SAS® Machine Learning for Pharmaceuticals

PhUSE 2017

Aneshan Ramaloo

Advanced Analytics Practise Lead - Africa

THE POWER TO KNOW®

sas

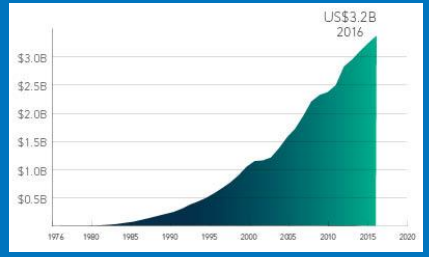
41

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.

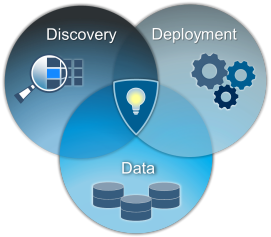
- SAS became in 2016 an **official partner in research for IMI/EFPIA** demonstrating the **continuous level of investment in Life Sciences industry**. IMI2 Activities in 2017 will address the insufficient level of medication compliance and adherence by creating a multi-stakeholder network which will 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

SAS Analytics in Action



Machine Learning

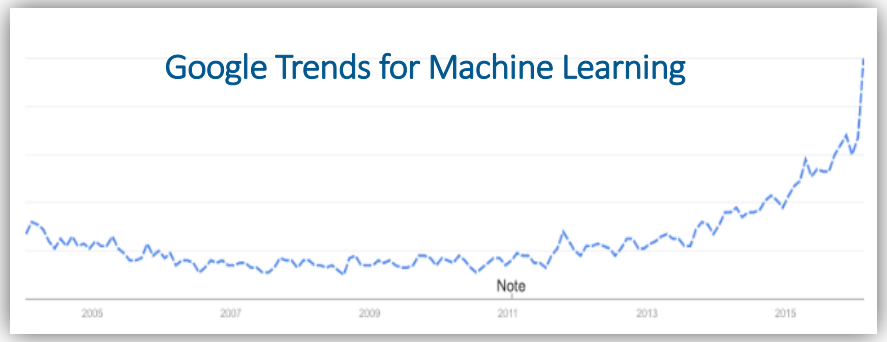
Automate
- Provide automation to the model building process by minimizing human intervention

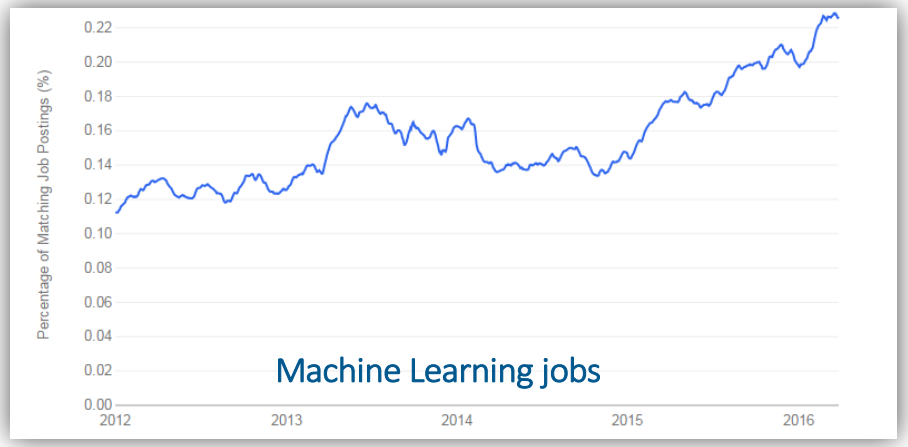
Customize
- Build powerful models using SAS's state-of-the-art algorithms in conjunction with open source tools

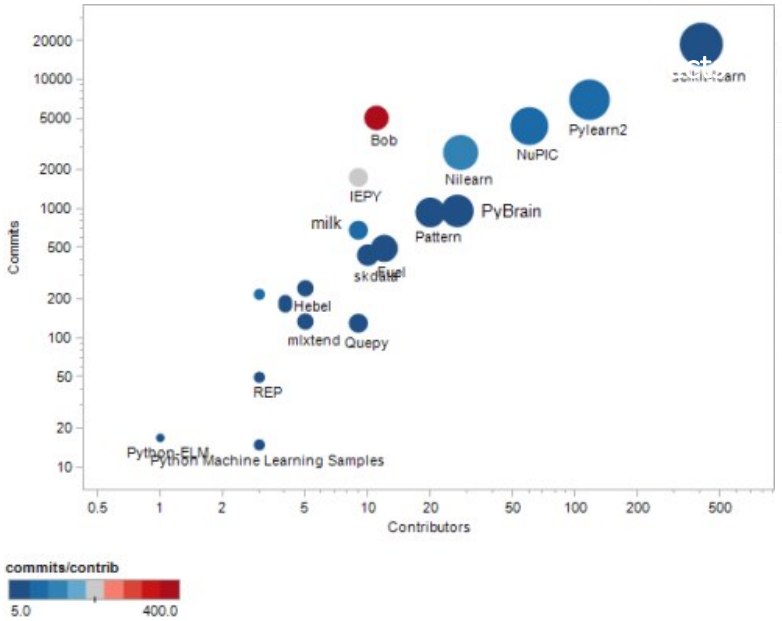
Speed
- Fast response time for sophisticated analytics applied to data of any size or complexity

Machine Learning

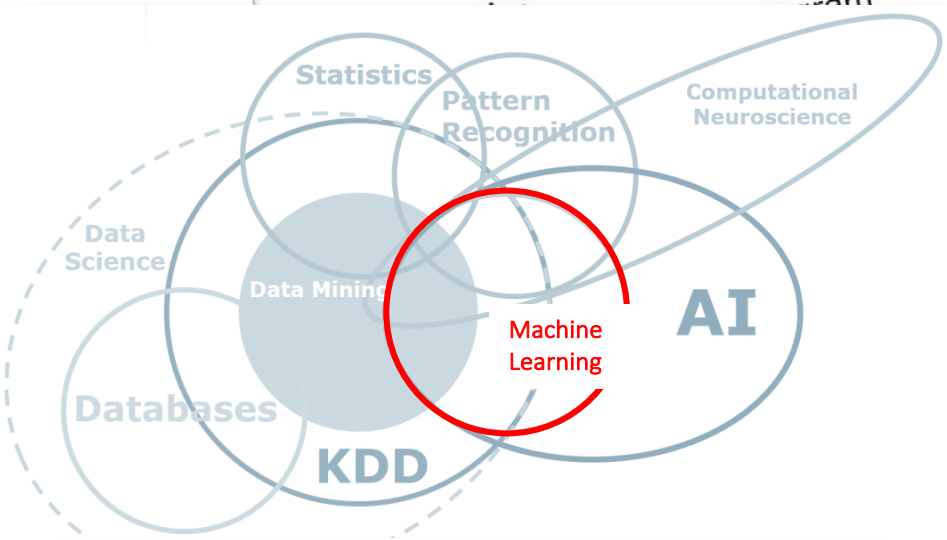
Machine Learning Trends







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



SAS Data Mining Primer Course
SAS Institute, 1998

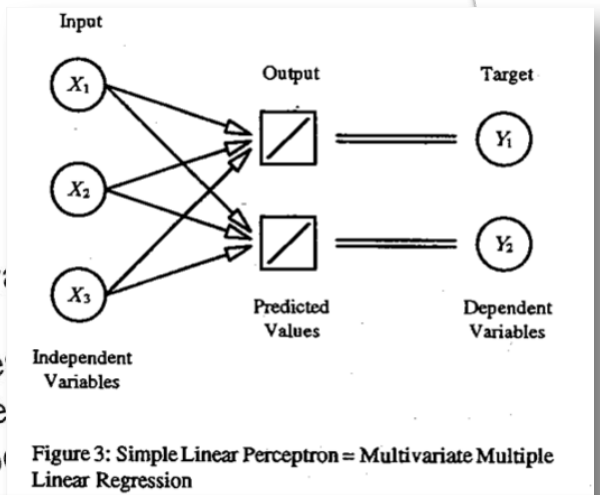


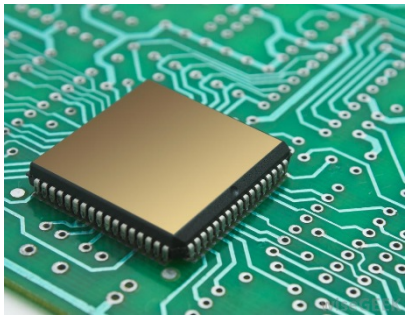
Figure 3: Simple Linear Perceptron = Multivariate Multiple Linear Regression

Neural Networks and Statistical Models,
SAS Institute, 1994

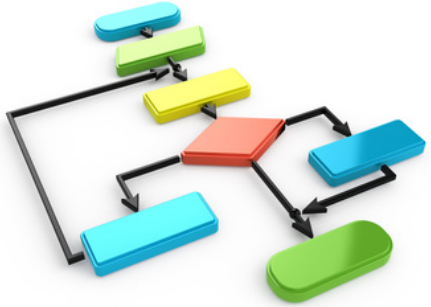
Machine learning

WHY IS IT SO IMPORTANT NOW?

