

Data science for banking

R&D in the financial industry

The financial services industry is facing 4 major data science challenges for which Chappuis Halder has been addressing

“Statistics are ubiquitous in life, and so should be statistical reasoning.”

Alan Blinder, former Federal Reserve vice chairman, NYTimes

Machine Learning Challenges

1. Storage and processing	2. Data governance	3. Data visualization	4. Data Science	Data science is . . .
<ul style="list-style-type: none"> › IT infrastructure › Storage and processing (Hadoop, Spark) › Data collection 	<ul style="list-style-type: none"> › Data quality › Data Compliance › Data management 	<ul style="list-style-type: none"> › Strategic axis / indicators › Reporting format / content › Process / Tools / Data 	<ul style="list-style-type: none"> › Prediction / Anticipation & simulation › Estimation › Ranking/Discrimination › Behavioural analysis › Self-learning models 	<ul style="list-style-type: none"> › . . . driven by the Big Data revolution, the emergence of new technologies, the development of “new” techniques and new business strategies › . . . an interdisciplinary field whose purpose is to make the data speak. It can be used for prediction, rating & discrimination, anticipation & simulation, behavioural analysis, etc.
<p>How CH&Co. sees a data scientist in the FS ?</p>		<p>The financial services industry needs to address 4 key issues:</p> <ul style="list-style-type: none"> › How to extract the value of the data? › What techniques and methodologies should be used and for what? › How can machine learning better support the business strategy? › How to organise / structure internally to address this challenge? 		

→ Deliver customized solutions through our R&D results straddled with our IT capacity held by CH&Co Lab

R&D

What R&D is according to the GRA.

“Data science is the transformation of data using mathematics and statistics into valuable insights, decision and products”

John W. Foreman

Capturing and analysing data, building predictive models and running simulations of financial events is complex and important but there is an even bigger question.

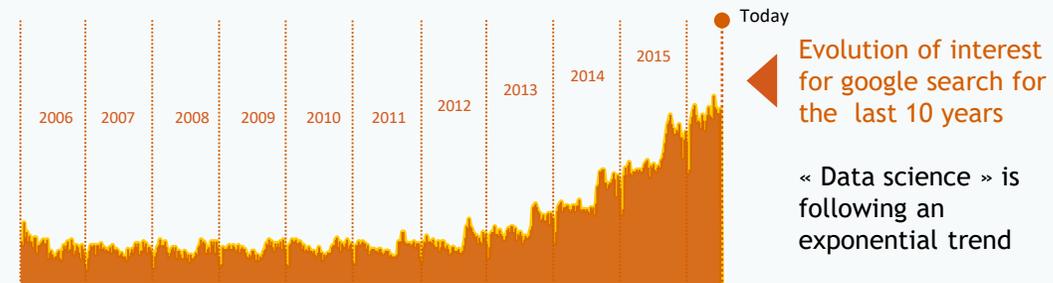
The first priority and biggest challenge is to find the question “What do our customers care about?”

That is why the CH&Co.’s data science offer is based on the following key streams:

- 1 Expertise in statistics is not enough.** To make sense of data sets experience and knowledge of the financial industry is key
- 2 Data does not need to be “Big”.** Data intelligence is not size dependent
- 3 Data science is not a magic formula.** Knowledge of how, when and in what context to apply data science to what ends is necessary to extract insight

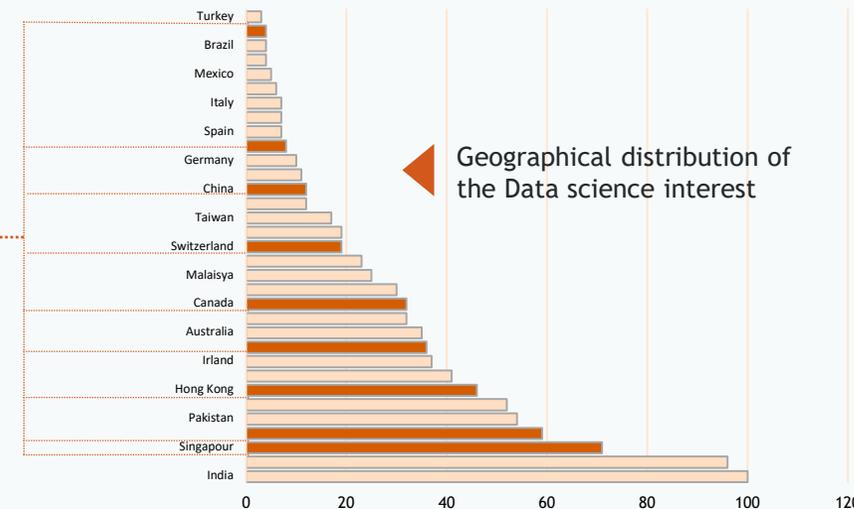
Data Science in figures | Trends & Overview

