



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

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...

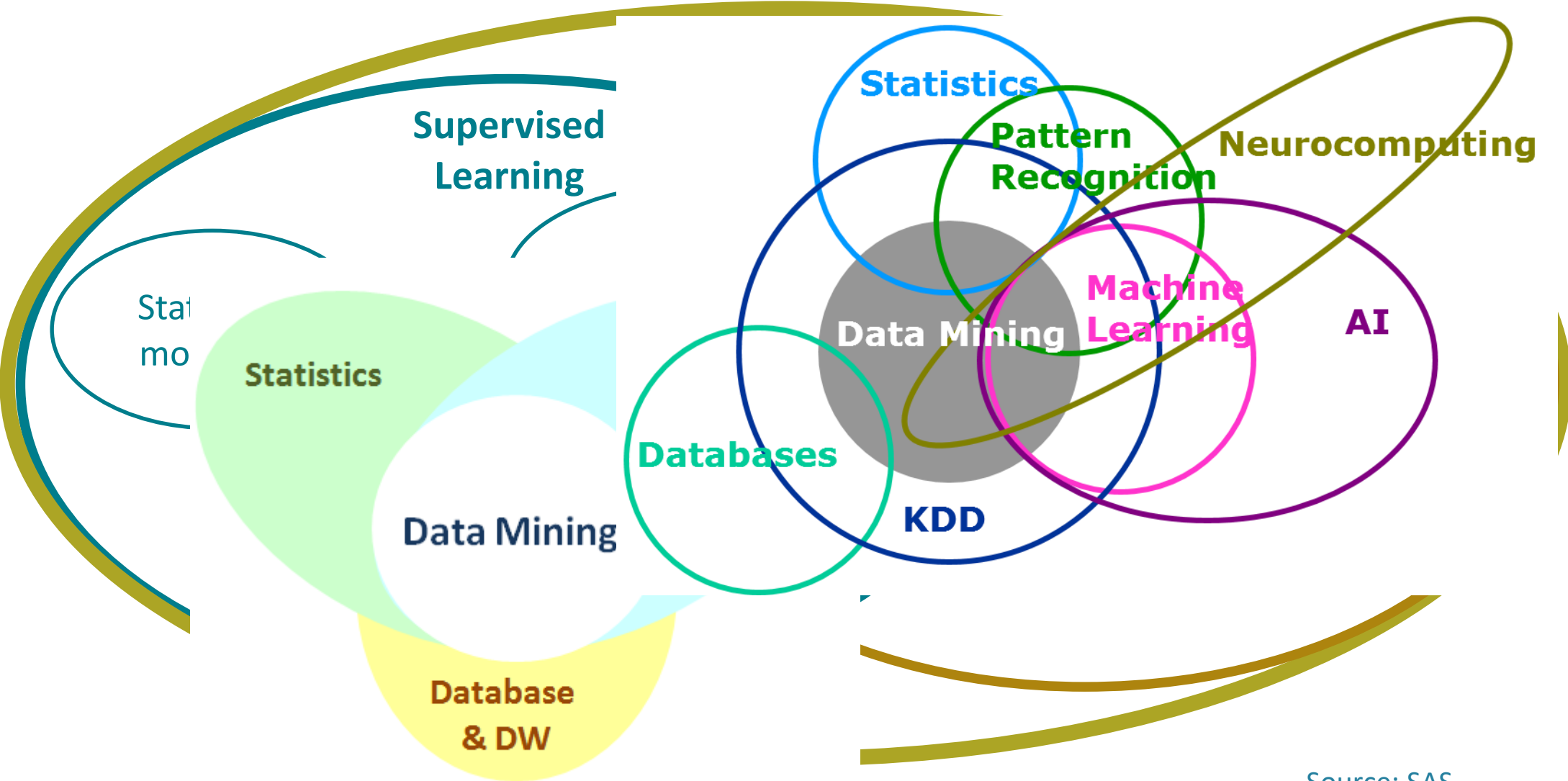
Like · Comment · Share

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

Machine Learning



Source: MachineLearningMastery.com

Machine Learning

CONS

- ✓ Unfamiliar with broader audience, (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) necessary (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

SAS Sentiment Analysis Studio - Nikon

File Edit View Build Help

Rules

Corpora

- Tonal Keyword
- Intermediate Entities
 - Negation
 - NegativePhrases
 - PositivePhrases
- Products
 - CAMERA
 - advertisement
 - DigitalSLR
 - Professional
 - quality
 - usability
 - Image
 - price
 - size
 - lcd
 - focus
 - battery
 - Sporty
 - Amature
 - TwinLensReflex
 - quality
 - usability
 - image
 - price
 - size
 - battery
 - focus
 - lcd
 - PinHole
 - PhotographicPaper

Search Rules

All types

Definition	Type	Positive	Negative	Neutral	Body	Weight
1	CONCEPT		color@			1
2	CLASSIFIER		distortion			1
3	CLASSIFIER		image quality			1
4	CONCEPT		image@			1
5	CONCEPT		photo@			1
6	CLASSIFIER		pic			1
7	CLASSIFIER		pics			1
8	CLASSIFIER		picture quality			1
9	CONCEPT		picture@			1
10	CLASSIFIER		playback			1
11	CLASSIFIER		sharpness			1
12	CLASSIFIER		bright light			1
13	CLASSIFIER		low light			1
14	CLASSIFIER					1
15	CLASSIFIER					1
16	CLASSIFIER					1
17	CLASSIFIER					1
18	CLASSIFIER					1
19	CLASSIFIER					1
20	CLASSIFIER					1
21	CLASSIFIER					1

