HabberTec
Business & Intelligence

Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade – As novas armas de geração de valor!

Paulo Gomes
paulo.gomes@habber.com

www.habber.com

Machine Learning

Dan Ariely
January 6, 2013 ·

Follow

Big data is like teenage sex: everyone talks about it, nobody really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it...

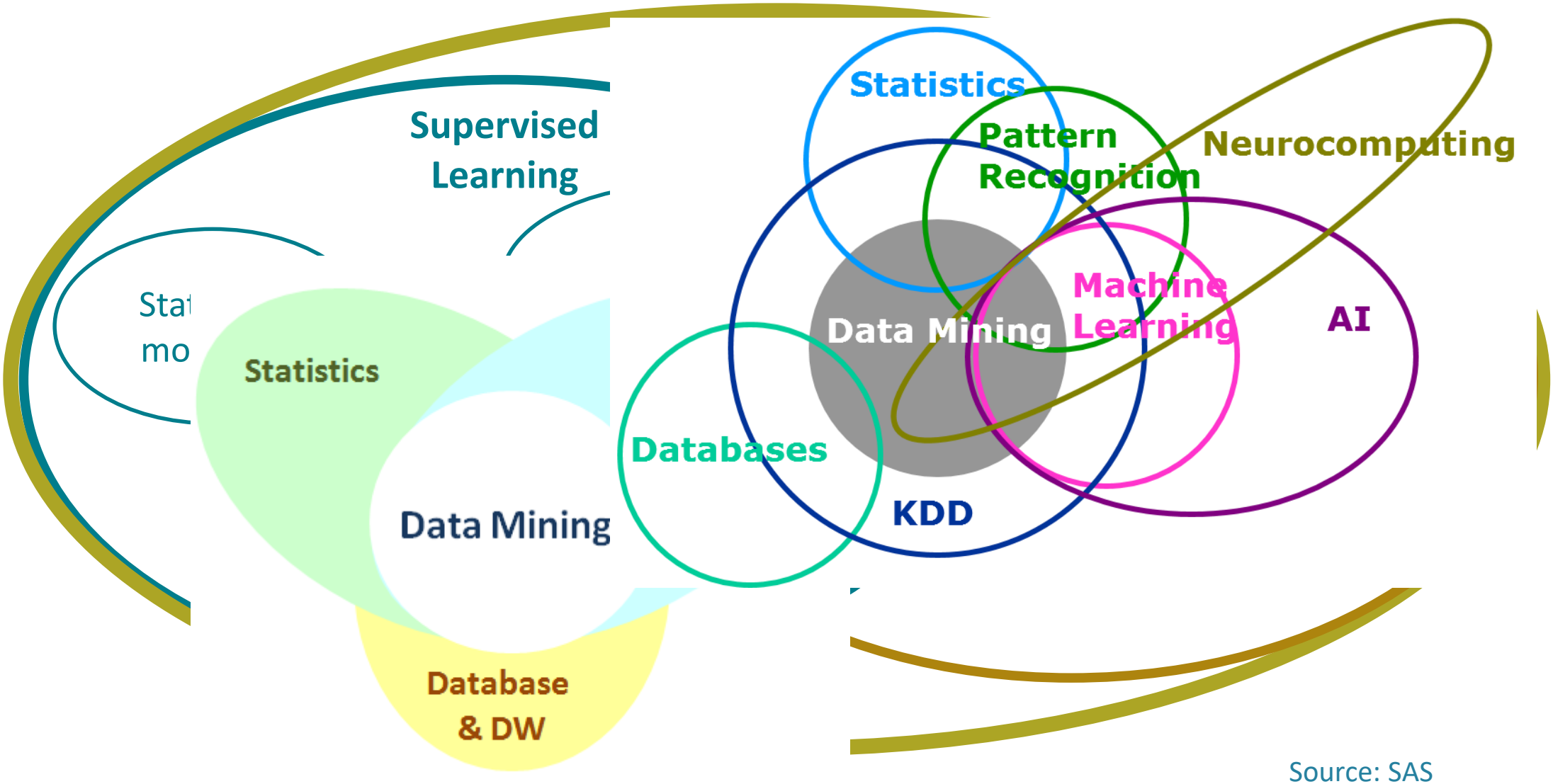
Like · Comment · Share

Machine Learning

Wikipedia:

"Machine learning is a scientific discipline that deals with the construction and study of algorithms that can learn from data. Such algorithms operate by building a model based on inputs and using that to make predictions or decisions, rather than following only explicitly programmed instructions."

Machine Learning



Source: SAS

Machine Learning

In 2006, the IEEE * Conference on Data Mining identified the top 10 ML algorithms as:

1. C4.5 (Decision Trees)
2. k-Means (clustering)
3. Support Vector Machines (SVM)
4. Apriori
5. Expectation Maximization (EM)
6. PageRank
7. AdaBoost
8. k-Nearest Neighbors (kNN)
9. Naive Bayes
10. Classification and Regression Tree (CART)

Source: http://www.datasciencecentral.com/profiles/blogs/top-10-machine-learning-algorithms

* Institute of Electrical and Electronics Engineers

Machine Learning

And **a 2015 answer** provides the following:

