



# Introdução a Data Science e ao ambiente Python

Anderson Cordeiro



# Professor



## Anderson Cordeiro

**#[UFRJ] Doutorado em Informática**

Análise de Redes Complexas

**#[Globo] Cientista de Dados Sr.**

Gato Mestre - Cartola FC

**#[UERJ] Mestre em Modelagem Computacional**

Matemática Aplicada e Computação Científica

# Analista de Sistemas

# Especialista em Infosec/Pentest

# Python/Django Developer



# Manda um oi

---

- Teams
- [linkedin.com/in/andercordeiro/](https://www.linkedin.com/in/andercordeiro/)
- [github.com/andersoncordeiro](https://github.com/andersoncordeiro)
- [pyratalab.org/pyrata/andersoncordeiro/](https://pyratalab.org/pyrata/andersoncordeiro/)
- [andcordeiroc@gmail.com](mailto:andcordeiroc@gmail.com)
- 22 99875 7047

# Nossas aulas

---

- Interação
- Falar com pessoas
- Interrompa SEMPRE
- 7:12
- Aprender Fazendo
- Tudo faz sentido

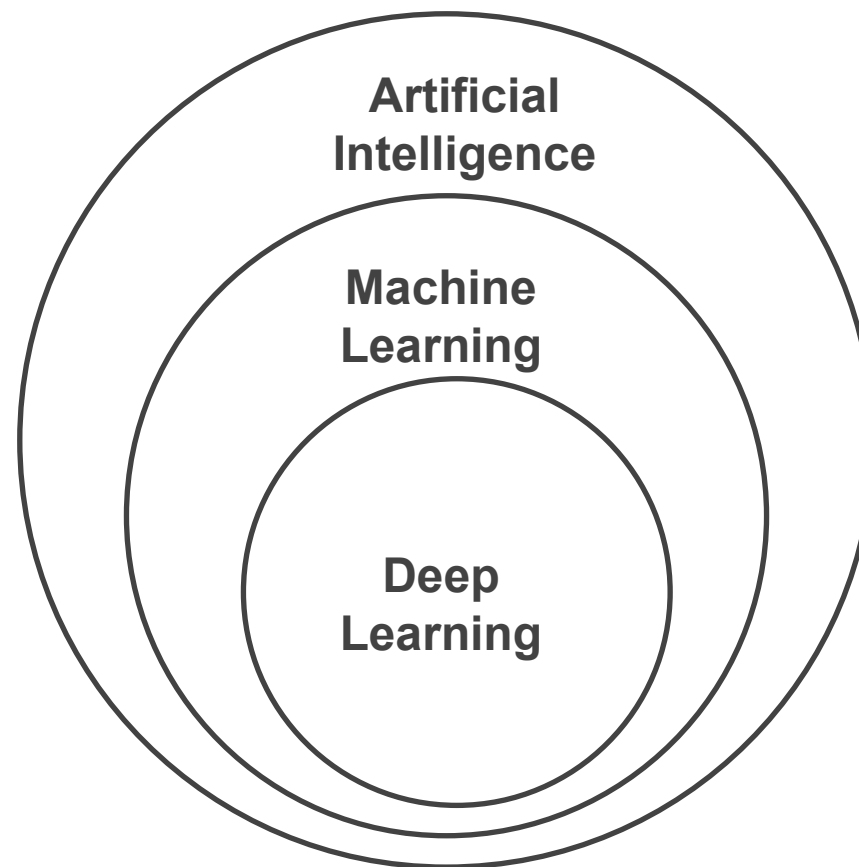
# Algumas ferramentas

---

- Colab (Jupyter)
- Github
- VSCode / PyCharm / Spyder
- Criando o ambiente perfeito