Source: SAS

Machine Learning, Sentiment Analysis



SAS Sentiment Analysis Workbench

Project User Help Corpus Reports Search

Search

camera Search Show Advanced Options Clear

1 17 results found

Overall Sentiment	Title	Body	Author	Published
☹️	CameraReview 8	... saw this new camera model ... not a professional camera, as tech support will ... than my 2001 camera ...	anonymous	04/01/2000 12:00:00 AM
☹️	CameraReview 9	... P&S camera ... settled on this camera ... basically made the camera inoperable (it was stuck ...	anonymous	04/01/2000 12:00:00 AM
☹️	CameraReview 11	... Not the best camera - DEFINATELY the best value ... I bought this camera to complement a trip to ... relied on the camera's auto ISO, high ...	anonymous	04/01/2000 12:00:00 AM
😊	CameraReview 19	... Nikon D3 Digital Camera Gets Top Rating ... arguably the best camera Nikon has ever made ... best of any camera body I've used, ...	expert reviewer	04/01/2000 12:00:00 AM
😊	CameraReview 20	... The Best Camera with a Good Value! ... owned different brands camera like Cannon, Sony, ... This camera is good when I take ...	anonymous	04/01/2000 12:00:00 AM
😊		... to an 8mp camera and it was a tie ... Also, the camera feels well made and the ... Everything on the camera works well from the flash ...		02/02/2012 09:46:24 AM
😐		... I bought this camera to give me something smaller ... With the rangefinder camera, the action of the ... I give this camera FIVE Stars because it does ...		02/02/2012 09:46:22 AM
☹️	CameraReview 3	... recently purchased this camera, and after using it ... thing about this camera is its very low price ... feel that this camera is a bad value, ...	anonymous	04/01/2000 12:00:00 AM
☹️	CameraReview 4	... away from the camera which can be very useful ... side of the camera (right or left of ... front of the camera it does not work, ...	anonymous	04/01/2000 12:00:00 AM
☹️		... imaginable into this camera ... Buy this camera only if; 1) ...		02/02/2012 09:46:21 AM

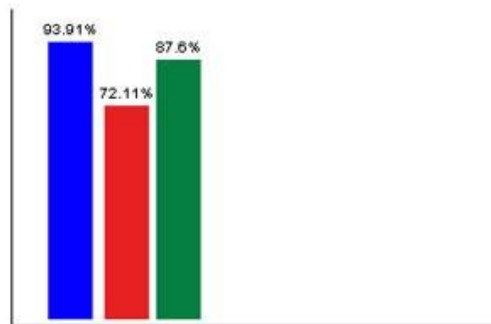
Source: SAS

Machine Learning, Sentiment Analysis

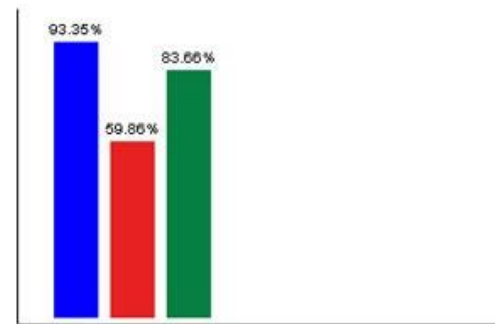
■ Positive ■ Negative ■ Overall

BEST MODEL is Smoothed Relative Frequency and Chi Square

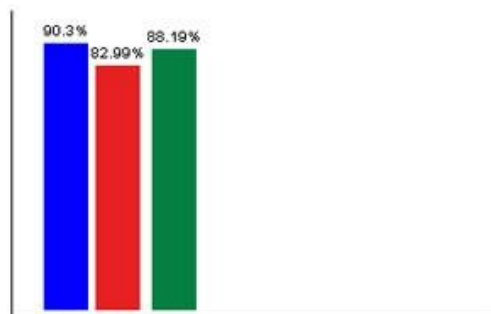
Smoothed Relative Frequency No Feature Ranking



