VDMML Enablement Session

Data Science Jam Sessions

May 3rd, 2018

Joline Jammaers, Pre-Sales Analytical Consultant
Véronique Van Vlasselaer, Pre-Sales Analytical Consultant

Your presenters

Véronique Van Vlasselaer

Pre-Sales Analytical Consultant
Veronique.Van.Vlasselaer@sas.com

Joline Jammaers

Pre-Sales Analytical Consultant
Joline.Jammaers@sas.com

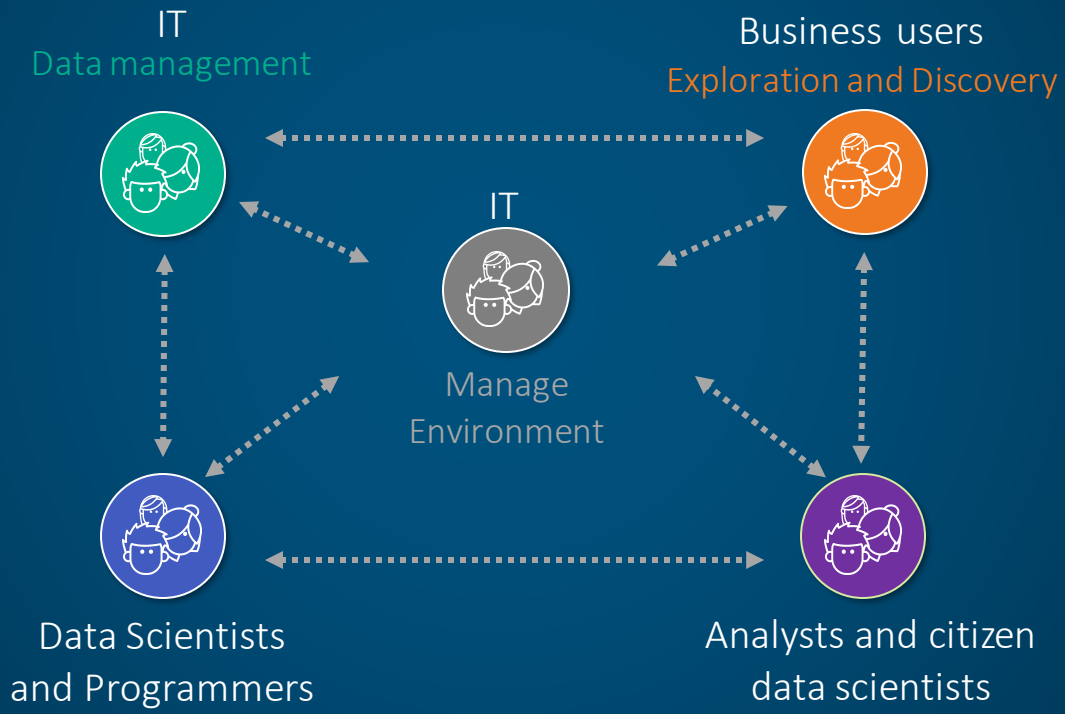
Overview

1. Introduction to Visual Data Mining and Machine Learning (VDMML)
   - Supervised Learning
   - Unsupervised Learning
2. Analytics Lifecycle
3. A Unified Environment
   - Demo

1. Introduction to VDMML

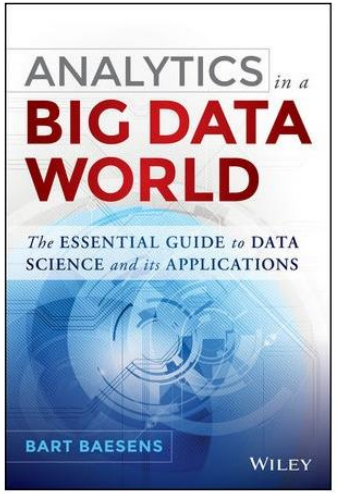
Visual Data Mining and Machine Learning

Introduction to VDMML
User profiles?



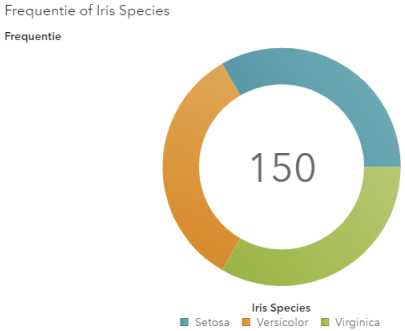
Analytics?

