VINI CARIDÁ

Human-Centered Data & AI

Pesquisas
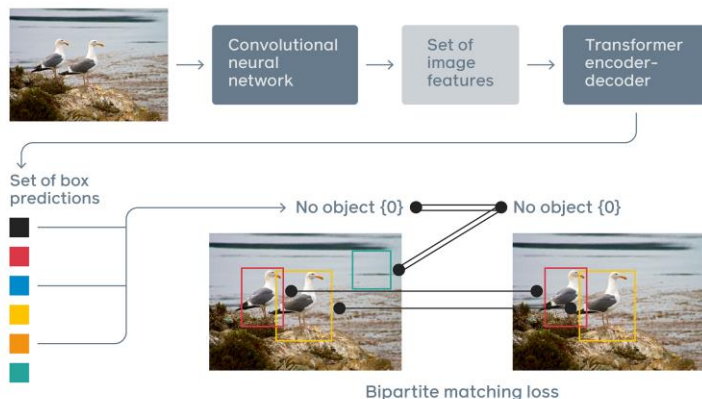
[Detection Transformers (DETR)](#)
[Facebook AI](#)
[https://lnkd.in/ehqeAFe](#)

Paper: **[https://t.co/EdFKftWRpv](#)**
Video: **[https://t.co/PmQ6PA5eWR](#)**
Post: **[https://lnkd.in/en4Ty2i](#)**

[https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html](#)
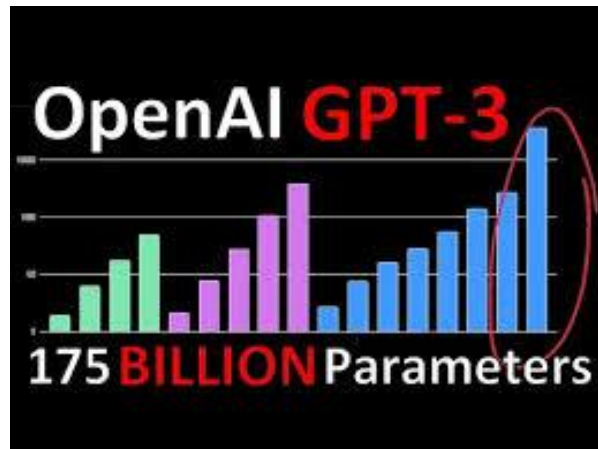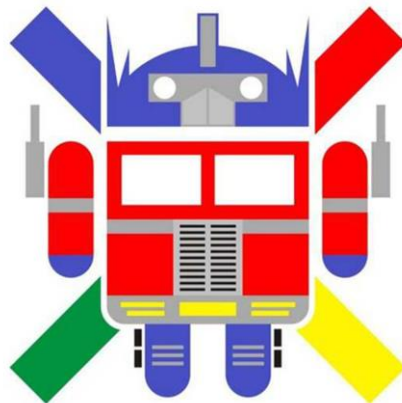
> As ML matures from research to applied business solutions, so do we need to improve the maturity of its operation processes

Vinicius F. Caridá

Launching is easy,
Operating is hard.

"The real problems with a
ML system will be found
while you are continuously
operating it for the long term"

Vinicius F. Caridá

How can we
transform our business
with Artificial Intelligence?

Machine learning é uma tecnologia central e transformadora que nos faz repensar como fazemos tudo "Sundar Pichai, CEO – Google"