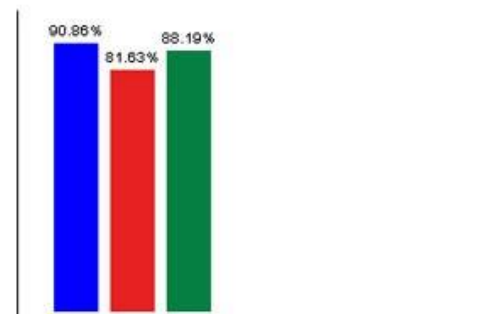
Smoothed Relative Frequency Risk Ratio



Smoothed Relative Frequency Chi Square

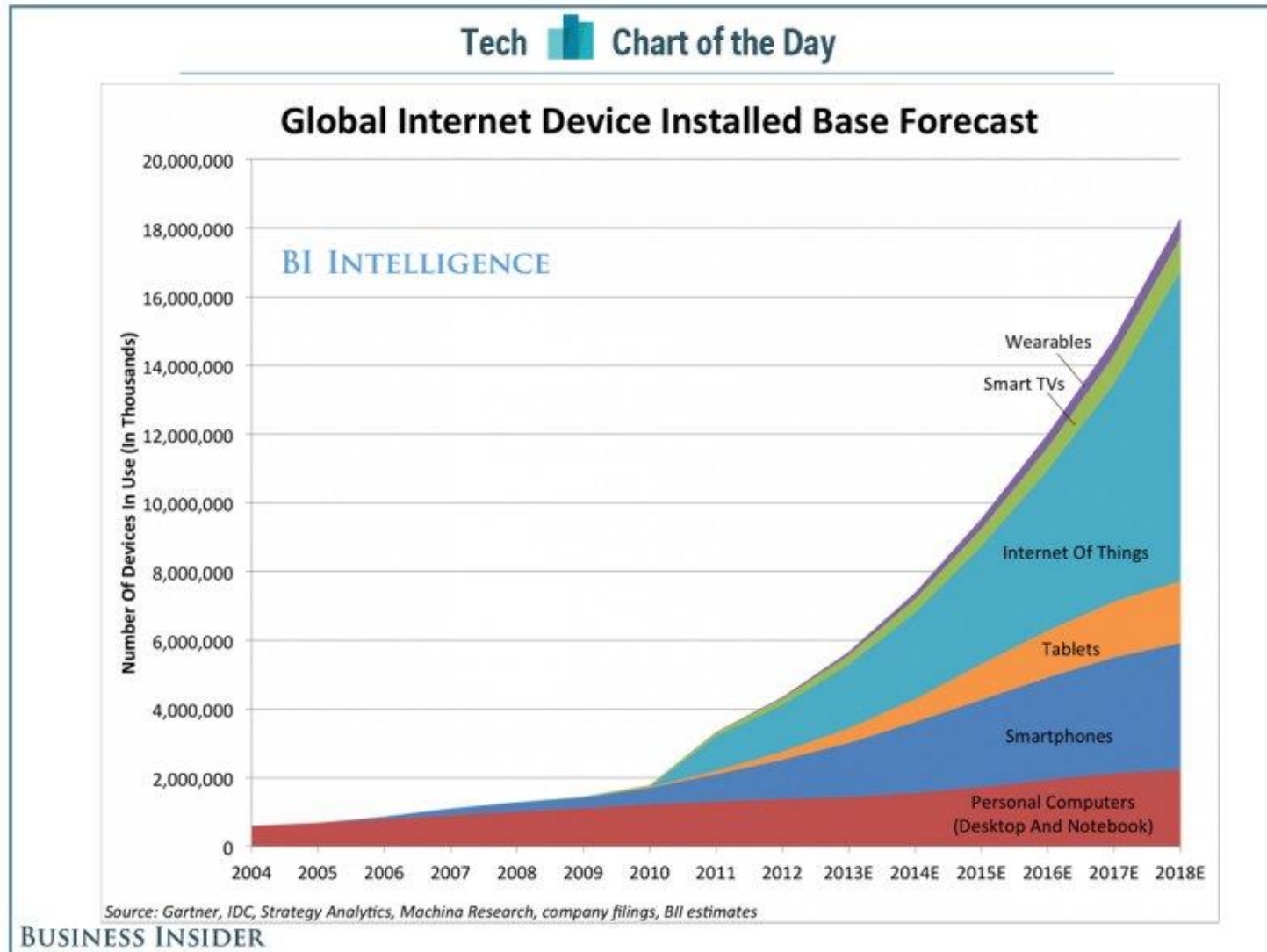