- Analytics objective: Extracting useful business **patterns** or mathematical decision **models** from data (Baesens, 2014).
  - Many synonyms: analytics, data mining, data science, knowledge discovery…
- **Descriptive analytics** or unsupervised learning
  - No target variable is available
  - Example: Which customers are behaving similarly?
- **Predictive analytics** or supervised learning
  - Target variable is available
  - Example: Which transactions are likely to be fraudulent?



Analytics?

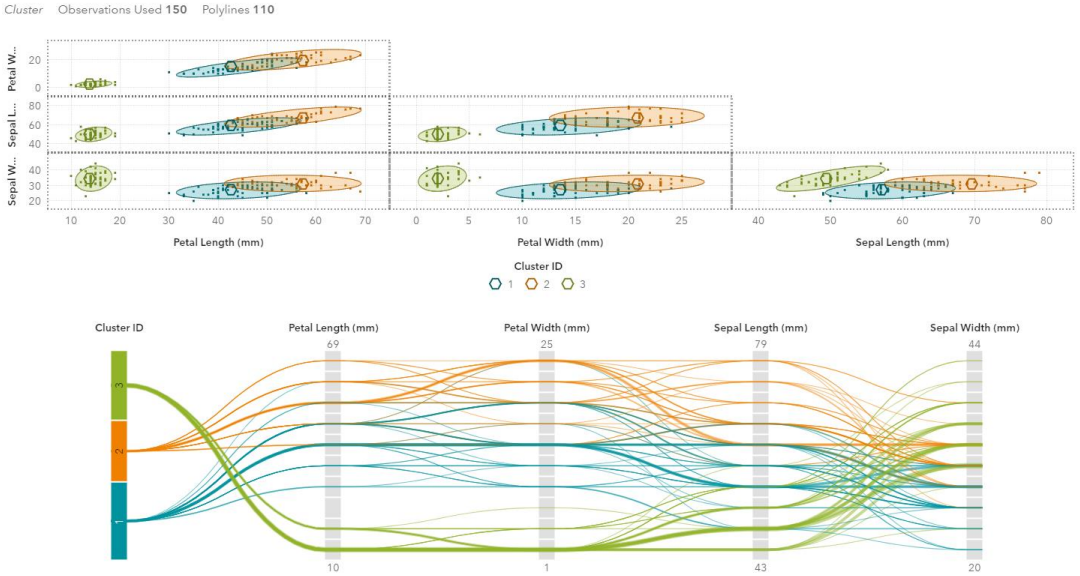
- Descriptive analytics or unsupervised learning
  - No target variable is available
  - Example: Which customers are behaving similarly?
  - Example: Which flowers are likely to be similar?



Petal Length (mm)	Petal Width (mm)	Sepal Length (mm)	Sepal Width (mm)	Iris Species
14	2	46	32	Setosa
14	2	55	42	Setosa
16	2	48	34	Setosa
15	2	46	31	Setosa
15	1	52	41	Setosa
14	3	46	34	Setosa
16	4	50	34	Setosa
14	2	49	30	Setosa
15	2	54	37	Setosa
12	2	50	32	Setosa

Analytics?

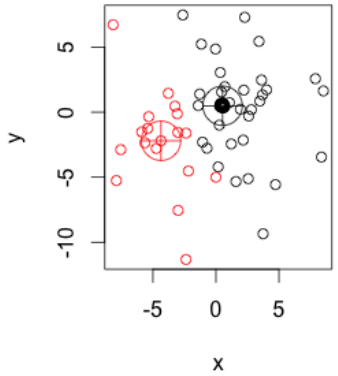
- Descriptive analytics or unsupervised learning
  - No target variable is available
  - Example: Which customers are behaving similarly?
  - Example: Which flowers are likely to be similar?