Data

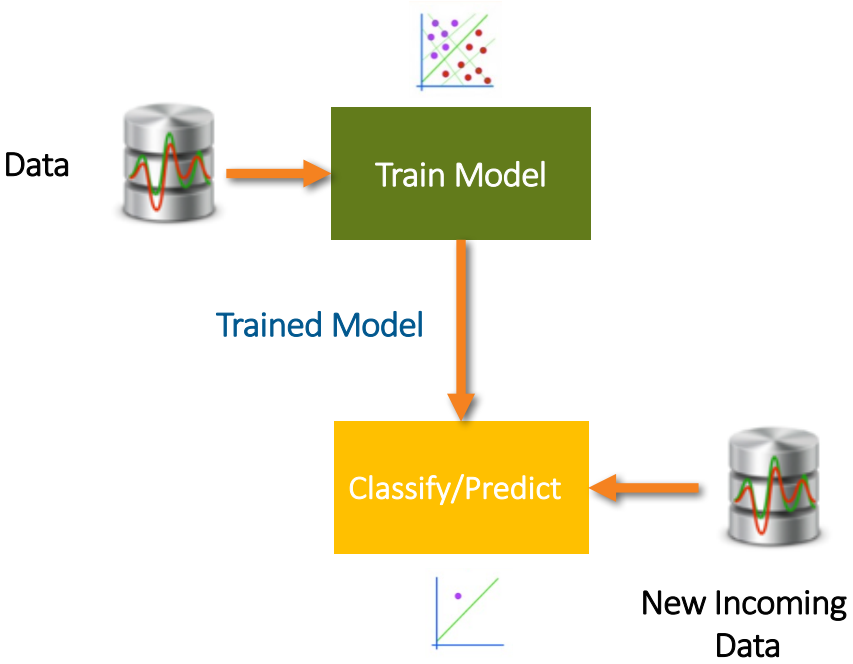


Computing Power



Algorithms

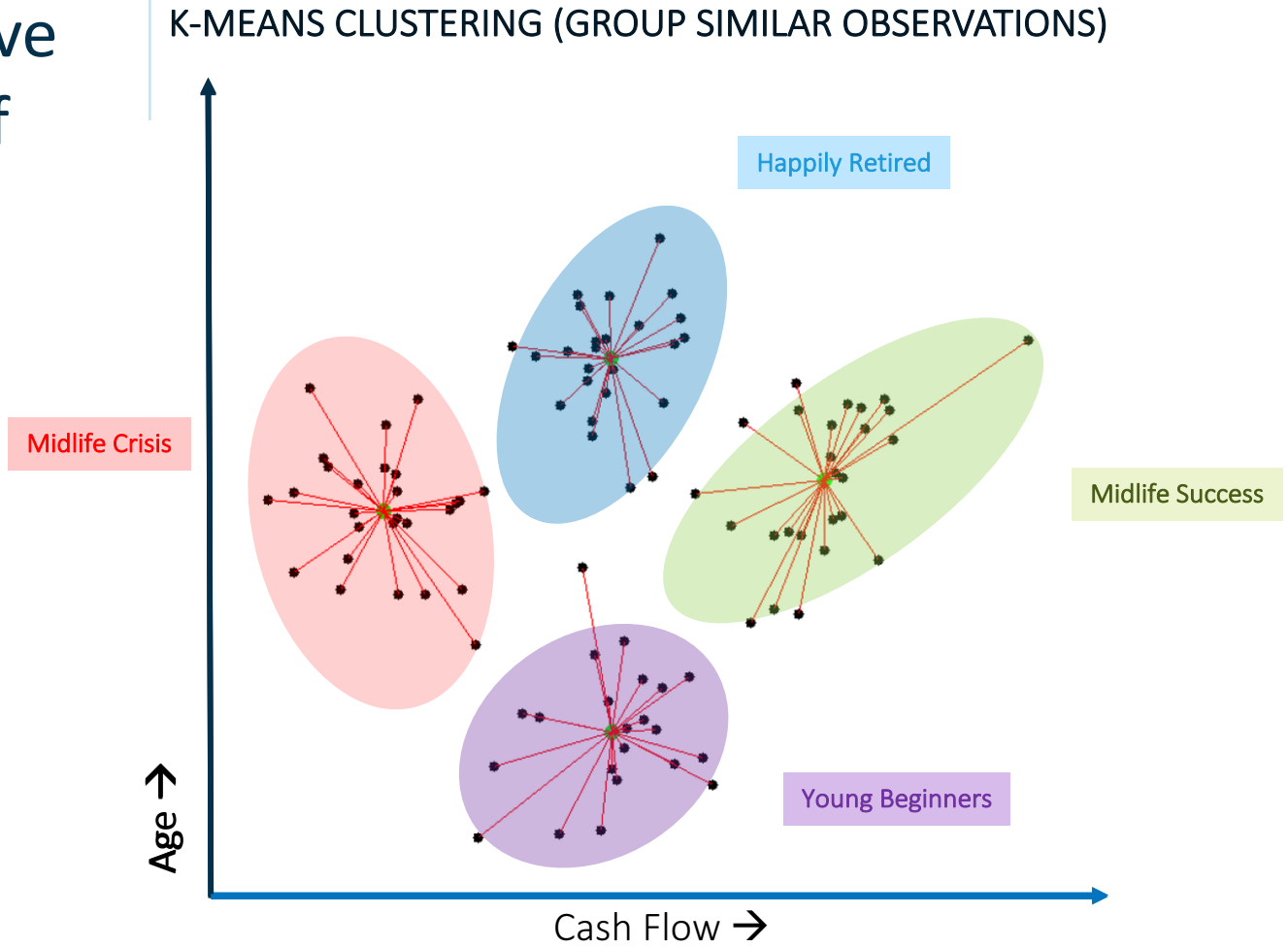
WHAT IS IT?



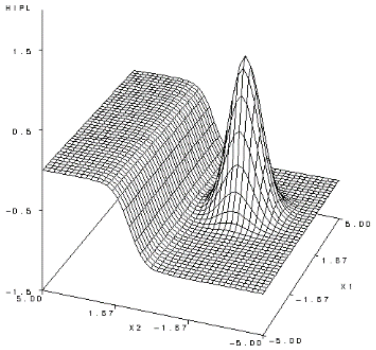
- Machine learning automates the model building process by
  - learning iteratively from data to identify patterns and
  - predicting future results

  - with minimal human intervention
- Machine Learning emphasizes more on **predictive accuracy** rather than interpretability of the models

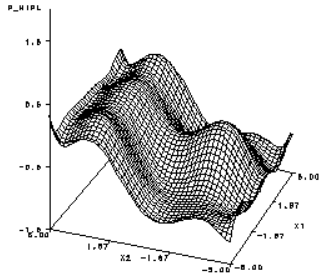
The iterative nature of Machine learning