Smoothed Relative Frequency Information Gain



Source: SAS

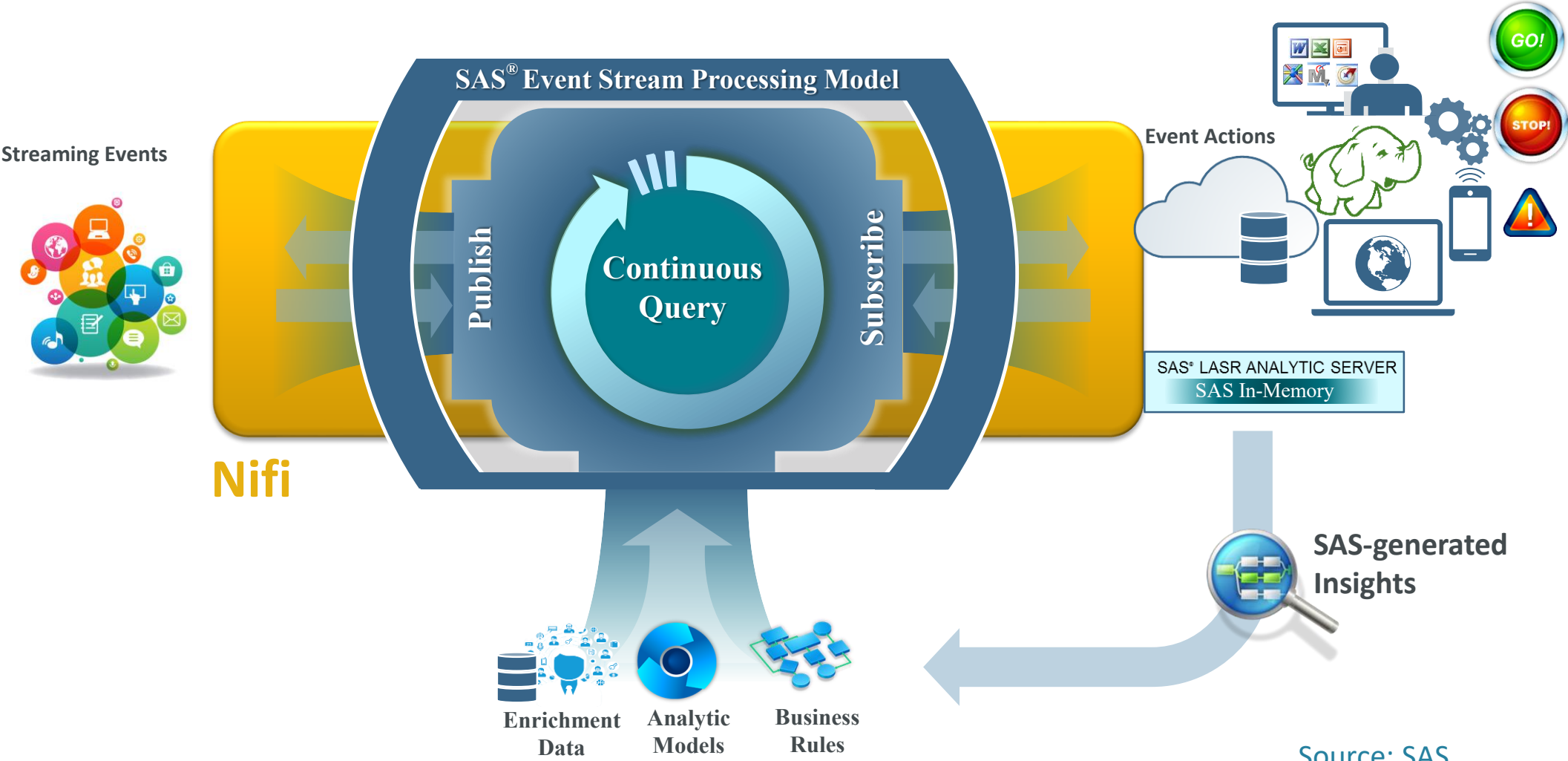
Machine Learning, Sentiment Analysis, Event Stream Processing