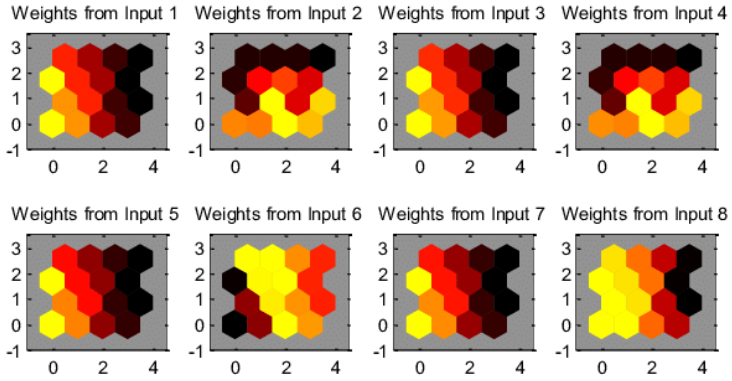


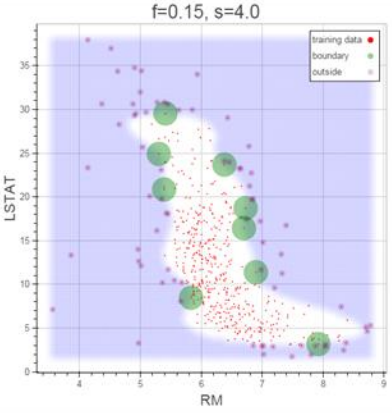
Quiz

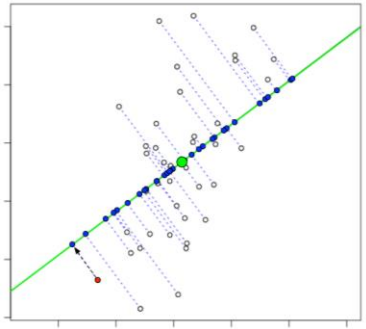
Unsupervised Learning Algorithms

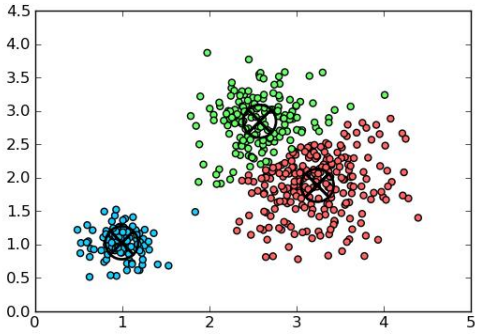
k-means clustering – k-mode clustering – Kohonen Maps – One-Class SVM – Principal Component Analysis











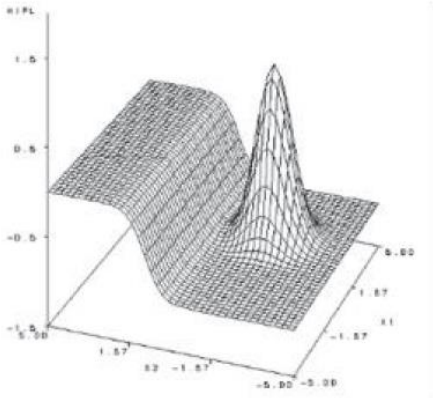
Analytics?

- Predictive analytics or supervised learning
  - Target variable is available
  - Example: Which transactions are likely to be fraudulent?
  - Example: Given a new flower, to which iris family does it belong?