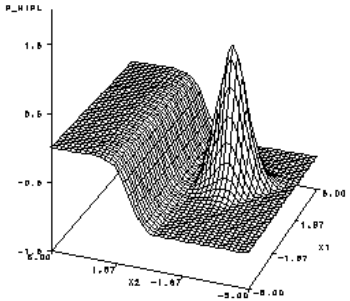
ACCURACY AND INTERPRETABILITY



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

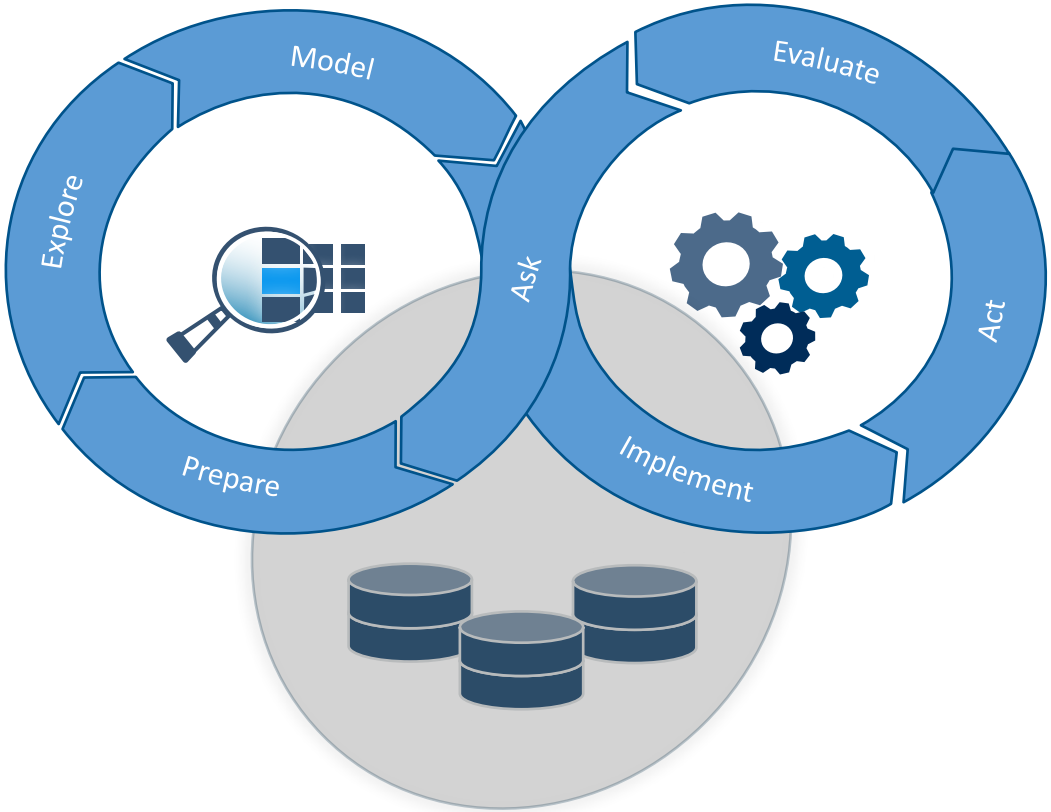
Why SAS Visual Data Mining and Machine Learning

