



SAS® Machine Learning for Pharmaceuticals

PhUSE 2017

Aneshan Ramaloo

Advanced Analytics Practise Lead - Africa



41

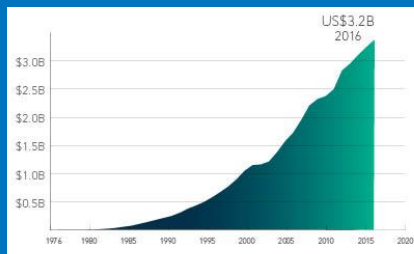
Years of



BUSINESS
ANALYTICS

#1

World's
LARGEST
privately held
software company



2016 Revenue

US \$ 3.2 b

Continuous Revenue
Growth since 1976

13,919



SAS employees worldwide

TOP

World's Best
Multinational
Workplaces list

GREAT
PLACE
TO
WORK®



94%

Annual customer
retention rate

94

of the top
100
companies

on the 2016

FORTUNE®

GLOBAL **500®** LIST



26%

Annual reinvestment in
R&D

83,000+

Customer sites in 149 countries

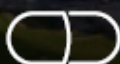


100%

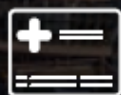
of Fortune 500
companies in these
areas rely on SAS®:



Commercial
banking



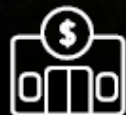
Pharmaceuticals



Health
Insurance



Electricity



Retail
Banking



Telco and
service providers



Hotels



Food
services

SAS in Life Sciences

Current Focus Areas

R&D
**Computational Clinical
Trial Environment**

Real World Evidence
Big Data analytics

Regulatory Compliance
IDMP and Clinical Trial
Data Transparency

Commercial Analytics
Price Optimization,
Simulation, Marketing
Campaigns

Clinical Trial
Operations Analytics
Risk Based Monitoring,
Clinical Data Surveillance

IOT
Analyze data from
devices/wearables

Unstructured Data
Analytics
Patient reported
outcomes, Safety Signal
Detection

Supply Chain
Forecasting,
Production Quality

SAS in Life Sciences

Partnering for better outcome (Data4Good)

- **Cancer Research** - The **Project Data Sphere platform** provides one place where the research community can broadly share, integrate and analyze historical, patient-level, comparator arm data from academic, government, and industry phase III cancer clinical trials. Advanced SAS® tools are freely available to all users.



- SAS became in 2016 an **official partner in research for IMI/EFPIA** demonstrating the **continuous level of investment in Life Sciences industry**. IMI2 Activities in 2017 will address the insufficient level of medication compliance and adherence by creating a multi-stakeholder network which will establish common processes, standards and guidelines for a digital patient platform with approved medicines' information and map trusted sources and needs for additional information.

Current involvement in Big Data 4 Better Outcome initiative: <http://bd4bo.eu/>



- In 2016, **Duke Clinical Research Institute** has agreed to **share the cardiovascular data set** part of their Duke Databank for Cardiovascular Disease (DDCD) **with SAS for the data transparency** initiative.



SAS® Visual Data Mining And Machine Learning

Machine Learning

- Automate**
 - Provide automation to the model building process by minimizing human intervention
- Customize**
 - Build powerful models using SAS's state-of-the-art algorithms in conjunction with open source tools
- Speed**
 - Fast response time for sophisticated analytics applied to data of any size or complexity

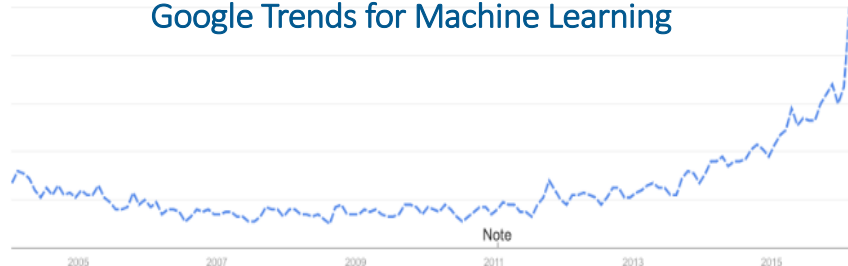


Machine Learning

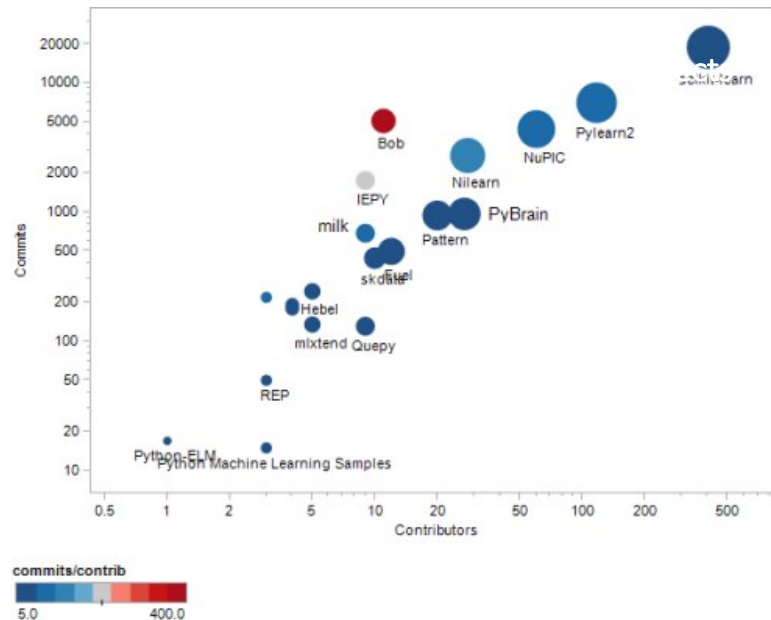


Machine Learning Trends

Google Trends for Machine Learning

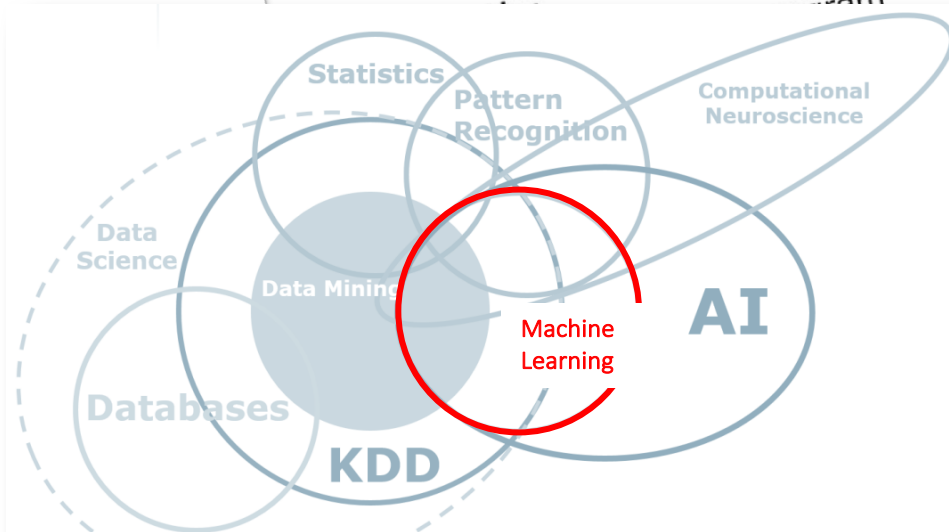


Machine Learning jobs



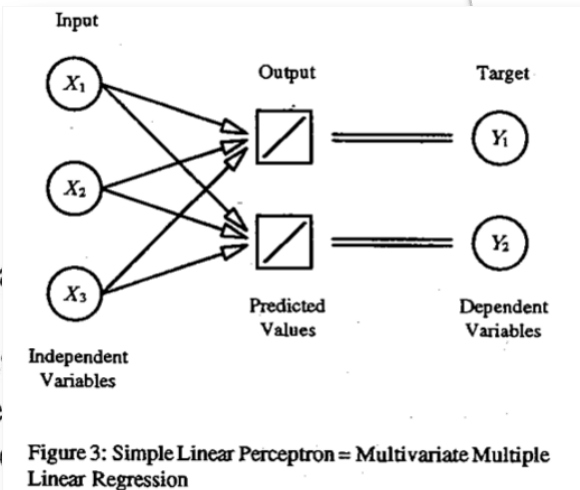
PROC DISCRIM (K-nearest-neighbor discriminant analysis)

– James Goodnight, SAS founder and CEO, **1979**



SAS Data Mining Primer Course

SAS Institute, 1998



Neural Networks and Statistical Models,

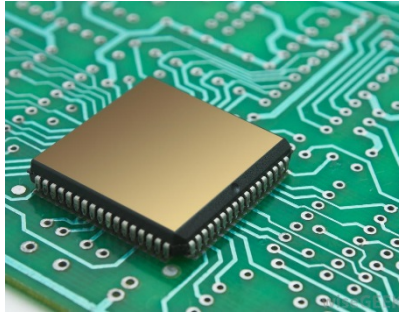
SAS Institute, 1994

Machine learning

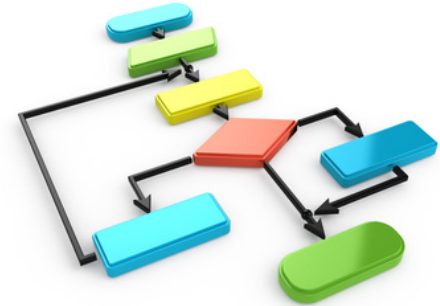
WHY IS IT SO IMPORTANT NOW?



