What is a CASLIB and How Do I Use Them

SAS Viya is the next big thing, but before you can get started you need to know what a CASLIB is and how it can help you get access to your data with Cloud Analytic Services.

Register for live or on-demand event

Model Selection

Machine Learning Algorithms Cheat Sheet



Access Here

Recommended Resources

An Overview of SAS® Visual Data Mining and Machine Learning on SAS® Viya
https://support.sas.com/resources/papers/proceedings17/SAS1492-2017.pdf

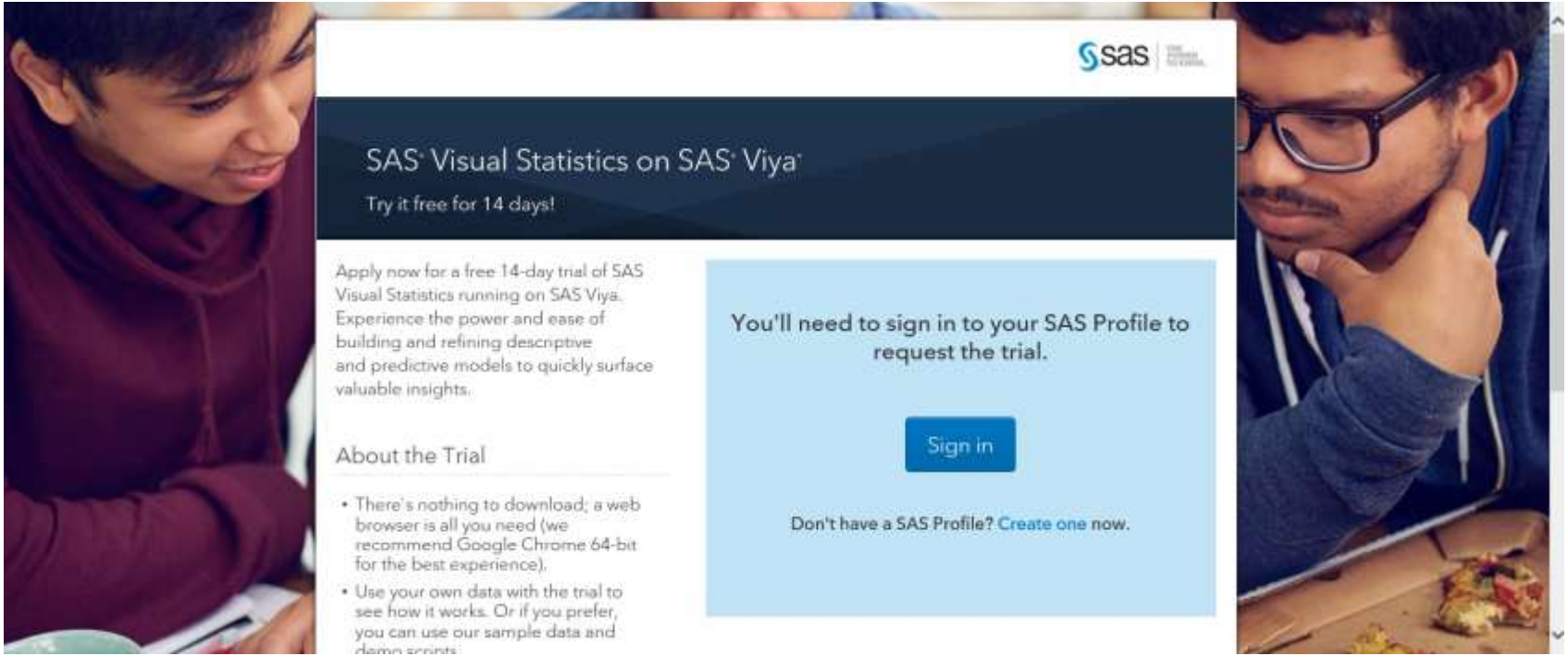
Video - Automated Machine Learning at Scale
http://www.sas.com/en_us/webinars/automated-machine-learning-scale.html

Machine learning - what it is and why it matters (reading)
http://www.sas.com/en_us/insights/analytics/machine-learning.html

Live web and classroom training - Big Data, Data Mining, and Machine Learning
Big Data course