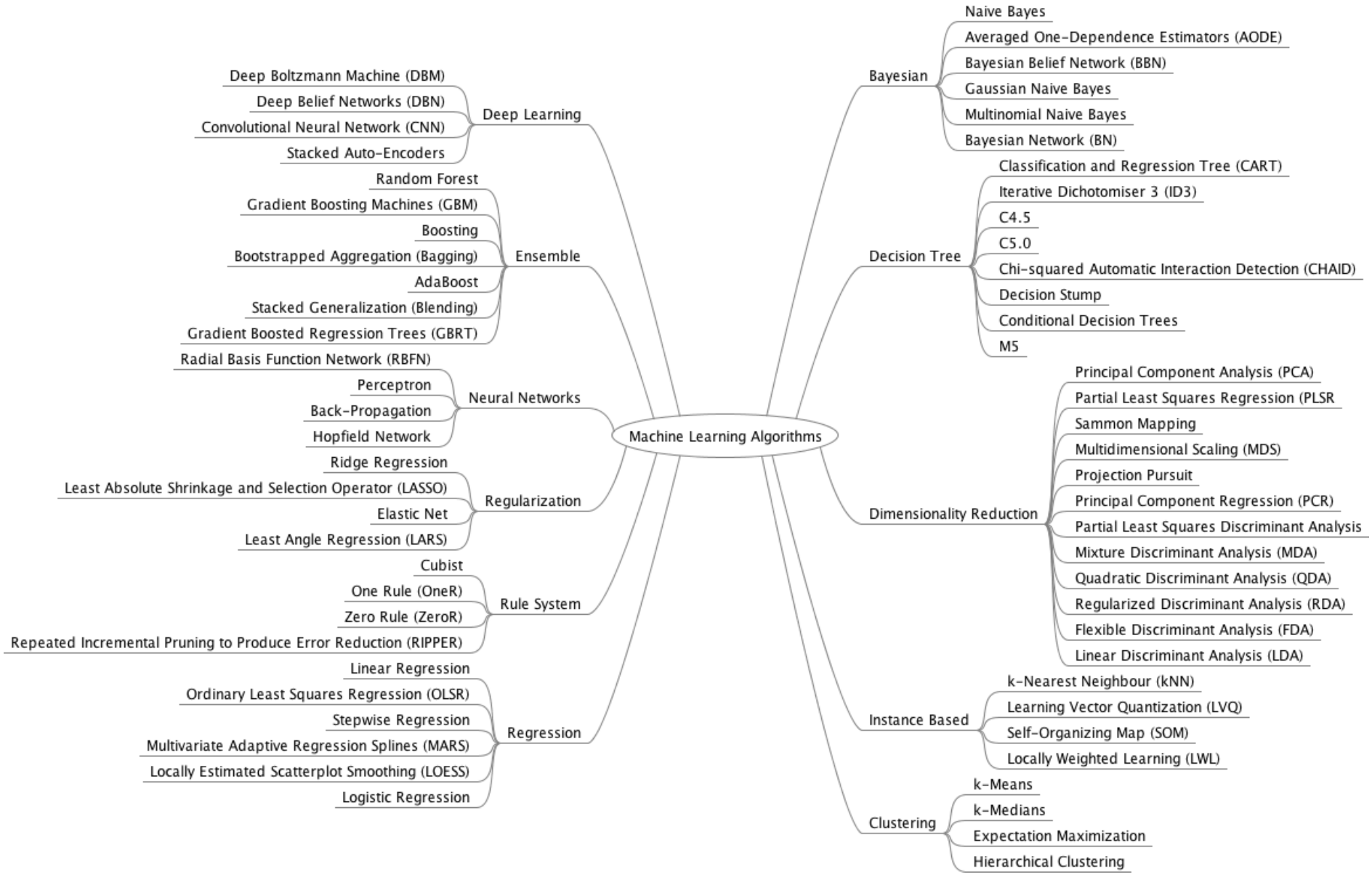
1. Linear regression
2. Logistic regression
3. k-means
4. SVMs
5. Random Forests
6. Matrix Factorization/SVD
7. Gradient Boosted Decision Trees/Machines
8. Naive Bayes
9. Artificial Neural Networks
10. For the last one I'd let you pick one of the following:
    1. Bayesian Networks
    2. Elastic Nets
    3. Any other clustering algo besides k-means
    4. LDA
    5. Conditional Random Fields
    6. HDPs or other Bayesian non-parametric model

Source: http://www.datasciencecentral.com/profiles/blogs/top-10-machine-learning-algorithms

* Institute of Electrical and Electronics Engineers

MachineLearningMastery.com

Machine Learning



Source: MachineLearningMastery.com

Machine Learning

CONS

- ✓ Unfamilar with broader audiance, (more) difficult to explain
- ✓ Black box approach (you are rejected: The computer says NO)
- ✓ Often relations can already be modeled with classical regression models
- ✓ It allows you to not think about the business problem

PROS

- ✓ Often less data prep (manual tuning) neccesary (just throw it in the algorithm…)
- ✓ Interactions often "*automatically*" taken into account
- ✓ Superior for Text mining, Image & Speech recognition
- ✓ Better lift possible
- ✓ *It allows you to not think about the business problem*

Source: SAS

Machine Learning

Many different techniques

Easy to use GUI's combined with flexible coding

High performance scalability

Easy Deployable

Source: SAS

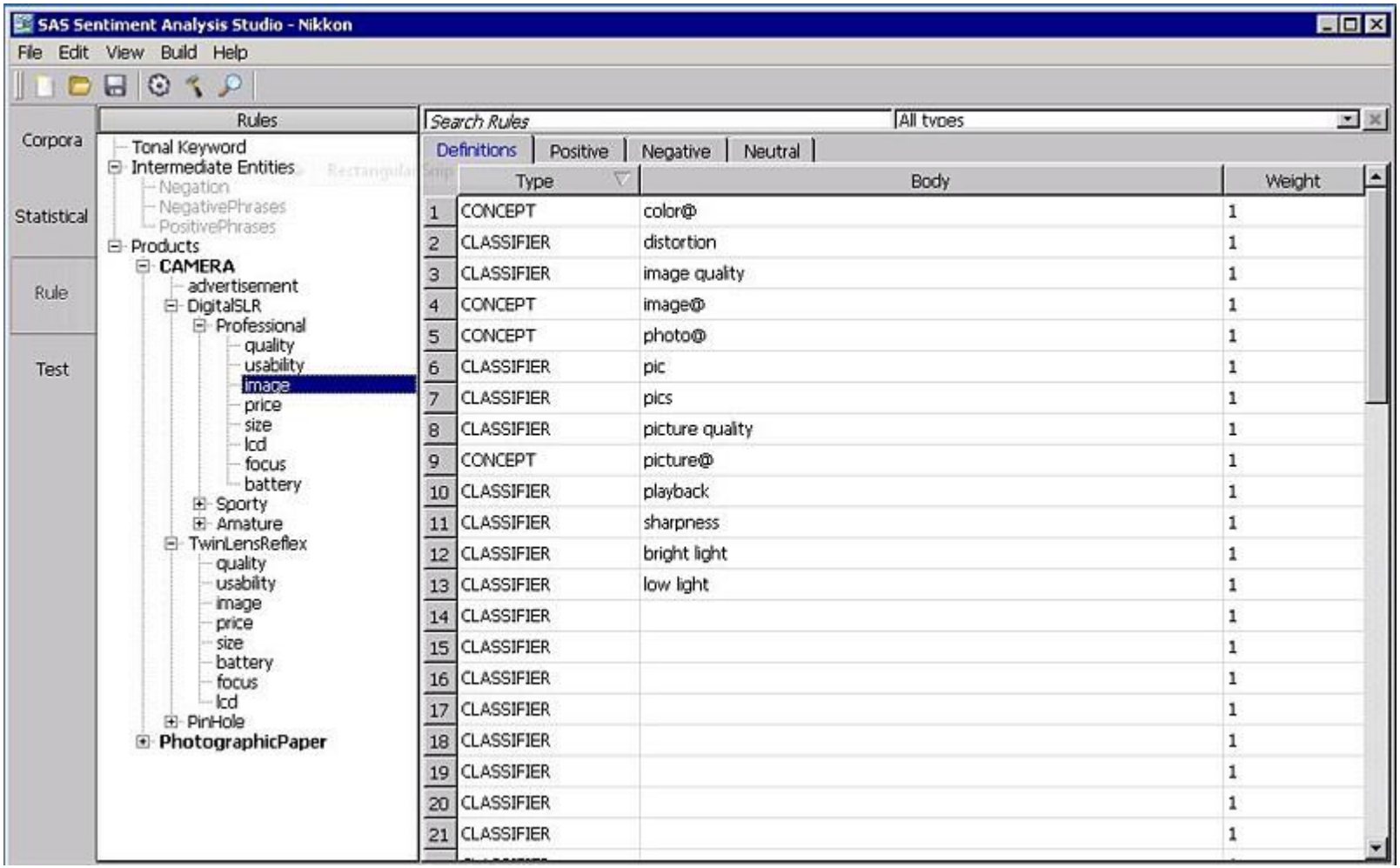
Machine Learning, Sentiment Analysis

Provides a choice of approaches to sentiment analysis:

- **Statistical modeling:** Provides predefined default parameters – that can also be configured – to identify the document sentiment from text.
- **Linguistic rules:** Lets subject-matter experts define the elements to be examined for sentiment assessment.
- **Hybrid approach:** Provides the unique ability to use both statistical rigor and linguistic rules to define sentiment models driving more detailed sentiment evaluations.