From Google trend - 2016



CH&Co. offices are clearly following the « Data science » trend.

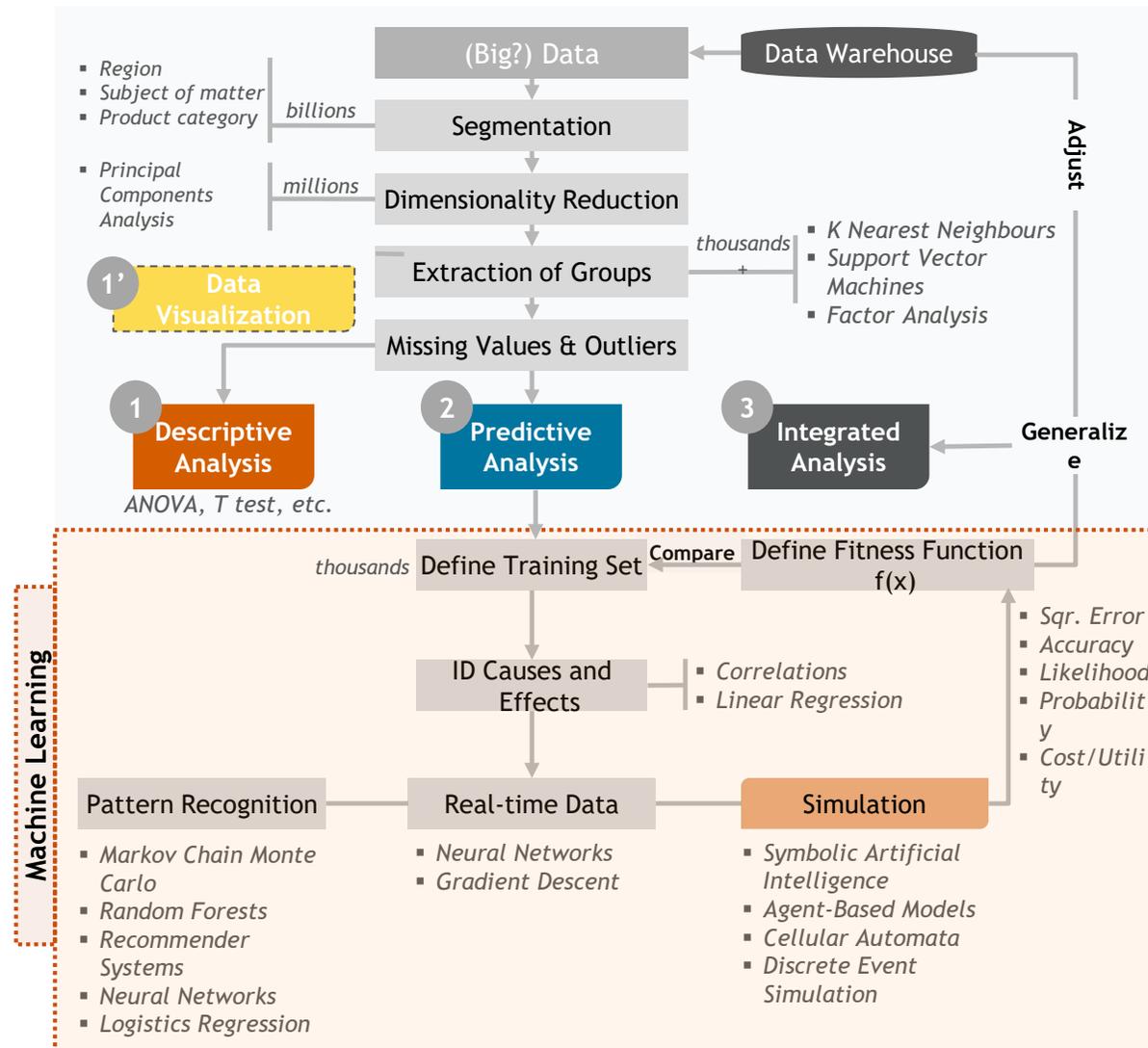
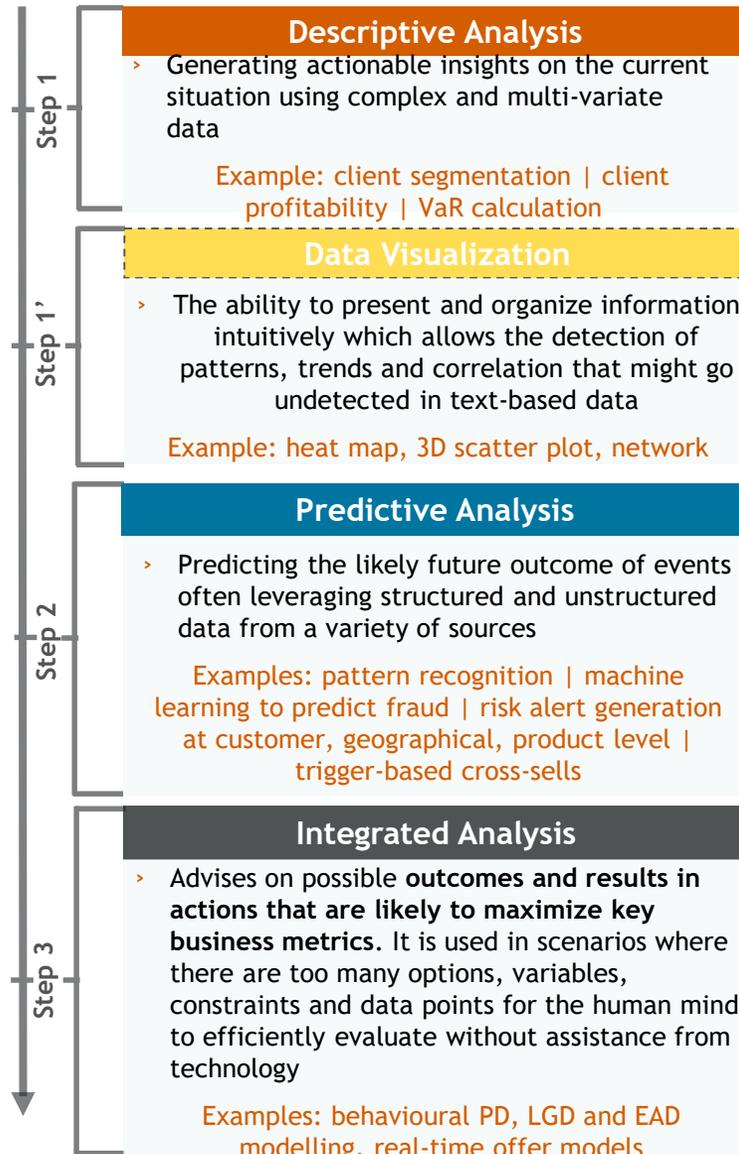
Financial cities play a key role



→ Deliver customized solutions through our R&D results straddled with our IT capacity held by CH&Co Lab

R&D in the financial industry

Analytics leveraged in the banking sector



R&D in the financial industry

Data science 3.0 | What is next ?