SAS® Visual Statistics

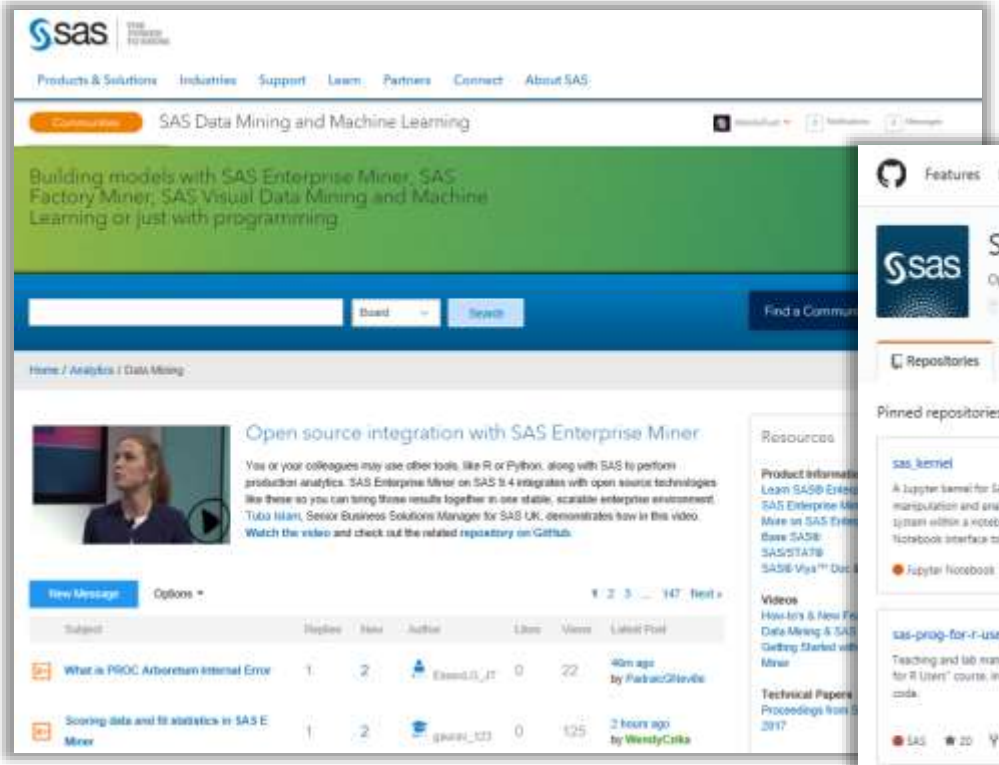
Try it before you buy!

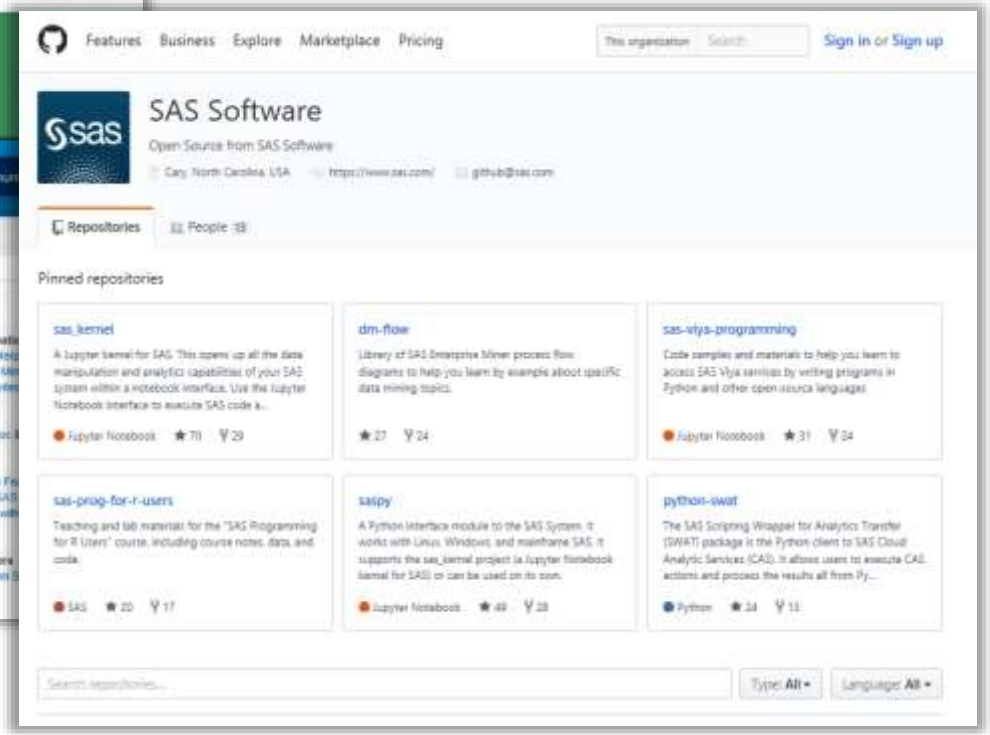


SAS® Visual Data Mining and Machine Learning

Try it before you buy!

Communities





Questions?

Thank you for your time and attention!

Connect with me:
LinkedIn: https://www.linkedin.com/in/melodierush
Twitter: @Melodie_Rush

sas.com

THE POWER TO KNOW.