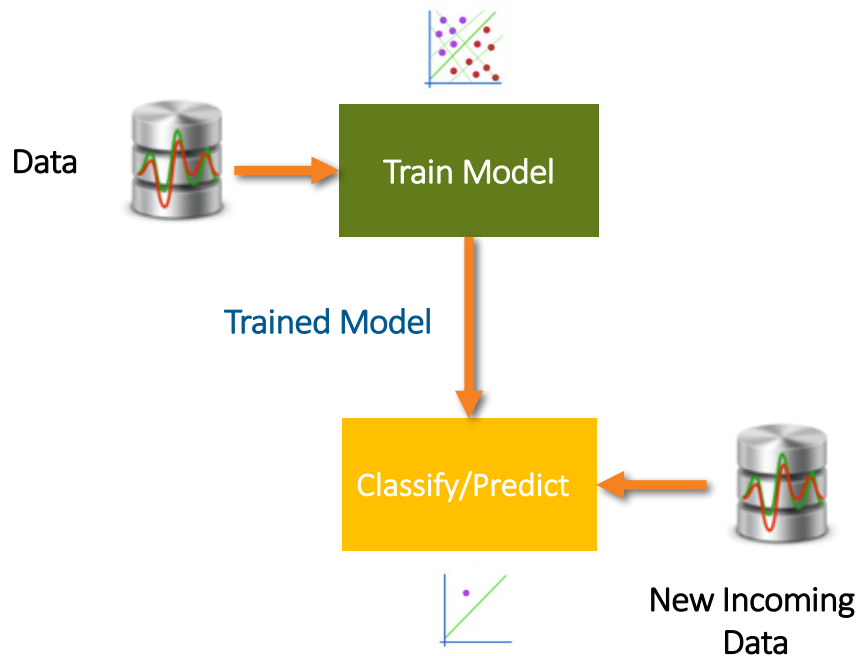
Data



Computing
Power



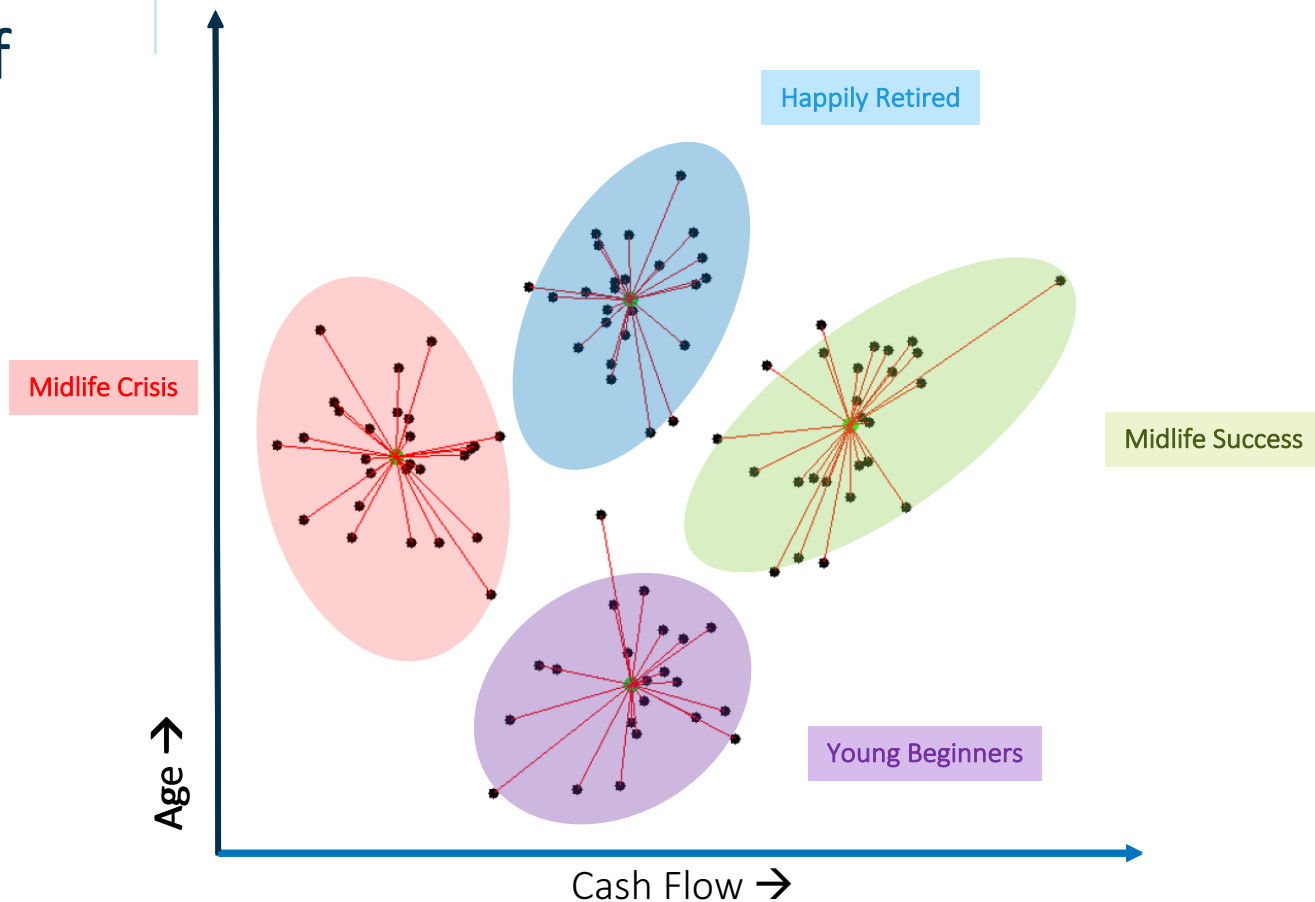
Algorithms

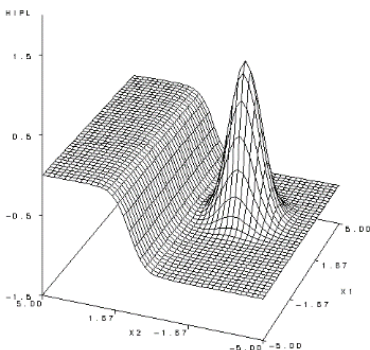


- Machine learning **automates** the model building process by
 - learning **iteratively** from data to identify **patterns** and
 - **predicting** future results
- with minimal human intervention
- Machine Learning emphasizes more on **predictive accuracy** rather than interpretability of the models

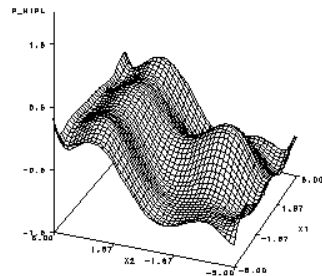
The iterative nature of Machine learning

K-MEANS CLUSTERING (GROUP SIMILAR OBSERVATIONS)

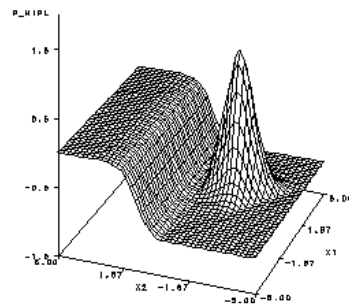




Sample data



Traditional regression



Neural Network

- Machine learning automates the model building process by
 - learning iteratively from data to identify patterns and
 - predicting future results
- with minimal human intervention
- **Machine Learning emphasizes more on predictive accuracy rather than interpretability of the models**