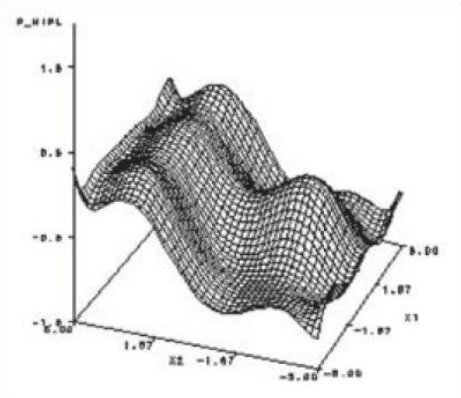
Quiz

Supervised Learning Algorithms

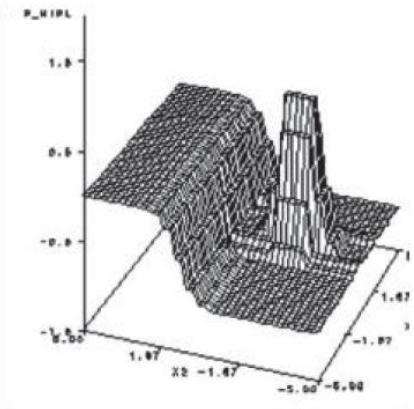
Sample data



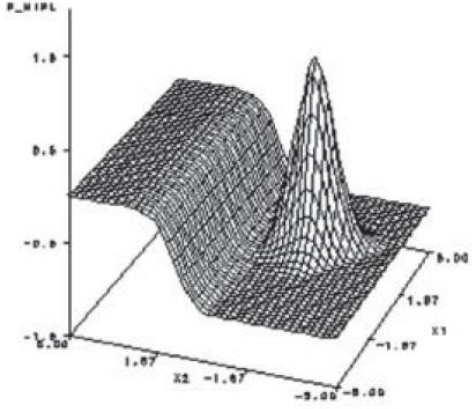
Model A



Model B



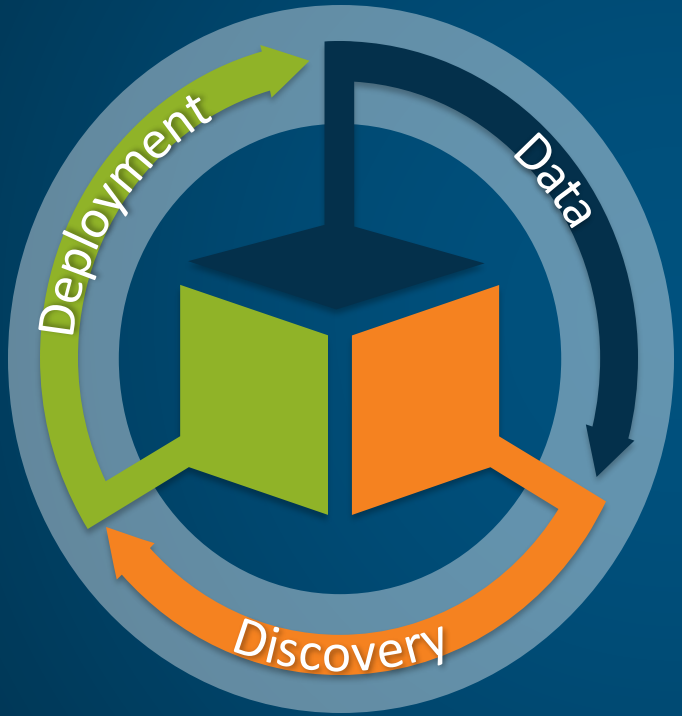
Model C



2. The Analytics Lifecycle

From data to deployment

The Analytics Lifecycle



DATA

VISUALIZATION

MACHINE LEARNING

TEXT ANALYTICS

FORECASTING