Vinicius F. Caridá

THE DATA SCIENCE BUSINESS MAP

Vinicius F. Caridá

Dados   **Técnica de ML**   **Técnica + Dados**   **Treinamento**

Cientista de Dados   **Modelo em produção**   **Modelo**

Vinicius F. Caridá
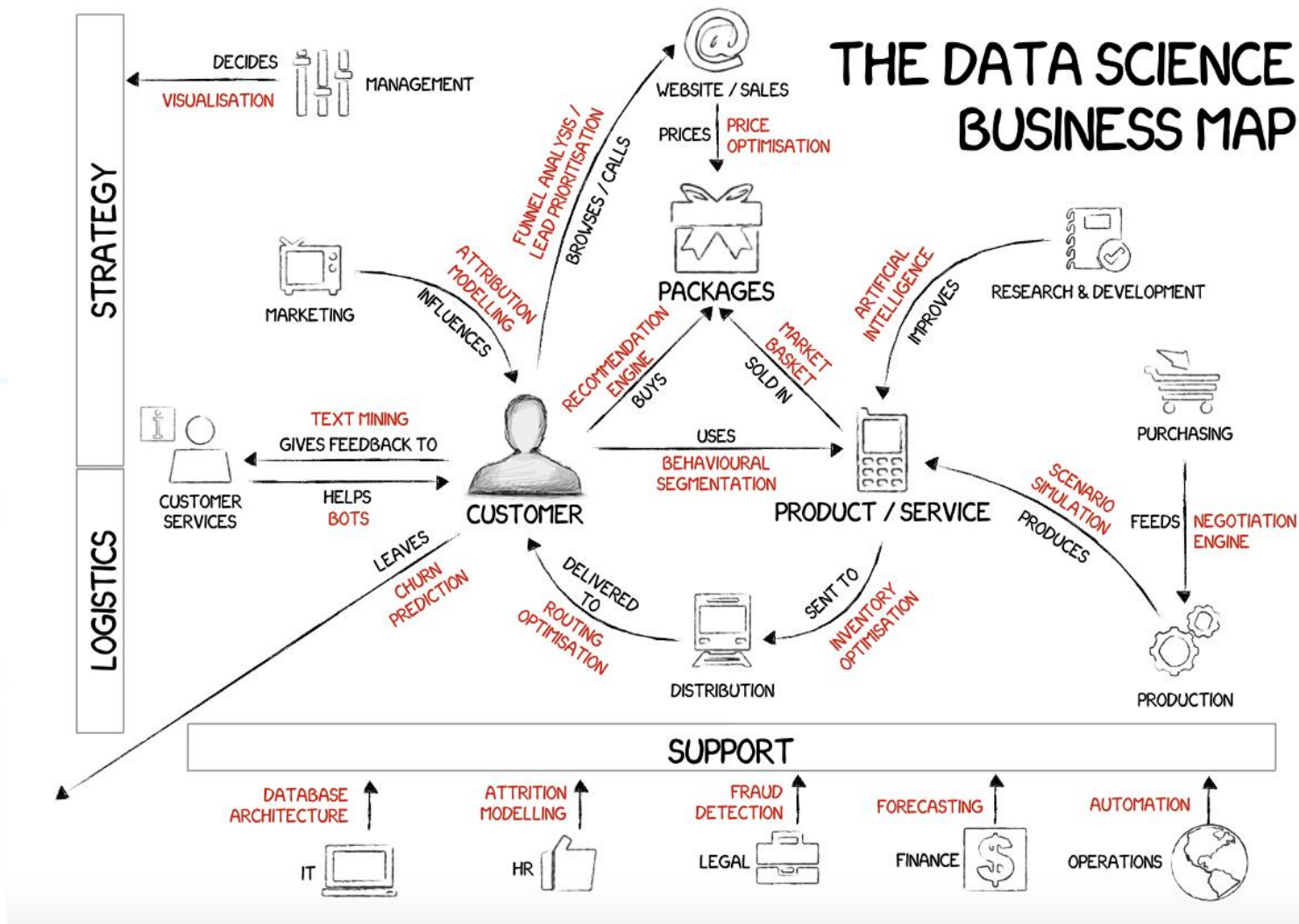
Além de treinar um modelo incrível...

