

Machine Learning e Data Science no Serasa Experian DataLab

Renato Vicente

Chief Scientist no LatAm Experian DataLab

Professor Associado na Universidade de São Paulo

Ago/2019

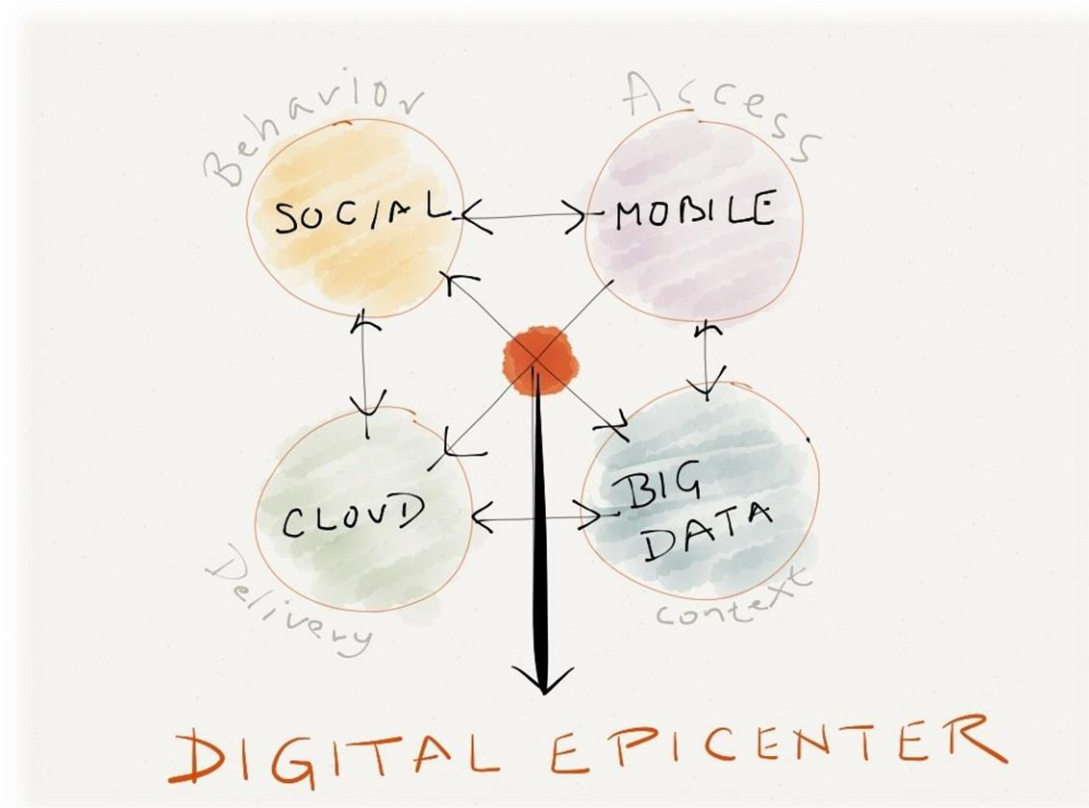


DataLab

Contexto

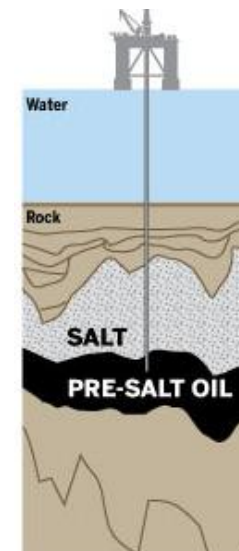
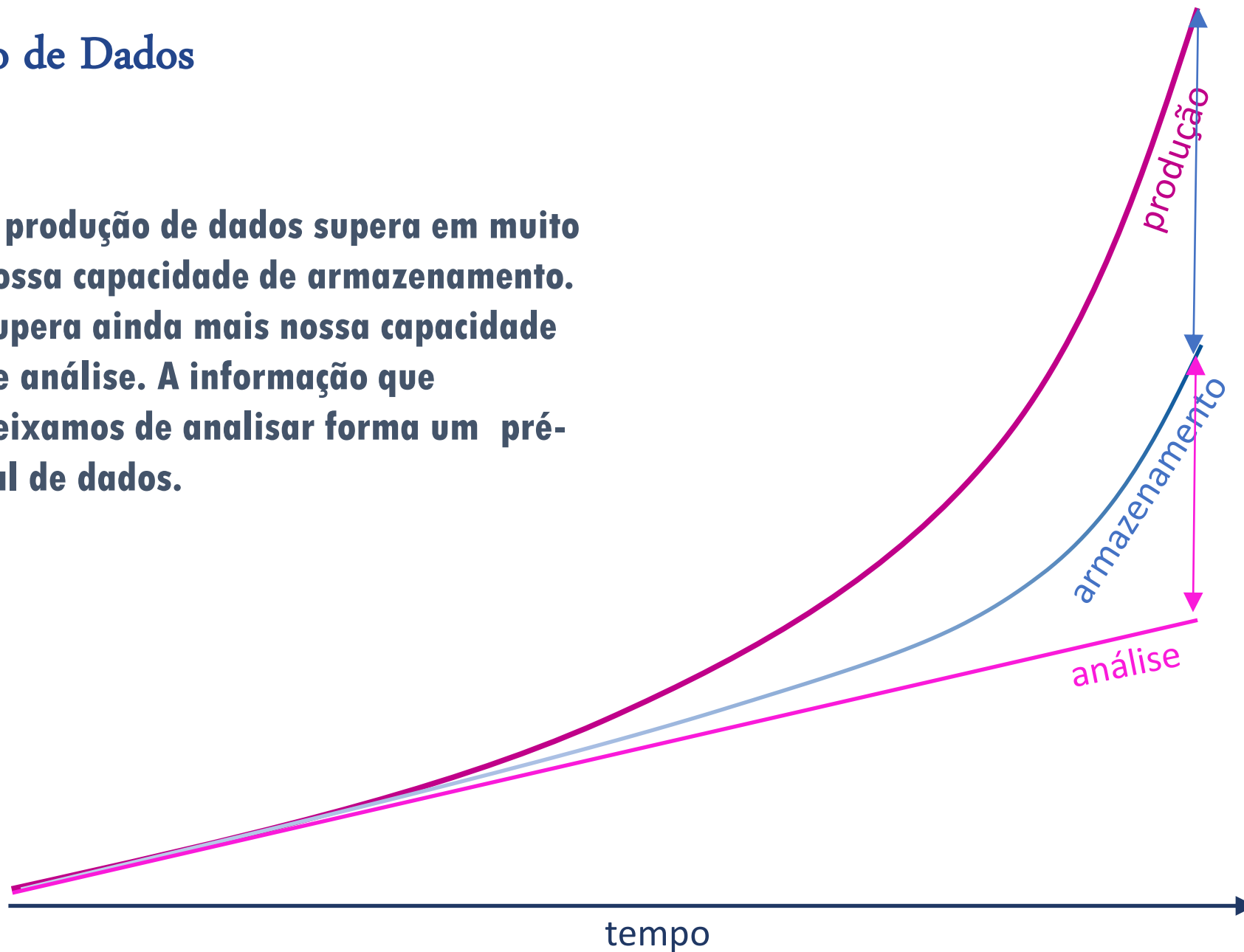
Infosfera

Tecnologias de rede multiplicam novas fontes de dados em ritmo exponencial. Cada usuário se tornou um produtor de informação.