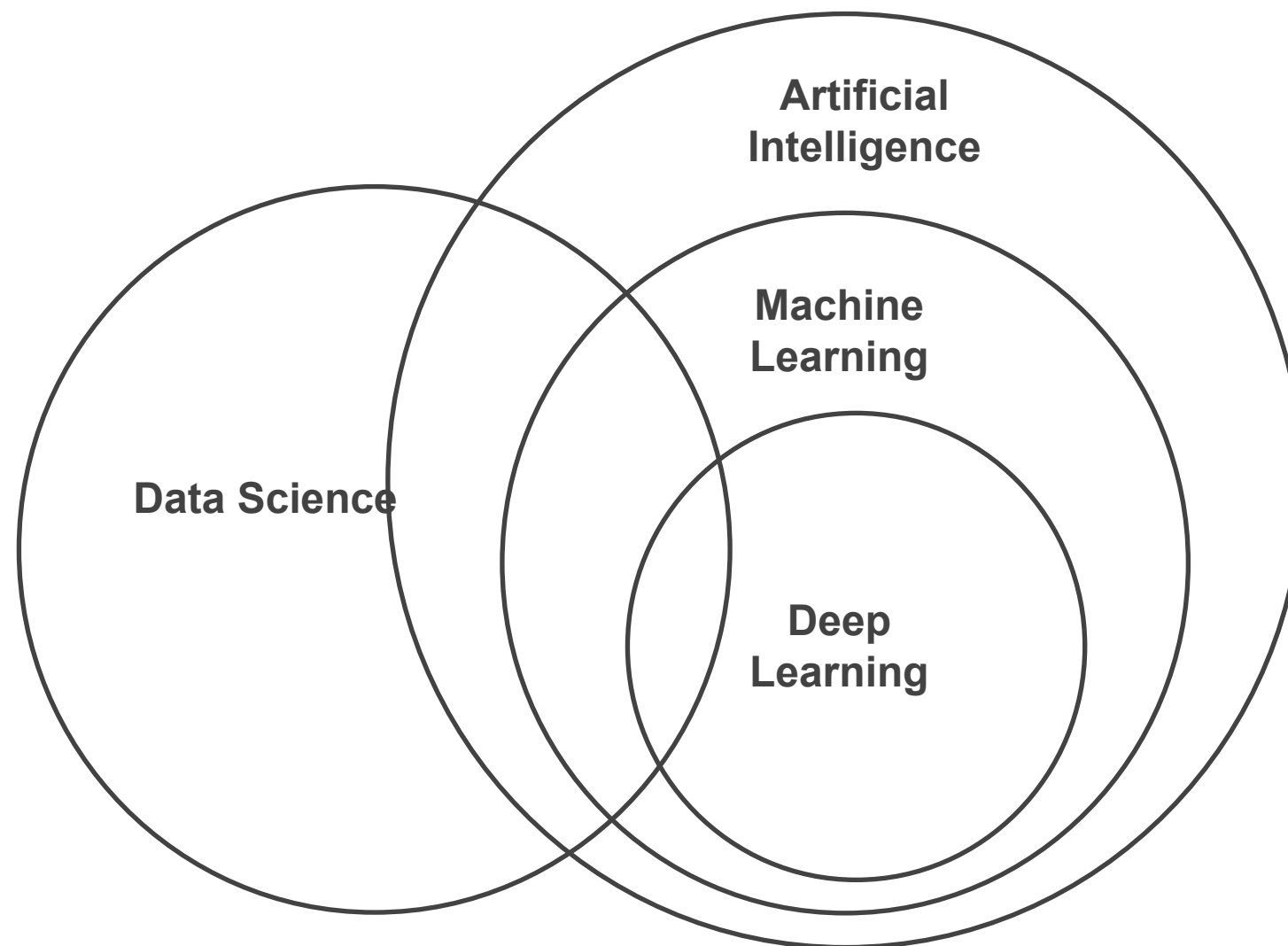
# Onde estou?