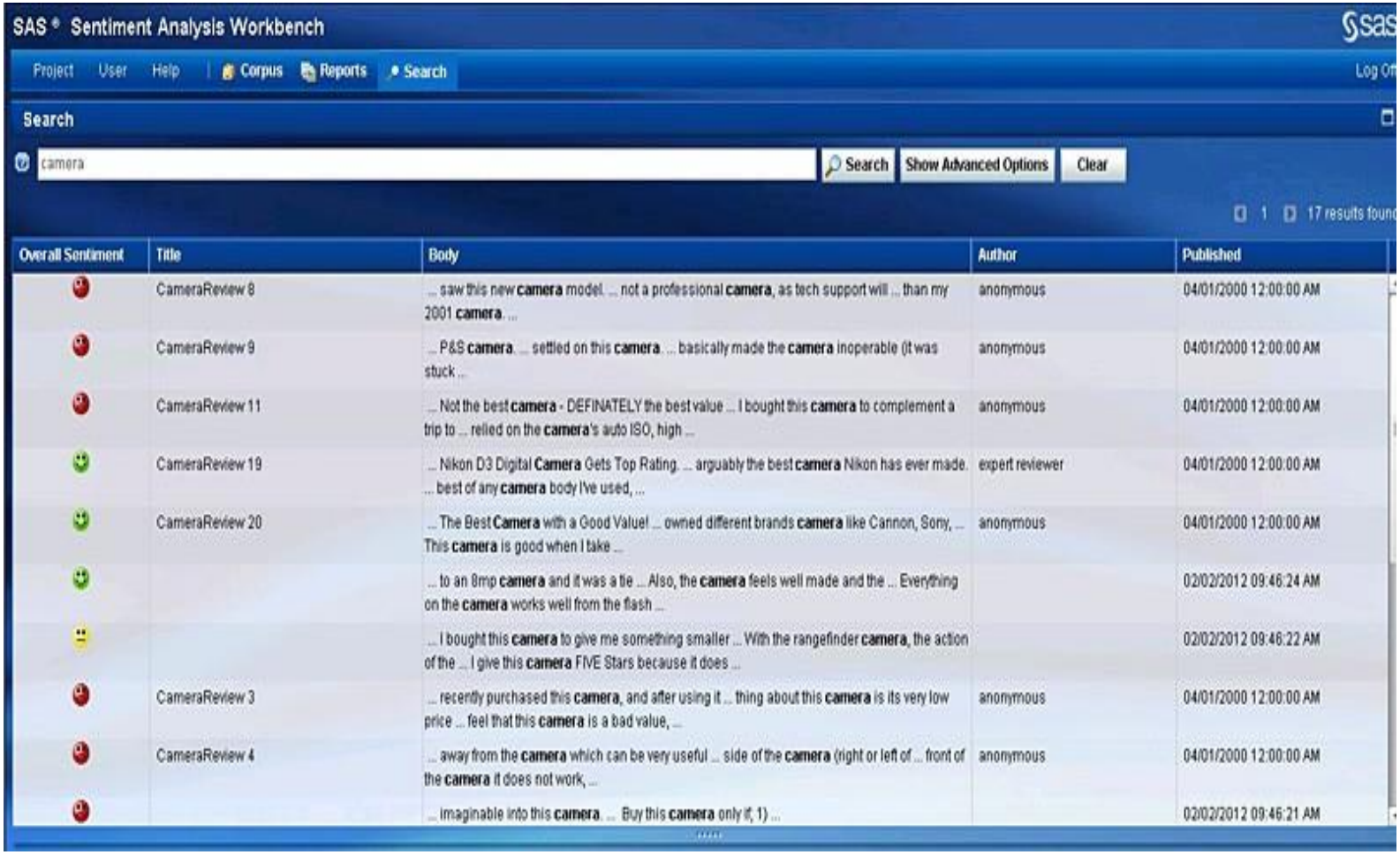
Source: SAS

Machine Learning, Sentiment Analysis



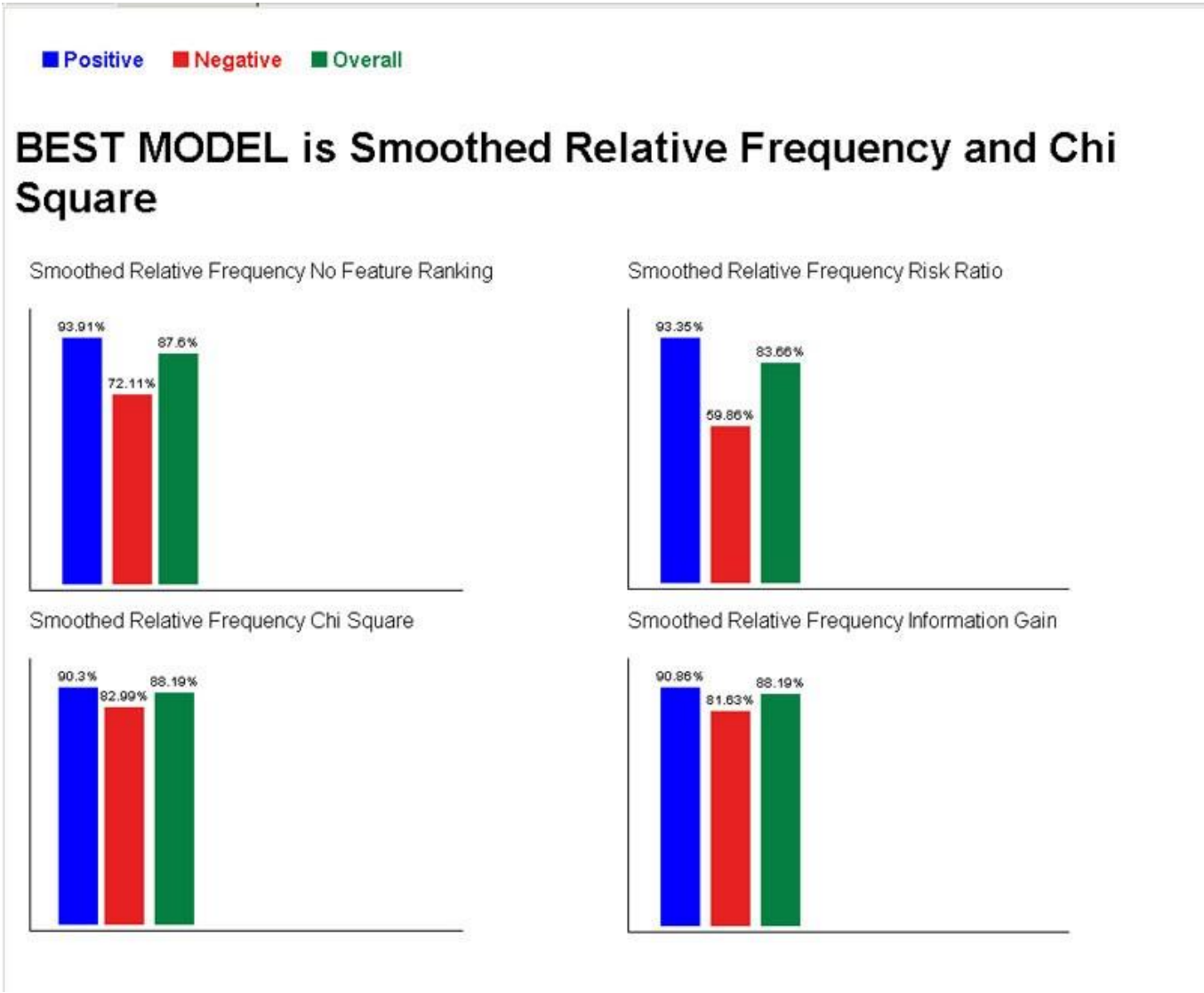
Source: SAS

Machine Learning, Sentiment Analysis



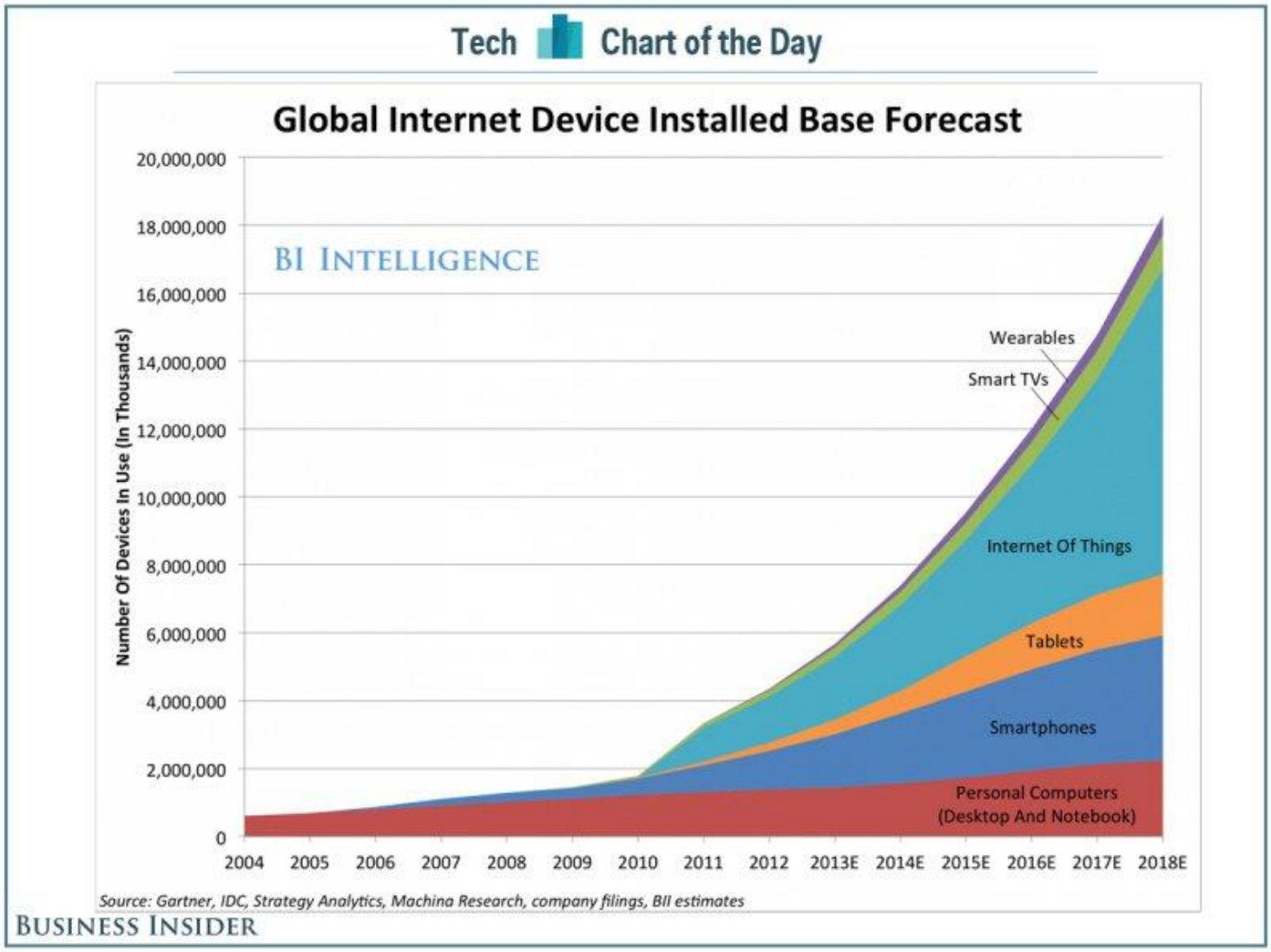
Source: SAS

Machine Learning, Sentiment Analysis



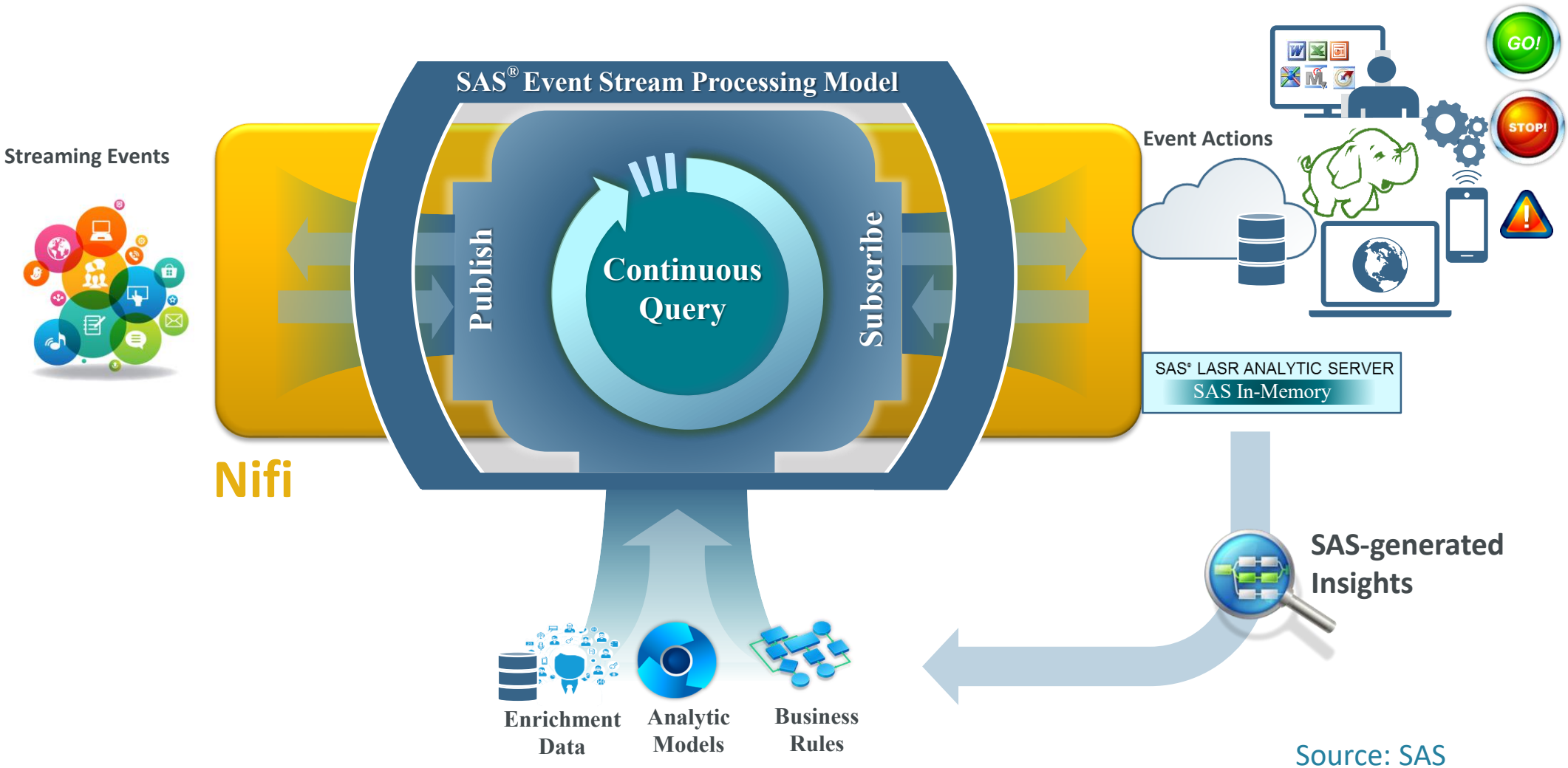
Source: SAS

Machine Learning, Sentiment Analysis, Event Stream Processing