Dilúvio de Dados

A produção de dados supera em muito nossa capacidade de armazenamento. Supera ainda mais nossa capacidade de análise. A informação que deixamos de analisar forma um pré-sal de dados.



Ciência de Dados

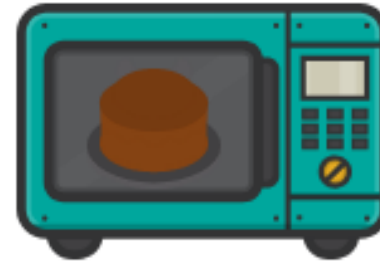
A Ciência de Dados é a prática que permite a industrialização do processo de extração de valor do pré-sal dos dados.



Coleta de dados



Pré-processamento



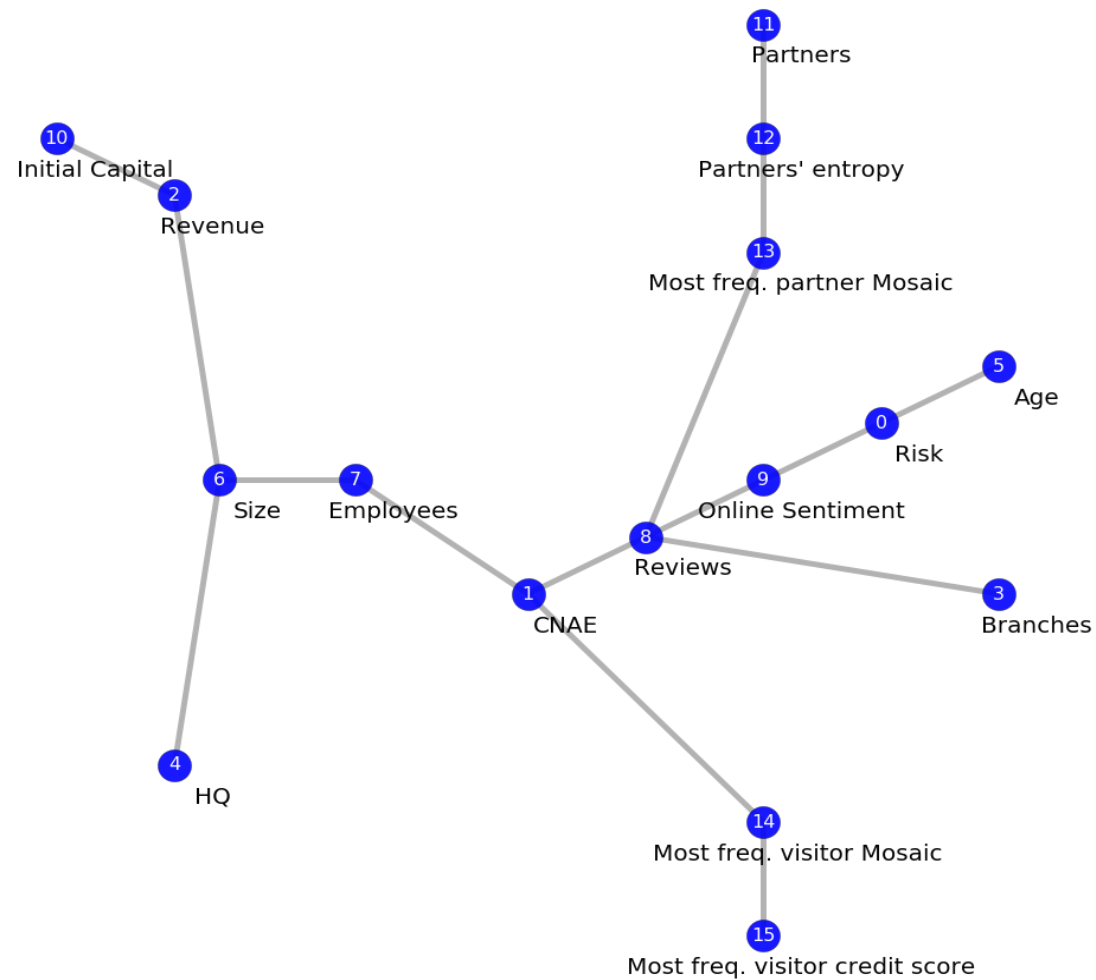
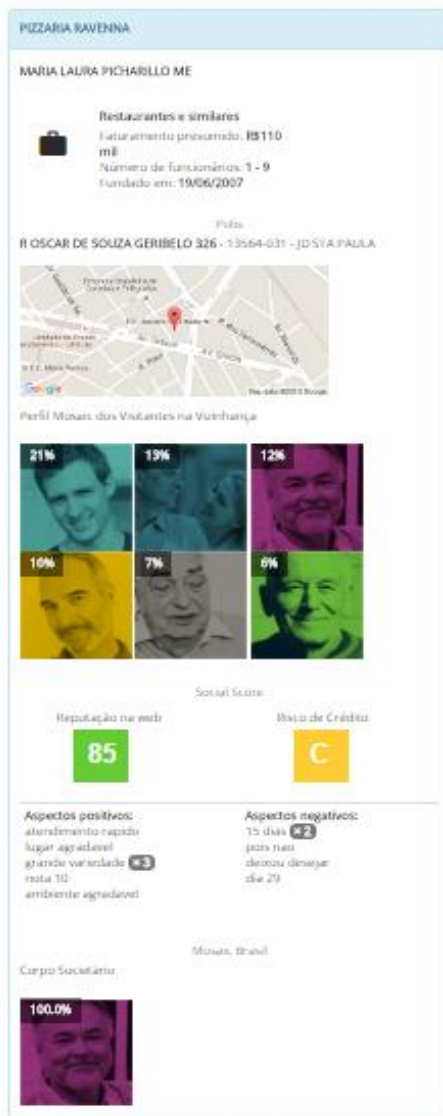
Machine Learning