"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**

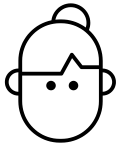


Visual Interfaces

Programming Interfaces

API Interfaces

GUI user

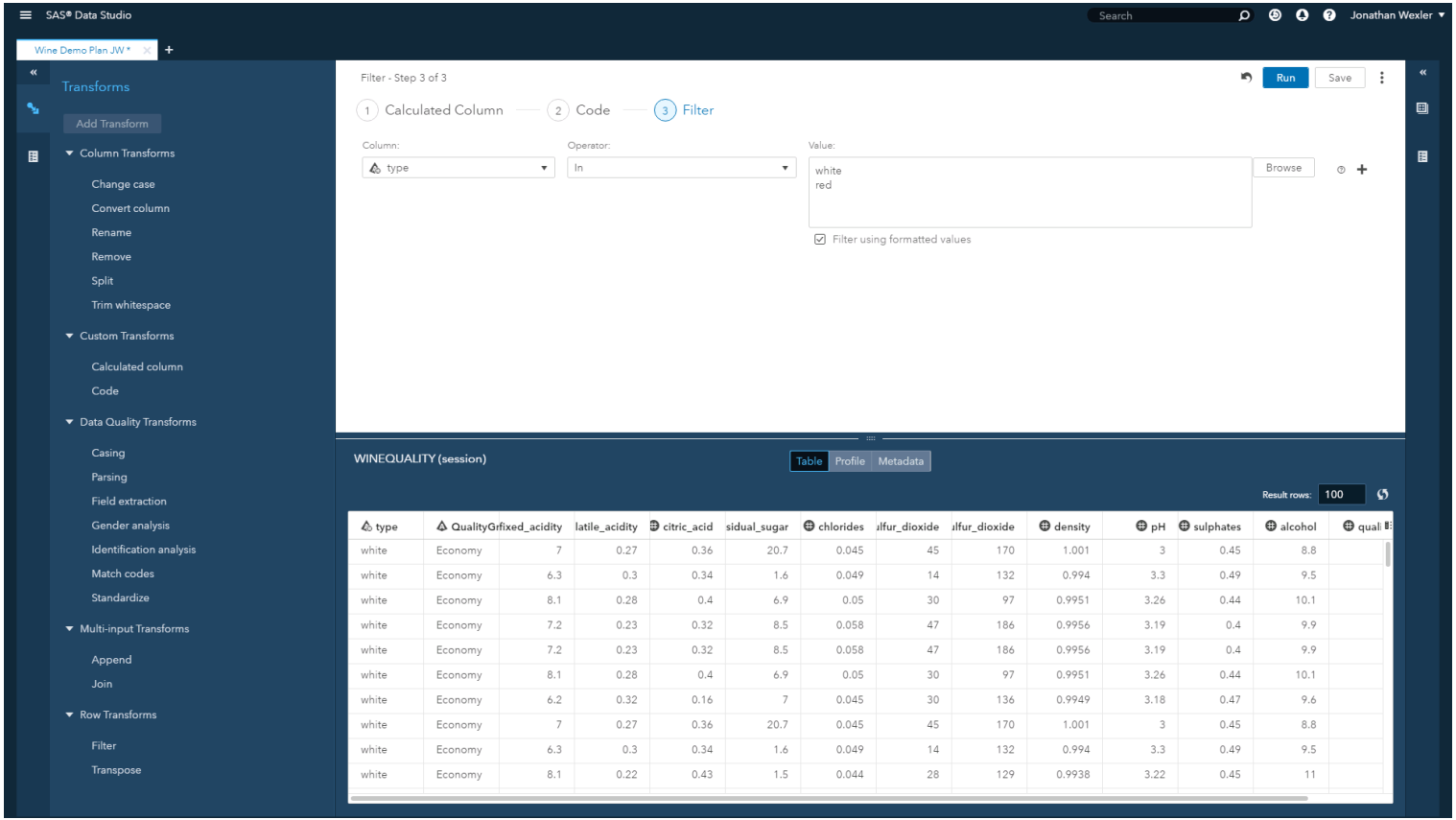


SAS Coder

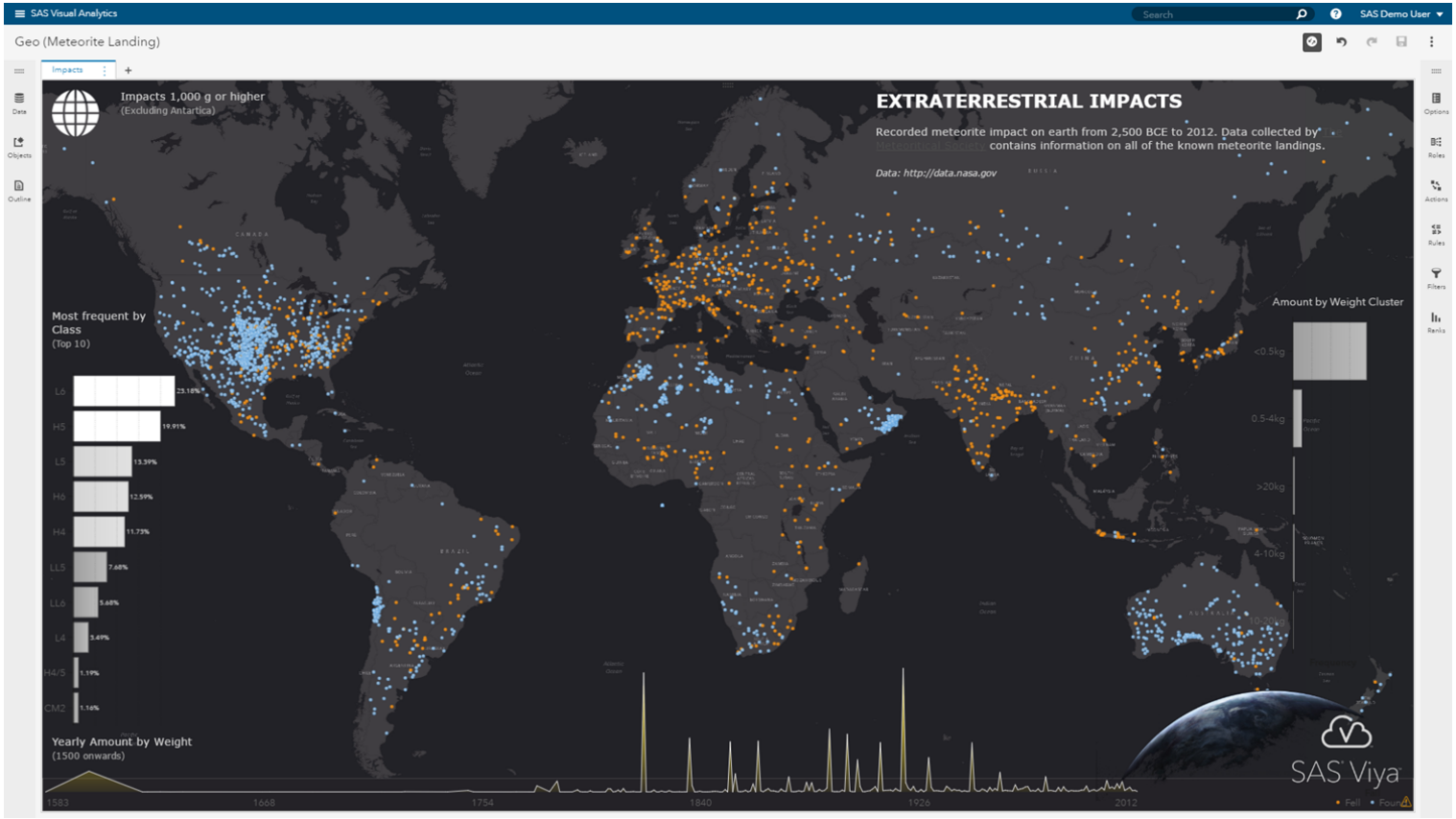
Open Source Coder

Developer

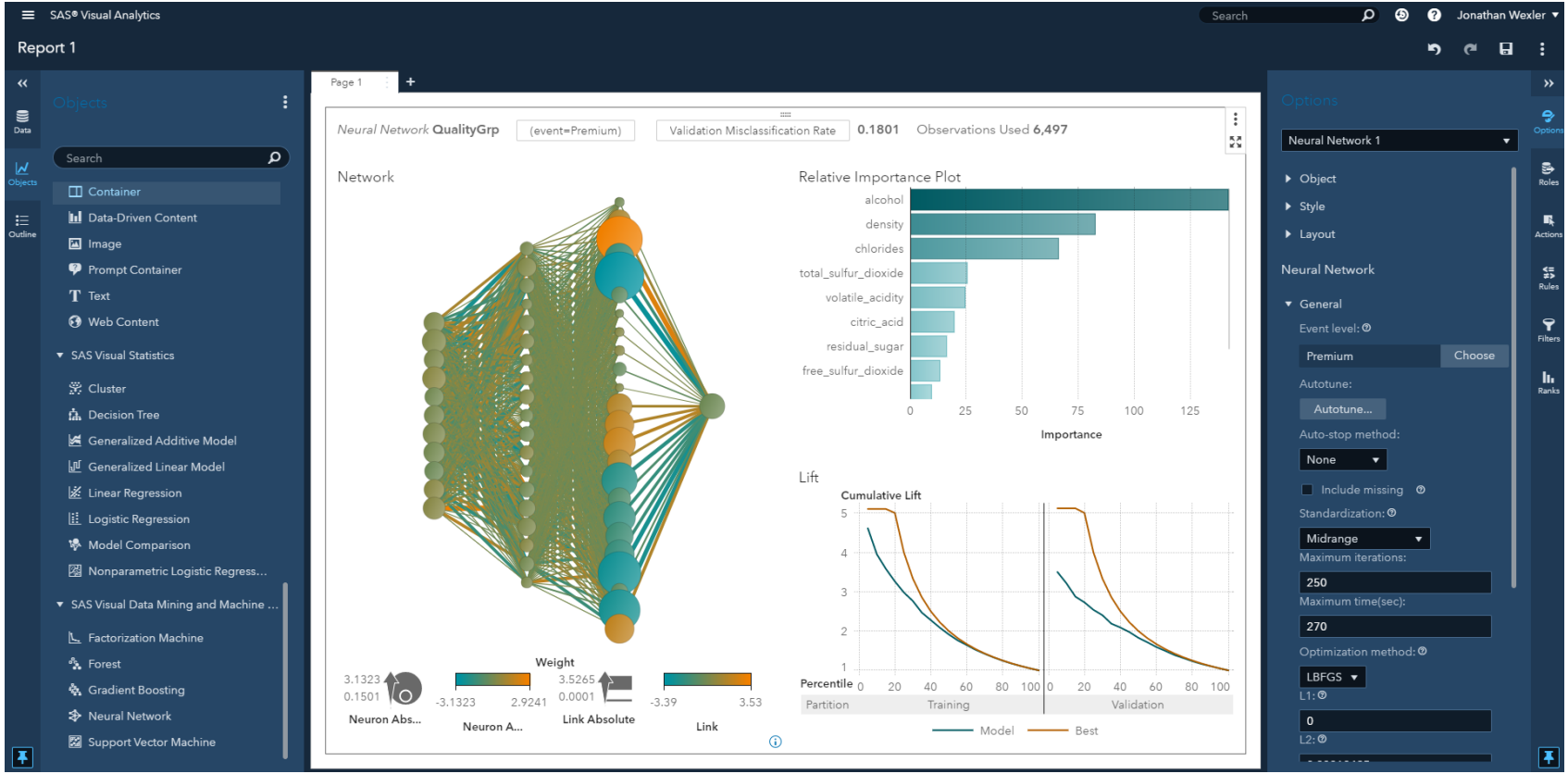
Analytic Data Preparation



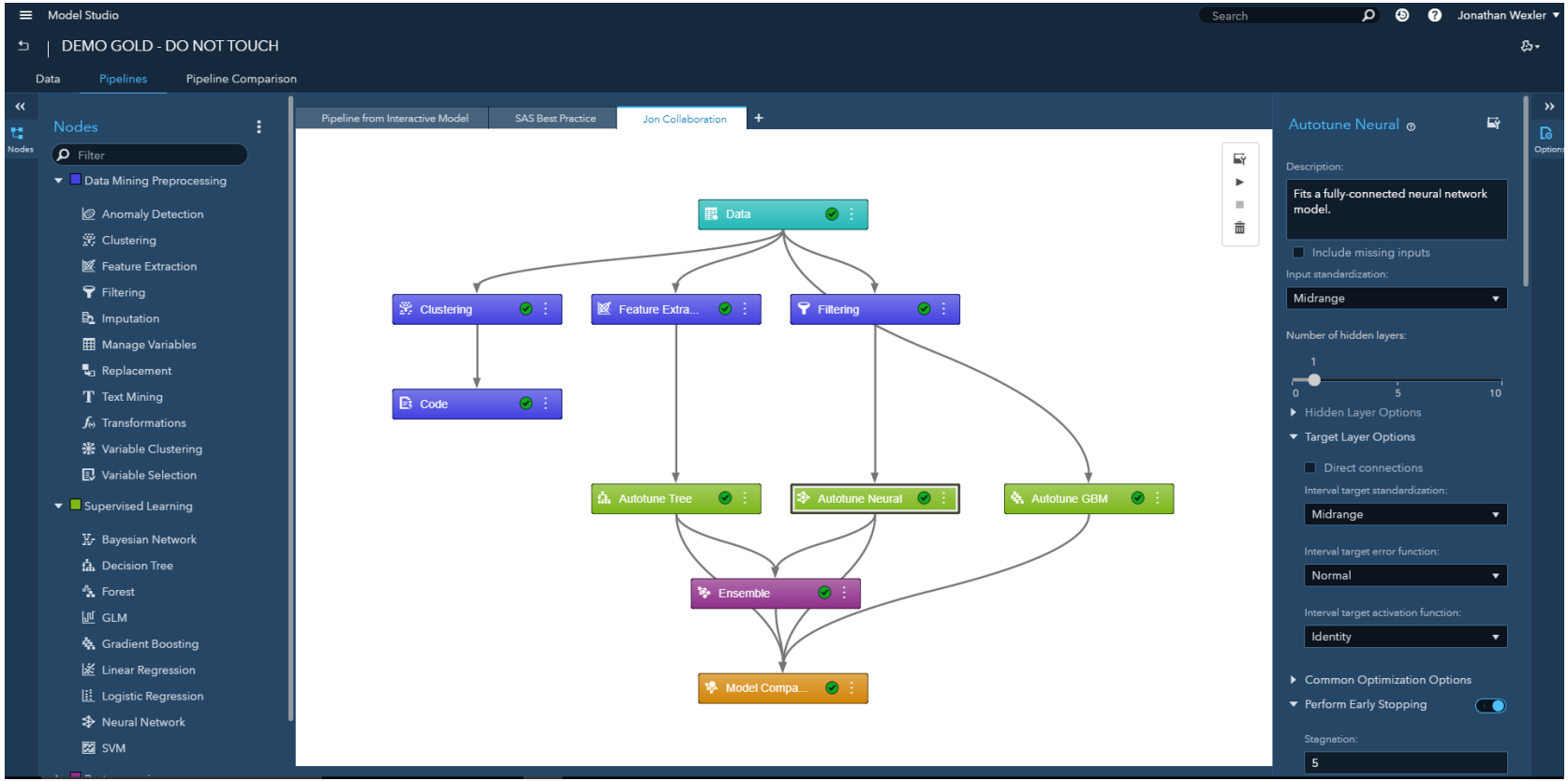
Data Visualisation



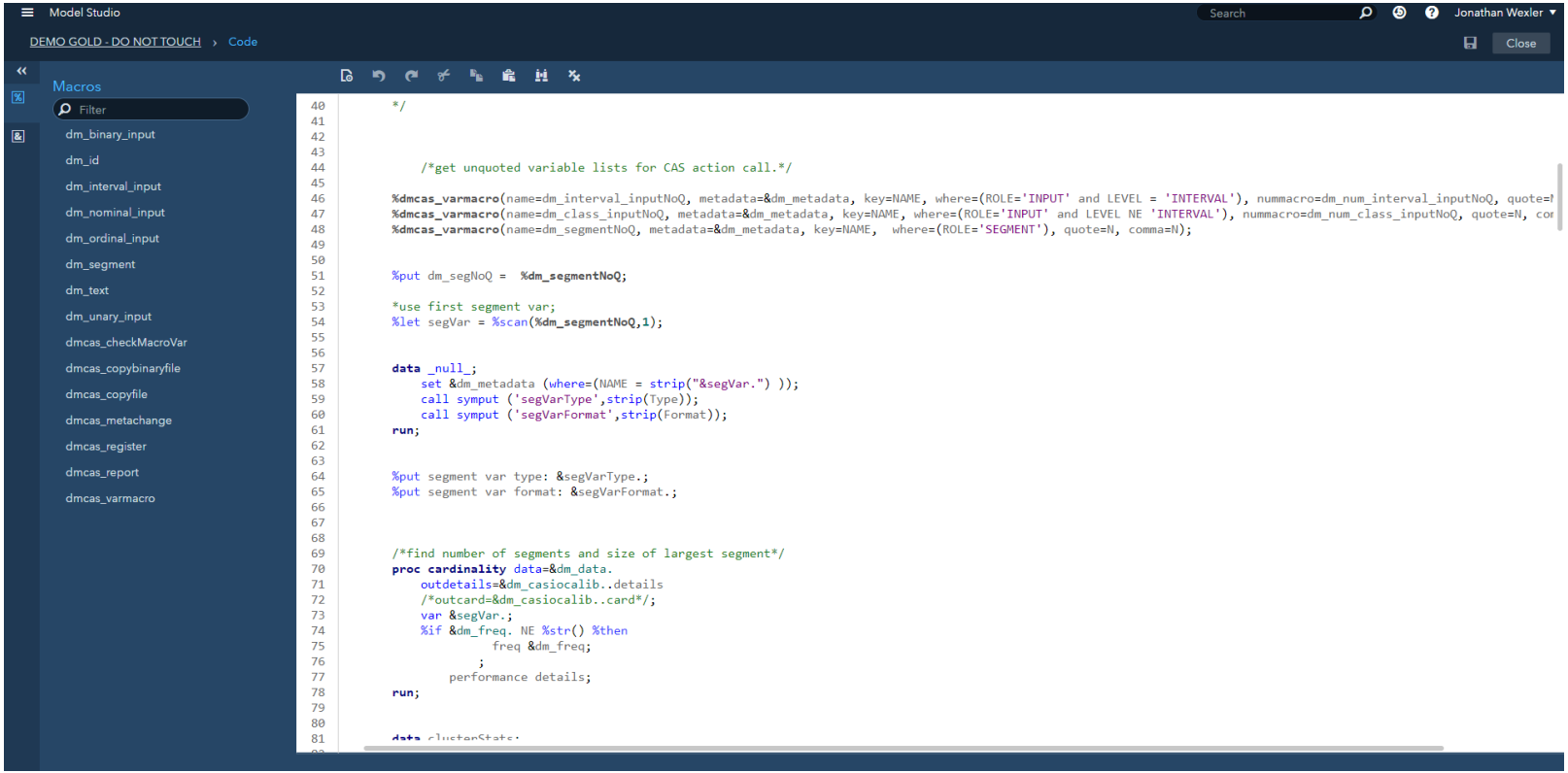
Interactive Discovery and Modeling



Modernized Pipelines and Automation



Embedded Coding



Open Access between SAS, Python & R

jupyter Tree_RF_GBM_NN_SVM Example Last Checkpoint: an hour ago (autosaved)

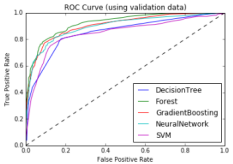
File Edit View Insert Cell Kernel Help Python 2

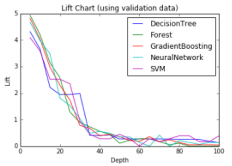
Markdown CellToolbar

Draw ROC and Lift plots

In [34]:
```
#/* Draw ROC charts */
plt.figure()
for key, grp in all_rocinfo.groupby(["model"]):
    plt.plot(grp["FPR"], grp["Sensitivity"], label=key)
plt.plot([0,1], [0,1], 'k--')
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.legend(loc="best")
plt.title("ROC Curve (using validation data)")
plt.show()

#/* Draw Lift charts */
plt.figure()
for key, grp in all_liftinfo.groupby(["model"]):
    plt.plot(grp["Depth"], grp["Lift"], label=key)
plt.xlabel("Depth")
plt.ylabel("Lift")
plt.legend(loc="best")
plt.title("Lift Chart (using validation data)")
plt.show()
```