The tendency, which seems to be converging towards a somewhat standardised, cross-industry risk framework, is forcing financial industry in the direction of automated straight through processes and intraday business monitoring. According to us, the emerging trend for treatment in banks is based upon four drivers:

Regulation	Transparency	Risk/Reward	Cost reduction
<ul style="list-style-type: none"> › Causing convergence across the industry with a narrowing freedom of interpretation and back stop models that are becoming the new standard 	<ul style="list-style-type: none"> › Shareholders, Stakeholders and customers all demand clear visibility. Clear and ordered data across the institution that generates coherent reports 	<ul style="list-style-type: none"> › Better understanding and management of risks; capturing, modelling, monitoring and optimizing all risk types 	<ul style="list-style-type: none"> › New technology and pressure to reduce costs drive automated solution and replacement of manual interference wherever possible

The Data science trends in brief | Future concepts & techniques

Algo Hedging - How AI and ML can influence or determine hedging strategy

- › A complex portfolio can be hedged in many ways. Here an algorithm using sensitivities and a library of hedging products would be able to construct alternative and improved ways to hedge a portfolio. Ways that are not intuitive to a trader and may be more cost effective. By combining the hedging tool with Machine Learning (ML) technics calibrated on past data, alerts for optimal hedge at optimal time could be generated, as could recommendations for switching to new hedging strategies

New ways of capturing risks - Supply Chain Finance and Complexity Networks

- › Embracing new methodologies to capture and view risk. For example consider Supply Chain Finance (SCF), this could be viewed as a closed system from a credit risk perspective. Which would open up new ways for raising capital, engaging with clients and offer services. Similarly one could use complexity networks to model market interactions and improve the understanding of various factors to enable impact analysis.

The old oldest trick in the book... to get off the books

- › Transferring risk through new products such as Credit Suisse's bond issue earlier this year. Could this be taken one step further and be tranching against the proposed buckets of operational risk losses in proposed in BIS new operational risk framework?

The full picture with the technology and techniques of tomorrow

- › It is an overwhelming task to find the hedges or best collateral solution for a complex portfolio. Even more so to understand the future margin requirements and align this with other liquidity strains. To bring the full picture of outflows and inflows, risk and capital requirements an integrated system is required. What would the dashboard, an insightful overview with alerts look like? How can this be achieved? What can be monitored in real-time vs time-slicing?

Machine learning and Artificial Intelligence to increase effectiveness & efficiency

- › Leveraging new technology and methods by developing machine learning for back and stress testing (automation). This would reduce the cost of resources and free up quants to analyze and improve models.

Data Science with CH&Co.

Our knowledge | The holy grail

“There are two kinds of statistics, the kind you look up and the kind you make up.”
Rex Stout, Death of a Dox

Everyone at CH&Co. are believers in science. This is why we invest in a dedicated research team that drives new initiatives and explores new techniques and areas of application.

- Practitioners of statistics are well aware that
- A handful of techniques, approaches and formulas provides solutions for 99% of your problems
 - Choosing the right formula is crucial and based on experience only

Just like other data experts, CH&Co. keeps developing expertise and knowledge by applying and testing recurrent statistical equations

TOP 10 FORMULAS | What the Financial industry is recurrently using ... From Rubens Zimbres - 2016

1

Naive Bayes

$$P(a|c) = \frac{P(c|a) \cdot P(a)}{P(c)}$$

$$Prob = \pi P(a|c)$$

2

Perception

$$f(x) = \text{sign} \left[\sum_{i=1}^n w_i x_{ij} \right]$$

3

Linear Regression

$$f(x) = \text{sign} \left[\sum_{i=1}^n m_i x_i + b \right]$$

4

K. Nearest Neighbor

$$d(x_i, x_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

5

Neural Network

$$f(x) = w_0 + K \cdot \sum_{i=1}^n w_i x_i$$

6

PCA

$$x_i = x_i - \bar{x}$$

Eingenvector = Eigenvalue $[x_1 \dots x_n]$

$$f(x) = \text{Eingenvector}^r [x_{j1} \dots x_{jn}]$$

7

Support Vector Machine

$$f(x) = \text{sign}[\text{weight} \cdot y \cdot K(x_i \cdot x_j)]$$

$$K(x_i \cdot x_j) = \frac{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\text{width}}$$