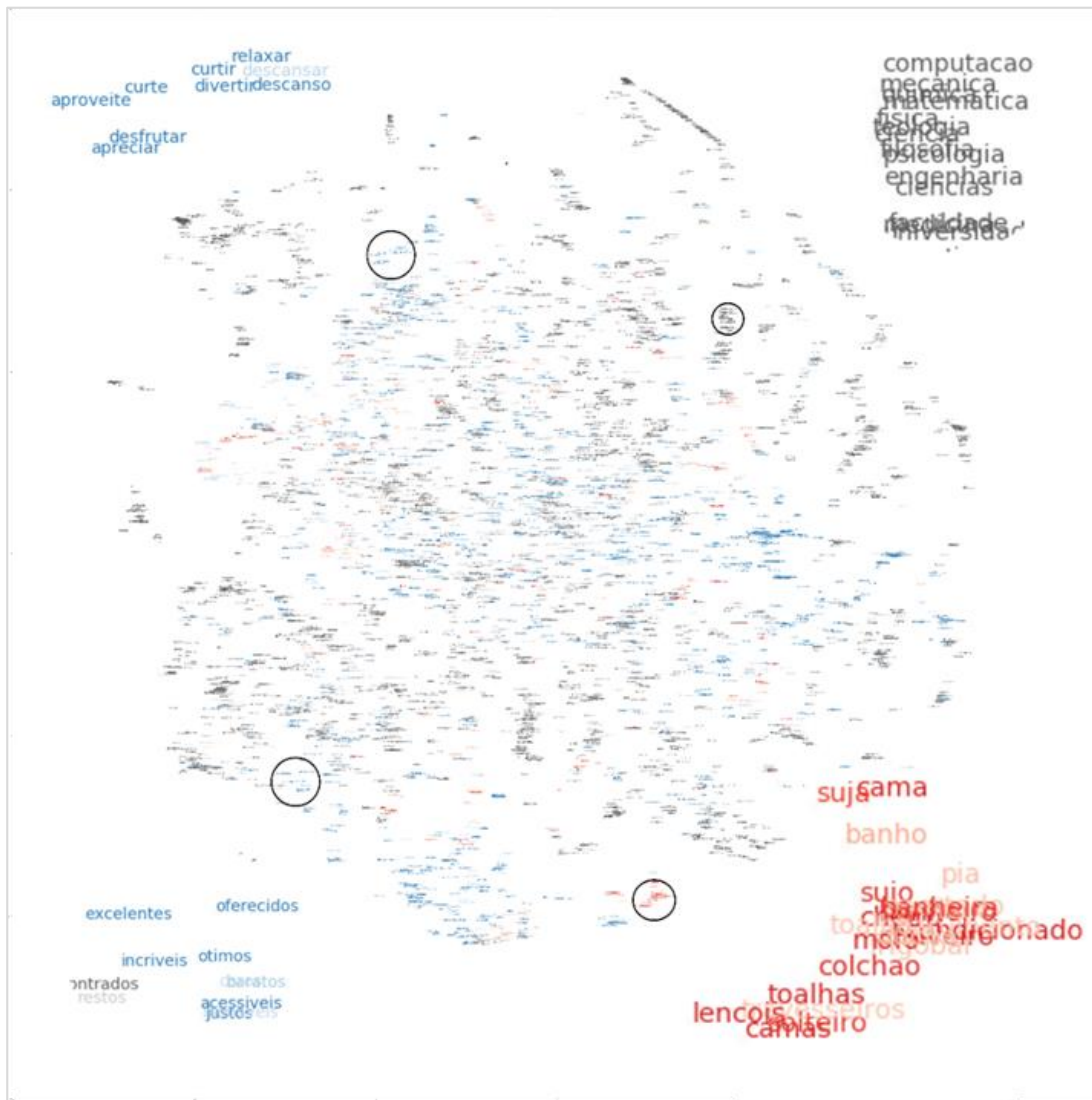
**Visualização dos
resultados**

Questões-chave em Ciência de Dados

Relevance



Contexto

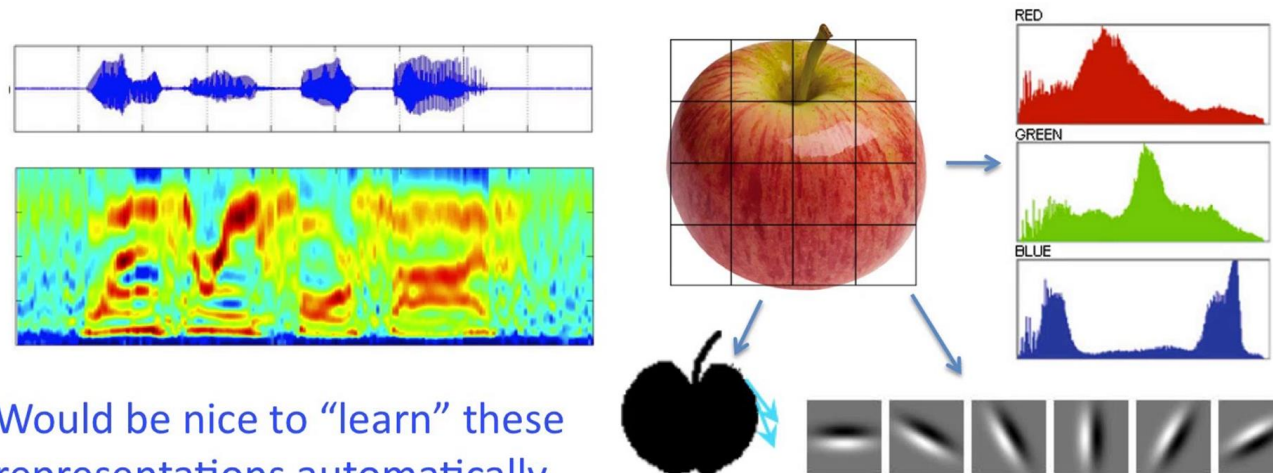


I eat an **apple** every day.

I eat an **orange** every day.