OPTIMIZATION

REPORTING

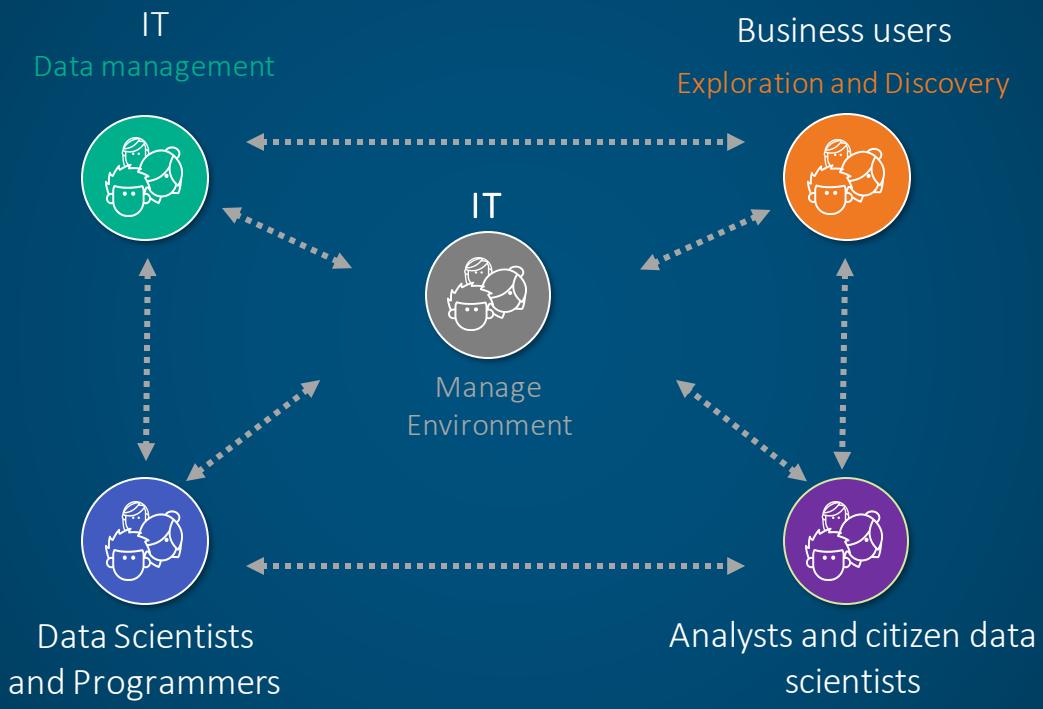
MODEL MANAGEMENT

DECISION MANAGEMENT

3. A Unified Environment

Demo

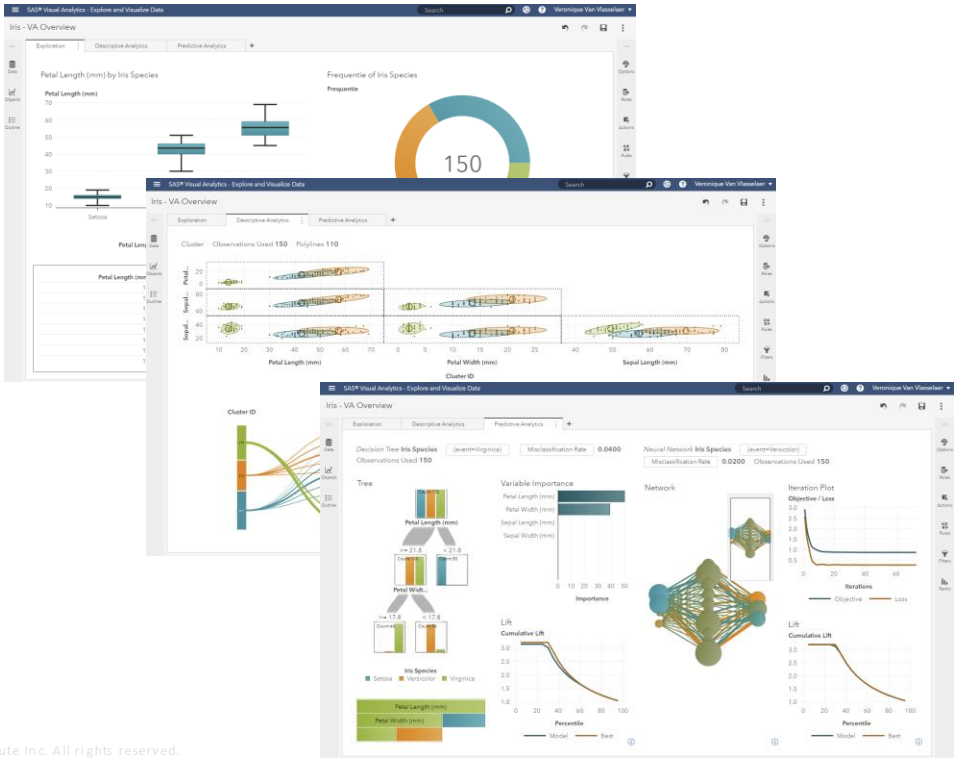
One Environment for All Users



One Environment for All Users

Business users

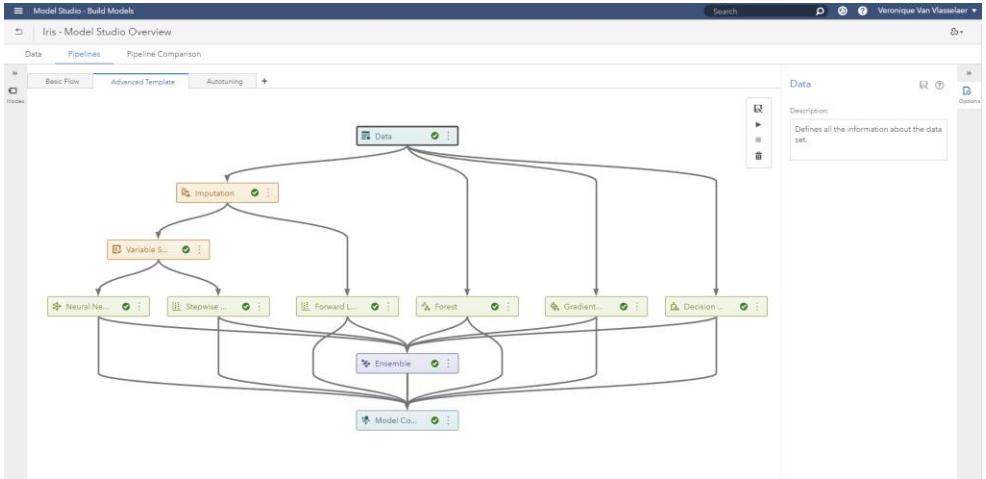
- Visual Analytics - Approachable analytics
  - Scalable to all your data
  - Interactive and Visual Reporting