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')
```

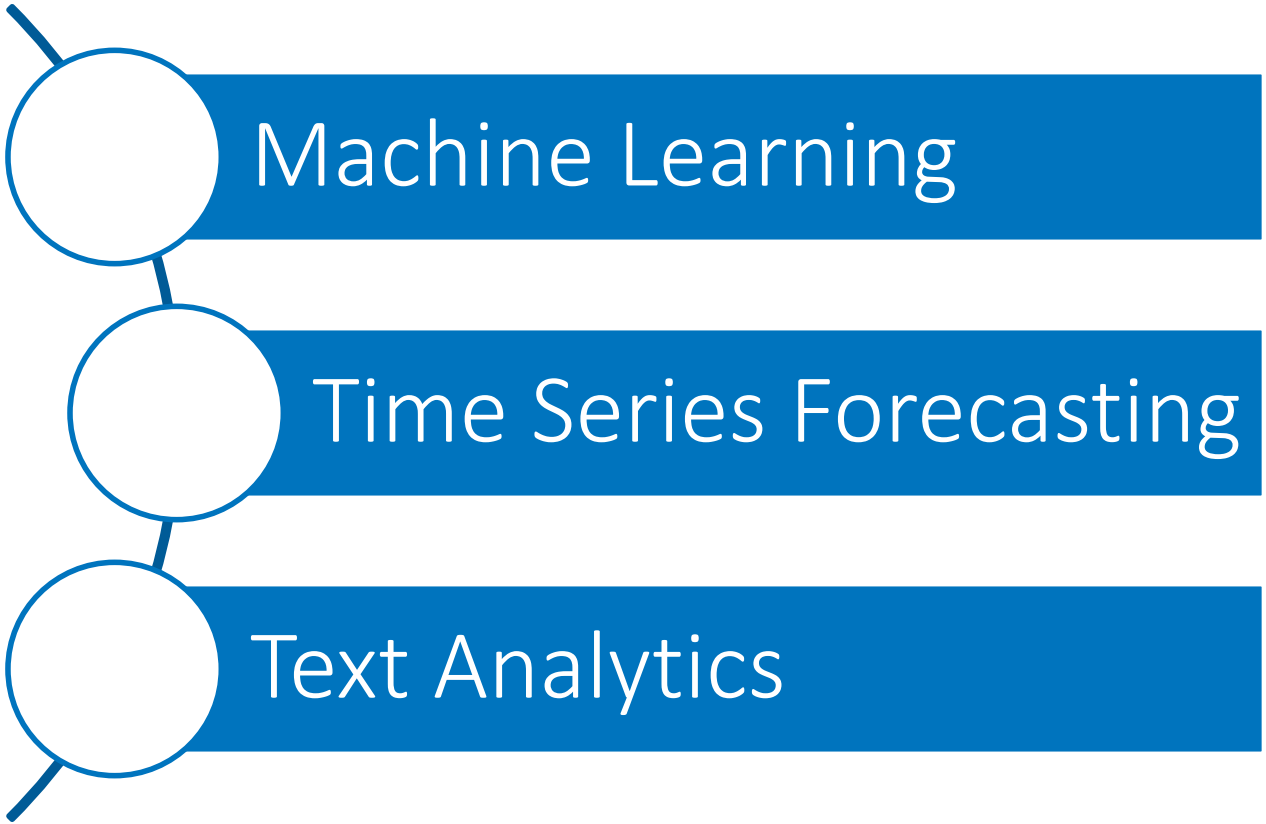
```
## Connecting to CAS and generating CAS action functions for loaded action sets...
```

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

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

View Data

Different Flavours of Advanced Analytics



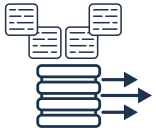
Full-Scale Modeling Support

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

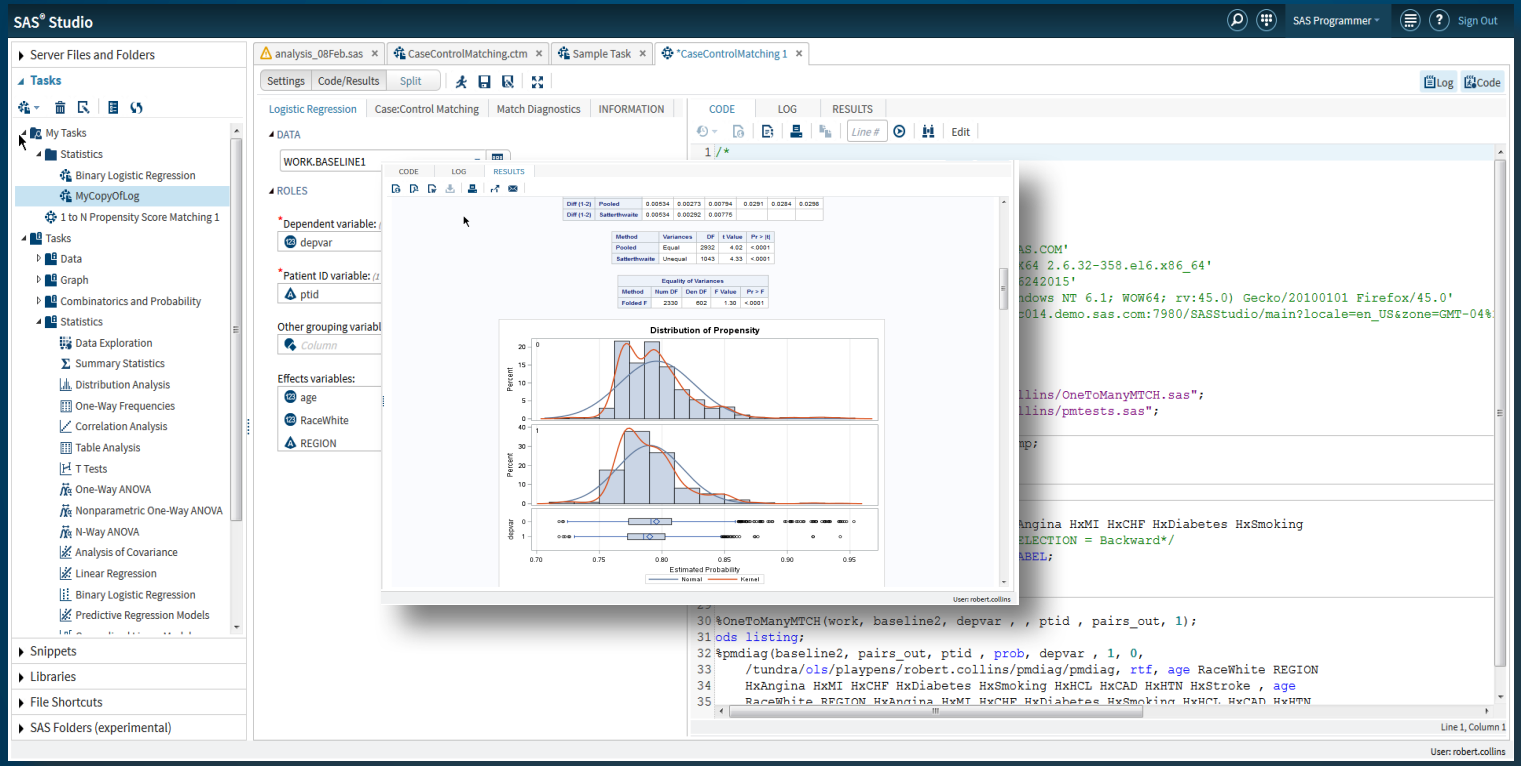
Cox Proportional Hazards
Decision Trees
Design Matrix
General Additive Models
Generalized Linear Models
K-means and K-modes Clustering
Linear Regression
Logistic Regression
Nonlinear Regression
Ordinary Least Squares Regression
Partial Least Squares Regression
Pearson Correlation