Welcome to the
analytics economy



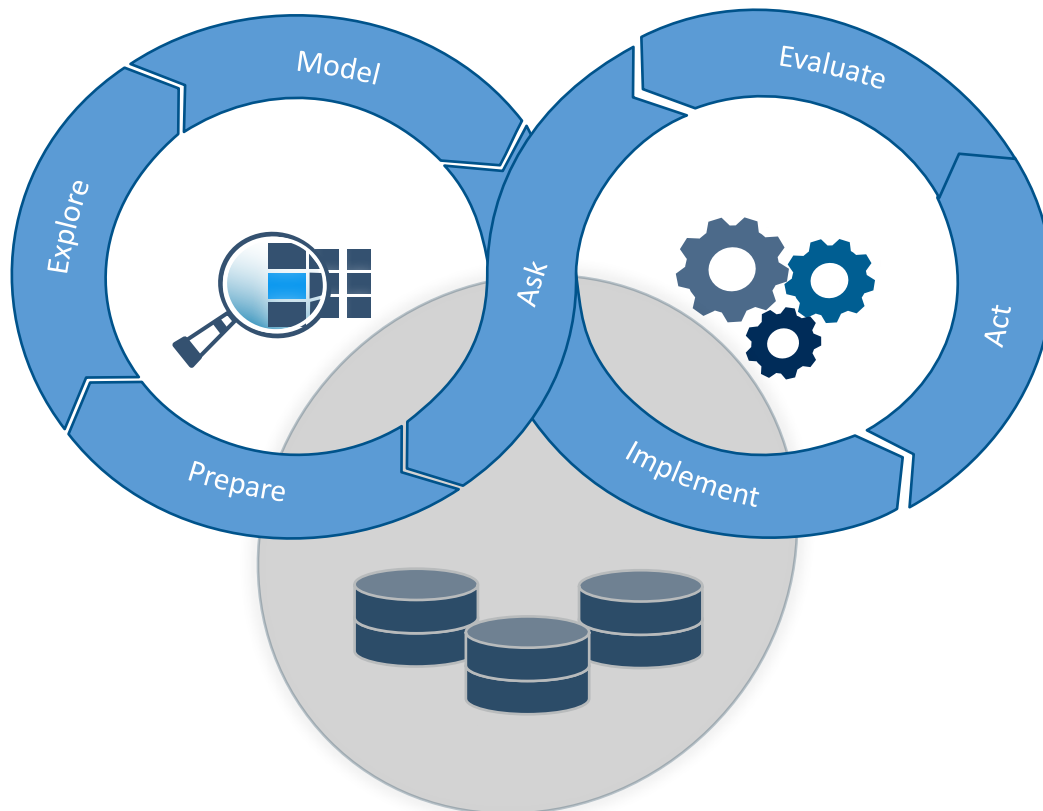
Why SAS Visual Data Mining and Machine Learning

“I want an Analytics Platform that supports **diverse** skillsets, processes at **scale**, and empowers **trust...**”

- Analytic leaders

Linking **Data** to **Discovery** to **Deployment**

For **All Your Data** at **High Speed** in one platform you can **Trust**





SAS Viya



Analytic Data Preparation

SAS® Data Studio

Search

Jonathan Wexler

Wine Demo Plan JW

Transforms

Add Transform

Column Transforms

- Change case
- Convert column
- Rename
- Remove
- Split
- Trim whitespace

Custom Transforms

- Calculated column
- Code

Data Quality Transforms

- Casing
- Parsing
- Field extraction
- Gender analysis
- Identification analysis
- Match codes
- Standardize

Multi-input Transforms

- Append
- Join

Row Transforms

- Filter
- Transpose

Filter - Step 3 of 3

1 Calculated Column — 2 Code — 3 Filter

Column: type Operator: In Value: white red

Browse

☒ Filter using formatted values

Run Save

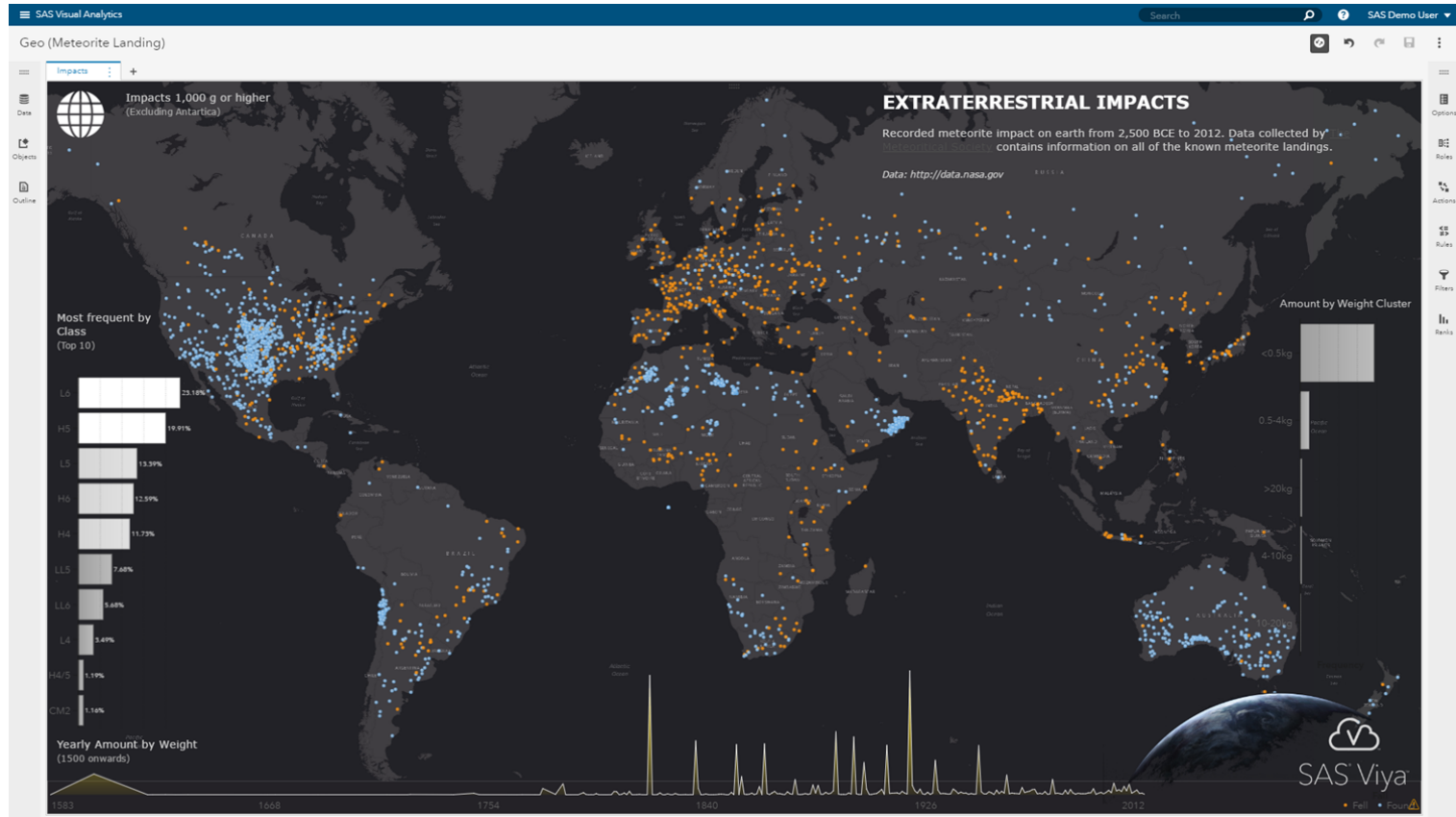
WINEQUALITY (session)