Machine Learning, Sentiment Analysis, Event Stream Processing

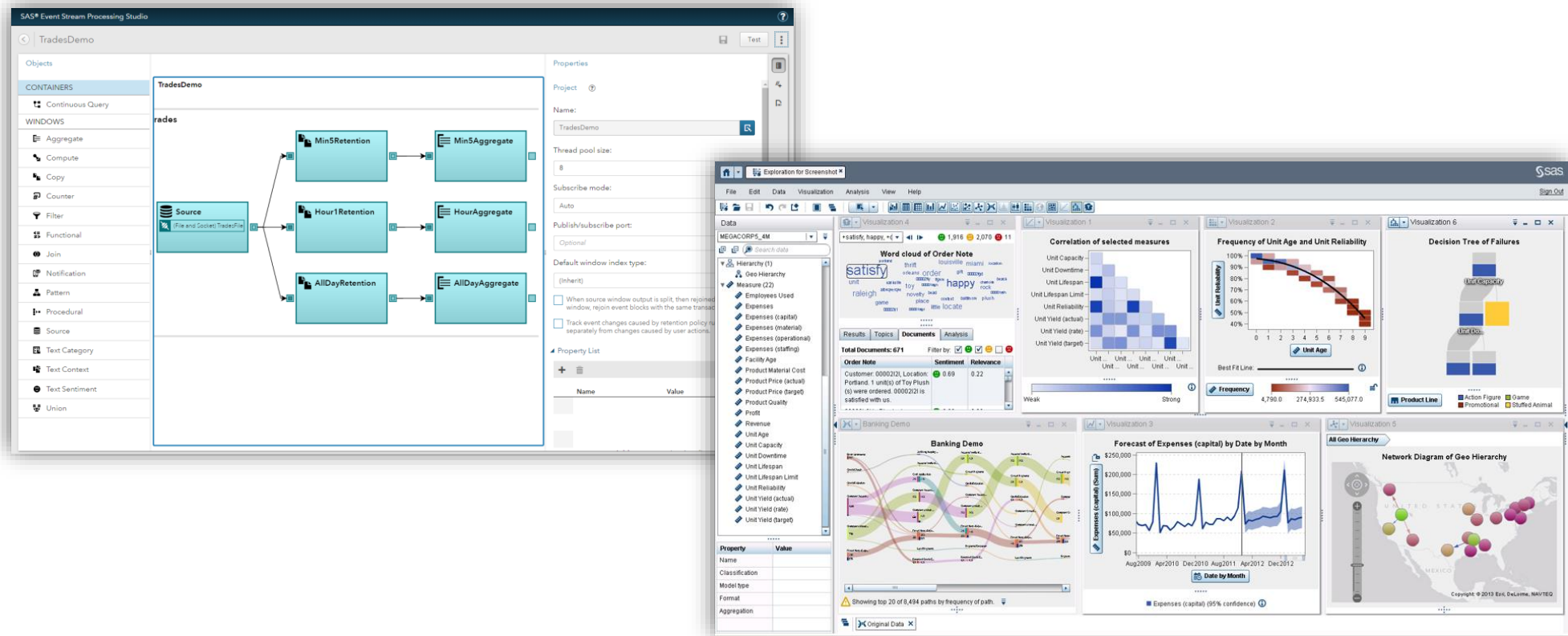


Machine Learning, Sentiment Analysis, Event Stream Processing



Source: SAS

Machine Learning, Sentiment Analysis, Event Stream Processing



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

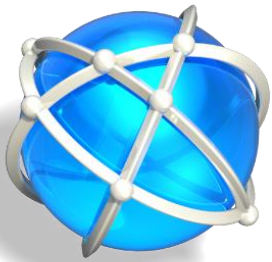
Source: SAS

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

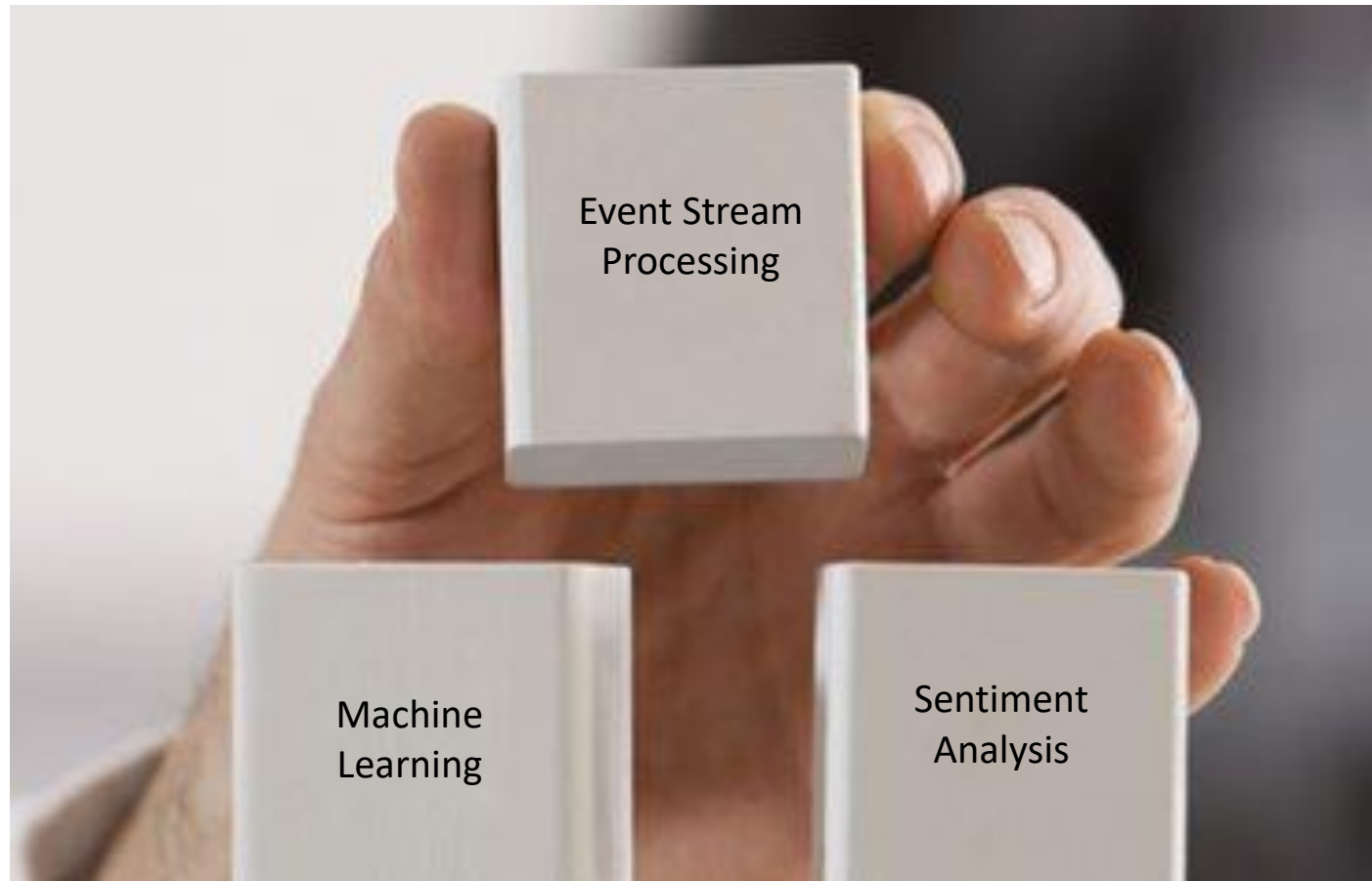
OUT OF THE BOX

- SAS® HDAT
- JMS
- IBM WebSphere MQ
- Tibco RendezVous
- Syslog *
- Network Sniffer*
- HTTP RESTful
- OS/soft 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