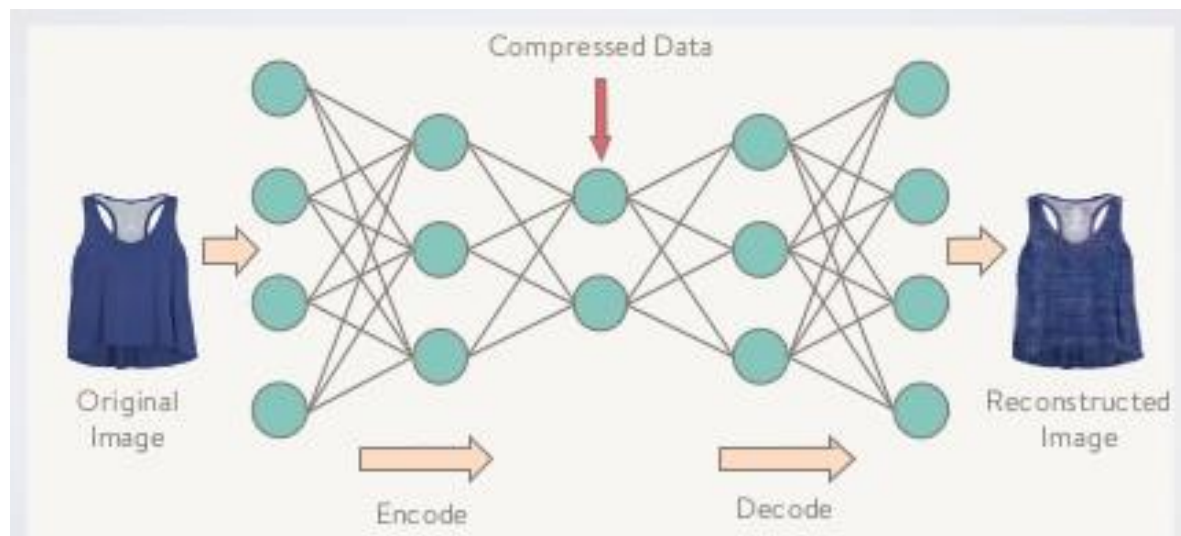
I like driving my **car** to work.

Representação

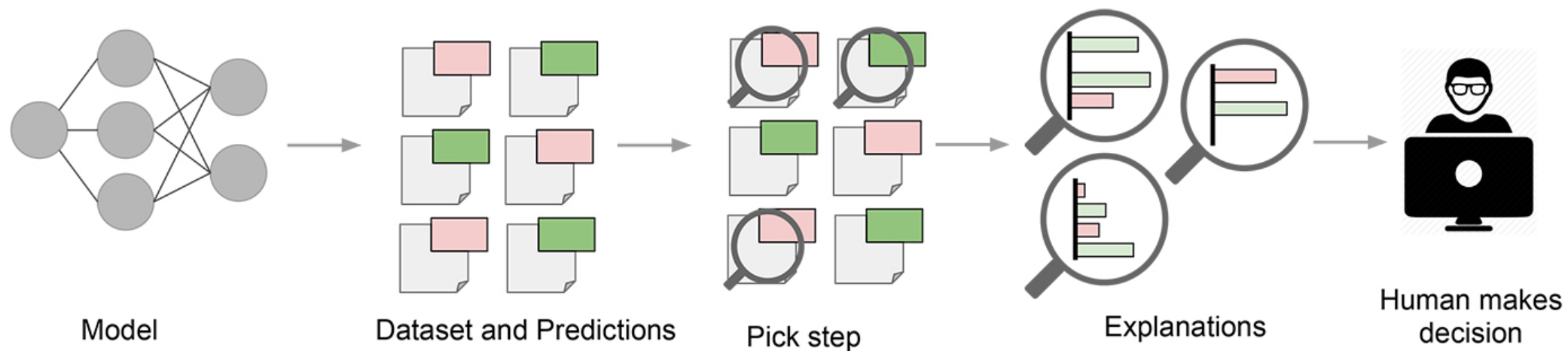


Would be nice to “learn” these representations automatically...