Table Profile Metadata

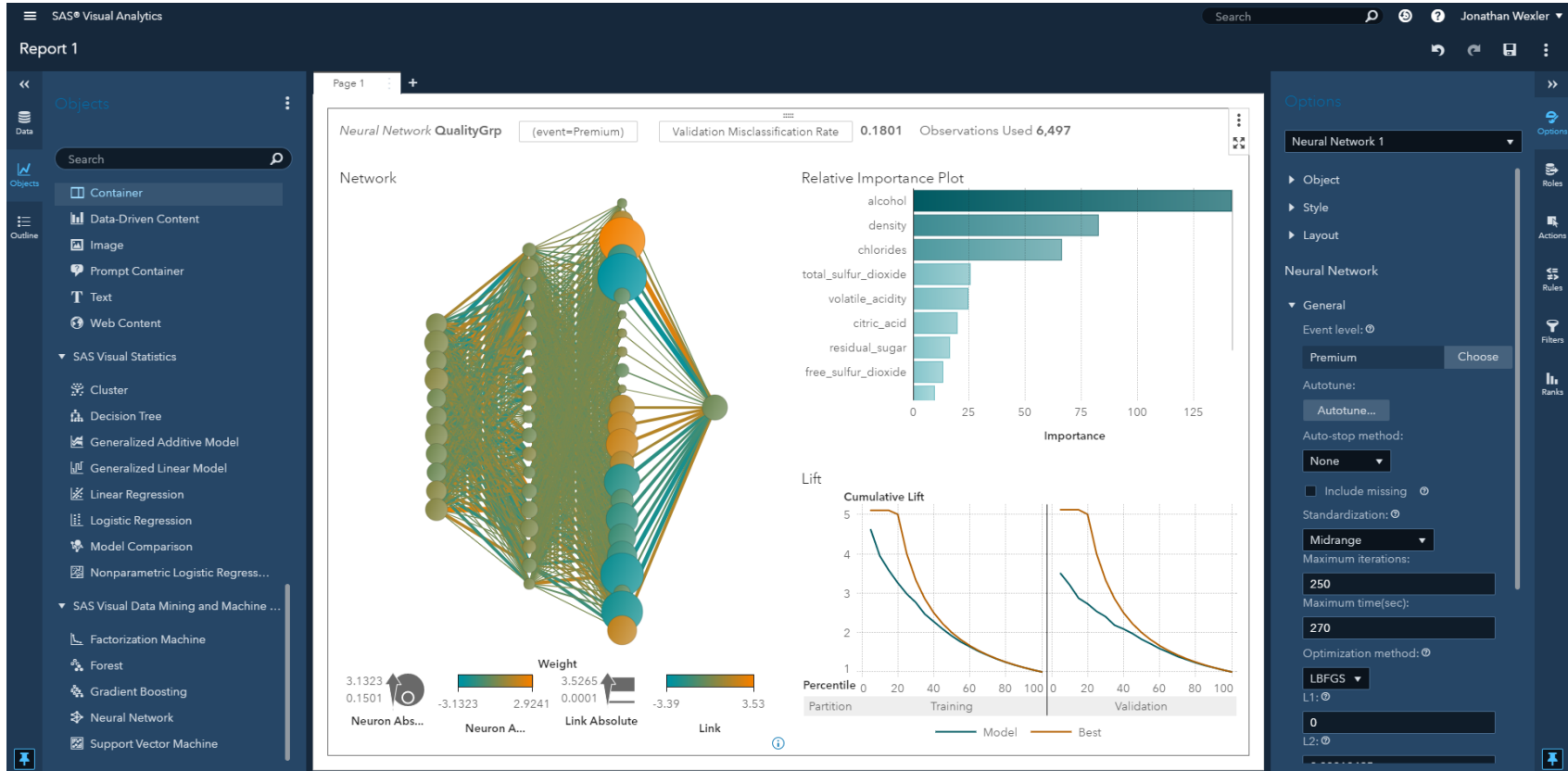
Result rows: 100

type	QualityGrfixed_acidity	lattle_acidity	citric_acid	sidual_sugar	chlorides	alfur_dioxide	alfur_dioxide	density	pH	sulphates	alcohol	quali B
white	Economy	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8
white	Economy	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5
white	Economy	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1
white	Economy	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9
white	Economy	7.2	0.23	0.32	8.5	0.058	47	186	0.9956	3.19	0.4	9.9
white	Economy	8.1	0.28	0.4	6.9	0.05	30	97	0.9951	3.26	0.44	10.1
white	Economy	6.2	0.32	0.16	7	0.045	30	136	0.9949	3.18	0.47	9.6
white	Economy	7	0.27	0.36	20.7	0.045	45	170	1.001	3	0.45	8.8
white	Economy	6.3	0.3	0.34	1.6	0.049	14	132	0.994	3.3	0.49	9.5
white	Economy	8.1	0.22	0.43	1.5	0.044	28	129	0.9938	3.22	0.45	11