Código ML

Machine Learning | Produção

Verificação de dados

Monitoramento

Coleção de dados

Gestão de recursos de máquina

Código ML

Configuração

Infraestrutura de atendimento

Extração de recursos

Ferramentas de analises

Ferramentas de gerenciamento de processos

Machine Learning | Produção

Realidade: ML requer DevOps
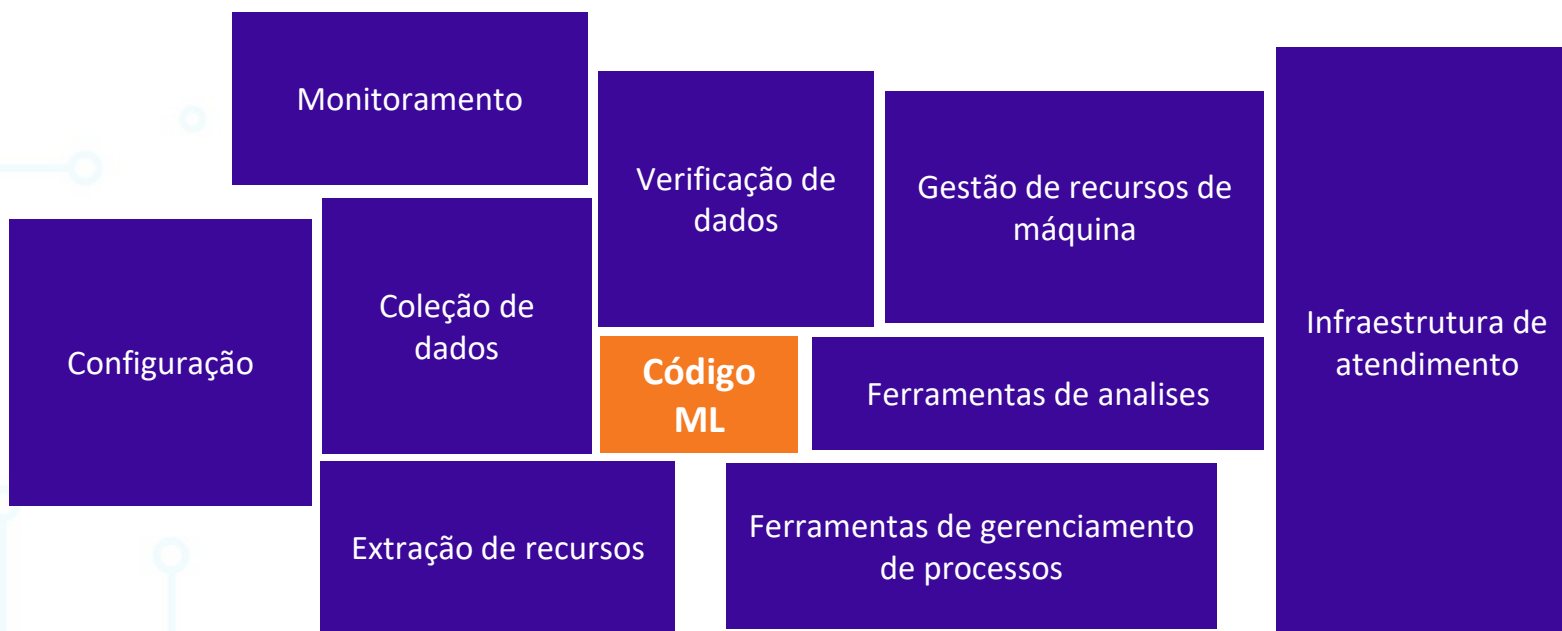
Monitoramento

Verificação de dados

Gestão de recursos de máquina

Configuração

Coleção de dados

Código ML

Ferramentas de analises

Infraestrutura de atendimento

Extração de recursos

Ferramentas de gerenciamento de processos

46x Deployment Frequency

2.555x Faster Lead Time For Changes

Faster Time to Market

DevOps

code
plan
build
test
release
deploy
operate
monitor

Increased Revenue

$$

7x Lower Change Failure Rate

2.604x Faster Mean Time to Recover

VINI CARIDÁ
Human-Centered Data & AI

Vinicius F. Caridá

Accelerate: State of DevOps

Machine Learning | Produção

Modern Software Development

- Scalability
- Extensibility
- Configuration
- Consistency & Reproducibility
- Modularity
- Best Practices
- Testability
- Monitoring
- Safety & Security