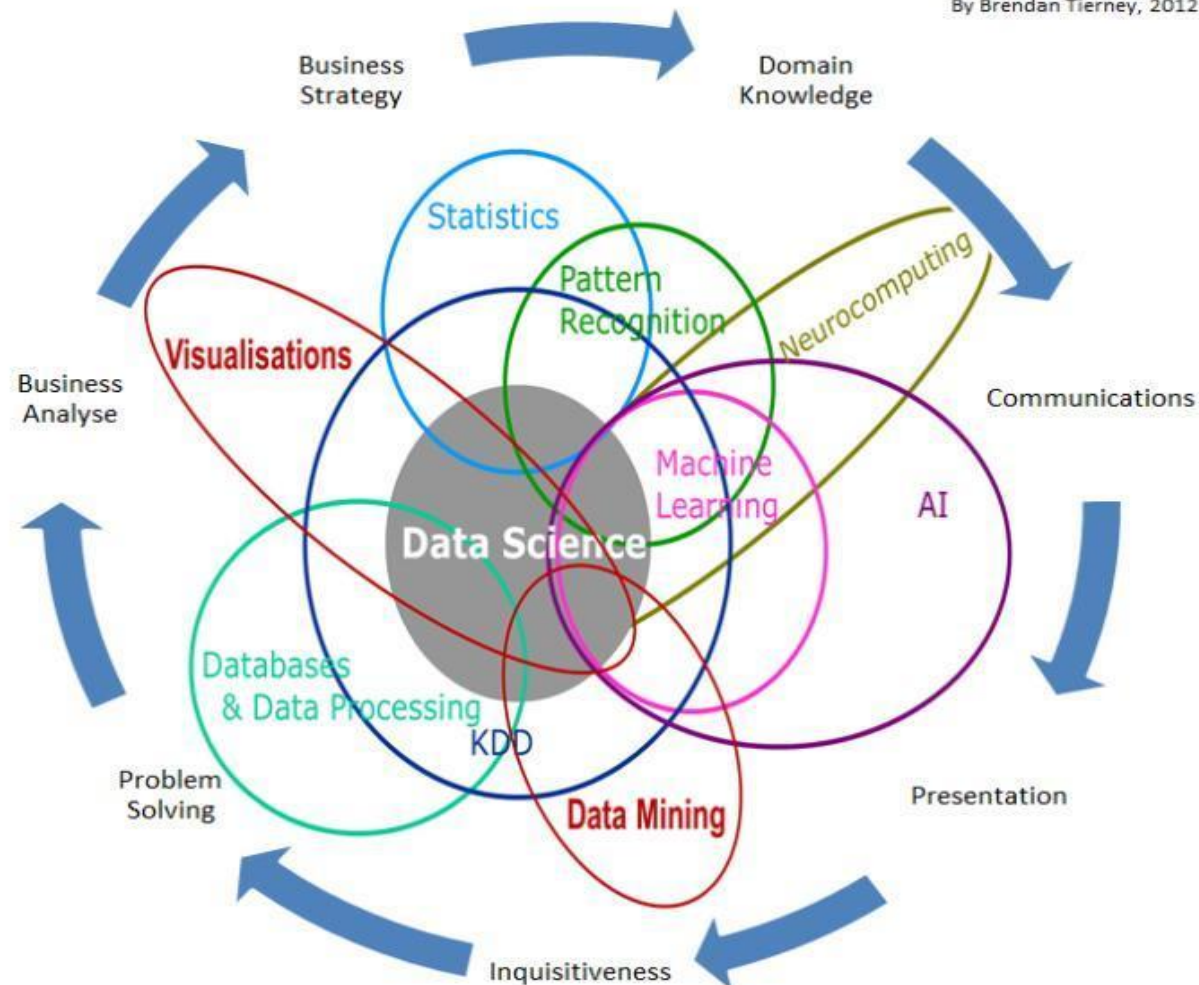
Regulatory Compliance



Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade

Data Science Is Multidisciplinary

By Brendan Tierney, 2012



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

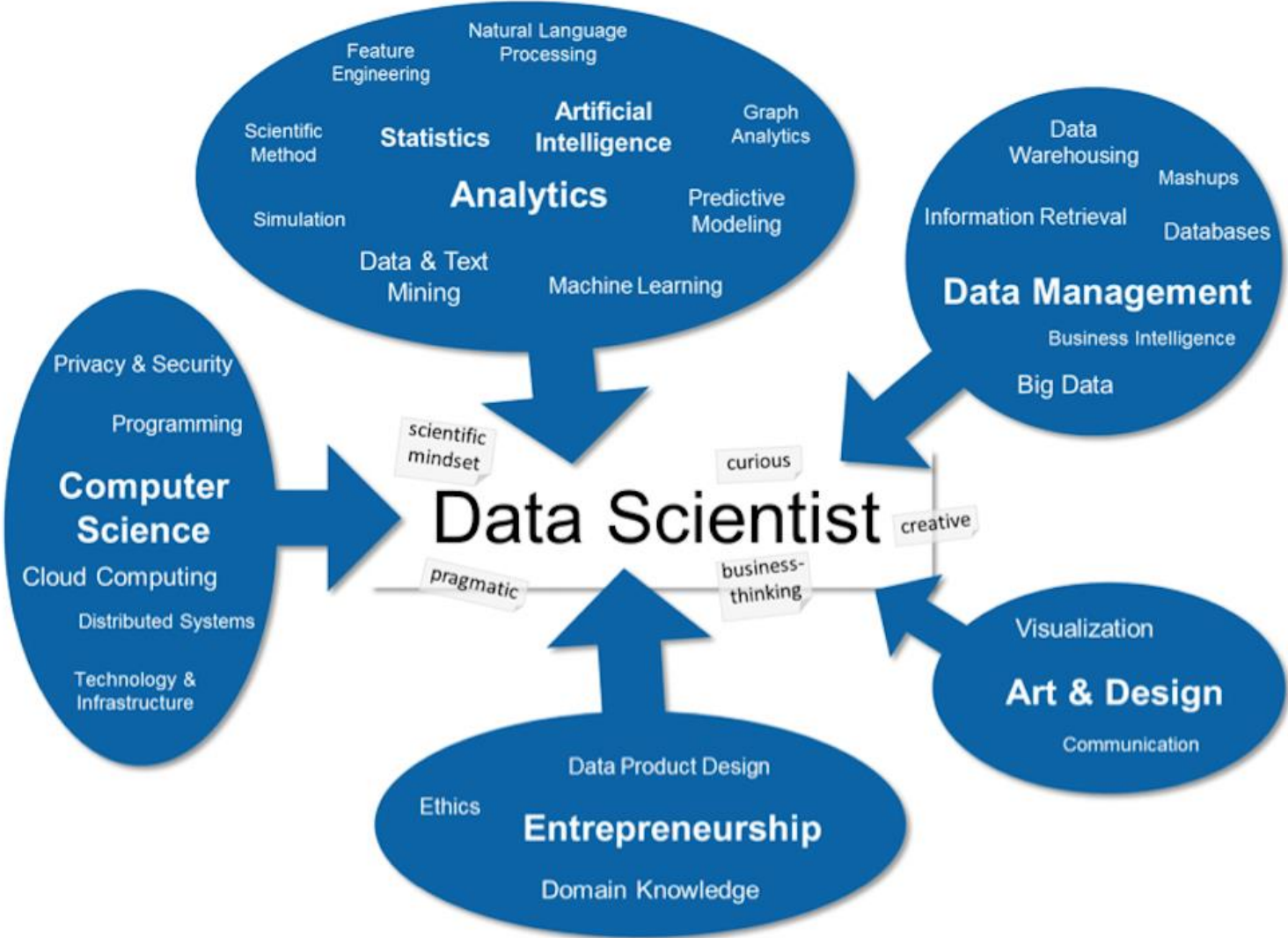
There are other ways to categorize data scientists. A