Copyright © 2014 Victor Lavrenko



Interpretabilidade



Os DataLabs

Rede Global de DataLabs



Lançamento do
DataLab NA

2010



Lançamento do
DataLab EMEA

2014



Lançamento do
DataLab LatAm

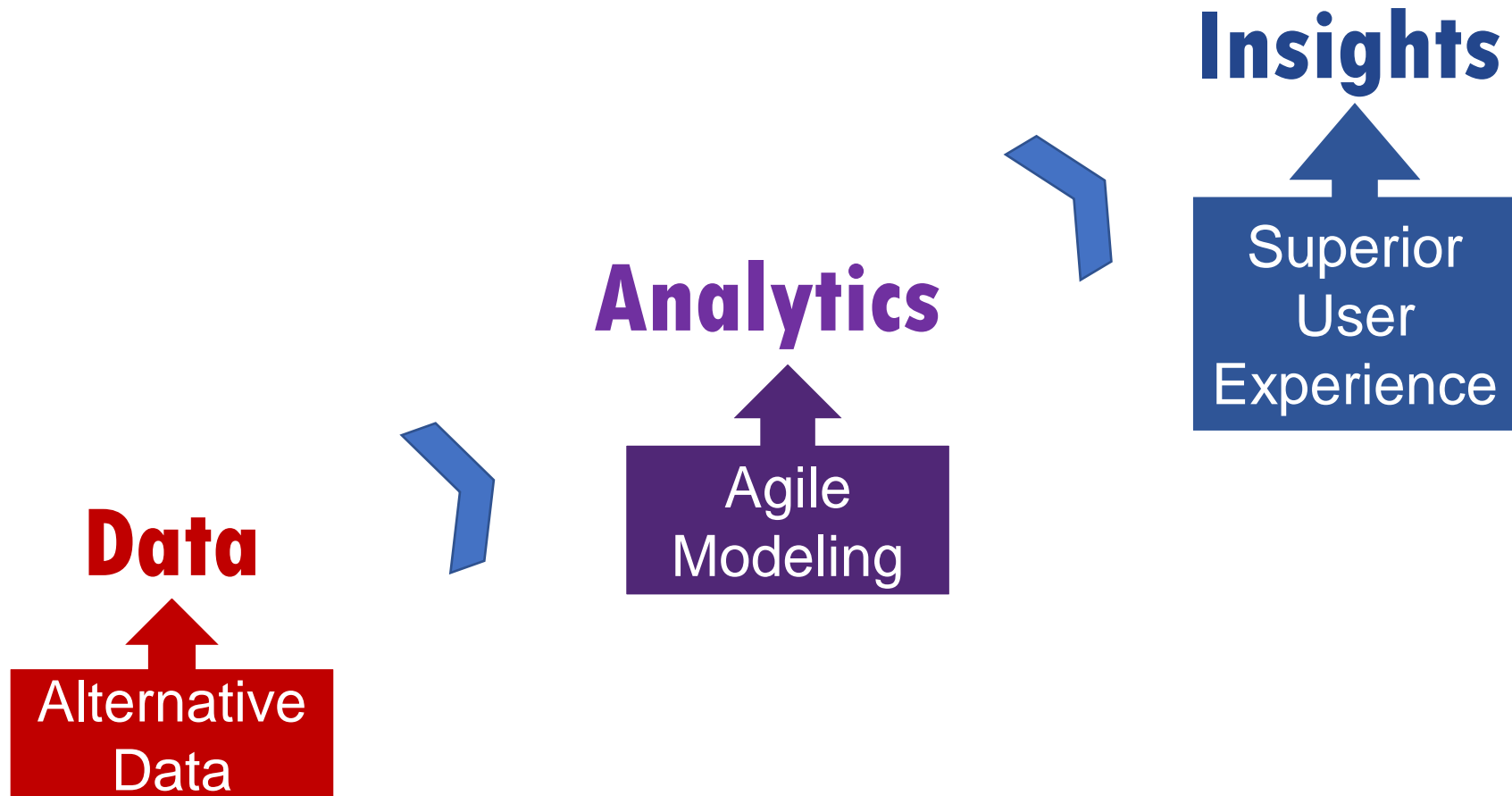
2016




Lançamento do
DataLab APAC

2018

Cadeia de Valor em Analytics



O mercado tem validado nossa estratégia

CORE				
		Equity Research Instant Insights		
		10 September 2018		
Paul Sullivan	+44 (0)20 3134 3307	paul.sullivan@barclays.com	Barclays, UK	Completed: 09-Sep-18, 17:38 GMT
Paul Checketts	+44 (0)20 3134 0522	paul.checketts@barclays.com	Barclays, UK	Released: 10-Sep-18, 04:00 GMT

<mailto:paul.checketts@barclays.com>

Experian PLC Beyond credit

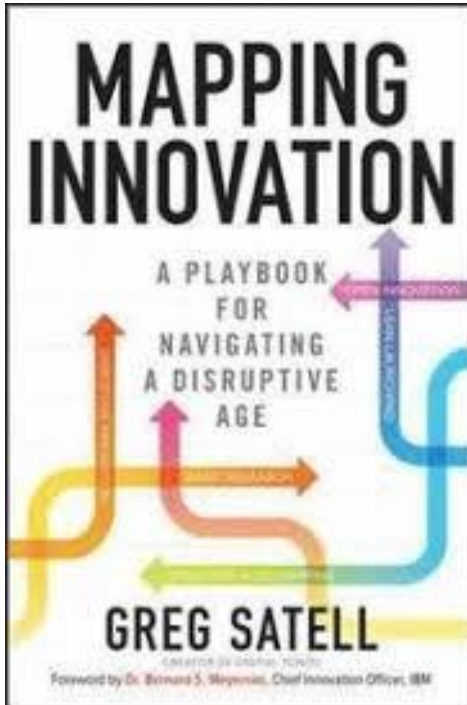
Stock Rating/Industry View: Overweight/Positive
Price Target: GBp 2,000
Price (07-Sep-2018): GBp 1,890
Potential Upside/Downside: 6%
Tickers: EXPN LN / EXPN.L

We recently hosted senior management as part of our Business Services Forum in New York. The message is clear: Experian is at the early stage of an elevated growth phase driven by new products and favourable market conditions, which should sustain annual organic revenue growth in the region of 6-8% in our view.

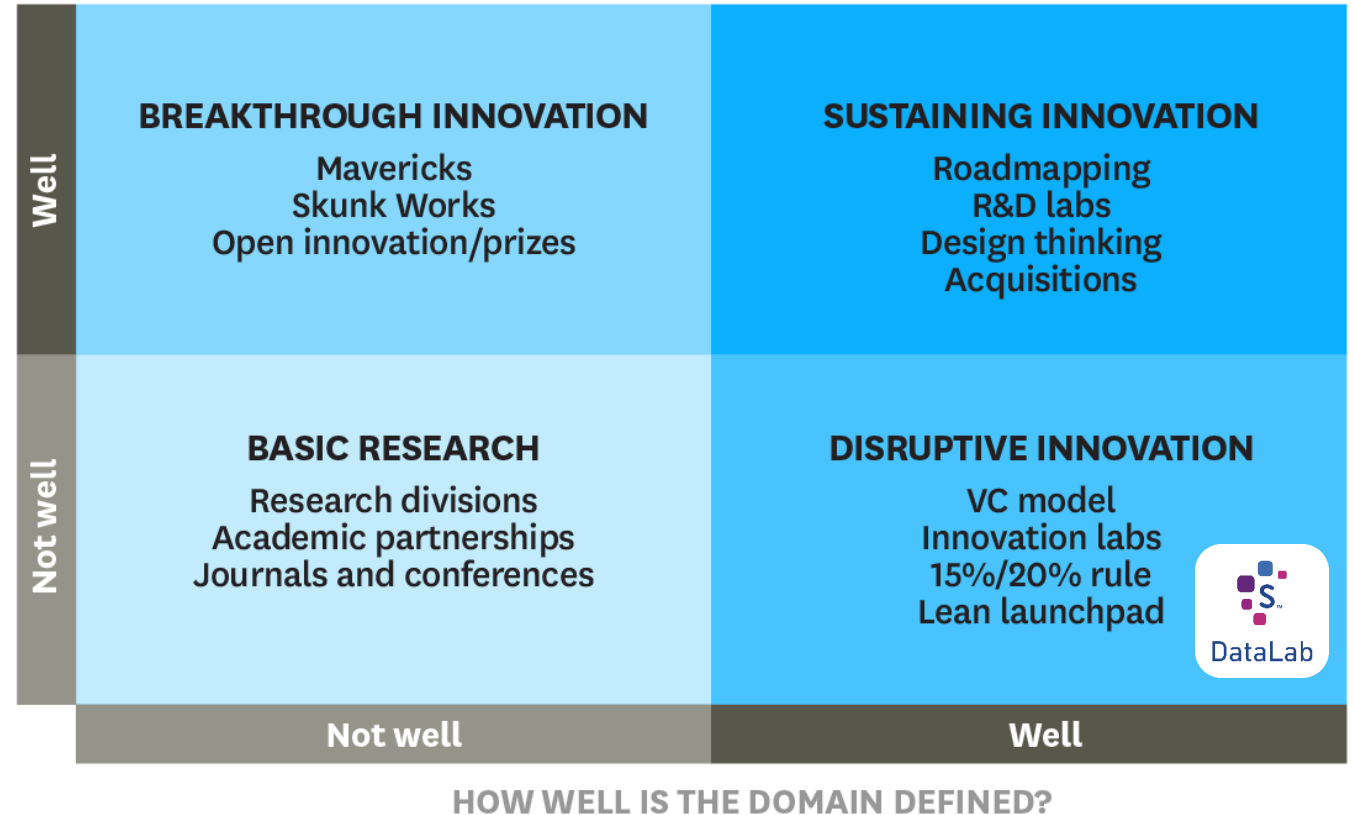
“ The pace of **product innovation** has intensified: The **pace of change** has undoubtedly quickened, with more new products coming to the market than ever before. We met with **Eric Haller**, EVP and Global Head, **Experian Data Labs** – which is an experimental R&D centre with **approximately 60 PhD data scientists**, tasked with discovering new ways to exploit linkages between Experian and client data. ”