  - Analytical Data Exploration
  - Analytical Modeling



One Environment for All Users

Analysts and citizen data scientists

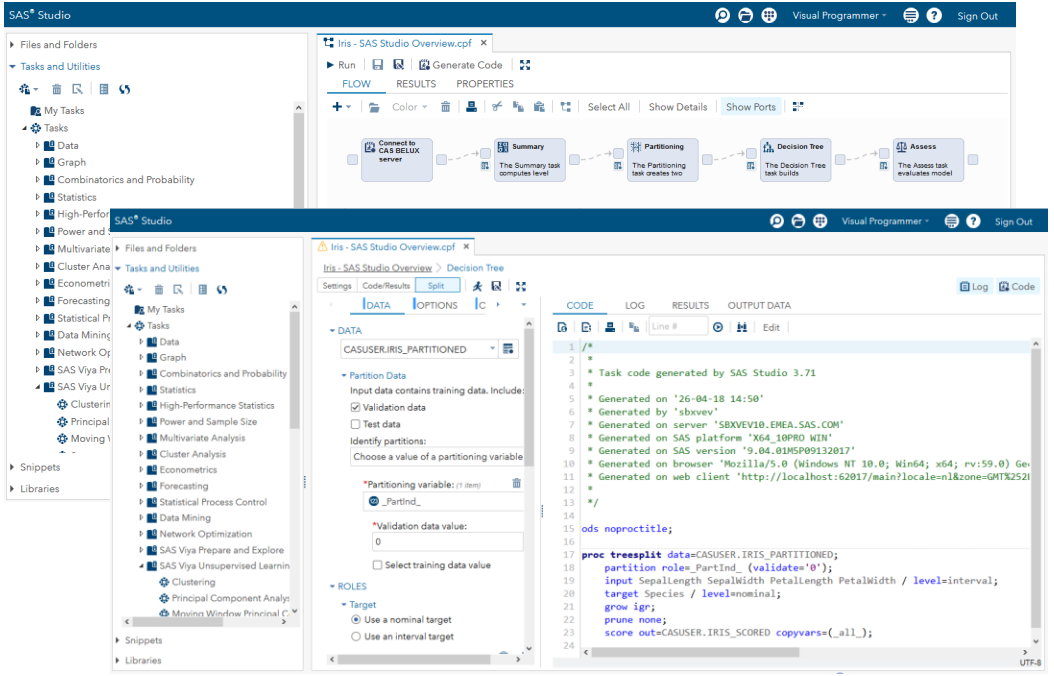
- Model Studio - Building DMML models
  - Pipeline of activities
  - Drag and drop and access to code
  - Node are run asynchronously
  - Reproducibility
  - SAS best practice toolkit
    - Templates
  - Focus on deployment