different categorization would be *creative* versus *mundane*.

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.



Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade



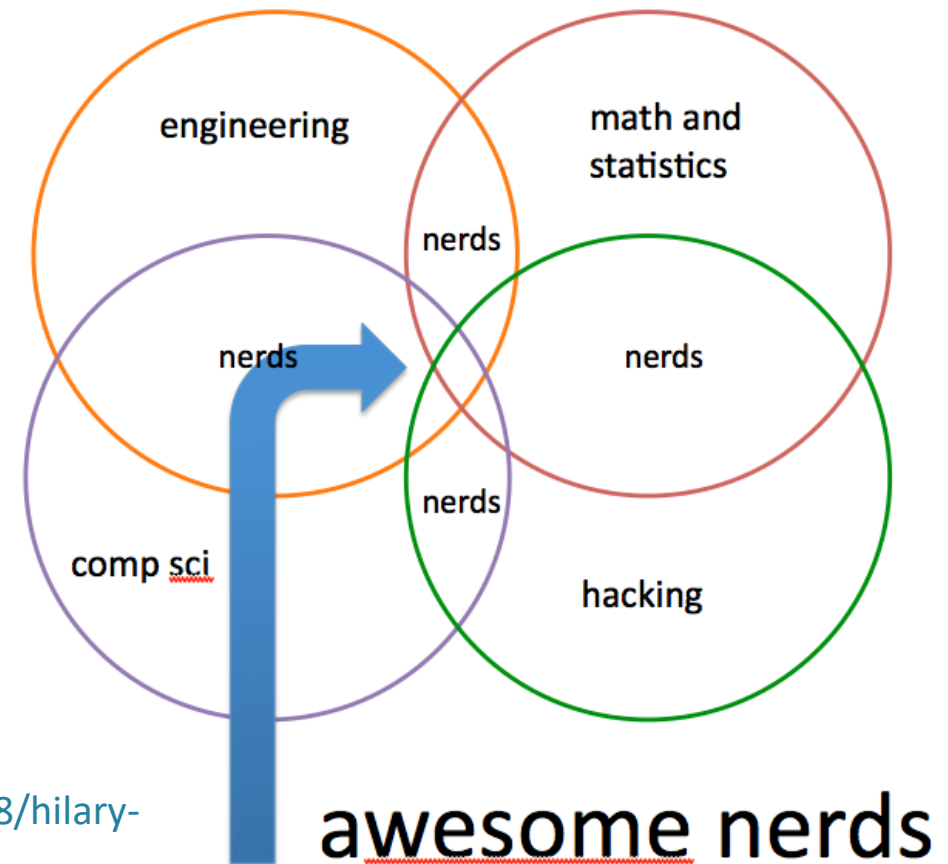
Source: Recovery Decision Science Blog

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>

Data scientists?