Praticamos inovação de ruptura

4 Types of Innovation



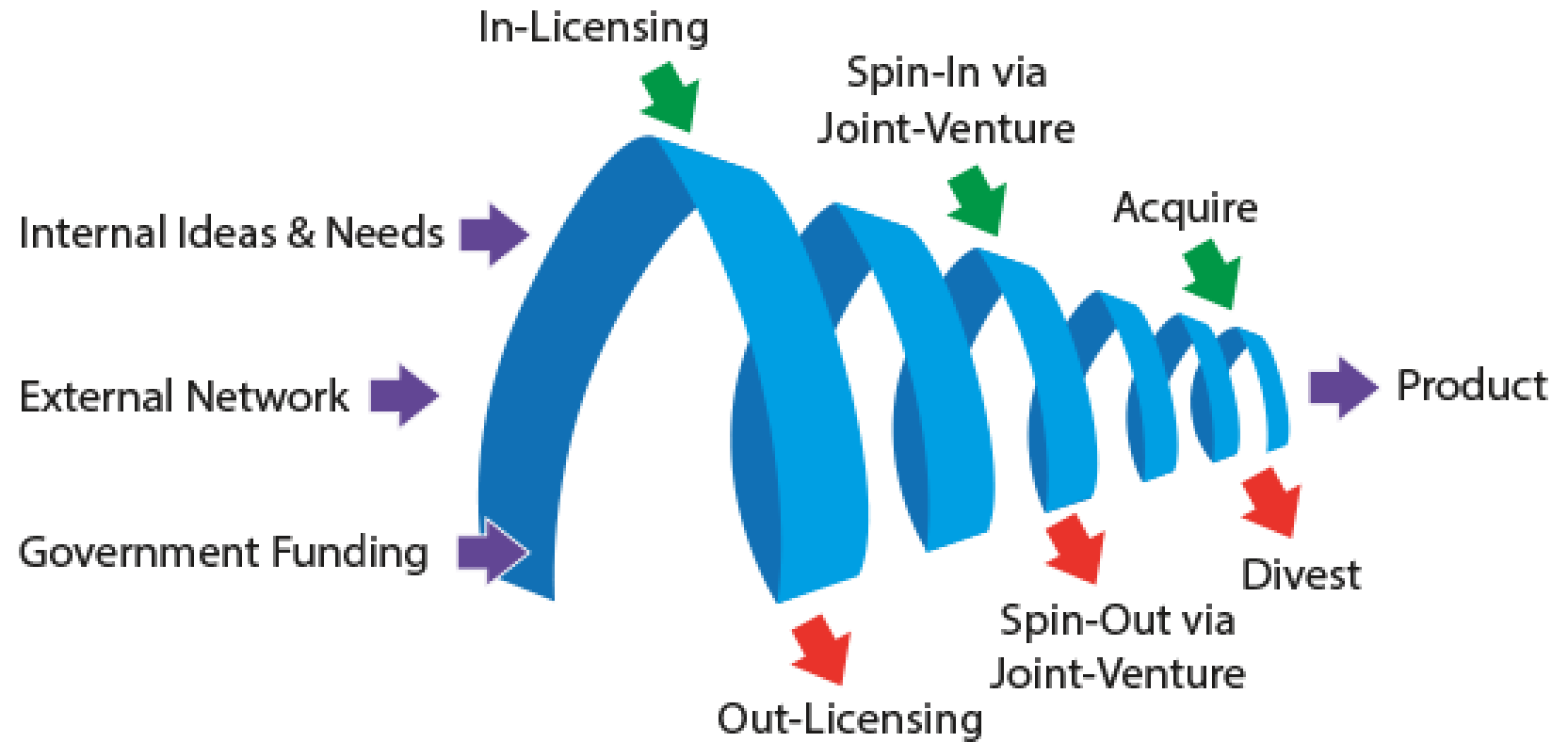
HOW WELL IS THE
PROBLEM DEFINED?