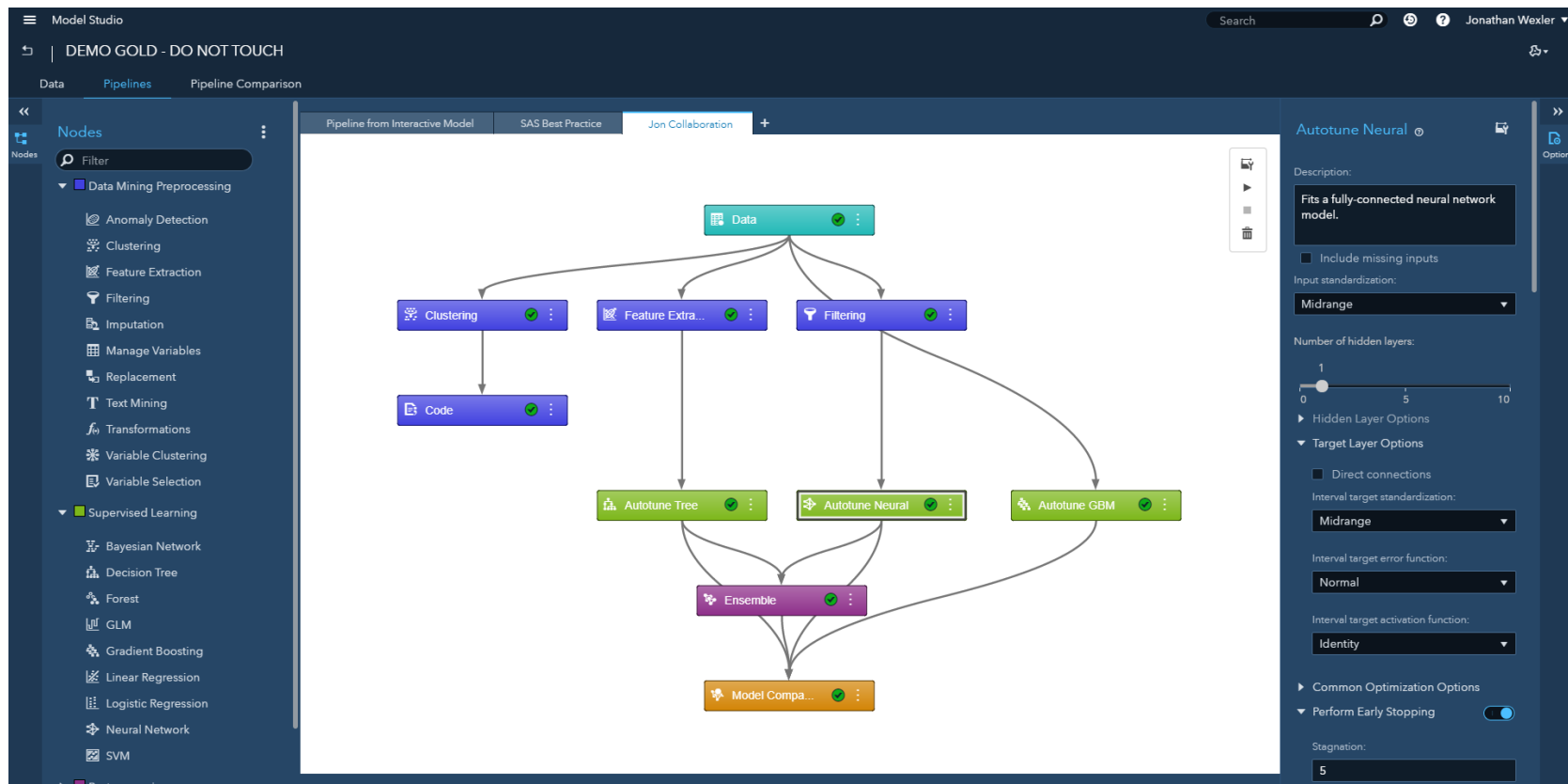
Data Visualisation



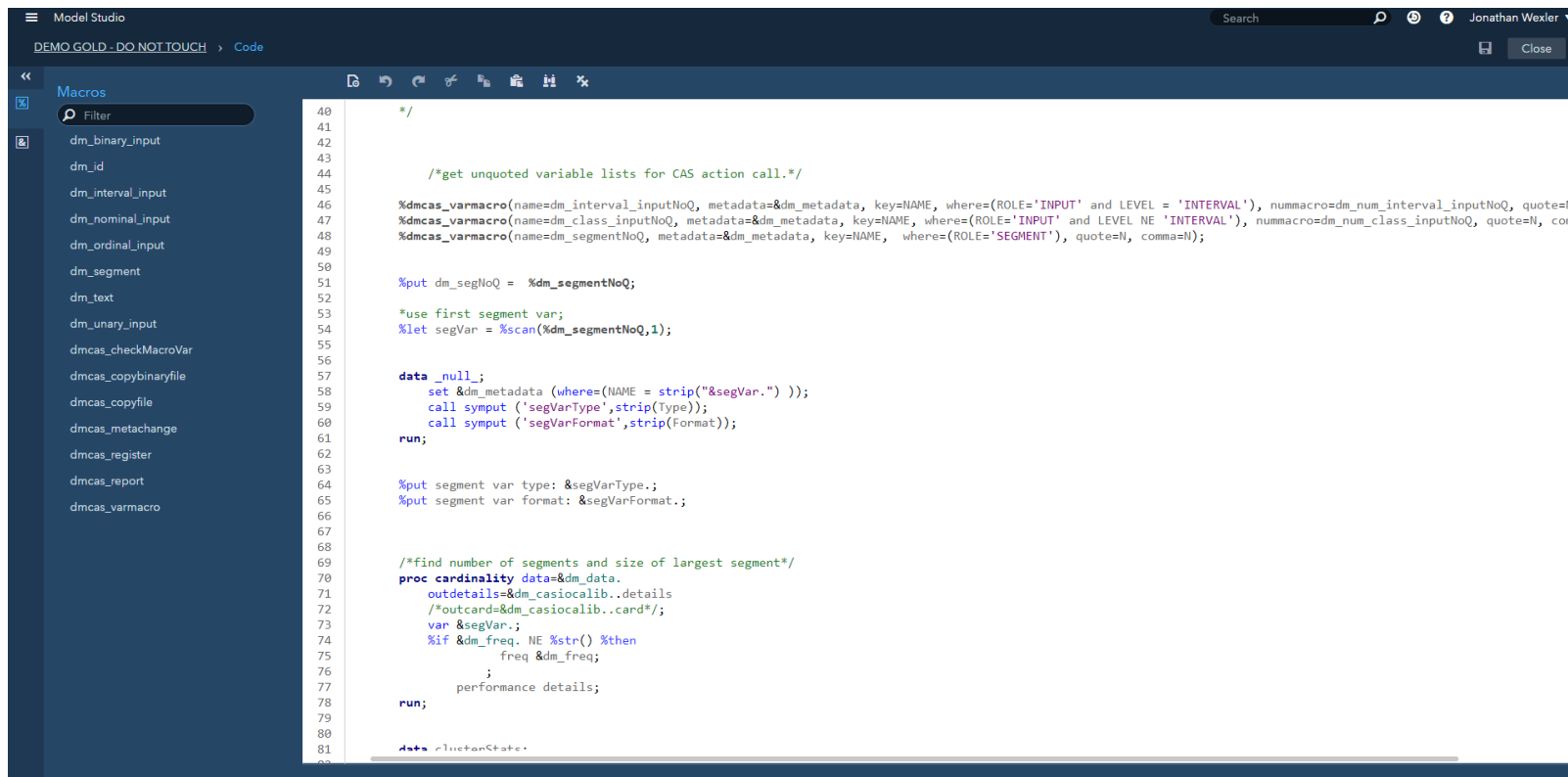
Interactive Discovery and Modeling



Modernized Pipelines and Automation



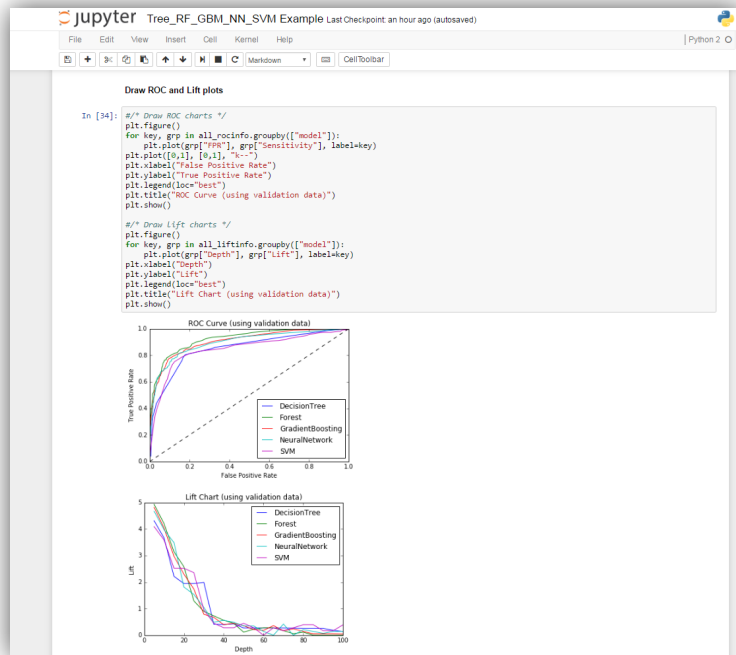
Embedded Coding



The screenshot displays the SAS Model Studio interface. On the left, a 'Macros' panel lists various macro names, including `dm_binary_input`, `dm_id`, `dm_interval_input`, `dm_nominal_input`, `dm_ordinal_input`, `dm_segment`, `dm_text`, `dm_unary_input`, `dmcas_checkMacroVar`, `dmcas_copybinaryfile`, `dmcas_copyfile`, `dmcas_metachange`, `dmcas_register`, `dmcas_report`, and `dmcas_varmacro`. The main editor area shows a SAS DATA step with embedded macro calls. The code is as follows:

```
40  */
41
42
43
44      /*get unquoted variable lists for CAS action call.*/
45
46      %dmcas_varmacro(name=dm_interval_inputNoQ, metadata=&dm_metadata, key=NAME, where=(ROLE='INPUT' and LEVEL = 'INTERVAL'), nummacro=dm_num_interval_inputNoQ, quote=|
47      %dmcas_varmacro(name=dm_class_inputNoQ, metadata=&dm_metadata, key=NAME, where=(ROLE='INPUT' and LEVEL NE 'INTERVAL'), nummacro=dm_num_class_inputNoQ, quote=N, cor
48      %dmcas_varmacro(name=dm_segmentNoQ, metadata=&dm_metadata, key=NAME, where=(ROLE='SEGMENT'), quote=N, comma=N);
49
50
51      %put dm_segNoQ = %dm_segmentNoQ;
52
53      *use first segment var;
54      %let segVar = %scan(%dm_segmentNoQ,1);
55
56
57      data _null_;
58          set &dm_metadata (where=(NAME = strip("&segVar.")));
59          call symput ('segVarType',strip(Type));
60          call symput ('segVarFormat',strip(Format));
61      run;
62
63
64      %put segment var type: &segVarType.;
65      %put segment var format: &segVarFormat.;
66
67
68
69      /*find number of segments and size of largest segment*/
70      proc cardinality data=&dm_data.
71          outdetails=&dm_casiocalib..details
72          /*outcard=&dm_casiocalib..card*/;
73          var &segVar.;
74          %if &dm_freq. NE %str() %then
75              freq &dm_freq;
76          ;
77          performance details;
78      run;
79
80
81      data _null_;
```

Open Access between SAS, Python & R



Data Mining Workflow

Set Up the R Notebook for Analysis

```
# Load necessary packages
library('swat')
```

```
## SWAT 0.1.4
```

```
# Hide credentials
login <- read.csv('login.csv')
hostname <- paste(login[1,])
username <- paste(login[2,])
password <- paste(login[3,])
```

```
# Data name
indata <- "hmeq_prepped"

# Hostname, port, username, password
conn = CAS(hostname, 8777, username, password, protocol = 'http')
```

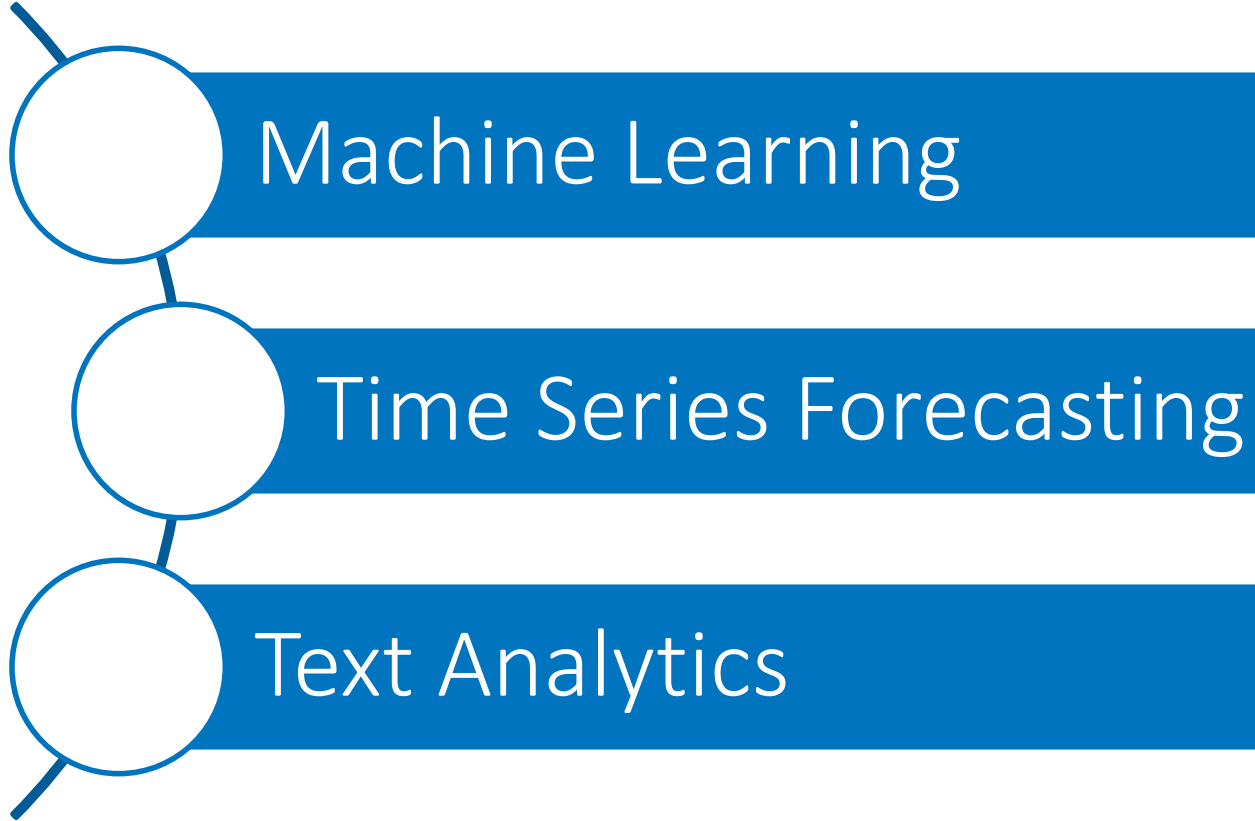
```
## Connecting to CAS and generating CAS action functions for loaded action sets...
```

```
## To generate the functions with signatures (for tab completion), add 'genActSyntax=TRUE' to your connection params.
```

```
# Change the active caslib to public
cas.sessionProp.setSessOpt(conn, caslib = 'public')
```