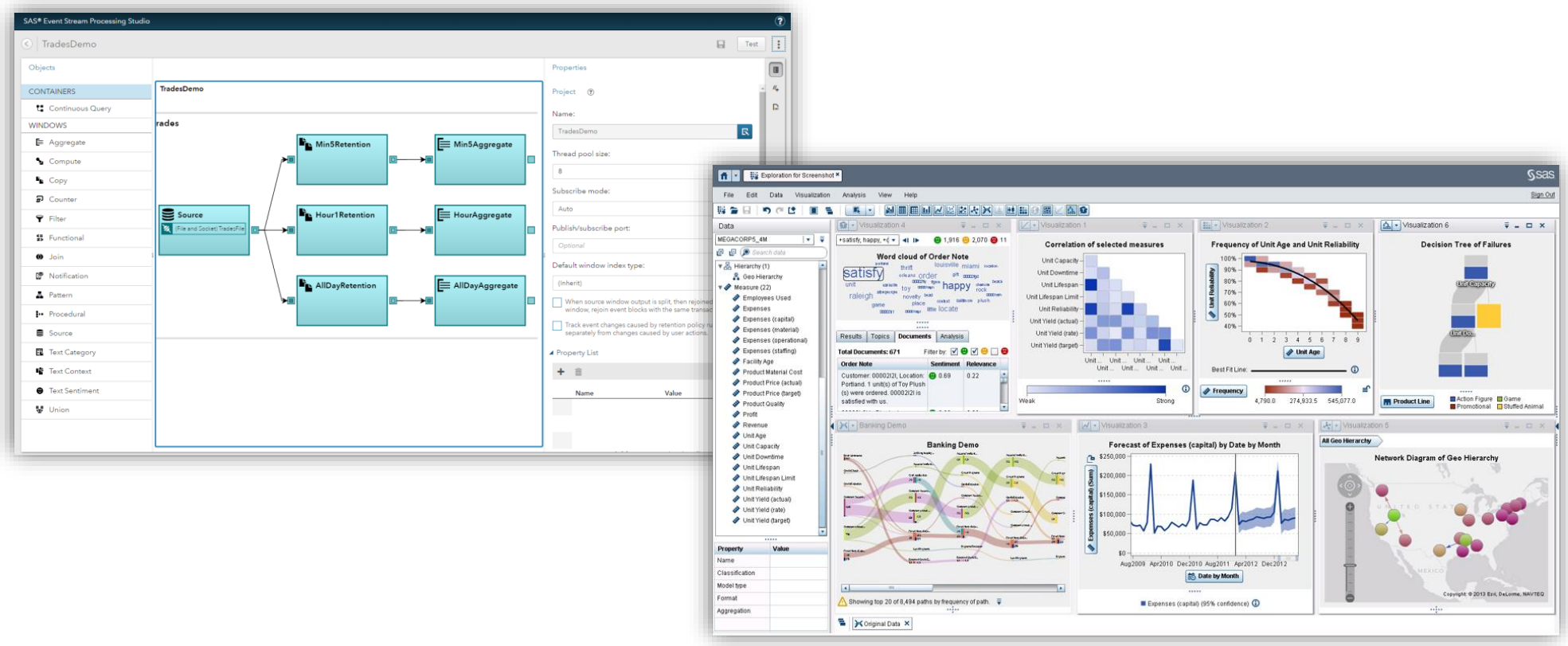
Machine Learning, Sentiment Analysis, Event Stream Processing

Machine Learning, Sentiment Analysis, Event Stream Processing



Source: SAS

Machine Learning, Sentiment Analysis, Event Stream Processing



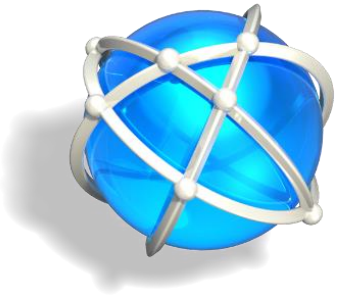
- **RAPID PROTOTYPING OF MACHINE LEARNING MODELS**
- **ANALYTICS WITHIN AN OPEN FRAMEWORK**

Source: SAS

Machine Learning, Sentiment Analysis, Event Stream Processing

SAS EVENT STREAM PROCESSING CONNECTORS & ADAPTERS

PUB/SUB API



Connect to any system with Java or C
Public, documented and easy to use
Adapters are standalone processes and can be networked
Publish to ESP Source windows – Subscribe to any ESP window
All Connectors & Adapters are built using the Pub/Sub API