Machine Learning | Produção
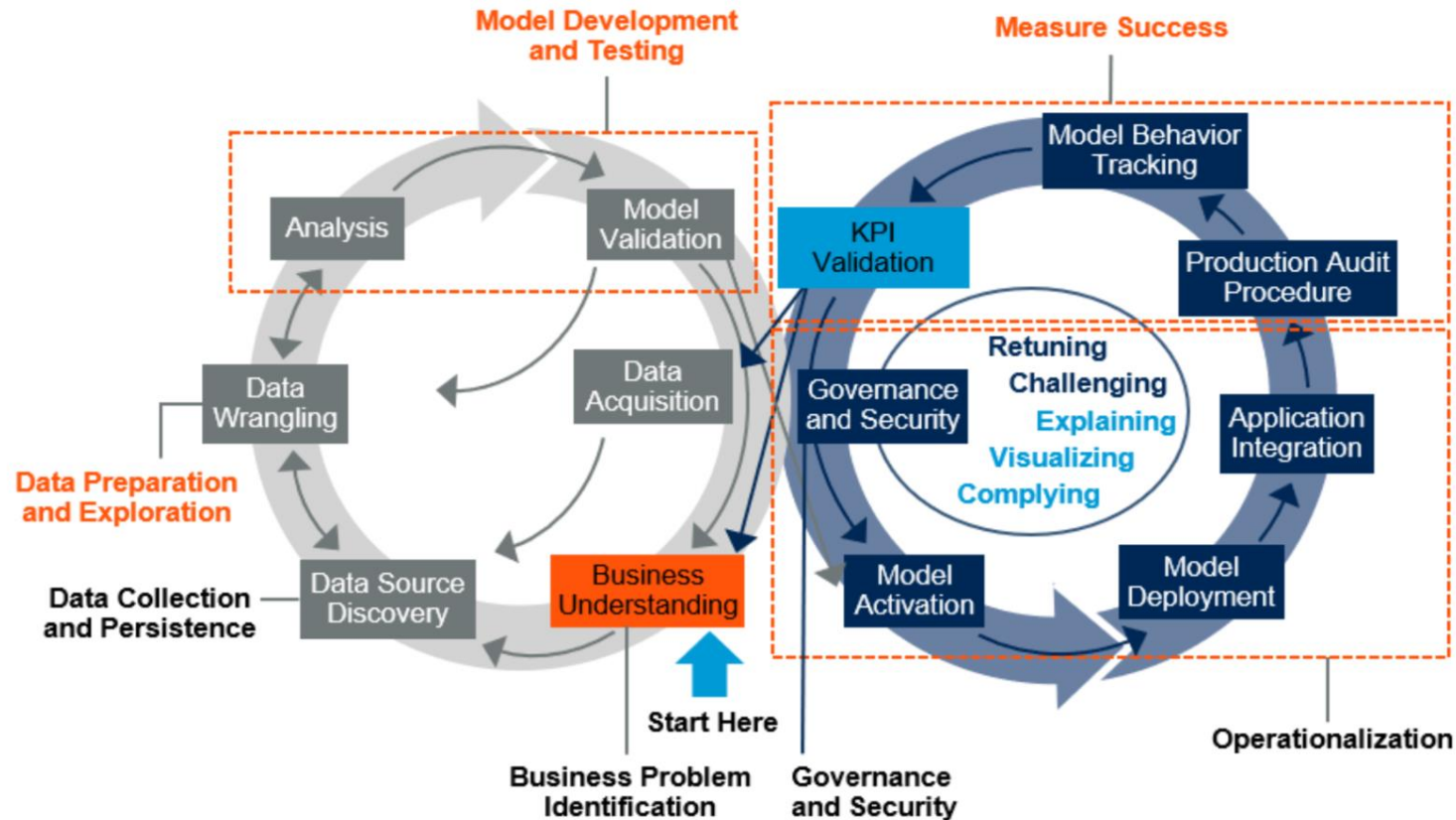
Modern Software Development

- Scalability
- Extensibility
- Configuration
- Consistency & Reproducibility
- Modularity
- Best Practices
- Testability
- Monitoring
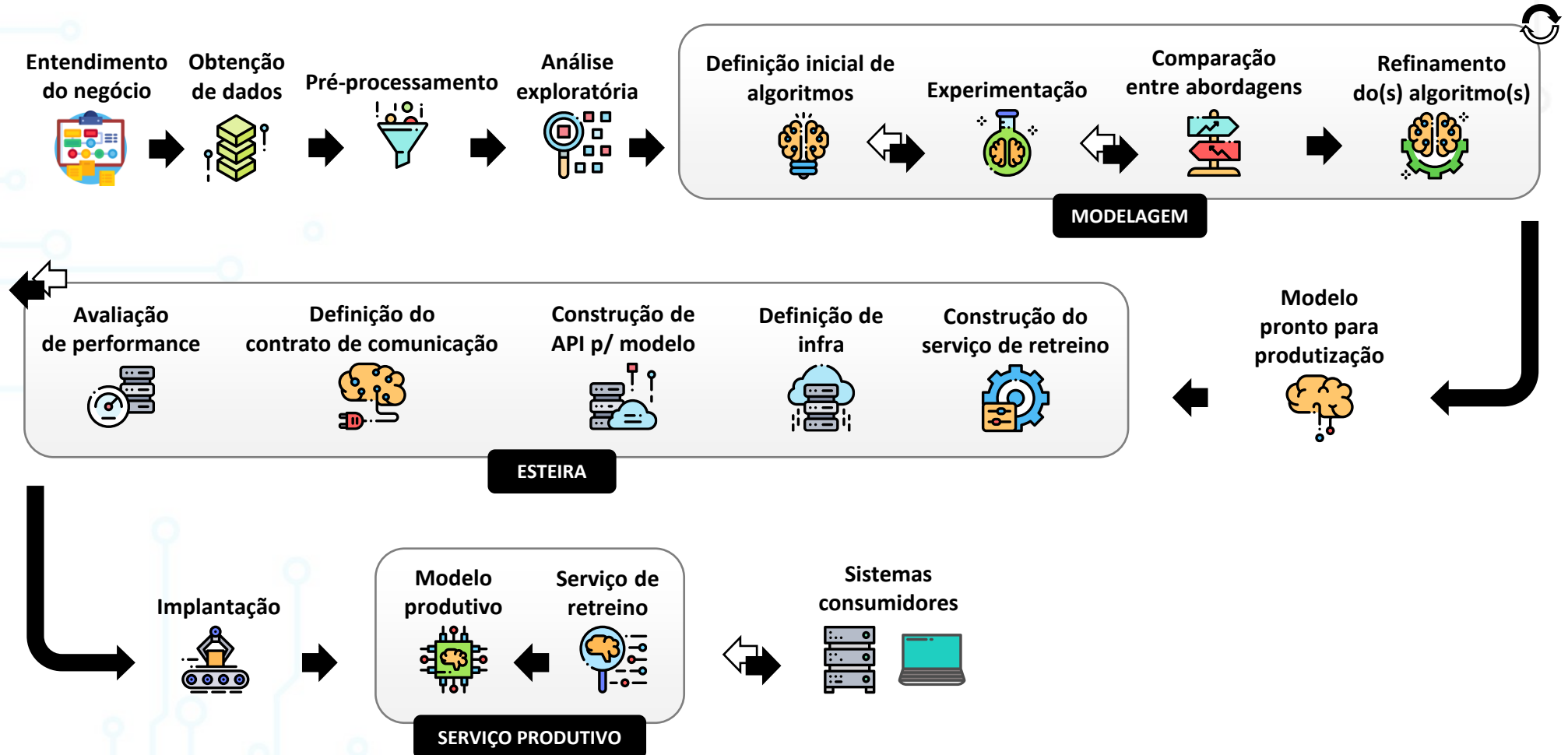- Safety & Security