View Data

Different Flavours of Advanced Analytics



Full-Scale Modeling Support



Data Manipulation

- In-Memory Data Step
- Frequency / Crosstab
- Data Transpose
- Variable Binning
- Variable Cardinality Analysis
- Variable Summary
- Sampling and Partitioning
- Missing Value Imputation
- Variable Selection
- Model Assessment
- DS2
- FedSQL
- Image Processing



Statistics

- Cox Proportional Hazards
- Decision Trees
- Design Matrix
- General Additive Models
- Generalized Linear Models
- K-means and K-modes Clustering
- Linear Regression
- Logistic Regression
- Nonlinear Regression
- Ordinary Least Squares Regression
- Partial Least Squares Regression
- Pearson Correlation
- Principal Component Analysis
- Quantile Regression
- Shewhart Control Chart Analysis



Machine Learning

- Bayesian Networks
- Boolean Rules
- Factorization Machines
- Frequent Item Set Mining
- Gradient Boosting
- K Nearest Neighbor
- Market Basket Analysis
- Moving Windows PCA
- Network Analytics/Community Detection
- Neural Networks / Deep Learning
- Random Forest
- Robust PCA
- Support Vector Data Description
- Support Vector Machines
- Text Mining
- Variable Clustering

Knowledge and Insights via Analytics - Propensity Score Matching



SAS® Studio

analysis_06Feb.sas x CaseControlMatching.ctm x Sample Task x *CaseControlMatching1 x

Settings Code/Results Split

Logistic Regression Case:Control Matching Match Diagnostics INFORMATION CODE LOG RESULTS

DATA WORK.BASELINE1

ROLES

Dependent variable: depvar

Patient ID variable: ptid

Other grouping variable: Column

Effects variables: age RaceWhite REGION

Distribution of Propensity

Percent

Estimated Probability

Normal Panel

30 %OneToManyMTCH(work, baseline2, depvar, , ptid, pairs_out, 1);

31 ods listing;

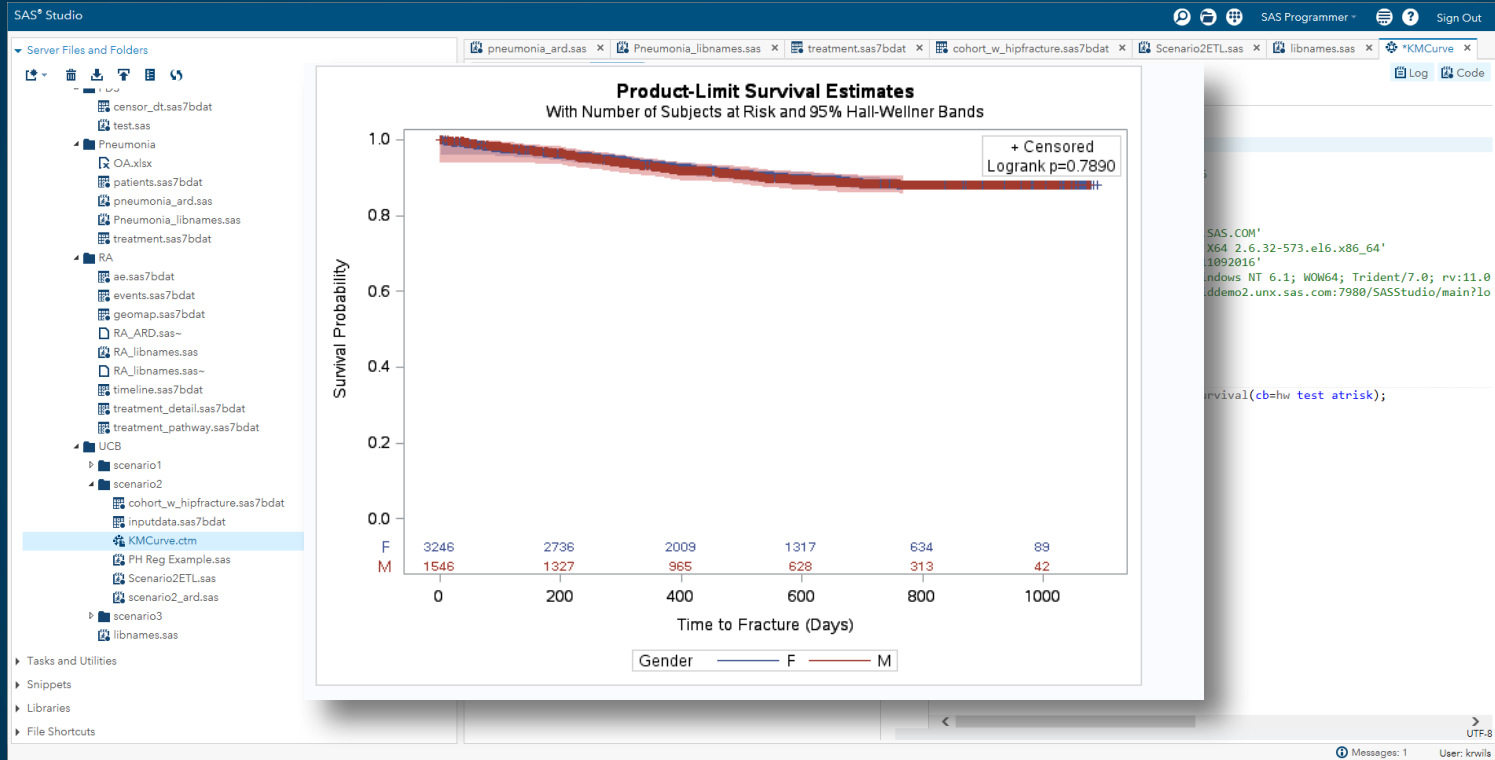
32 %pmdiag(baseline2, pairs_out, ptid, , prob, depvar, 1, 0,