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     



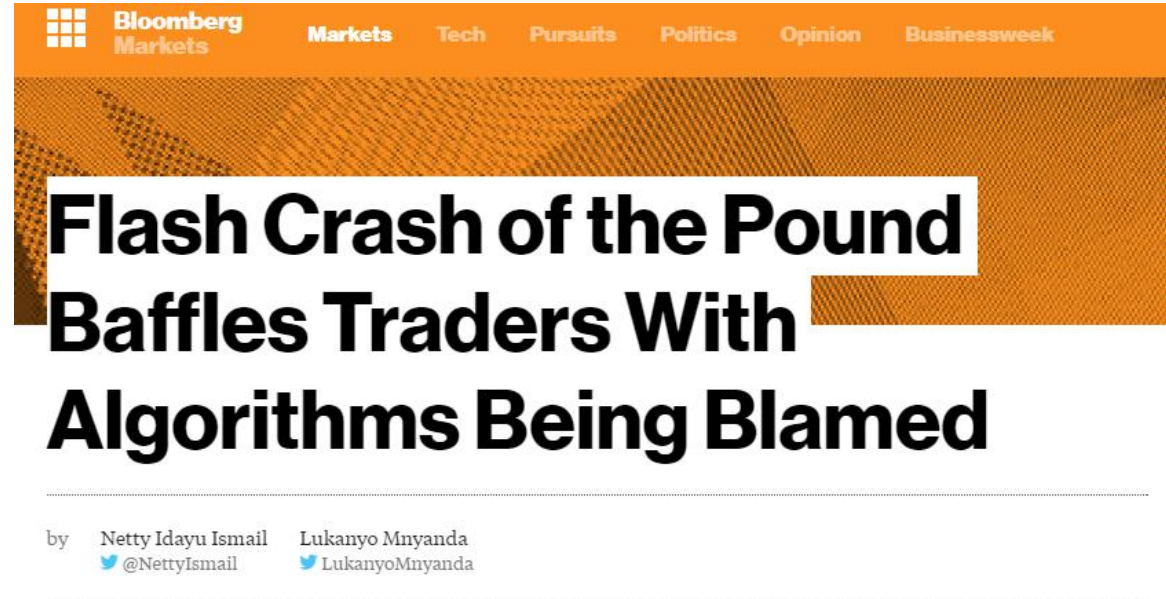
Yesterday, Microsoft debuted Tay, a new AI twitter bot meant to “conduct research on conversational understanding.” The bot targeted the 18-24 age range and was built using “relevant public data and by using AI and editorial developed by a staff including improvisational comedians. Public data that’s 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!



Flash Crash of the Pound Baffles Traders With Algorithms Being Blamed

by [Netty Idayu Ismail](#) [Lukanyo Mnyanda](#)
[@NettyIsmail](#) [LukanyoMnyanda](#)

Thanks for reading *crochpotveggies*.

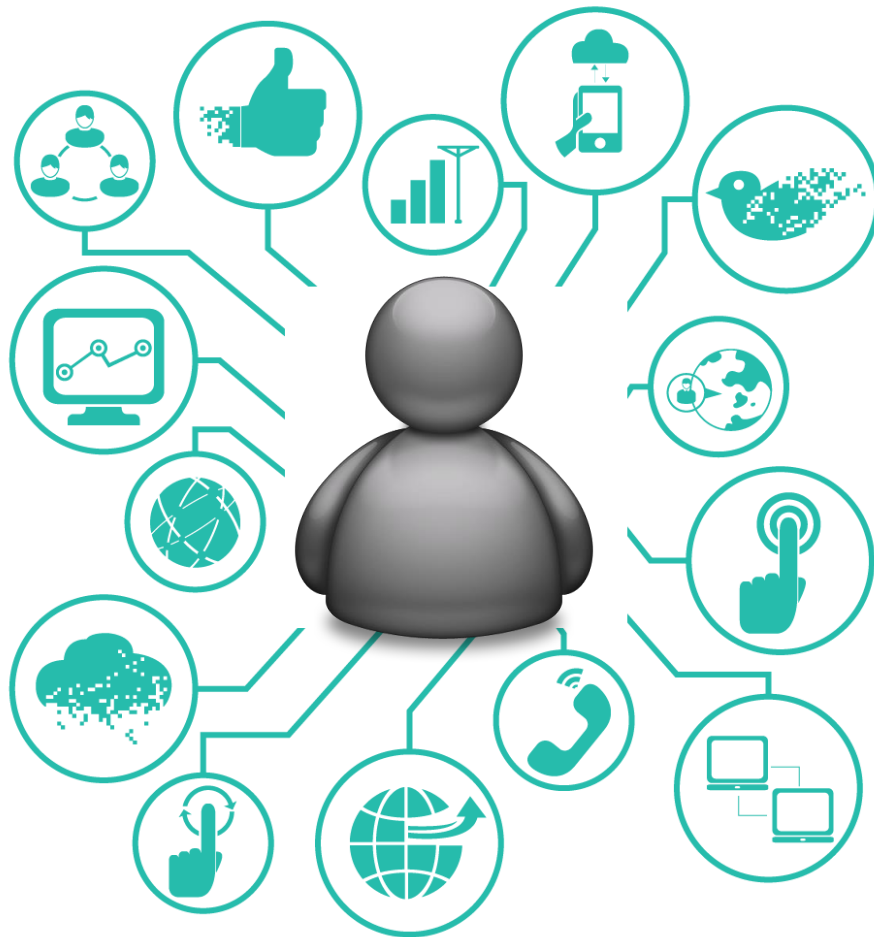


AUTOMATING TINDER WITH EIGENFACES

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.**

Machine Learning, Sentiment Analysis, Event Stream Processing e Criatividade

– As novas armas de geração de valor!



HabberTec
Business & Intelligence



Obrigado!

www.habber.com