+

Machine Learning Development

- Labeled data
- Feature space coverage
- Minimal dimensionality
- Maximum predictive data
- Fairness
- Rare conditions
- Data lifecycle management

Ciclo de vida de projetos de ML

Source: Gartner

Ciclo de vida de projetos de ML

Entendimento do negócio → **Obtenção de dados** → **Pré-processamento** → **Análise exploratória** →

MODELAGEM
- Definição inicial de algoritmos
- Experimentação
- Comparação entre abordagens
- Refinamento do(s) algoritmo(s)

Modelo pronto para produtização

ESTEIRA
- Avaliação de performance
- Definição do contrato de comunicação
- Construção de API p/ modelo
- Definição de infra
- Construção do serviço de retreino

SERVIÇO PRODUTIVO
- Implantação
- Modelo produtivo
- Serviço de retreino
- Sistemas consumidores

VINI CARIDÁ
Human-Centered Data & AI

Vinicius F. Caridá

Machine Learning

Freedom to experiment (and fail) is important

"Take your time and succeed"

IDEAS TRIED   SLOW FAILURES   FAST FAILURES   SUCCESSES

"Fail fast and iterate"

IDEAS TRIED   SLOW FAILURES   FAST FAILURES   SUCCESSES

Data Science Project Management says:

"Potentially Releasable Increment" Challenges: Scrum prescribes teams to deliver "potentially releasable Increment of 'Done' product" at the end of each sprint (Sutherland & Schwaber, 2017). However, Sutherland reports that "**80% of the hundreds of Silicon Valley teams I polled could not do this**" (Sutherland, 2015). **This challenge is particularly daunting for data science teams and might not even be necessary. Much of the data science process, especially in the early exploratory phases, may not be intended for release outside the data science team**. Moreover, testing requirements may extend beyond what is reasonable in a sprint, especially if a high degree of accuracy is needed (Jennings, 2017).

Vinicius F. Caridá

http://www.datascience-pm.com/scrum/

CRISP-DM

Cross Industry Standard Process for data Mining

Processo atual mais usado para Ciência de Dados

Destaca as naturezas cíclicas e interativas da Ciência de Dados

Vinicius F. Caridá

Production Machine Learning

"Hidden Technical Debt in Machine Learning Systems"

NIPS 2015

http://bit.ly/ml-techdebt

Rules of Machine Learning:

Best Practices for ML Engineering

Martin Zinkevich

https://developers.google.com/machine-learning/guides/rules-of-ml/

43 Rules

Rule #1: Don't be afraid to launch a product without machine learning.

Rule #23: You are not a typical end user.

Thanks !

Vinicius Fernandes Caridá

vfcarida@gmail.com

 @vinicius caridá   @vinicius caridá   @vfcarida   @vfcarida