33 /tundra/ols/playpens/robert.collins/pmdiag/pmdiag, rtf, age RaceWhite REGION

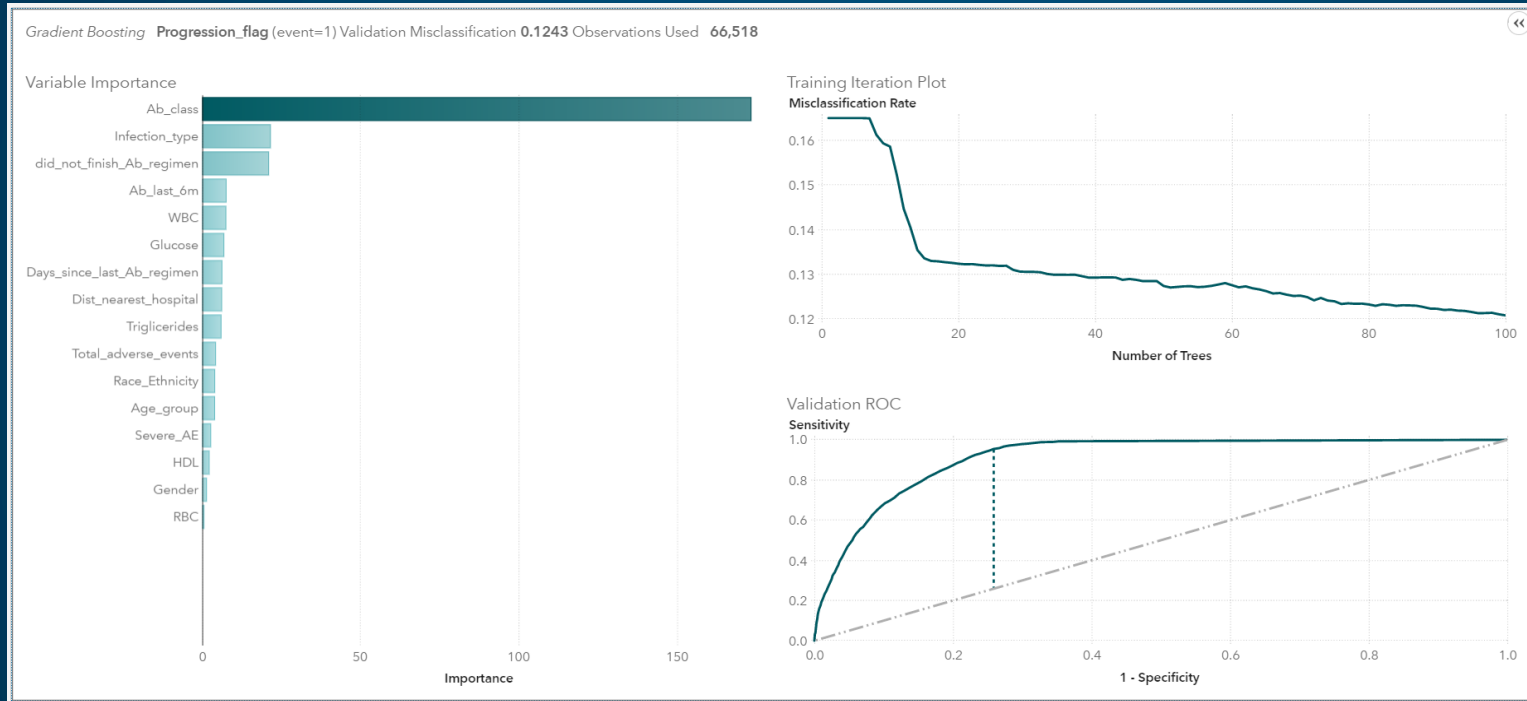
34 HxAngina HxMI HxCHF HxDiabetes HxSmoking HxHCL HxCAD HxHTN HxStroke, age

35 RaceWhite REGION HxAngina HxMI HxCHF HxDiabetes HxSmoking HxHCL HxCAD HxHTN

Knowledge and Insights via Analytics – Survival Analysis

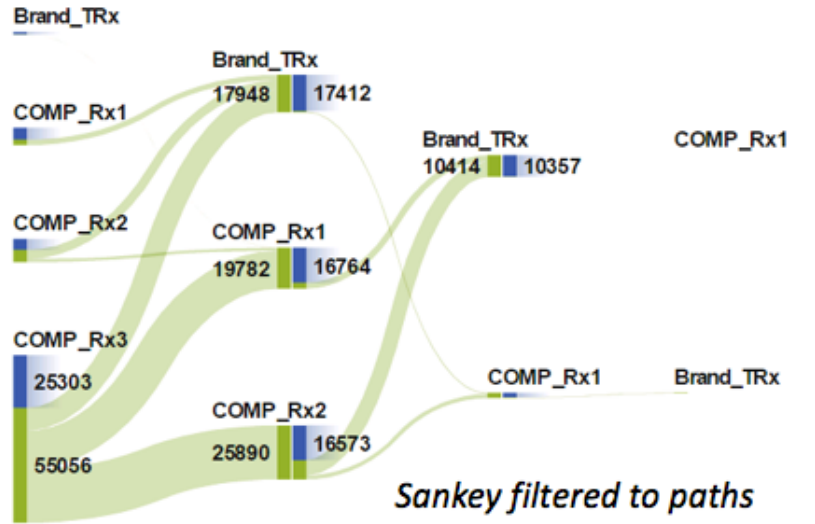


Knowledge and Insights via Analytics – Machine (deep) Learning



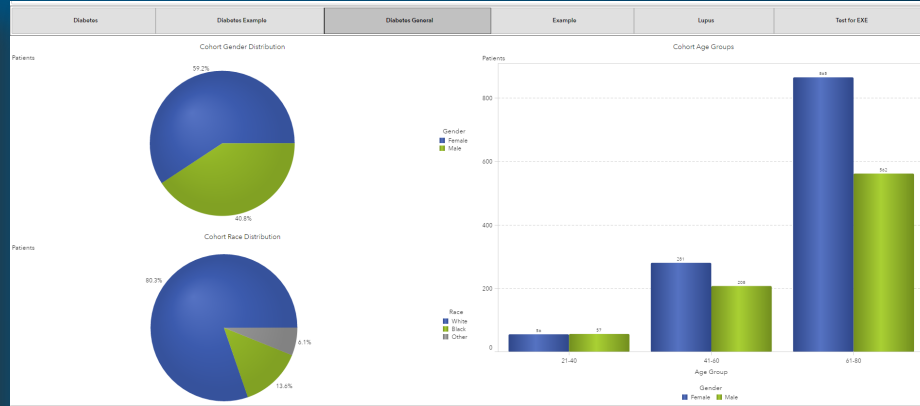
Real World Insights – Visualization & Reporting Examples (1/2)

PATIENT TREATMENT PATHWAYS



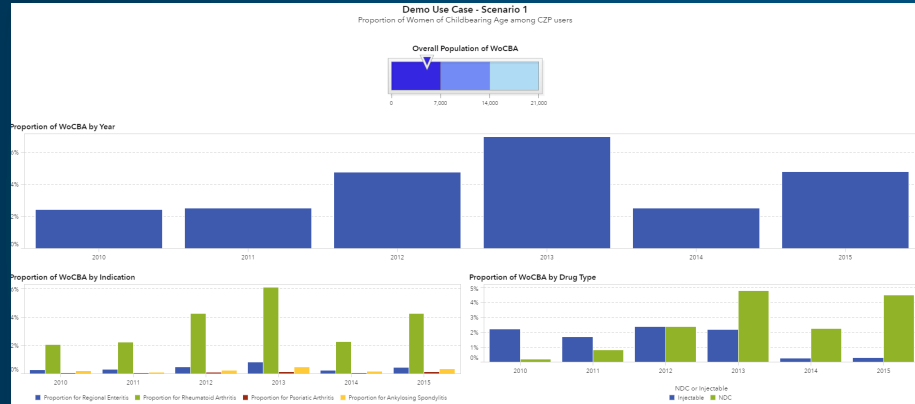
Sankey filtered to paths containing a specific node...

COHORT EXPLORATION and COMPARISON

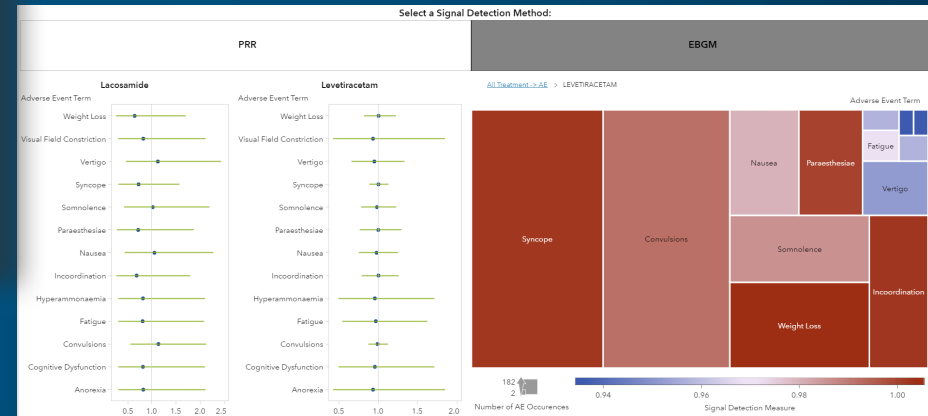


Real World Insights – Visualization & Reporting Examples (2/2)