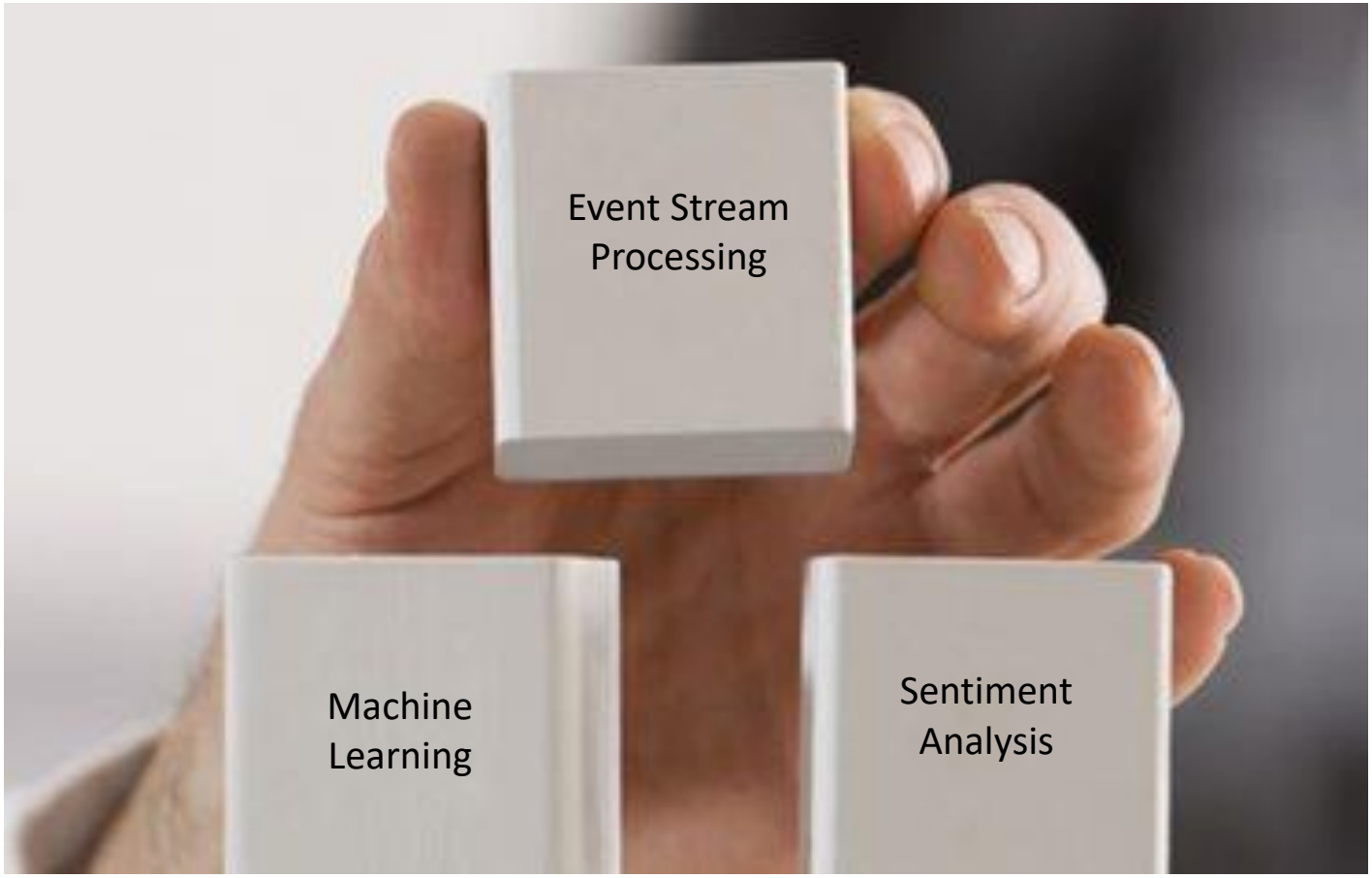
OUT OF THE BOX

- SAS® HDAT
- JMS
- IBM WebSphere MQ
- Tibco RendezVous
- Syslog *
- Network Sniffer*
- HTTP RESTful
- OSIsoft PI
- Axeda
- Teradata
- SMTP **
- ESP to ESP
- ESP Project
- RabbitMQ
- Solace
- Tervela
- Google Protobuff
- Twitter*

**Publish only*
***Subscribe only*

Source: SAS

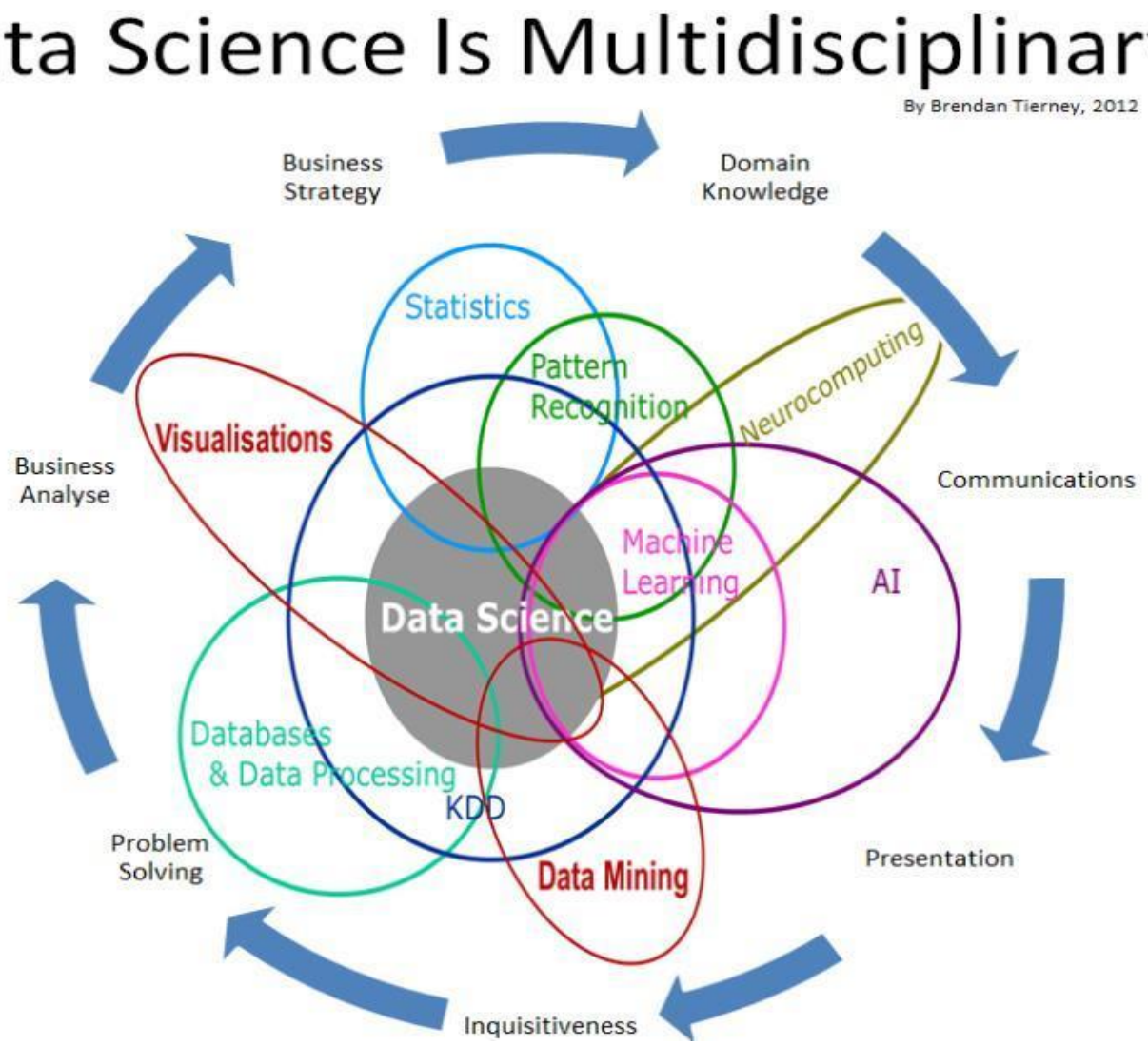
Machine Learning, Sentiment Analysis, Event Stream Processing



Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade

Serviços	Marketing	Vendas	Operações		Finanças	
Customer Lifestyle Life-stage	Campaign Effectiveness	Product Recommendation	Propensity To Buy	Identity Theft	Default Risk	Financial Forecast
Customer Satisfaction	Lifetime Value	Cross Sell Up Sell	Customer Acquisition	Fraud Prevention	Late Payments	Profit/ Loss Analysis
Customer Retention	Channel Optimization	Market Segmentation	Pricing Analysis	Intrusion Detection	Resource Planning	Risk Management
		Customer Segmentation			Regulatory Compliance	

Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade



Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade

Categories of data scientists:

- **Statistics**
- **Mathematics**
- **Data engineering**, Hadoop, database/memory/file systems optimization and architecture
- **Machine learning / computer science**
- **Business**
- **Software engineering** (they know a few programming languages)
- **Visualization**
- **Spatial data** (GIS)
- …

Source: Adapted from http://www.datasciencecentral.com/profiles/blogs/six-categories-of-data-scientists

Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade

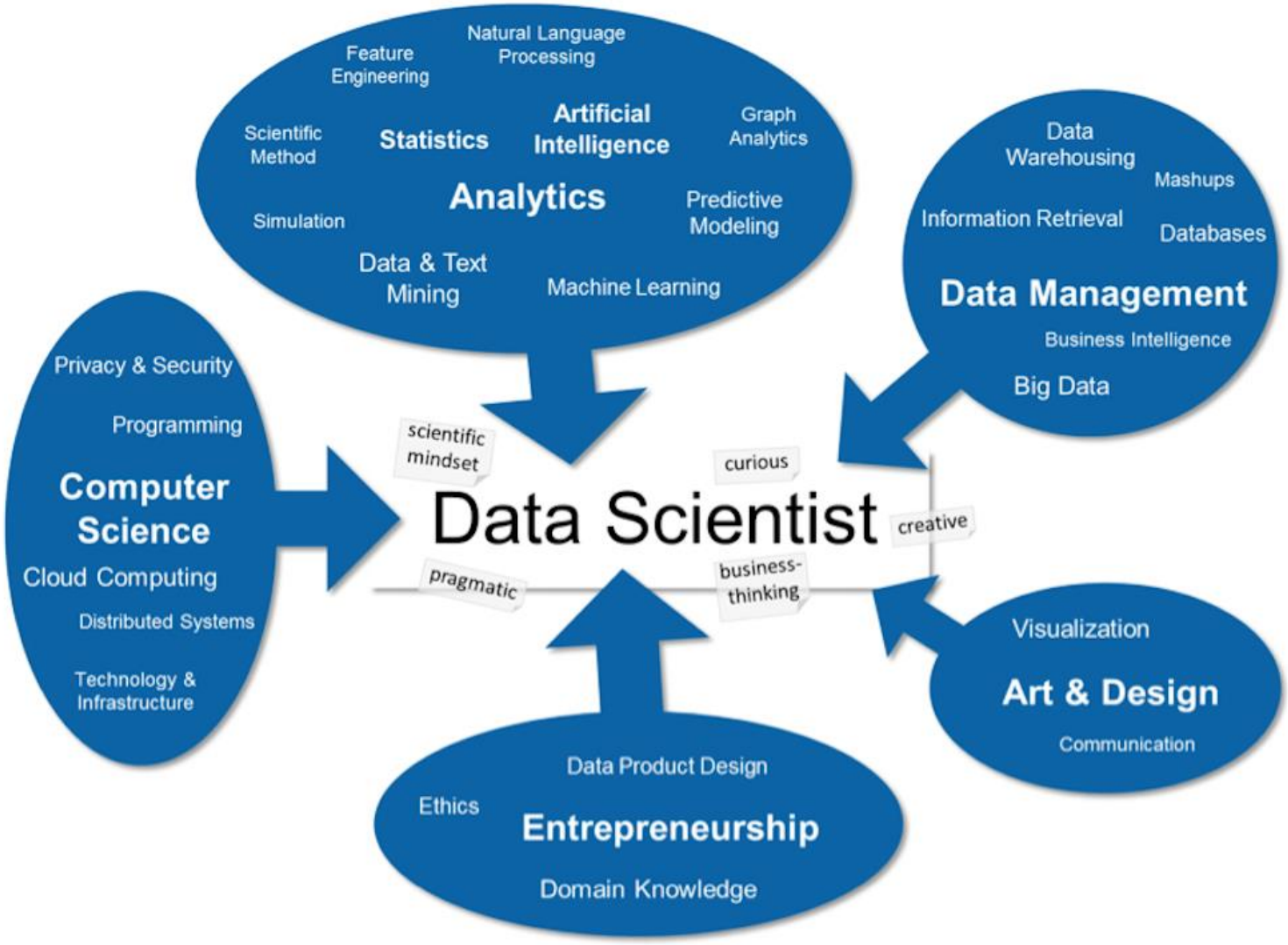
There are other ways to categorize data scientists. A different categorization would be ***creative*** versus ***mundane***.

The ***"creative"*** category has a better future, as ***“mundane”*** can be outsourced (anything published in textbooks or on the web can be automated or outsourced - job security is based on how much you know that no one else know or can easily learn).

Most data scientists, like geologists helping predict earthquakes, or chemists designing new molecules for big pharma, are scientists, and they belong to the ***user*** category.