One Environment for All Users

Data Scientists and Programmers

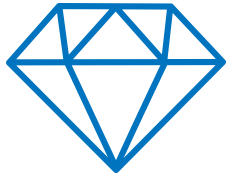
- SAS Studio – Coding environment
- Pre-built data science & machine learning tasks
- Ability to share code snippets among other data scientists



SAS Openness

1 Analytics Engine
Where the model is being built

2 Interface
Application or programming language to drive engine

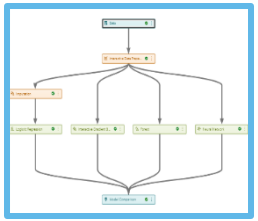


SAS® Model

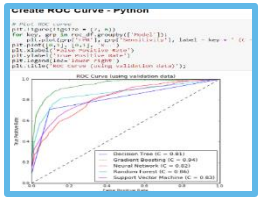
Open source Model

SAS® Open API (R, Python, etc.)

SAS® Open Source code node



SAS® Model Studio



R or Python coding environment

Loving R Studio?

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Iris - R Viya Overview.R

```
# EMPTY ENVIRONMENT
rm(list=ls())

# LOAD SWAT PACKAGES
library(swat)
library(ggplot2)

########################
### CONNECTION TO CAS ###
########################

# Connect to CAS (Typically port 8777 for REST API and port 5570 for binary connection)
conn <- CAS("                                    ", 8777, protocol='http')

# Load Iris data set into in-memory table
# iris.ct <- cas.read.sas7bdat(conn, "C:/Users/sbxvev/OneDrive - SAS/_PROJECTS_/_UTILITY_/Iris.sas7bdat",
#                              casOut=list(name="IRIS_sbxvev", caslib="casuser", replace=TRUE))
iris.ct <- cas.read.sas7bdat(conn, "C:/Users/sbxvev/OneDrive - SAS/_PROJECTS_/_UTILITY_/Iris.sas7bdat",
                             casOut=list(name="IRIS_orig", replace=TRUE))

# View Data
head(iris.ct)
names(iris.ct)
unique(iris.ct$Species)
unique(iris.ct$SepalLength)
summary(iris.ct)

####################
### SAMPLING DATA ###
####################

# Load the sampling actionset
loadActionSet(conn, 'sampling')
```

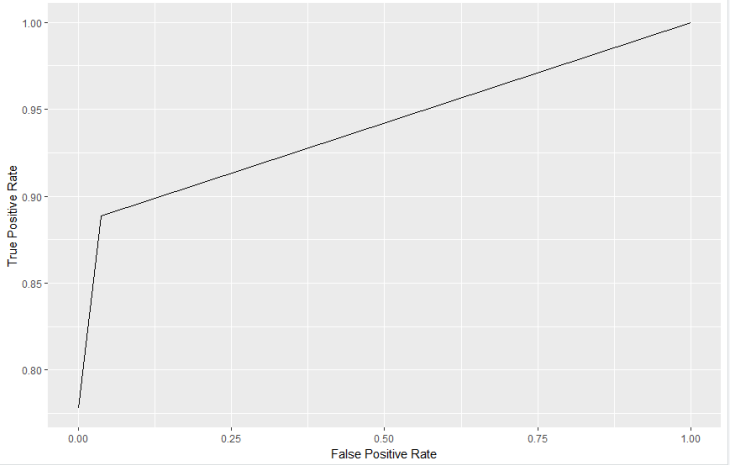
Environment | History | Connections

Global Environment

Data	
inputs	List of 4
iris.ct	Formal class CASTable
results	List of 2
train	Formal class CASTable
valid	Formal class CASTable
Values	
cas.accessControl.assumeR…	function (object, ...)
cas.accessControl.checkIn…	function (object, ...)
cas.accessControl.checkOu…	function (object, ...)
cas.accessControl.commitT…	function (object, ...)
cas.accessControl.complet…	function (object, ...)
cas.accessControl.createB…	function (object, ...)