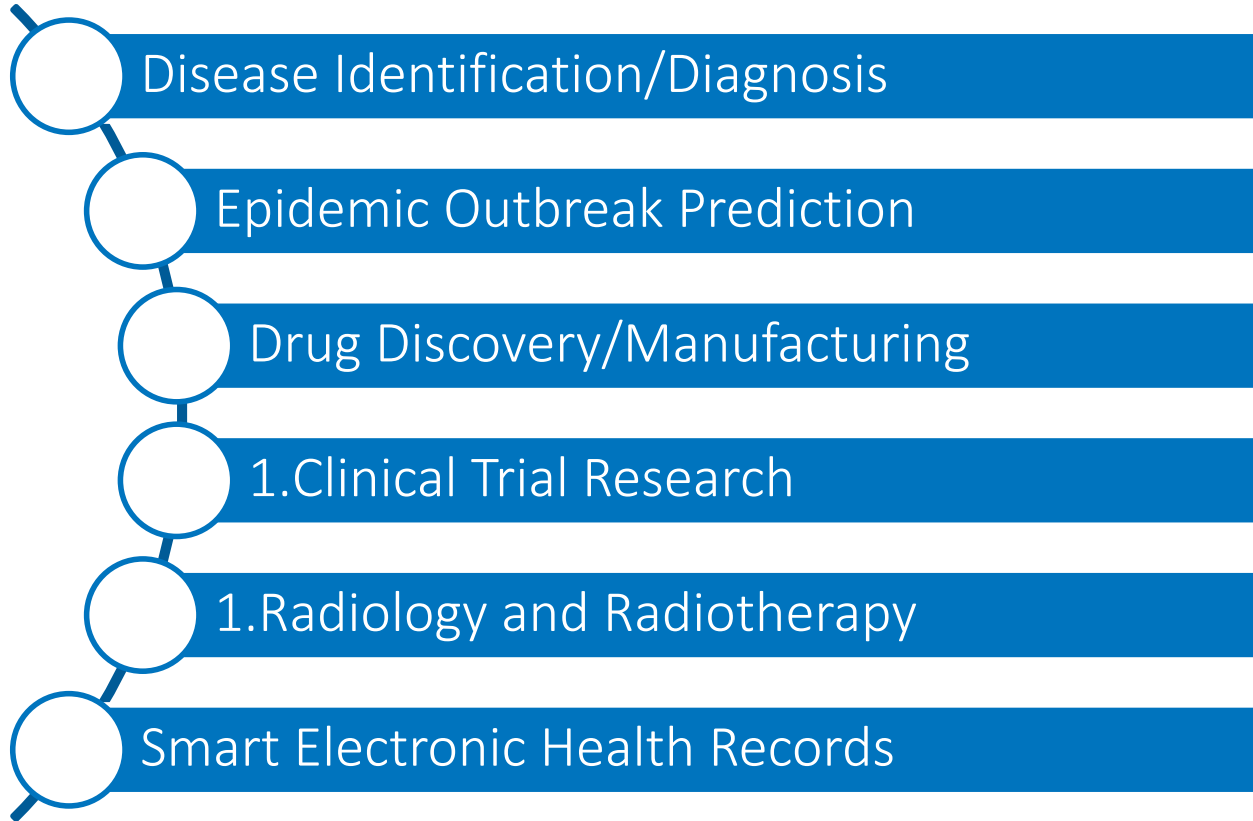
PROPORTION EXPOSED TO TREATMENT BY INDICATION



ADVERSE EVENTS and SIGNAL DETECTION



Use cases of Machine Learning



Patient streaming data as real world data

“Internet of Medical Things (IoMT)”

- Applications and hardware that can connect to healthcare information technology systems via networking technologies.
- Connected People, Environments and Machines facilitate the transfer of clinically relevant data over networks
- Integration of these data within monitoring, alerting and decision support systems will enable dramatic transformation in the development and manufacture of effective and safe medicinal products as well as accelerating the adoption of Value-Based Health Care.



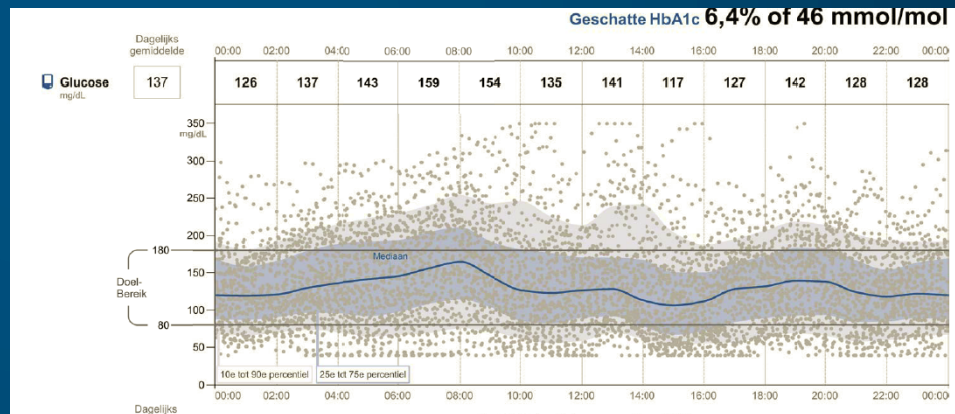
<http://www.gizbot.com>

Patient streaming data as real world data

Hidden Markov Models in type 1 diabetes



- Predicting the state of the patient (stable -> unstable) based on longitudinal flash glucose monitoring systems



Patient streaming data as real world data

SAS Use Case 1: Asthma detection on streaming data

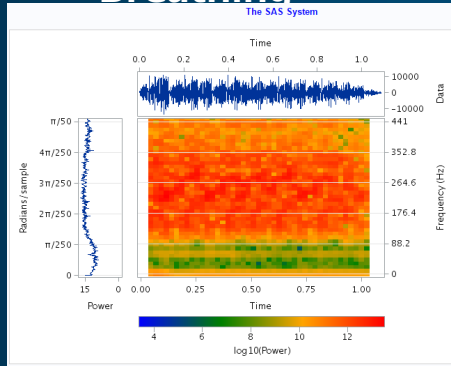
Background:

- Chronic disease with no cure
- Inflamed airways: difficulty moving air in and out of the lungs
- 242million people affected /489K deaths in 2013
- Symptoms: **coughing, wheezing, shortness of breath**
- Death and emergency room visits can be minimized **by early detection** and treatment (inhalers)

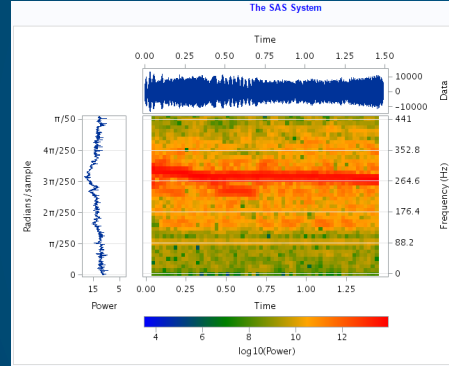
Patient streaming data as real world data

SAS Use Case 1: Asthma detection on streaming data

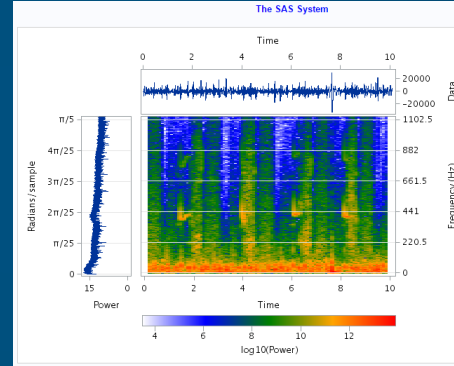
Normal Breathing



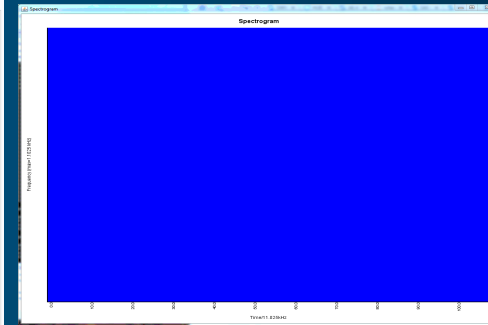
Wheezing



Running on Streaming Data



Detecting Onset Wheezing



Patient streaming data as real world data

SAS Use Case 2: Methods for Detecting Falls

Background:

- Application: Elderly Care
 - 1/3 of Americans 65 and older fall each year
 - In 2013, cost of fall injuries was \$34 billion
 - Every 19 minutes an older adult *dies* from a fall
- Implementation of **activity detection**



Entity extraction

DISEASE = epilepsy

SYMPTOM = impaired vision;
dizziness; difficulty breathing

DEVICE = vagal nerve
stimulator

SIDE-EFFECT = palpitations;
difficulty sleeping

I have been dealing with **epilepsy** for years now, many of which have included uncontrollable seizures and have been accompanied by symptoms including **impaired vision**, **dizziness**, and **difficulty breathing**. Since I have started using the **vagal nerve stimulator**, I have found these symptoms to have subsided. Though I have since experienced minor **palpitations**, such occurrences have been rare. I believe the worst is over. Based upon my experiences with the **VNS** device, I would recommend it to others as a possible solution.

Sentiment analysis

NEGATIVE (*impaired*)

NEGATIVE (*difficulty*)

NEUTRAL (*the worst is over*)

POSITIVE (*I would recommend*)

POSITIVE (*a possible solution*)

I have been dealing with epilepsy for years now, many of which have included uncontrollable seizures and have been accompanied by symptoms including **impaired** vision, dizziness, and **difficulty** breathing. Since I have started using the vagal nerve stimulator, I have found these symptoms to have subsided. Though I have since experienced minor palpitations, such occurrences have been rare. I believe **the worst is over**. Based upon my experiences with the VNS device, **I would recommend** it to others as **a possible solution**.

Summary



01

SAS enables a diverse set of users to collaboratively solve modern business problems.



02

SAS is unifying the analytics modeling experience.



03

SAS will continue to enhance and modernize algorithms in machine learning / forecasting / text .