SOURCE GREG SATELL

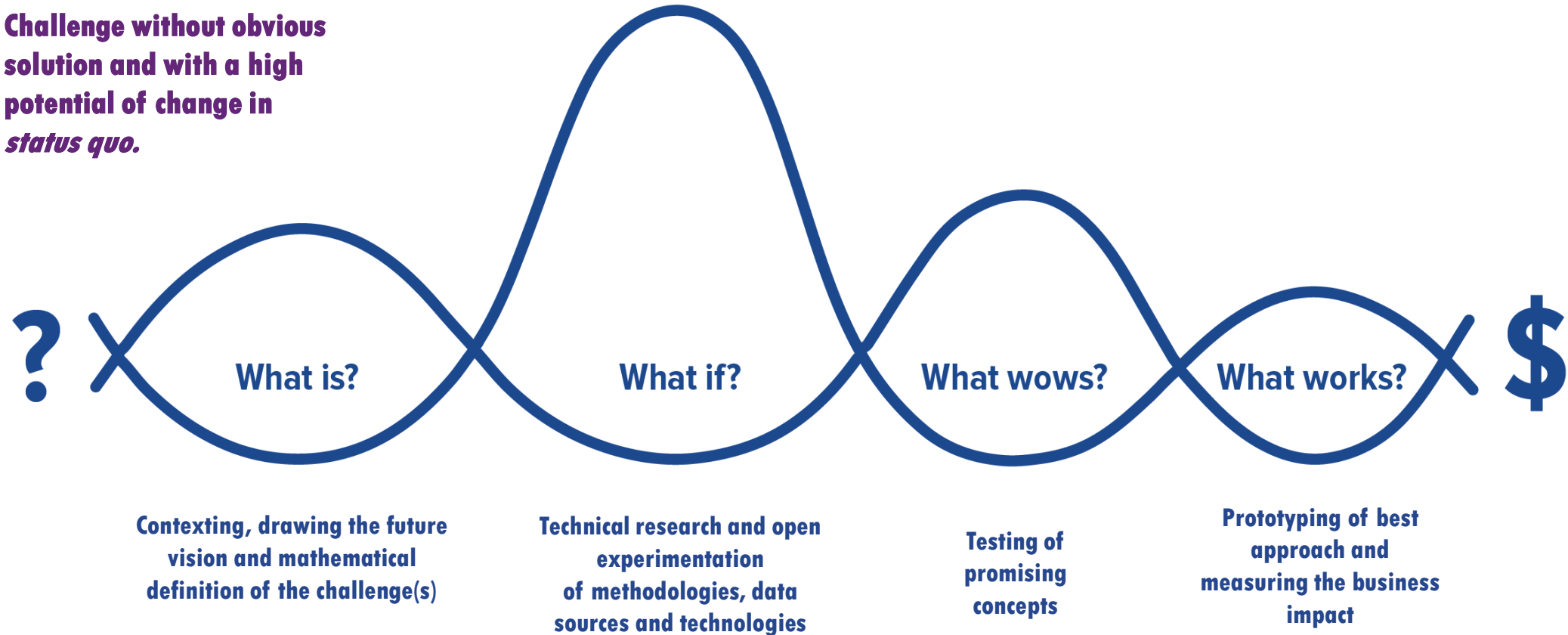
© HBR.ORG

Open Innovation

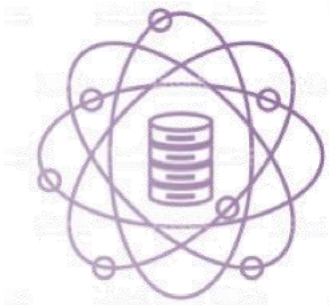


Engagement cycle for innovation

Challenge without obvious solution and with a high potential of change in *status quo*.



Estrutura dos times



Data Science



- 1) Pesquisa científica
- 2) Experimentação
- 3) Teste de novas idéias



Data Engineering



- 1) Prototipação
- 2) Construção de MVPs
- 3) Implantação em produção



BusDev



- 1) Dimensionamento de oportunidades
- 2) Definição de modelo de negócio
- 3) Estratégia de ida a mercado

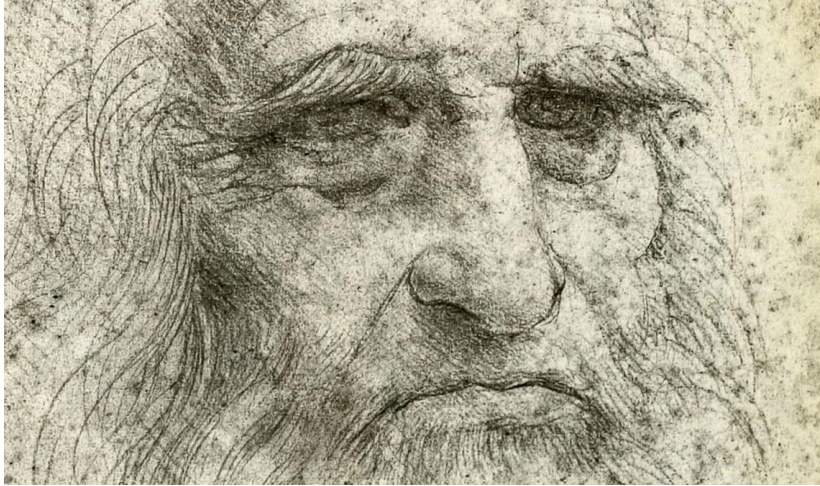
Atraímos jovens talentos



Cientistas do Datalab estão na final da competição internacional de Big Data



Cultura e melhores práticas



1. Seja curioso como uma criança
2. Observe
3. Comece pelos detalhes
4. Se concentre naquilo que ninguém parece dar atenção devida
6. Respeite os fatos
7. Pense visualmente
8. Evite silos
9. Crie para você não para seu chefe
10. Colabore

0. Be antifragile
1. Google it first
2. Test small first.
3. Check data quality
4. Use Bayes theorem



O espaço físico é expressão da cultura

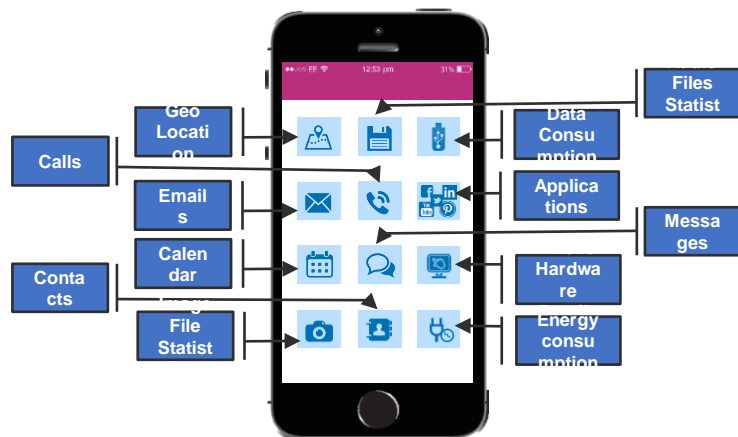


(a pedido dos próprios usuários) e são apenas um dos motivos que fazem do novo laboratório de inovação de 700m² da Serasa Experian, em São Paulo, uma instalação realmente “fora da caixa”.

