Big data for business

Value creation with data

Clément Levallois

2017-09-10

Table of Contents

last modified: 2017-09-27

Seven roads to data-driven value creation

Not a closed list, not a recipe!

Rather, these are essential building blocks for a strategy of value creation based on data.

1. PREDICT

Table 1. Predict

The data you need	**The hard part**
1. Domain specific historical data 2. Data mashups 3. Feedback mechanism on the accuracy	1. Collecting data ("cold start problem") 2. Risk missing the long tail, algorithmic discrimination, stereotyping 3. Neglect of novelty
The people you need	**The ones doing it**
1. Data scientists 2. Domain specialists	1. Predictive churn / default / … (banks / telco) 2. Predicting crime PredPol 3. Predicting deals TILKEE 4. Predictive maintenance

2. SUGGEST

Table 2. Suggest

The data you need	**The hard part**
1. A rich dataset and / or historical data on past transactions 2. A feedback mechanism providing data to improve on the suggestion procedure	1. Managing serendipity and bubble effects 2. Finding the value proposition which goes beyond the simple “you purchased this, you’ll like that”

The data you need	The hard part
The people you need	**The ones doing it**
1. Data scientists 2. Domain specialists for insights into the suggestion logic	1. Amazon's product recommendation system 2. Google's "Related searches…" 3. Retailer's personalized

3. CURATE

Table 3. Curate

The data you need	The hard part
1. Dirty (inaccuracies, missing values, bad formatting) 2. Cheap but hard to collect 3. Preferably unstructured	1. Slow progress 2. Must maintain continuity 3. Scaling up / right incentives for the workforce 4. Quality control
The people you need	**The ones doing it**
Large workforce ("digital labor") to manually inspect the data Data scientists used to work with crowd sourced data.	1. Clarivate Analytics curating metadata from scientific publishing 2. Nielsen and IRI curating and selling retail data 3. ImDB

4. ENRICH

Table 4. Enrich

The data you need	The hard part
1. Clean 2. Diverse	1. Knowing which cocktail of data is valued by the market 2. Limit replicability 3. Establish legitimacy
The people you need	**The ones doing it**

The data you need	The hard part
1. Specialists of APIs / data mashups / data integration 2. Possibly: specialists of linked data 3. Domain specialists for market research	Selling methods and tools to enrich datasets  Selling aggregated indicators

5. RANK / MATCH / COMPARE

Table 5. Rank / Match / Compare

The data you need	The hard part
Lots of data: a. with many features b. large dispersion values c. in large volumes	1. Finding emergent, implicit attributes 2. Insuring consistency of the ranking 3. Avoid gaming of the system by the users
The people you need	**The ones doing it**
1. Data scientists 2. Hacker mentality to imagine how unstructured data can contribute to ranking 3. Domain specialists for quality control	1. Search engines ranking results  2. Yelp, Tripadvisor, etc... which rank places  3. Any system that needs to filter out best quality entities among a crowd of candidates

6. SEGMENT / CLASSIFY

Table 6. Segment / classify

The data you need	The hard part
1. Any dataset, the richer the better 2. A feedback mechanism providing data to improve on the classification procedure	1. Evaluating the quality of the comparison 2. Dealing with boundary cases 3. Choosing between a supervised and unsupervised approach (how many categories?)
The people you need	**The ones doing it**
1. Data scientists 2. Domain specialists for insights into the classification logic	1. Tools for discovery / exploratory analysis by segmentation 2. Diagnostic tools (spam or not? buy, hold or sell? healthy or not?) 

7. GENERATE (experimental!)

Table 7. Generate (experimental!)

The data you need	**The hard part**
1. Unstructured data is more challenging but more interesting (pictures, sound, natural language…)	1. Should not create a failed product / false expectations 2. Both classic (think of ) and frontier science: not sure where it's going
The people you need	**The ones doing it**
1. Data scientists + academics 2. Excellent intrapreneurs to pilote this risky project	1. Intelligent BI aiden.ai 2. wit.ai, the chatbot by FB wit.ai 3. Virtual assistants CX company 4. Image

Combos!

Figure 1. Combinations

The end

Find references for this lesson, and other lessons, here.

This course is made by Clement Levallois.

Discover my other courses in data / tech for business: http://www.clementlevallois.net

Or get in touch via Twitter: @seinecle