

Seven roads to data-driven value creation

Clément Levallois

2017-09-10

Table of Contents

| | |
|--|---|
| 1. Predict | 1 |
| a. Examples of companies | 1 |
| b. Obstacles and difficulties | 1 |
| 2. Suggest | 1 |
| a. Examples of companies | 1 |
| b. Obstacles and difficulties | 1 |
| 3. Curate | 2 |
| a. Examples of companies | 2 |
| b. Obstacles and difficulties | 2 |
| 4. Enrich | 2 |
| Examples of companies | 2 |
| Obstacles and difficulties | 2 |
| 5. Rank / match / compare | 3 |
| Examples of companies | 3 |
| Obstacles and difficulties | 3 |
| 6. Segment / classify | 3 |
| Examples of companies | 3 |
| Obstacles and difficulties | 3 |
| 7. Generate / synthesize (experimental!) | 3 |
| Examples of companies | 4 |
| Obstacles and difficulties | 4 |
| Combos | 4 |
| The end | 4 |



Not a closed list, not a recipe! Rather, these are essential building blocks for a strategy of value creation based on data.

1. Predict

[prediction] | *prediction.jpg*

Figure 1. prediction

a. Examples of companies

1. Predicting crime [predpol]
2. Predicting deals [tilkee]
3. Predictive maintenance [cat]

b. Obstacles and difficulties

1. The [cold start problem](#)
2. Risk missing the long tail, algorithmic discrimination, stereotyping
3. Neglect of novelty

2. Suggest

[suggestion] | *suggestion.jpg*

a. Examples of companies

1. Amazon's product recommendation system [amazon]
2. Google's "Related searches..." [google]
3. Retailer's personalized recommendations [auchan]

b. Obstacles and difficulties

1. The cold start problem, managing [serendipity](#) and [filter bubble effects](#).

2. Finding the value proposition which goes beyond the simple “you purchased this, you’ll like that”

3. Curate

[curation] | *curation.jpg*

a. Examples of companies

1. Clarivate Analytics curating metadata from scientific publishing [crv logo rgb rev]
2. Nielsen and IRI curating and selling retail data [nielsen] [iri]
3. IMDb curating and selling movie data [imdb]
4. NomadList providing practical info on global cities for nomad workers [nomadlist]

b. Obstacles and difficulties

1. Slow progress: curation needs human labor to insure high accuracy, it does not scale the way a computerized process would.
2. Must maintain continuity: missing a single year or month hurts the value of the overall dataset.
3. Scaling up / right incentives for the workforce: [the workforce doing the digital labor of curation should be paid fairly](#), which is not the case yet.
4. Quality control

4. Enrich

[enrich] | *enrich.jpg*

Examples of companies

1. Selling methods and tools to enrich datasets [watson]
2. Selling aggregated indicators [edf]
3. Selling credit scores

Obstacles and difficulties

1. Knowing which cocktail of data is valued by the market
2. Limit duplicability
3. Establish legitimacy

5. Rank / match / compare

[rank] | *rank.jpg*

Examples of companies

1. Search engines ranking results [google]
2. Yelp, Tripadvisor, etc... which rank places [tripadvisor]
3. Any system that needs to filter out best quality entities among a crowd of candidates

Obstacles and difficulties

1. Finding emergent, implicit attributes (imagine: if you rank things based on just one public feature: not interesting nor valuable)
2. Insuring consistency of the ranking (many rankings are less straightforward than they appear)
3. Avoid gaming of the system by the users (for instance, [companies try to play Google's ranking of search results at their advantage](#))

6. Segment / classify

[muffin] | *muffin.jpg*

Examples of companies

1. Tools for discovery / exploratory analysis by segmentation
2. Diagnostic tools (spam or not? buy, hold or sell? healthy or not?) [medimsight]

Obstacles and difficulties

1. Evaluating the quality of the comparison
2. Dealing with boundary cases
3. Choosing between a pre-determined number of segments (like in the k-means) or letting the number of segments emerge

7. Generate / synthesize (experimental!)

[generate] | *generate.jpg*

Examples of companies

1. Intelligent BI with [Aiden](#) [aiden]
2. [wit.ai](#), the chatbot by FB [wit]
3. [Virtual assistants](#) [cx]
4. [Image generation](#) [deepart]
5. Close-to-real-life [speech synthesis](#) [google]
6. Generating realistic car models from a few parameters by [Autodesk](#): [autodesk]

A video on the generation of car models by Autodesk:

► <https://www.youtube.com/watch?v=25xQs0Hs1z0> (YouTube video)

Obstacles and difficulties

1. Should not create a failed product / false expectations
2. Both classic (think of [clippy]) and frontier science: not sure where it's going

Combos

[data driven value creation] | *data-driven-value-creation.png*

Figure 2. Combinations

The end

Find references for this lesson, and other lessons, [here](#).

[round portrait mini 150][align="center", role="right"] This course is made by Clement Levallois.

Discover my other courses in data / tech for business: <https://www.clementlevallois.net>

Or get in touch via Twitter: [@seinecle](#)