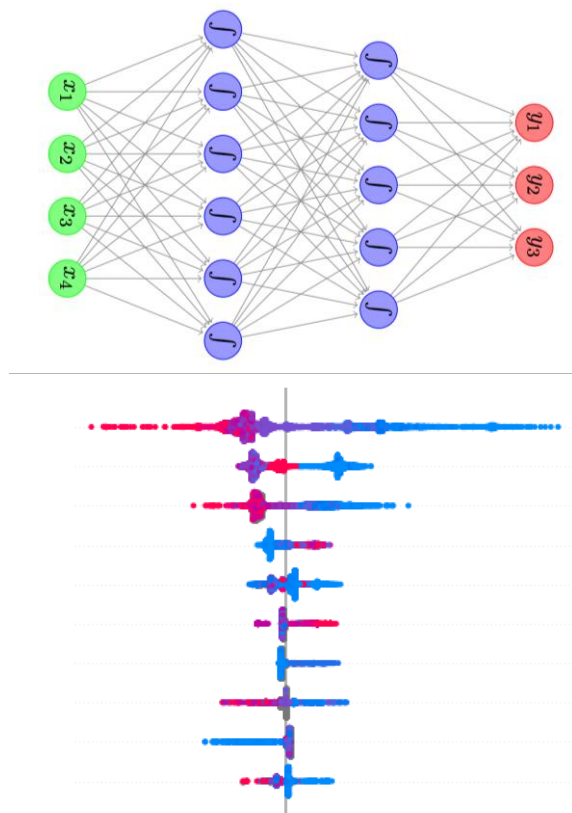
Localizado no 13º andar do Edifício Central Vila Olímpia (Alameda Vicente Pinzon, 51), o DataLab é uma unidade

Áreas de Pesquisa

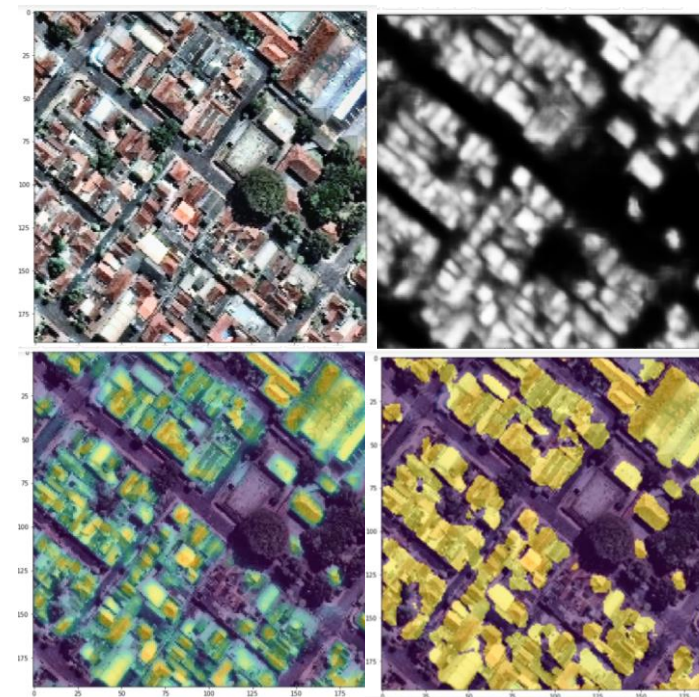
Mobile Insights



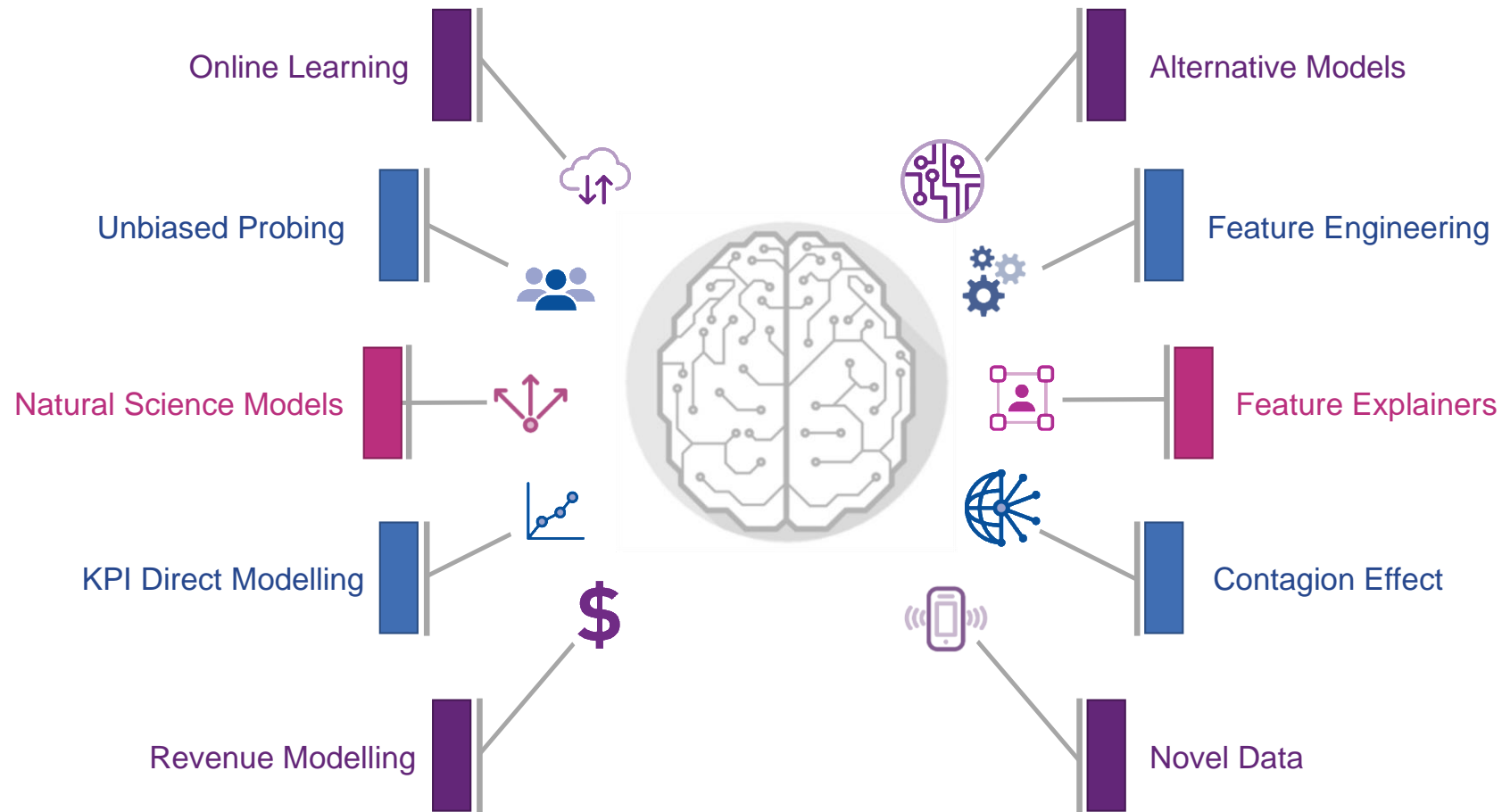
NextGen Credit



Remote Sensing Data



NextGen Credit



Mobile Insights: Score Digital