Principal Component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



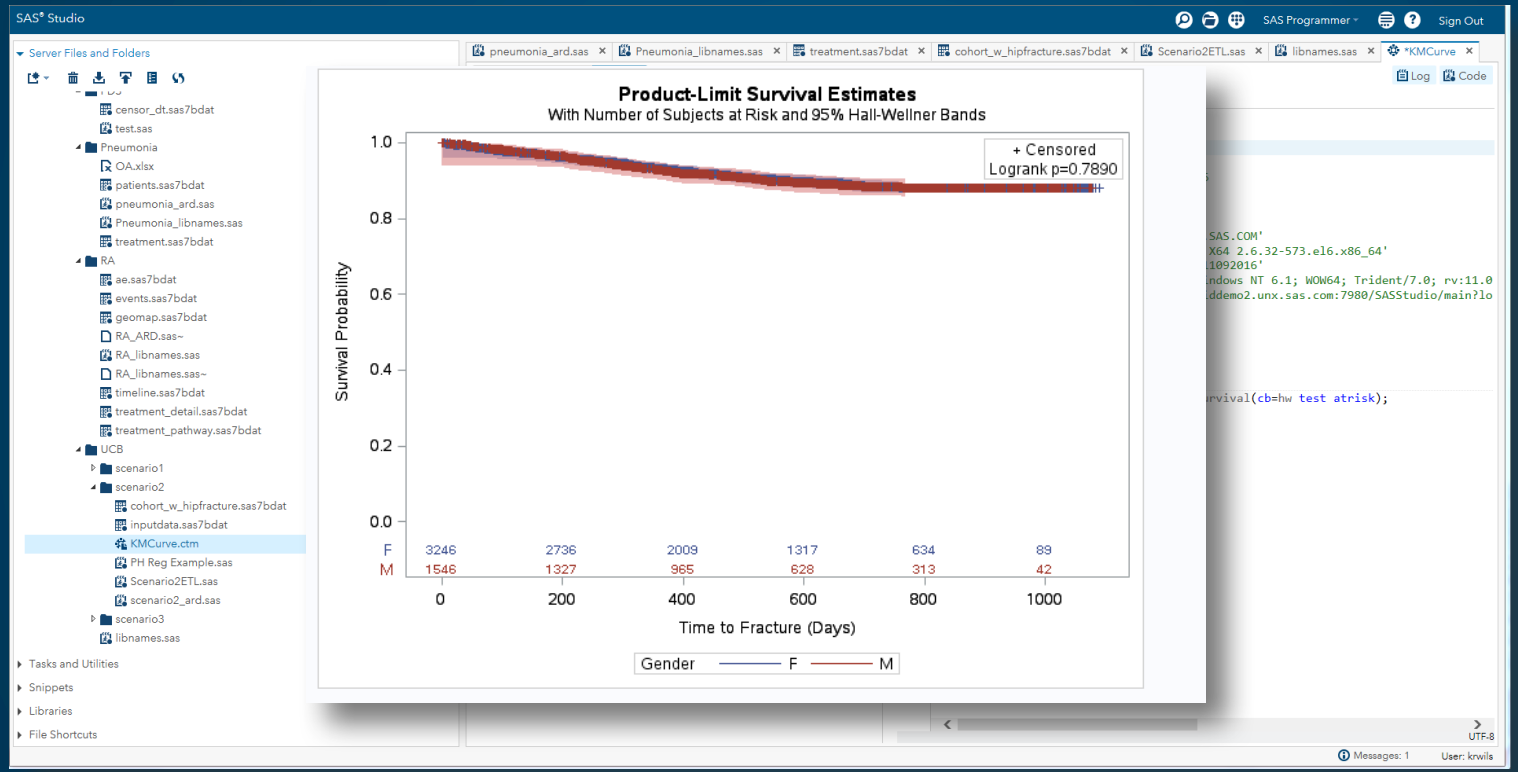
Machine Learning

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

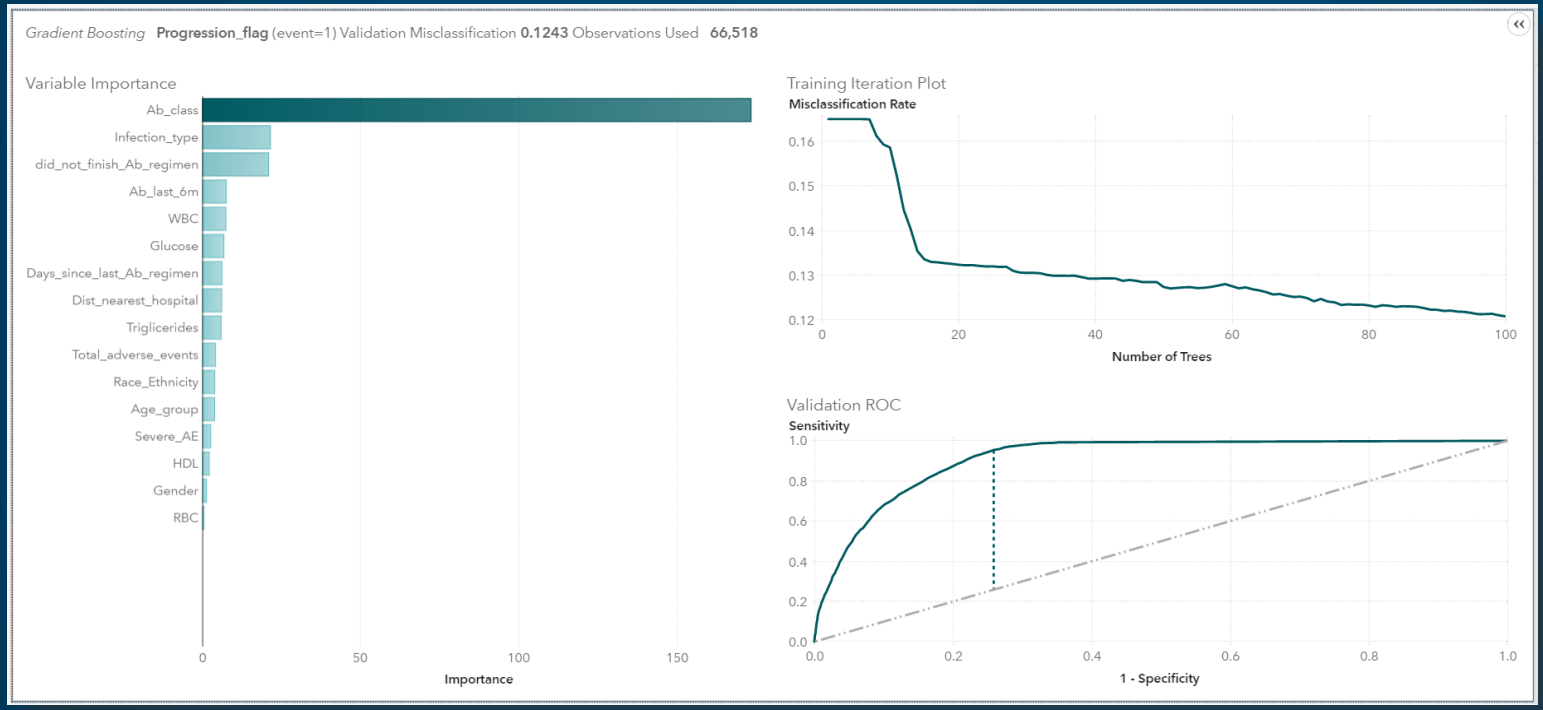
Knowledge and Insights via Analytics - Propensity Score Matching



Knowledge and Insights via Analytics – Survival Analysis

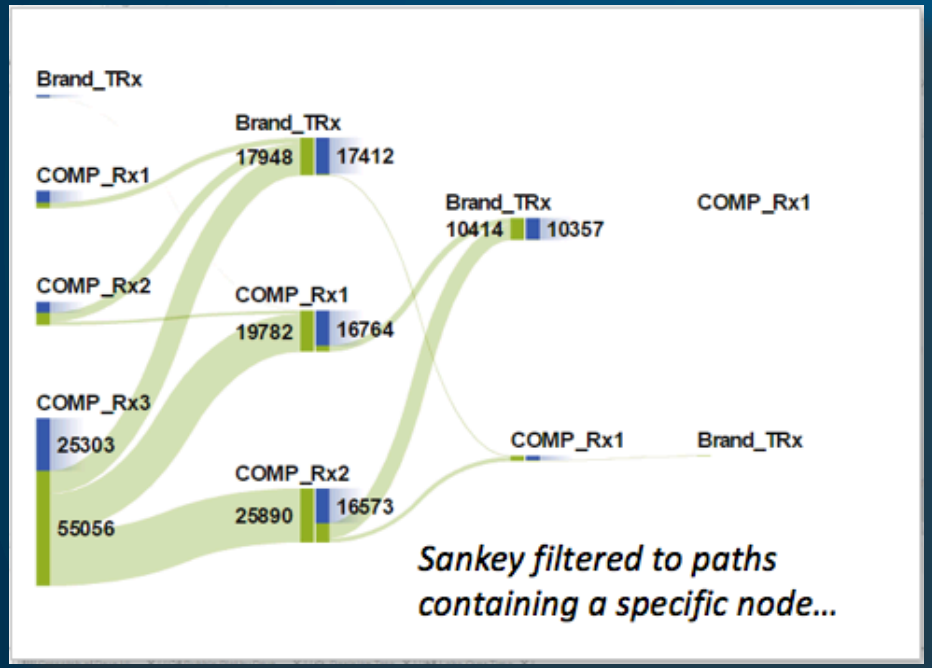


Knowledge and Insights via Analytics – Machine (deep) Learning

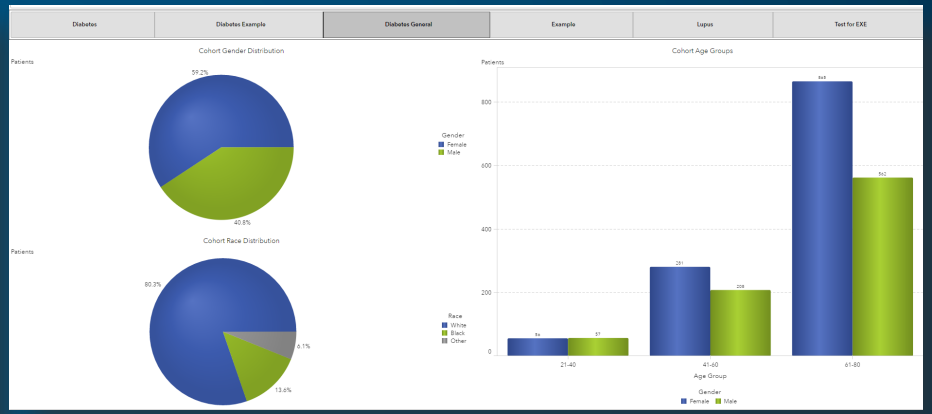


Real World Insights – Visualization & Reporting Examples (1/2)

PATIENT TREATMENT PATHWAYS

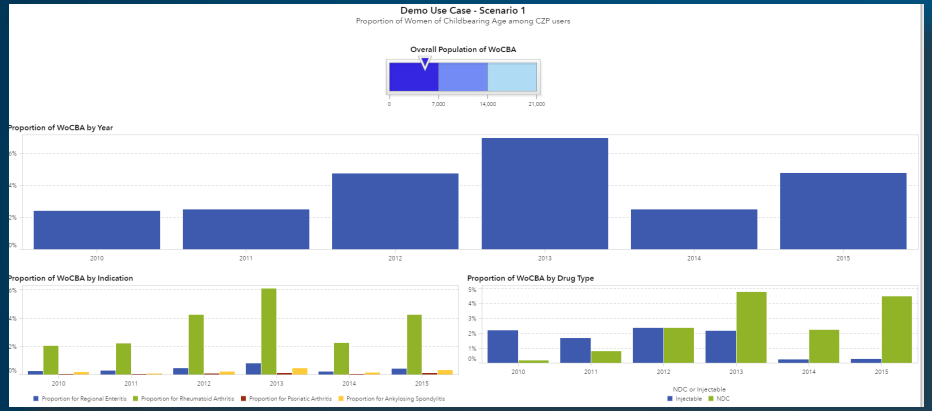


COHORT EXPLORATION and COMPARISON



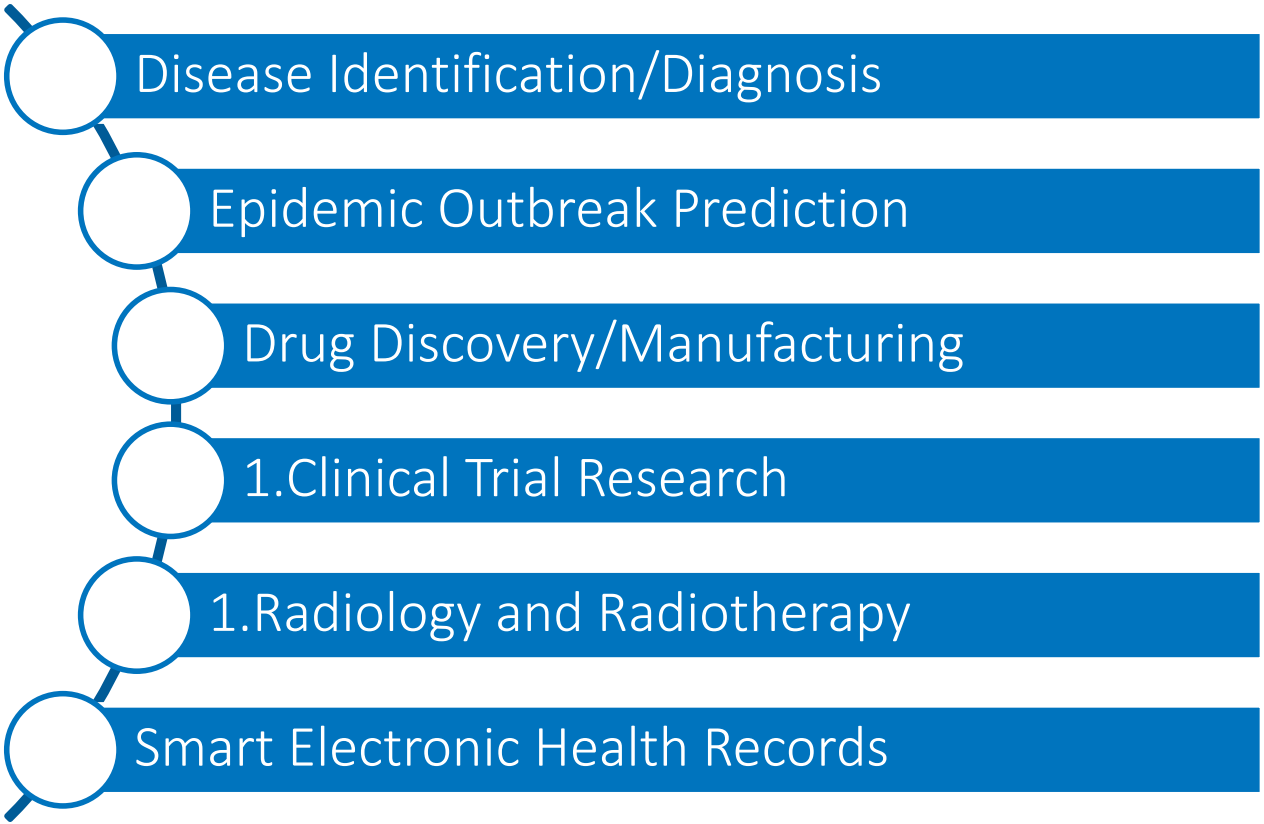
Real World Insights – Visualization & Reporting Examples (2/2)

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)"