Files | Plots | Packages | Help | Viewer

Zoom | Export | Publish



Console | Terminal

```
+                               table="dt_scored",
+                               inputs   = c('_DT_P_Versicolor'),
+                               response = target,
+                               event    = 'Versicolor'
+ )
>
> # View the 50% cutoff rate
> results$ROCInfo[results$ROCInfo$CutOff==0.5,
+                 c('CutOff', 'TP','FP','FN','TN')]
   CutOff TP FP FN TN
51    0.5 14  0  4 27
>
> # Plot the ROC curve
> ggplot(data = results$ROCInfo[c('FPR', 'Sensitivity')],
+        aes(x = as.numeric(FPR), y = as.numeric(Sensitivity))) + geom_line() +
+   labs(x = 'False Positive Rate', y = 'True Positive Rate')
> ########################
> ### CLOSE CONNECTION ###
> ########################
>
> cas.session.endSession(conn)
list()
>
```

Loving Python?

jupyter Giraffe_vs_Dolphine_Tutorial Last Checkpoint: 27-02-2018 (unsaved changes)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run | Markdown

In [1]:
```
%matplotlib inline
from swat import * # needed to start a CAS server.
sess = CAS('viya3-3-cc.ps-cp.sashq-d.openstack.sas.com','8777','user','pswrd', protocol='http')
```

In [2]:
```
import sys
sys.path.append("C:Users\sbxvev\Python Packages\dlpy")
from dlpy.images import ImageTable
```

```
my_images = ImageTable.load_files(sess, path='/opt/demodata/images', casout=None)
```

In [4]:
```
#my_images?
```

In [5]:
```
my_images.head()
```

Out[5]:

Selected Rows from Table IMAGEDATA_3TURHS

	image	_size_	_path_	_label_	_type_	_id_
0	b'\xff\xd8\xff\xe0\x00\x10JFIF\x00\x01\x01\x01...	539941	/opt/demodata/images/Dolphin/dolphin_10937.jpg	Dolphin	jpg	5
1	b'\xff\xd8\xff\xe0\x00\x10JFIF\x00\x01\x01\x01...	37426	/opt/demodata/images/Dolphin/dolphin_10924.jpg	Dolphin	jpg	7
2	b'\xff\xd8\xff\xe0\x00\x10JFIF\x00\x01\x01\x01...	567420	/opt/demodata/images/Dolphin/dolphin_10923.jpg	Dolphin	jpg	8
3	b'\xff\xd8\xff\xe0\x00\x10JFIF\x00\x01\x01\x01...	644416	/opt/demodata/images/Dolphin/dolphin_10919.jpg	Dolphin	jpg	10
4	b'\xff\xd8\xff\xe0\x00\x10JFIF\x00\x01\x01\x00...	36617	/opt/demodata/images/Dolphin/dolphin_10918.jpg	Dolphin	jpg	11

In [6]:
```
my_images.summary()
```

Out[6]:

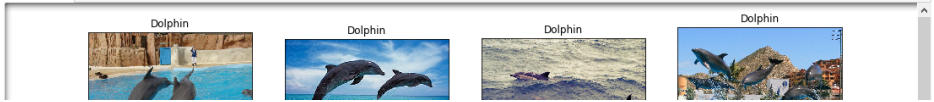
§ Summary

Descriptive Statistics for IMAGEDATA_3TURHS

	Column	Min	Max	N	NMiss	Mean	Sum	Std	StdErr	Var	USS	CSS	CV
0	_size_	9056.0	1075587.0	409.0	0.0	562709.784841	230148302.0	192272.744517	9507.275336	3.696881e+10	1.445900e+14	1.508327e+13	34.16907
1	_id_	1.0	409.0	409.0	0.0	205.000000	83845.0	118.212379	5.845226	1.397417e+04	2.288968e+07	5.701460e+06	57.66457

elapsed 0.0207s · user 0.00749s · sys 0.0232s · mem 6.31MB

In [7]:
```
my_images.show(nimages=8,ncol=4)
```



Free SAS Viya Trial
Environment

- SAS® Visual Data Mining and Machine Learning on SAS® Viya® Try it for Free *(Google: sas free trials)*

- Pure browser based, no tools to download

- Environment includes:
  - Example data (possible to load own data)
  - Introduction scripts
  - Links to tutorial videos
  - Possibility to invite up to 4 colleagues

- 14 days access, extendable with 14 days

Questions? Feedback? Comments?

veronique.van.vlasselaer@sas.com

joline.jammaers@sas.com