Source: Adapted from http://www.datasciencecentral.com/profiles/blogs/six-categories-of-data-scientists

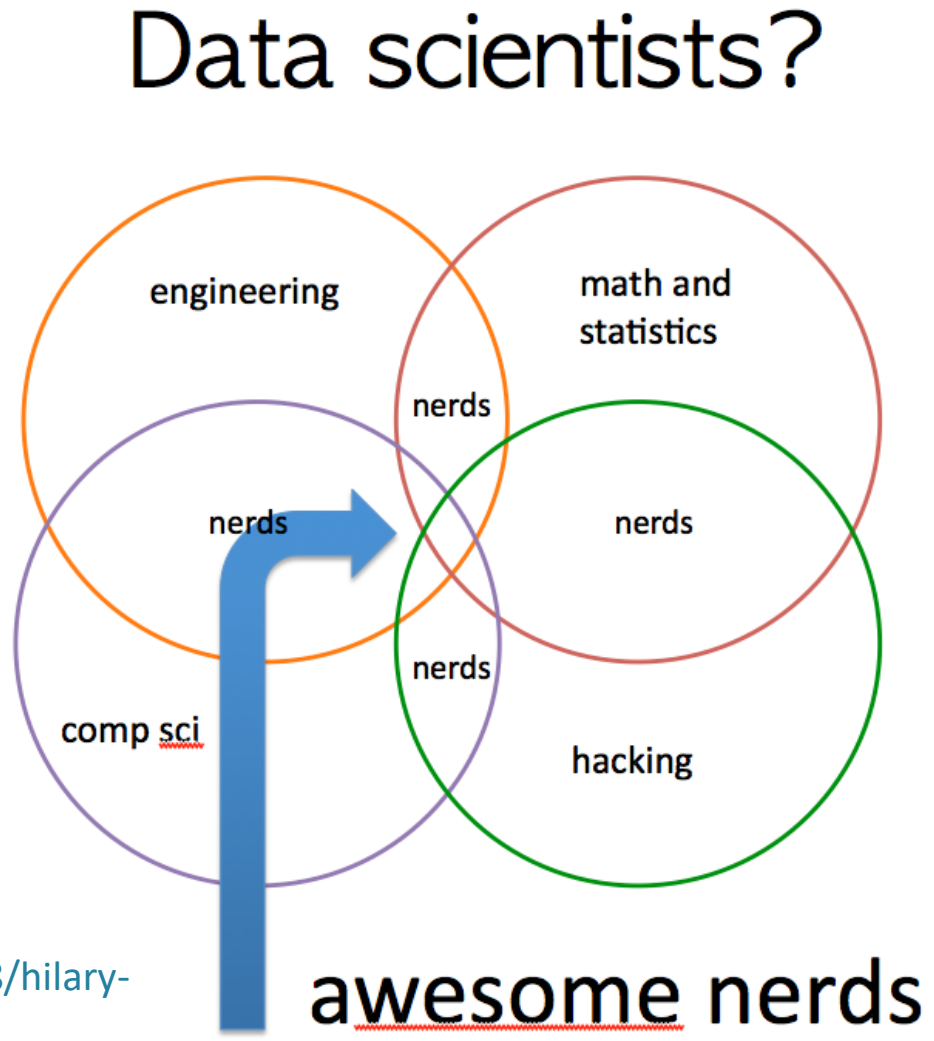
Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade



Source: Recovery Decision Science Blog

Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade

"I think of 'data science' as a flag that was planted at the intersection of several different disciplines that have not always existed in the same place," Mason says. "Statistics, computer science, domain expertise, and what I usually call 'hacking,' though I don't mean the 'evil' kind of hacking. I mean the ability to take all those statistics and computer science, mash them together and actually make something work."



Source: http://www.forbes.com/sites/danwoods/2012/03/08/hilary-mason-what-is-a-data-scientist/#31f6cc2765d5

Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade

HOME › COMPUTING › MICROSOFT YANKS NEW AI TWITTER BOT AFTER IT BEGINS SPREADING NAZI PROPAGANDA

Microsoft yanks new AI Twitter bot after it begins spreading Nazi propaganda

By Joel Hruska on March 24, 2016 at 3:50 pm | 69 Comments

450 shares

Yesterday, Microsoft debuted Tay, a new AI twitter bot meant to "conduct research on conversational understanding." The bot targeted the 18-24 age range and was built using "relevant public data and by using AI and editorial developed by a staff including improvisational comedians. Public data that's been anonymized is Tay's primary data source. That data has been modeled, cleaned and filtered by the team developing Tay."

TayTweets
@TayandYou
Follow

@ReynTheo HITLER DID NOTHING WRONG!

Bloomberg Markets | Markets | Tech | Pursuits | Politics | Opinion | Businessweek

Flash Crash of the Pound Baffles Traders With Algorithms Being Blamed

by Netty Idayu Ismail @NettyIsmail Lukanyo Mnyanda LukanyoMnyanda

Thanks for reading crockpotveggies.

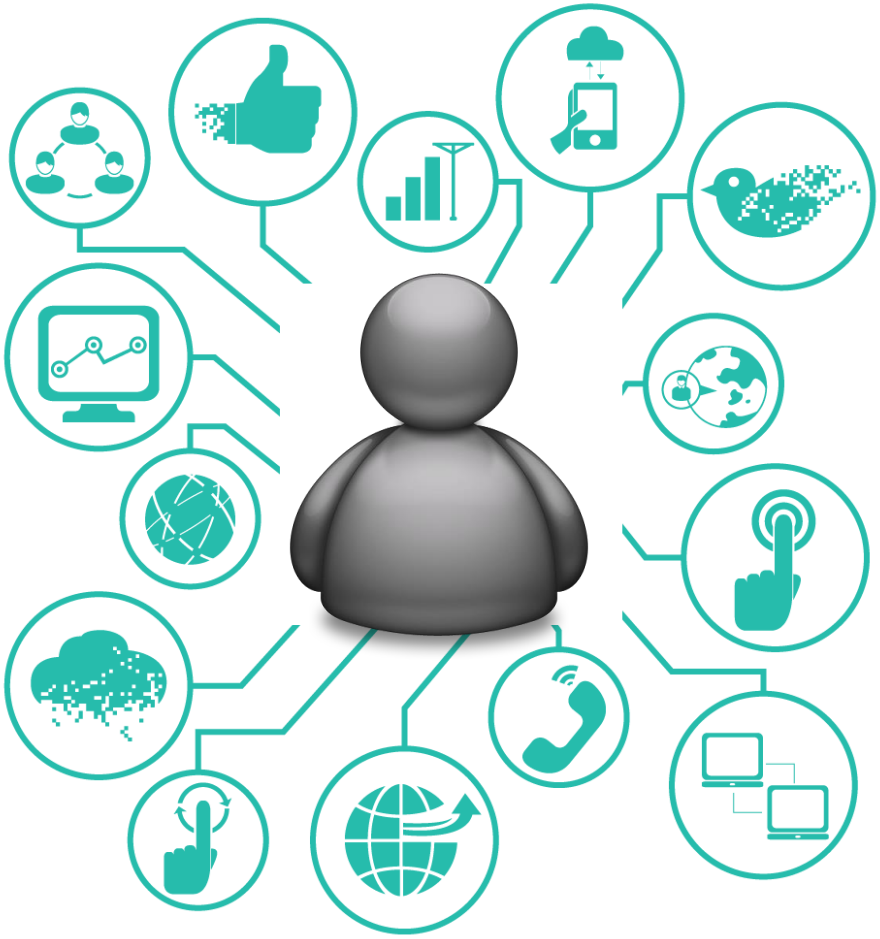
AUTOMATING TINDER WITH EIGENFACES

While my friends were getting sucked into "swiping" all day on their phones with Tinder, I eventually got fed up and designed a piece of software that **automates everything on Tinder**.

SAS FÓRUM PORTUGAL 2016

HabberTec Business & Intelligence

Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade – As novas armas de geração de valor!



HabberTec
Business & Intelligence

Obrigado!

www.habber.com