$$y = 1 \wedge y = -1$$

8

Backpropagation

$$\Delta w_{ij}(n) = n \delta_j x_{ij} + \alpha \Delta w_{ij}(n-1)$$

9

Gradient descent

$$\theta_j = \theta_j - \alpha \sum_{i=1}^n (h(x_i) - y) \cdot x_i$$

10

Logit Regression

$$\text{Odds ratio} = \log \left(\frac{P(a|c)}{1 - P(a|c)} \right)$$

$$\text{Prob}(y = 1) = \frac{1}{1 + e^{-\beta (\sum_{i=1}^n m_i x_i + b)}}$$

From data science to quantitative techniques

Data science & machine learning can be mapped across 5 dimensions

	What is it ?	Techniques used	Concrete examples
1 Prediction, Anticipation and Simulation	<ul style="list-style-type: none">› Modeling of a variable from existing data, enabling its prediction & anticipation according to several scenarios	<ul style="list-style-type: none">› Time series› Artificial Neural Network› Regressions	<ul style="list-style-type: none">✓ Predict future value of a stock✓ Estimate a variable of interest for new people✓ Detect risk periods
2 Estimation	<ul style="list-style-type: none">› Estimate the value of a variable of interest based on explanatory variables	<ul style="list-style-type: none">› Regression models› Classification models	<ul style="list-style-type: none">✓ Optimization of products offers✓ Estimation of credit interest rate
3 Ranking / Discrimination	<ul style="list-style-type: none">› Creating various homogenous classes making the ranking of individuals possible	<ul style="list-style-type: none">› Unsupervised / supervised Classification› PCA / MCA / FCA› K-means / Neural Networks / Random Forest / SVM	<ul style="list-style-type: none">✓ Homogeneous risk class in Credit risk (PD/LGD)✓ Risk segmentation for insurance products
4 Behavioural analysis	<ul style="list-style-type: none">› Interpreting and predicting behaviours using statistical data and text/sound/image mining	<ul style="list-style-type: none">› Text mining using sophisticated machine learning algorithms› Descriptive statistics	<ul style="list-style-type: none">✓ Behavioural analysis from the emails database of a company✓ Human resources digitalization✓ Client targeting
5 Self-learning models	<ul style="list-style-type: none">› Implementing models that automatically teach themselves how to optimise their parameters from available data	<ul style="list-style-type: none">› Computational statistics› Mathematical optimisation	<ul style="list-style-type: none">✓ Classification (SVM, clustering, logistic regression, k-means, PCA, etc.)

From data science to quantitative techniques

Data science & machine learning can be mapped across 5 dimensions

	Business application	CH&Co. Credentials	Other examples
1 Prediction, Anticipation and Simulation	<ul style="list-style-type: none">› Improve profitability› Assess & monitor› Predict and anticipate	<ul style="list-style-type: none">› Improving and optimizing the stress-testing exercise	<ul style="list-style-type: none">✓ Expected return✓ Natural catastrophes prediction✓ Terrorism anticipation
2 Estimation	<ul style="list-style-type: none">› Improve the quality of your services› Improve profitability	<ul style="list-style-type: none">› Marketing intelligence› Scoring / Risk Estimation	<ul style="list-style-type: none">✓ Optimal pricing of new products✓ Credit interest rate valuation
3 Ranking / Discrimination	<ul style="list-style-type: none">› Optimize your costs› Optimize your risk› Develop customized existing products or services	<ul style="list-style-type: none">› Scoring / Risk estimation› Client segmentation› Marketing intelligence	<ul style="list-style-type: none">✓ Fraud detection✓ Credit granting choice
4 Behavioural analysis	<ul style="list-style-type: none">› Understand your business and your customers› Develop new products or services	<ul style="list-style-type: none">› Proof of Concept “Highway to mail”	<ul style="list-style-type: none">✓ Products and services real time offering✓ Churn detection✓ HR Planning
5 Self-learning models	<ul style="list-style-type: none">› Improve process, including decision process (time, quality, information, etc)› Automatize task / process	<ul style="list-style-type: none">› Automatic and dynamic models backtesting	<ul style="list-style-type: none">✓ Real time insurance pricing✓ Automatic asset management re-allocation



Geneva



London



Paris



Niort



Budapest



New York



Montreal



Hong Kong



Singapore



Frankfurt

<http://www.chappuisalder.com>