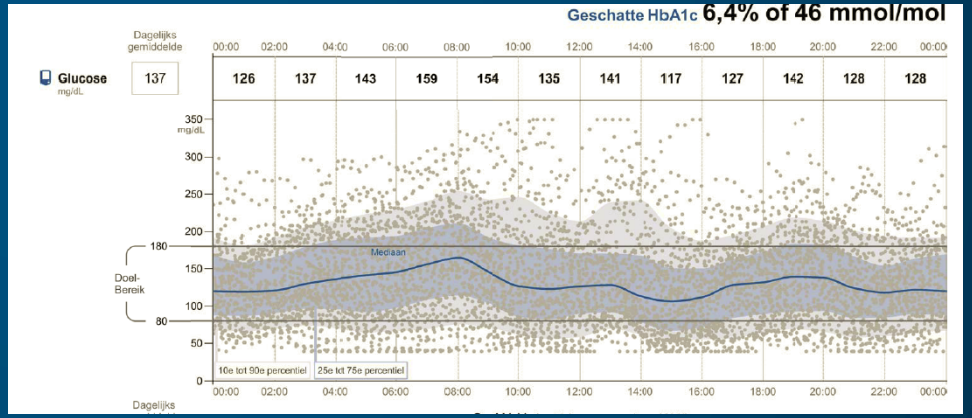
- Applications and hardware that can connect to healthcare information technology systems via networking technologies.
- Connected People, Environments and Machines facilitate the transfer of clinically relevant data over networks
- Integration of these data within monitoring, alerting and decision support 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

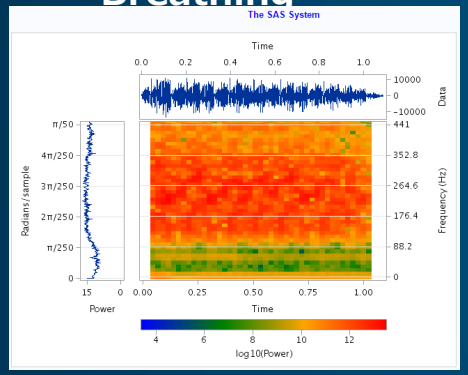
Background:

- 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)

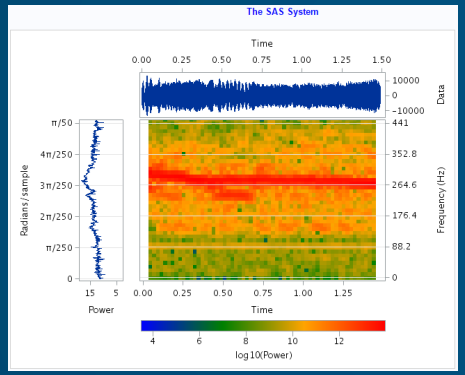
Patient streaming data as real world data

SAS Use Case 1: Asthma detection on streaming data

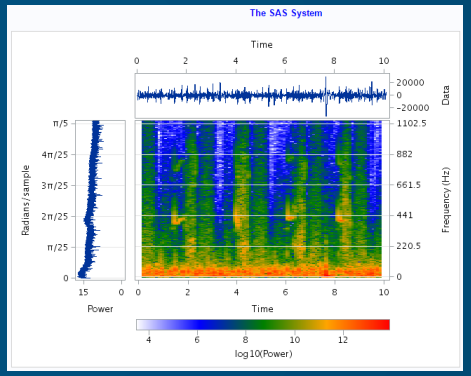
Normal Breathing



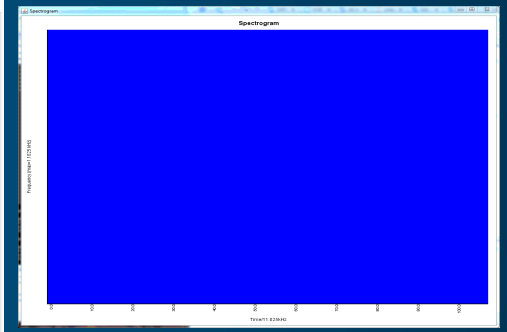
Wheezing



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 .