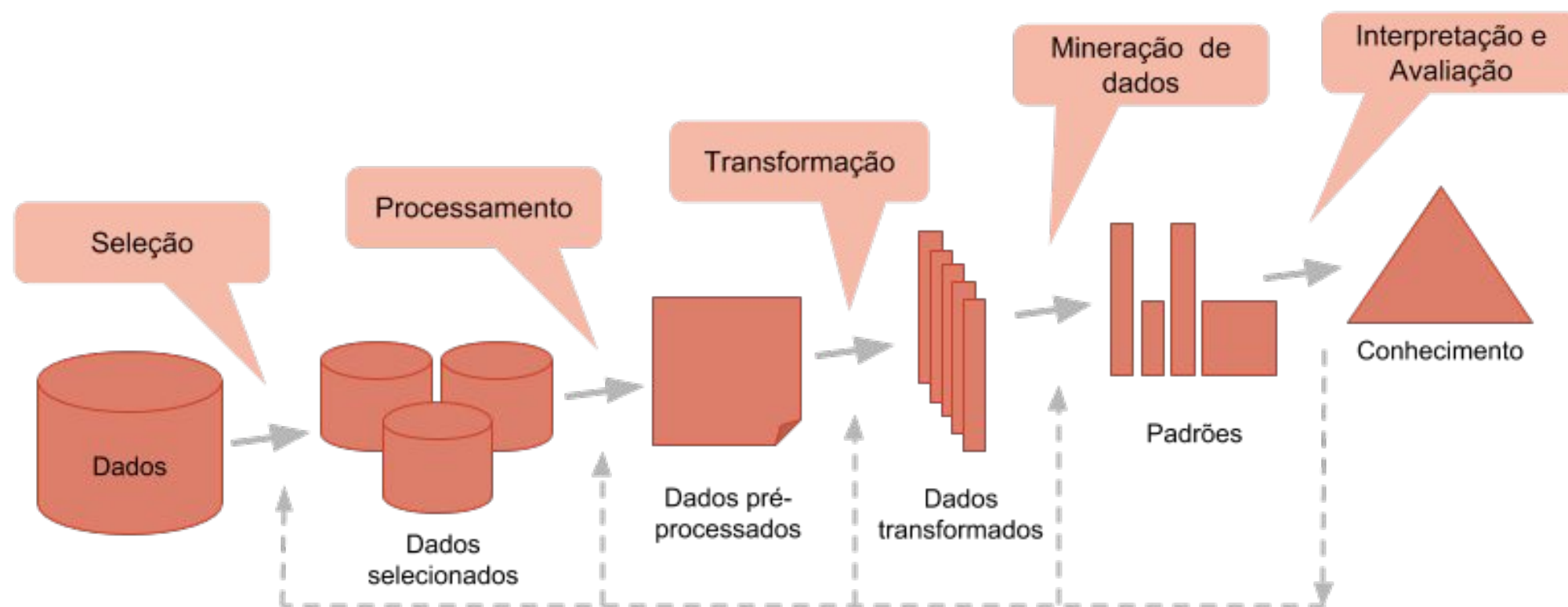
---



# Onde estou?

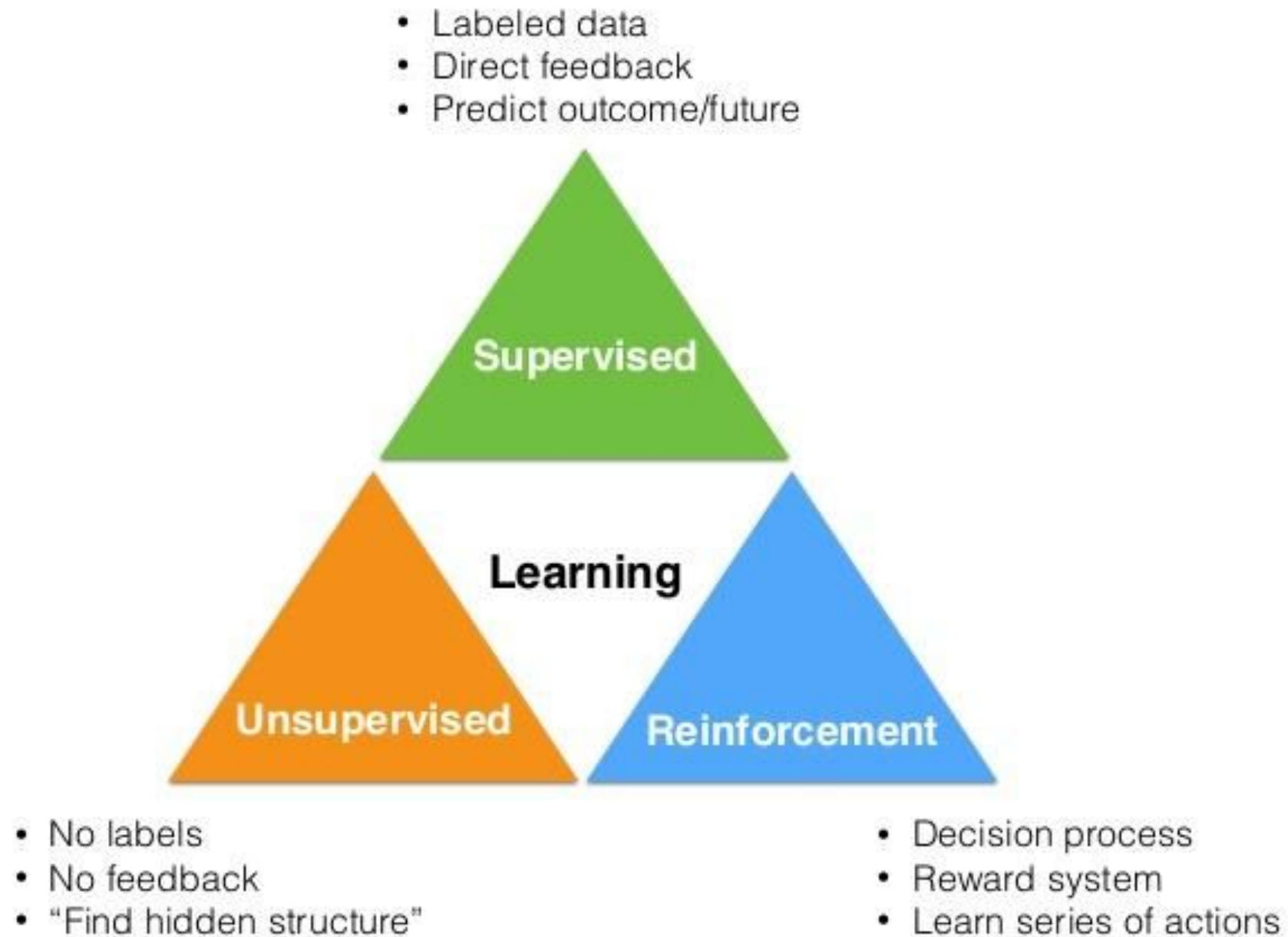
---







# Aprendizado



# Aplicações ML

---

- Uber – Preço dinâmico
- Magazine Luiza – Preço dinâmico
- Quinto Andar – Calculadora de Aluguel
- Nubank – Análise de Crédito
- Netflix – Recomendações Personalizadas

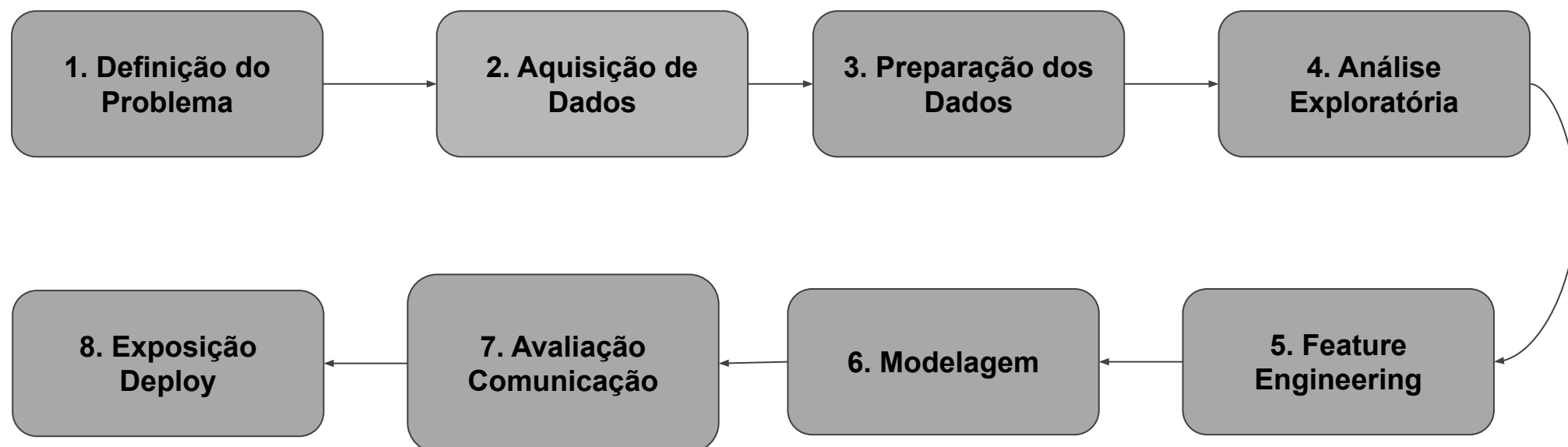
# Aplicações DL

---

- Google - Google Translate
- Android - Pesquisa por voz
- Facebook - Reconhecimento facial
- SkinVision App – Câncer de Pele
- My Heritage – Colorização de Fotos

# Fluxo de Trabalho

## versão 1



# Definição do Problema

---

- Quais dores quero resolver?
- Não conheço o problema?

# Análises

---

- **Descritiva:** O que aconteceu?
- **Diagnóstica:** O que está acontecendo?
- **Preditiva:** O que pode acontecer?
- **Prescritiva:** O que precisamos fazer?

# Aquisição de Dados

---

- Definição da coleta de Dados
- Data estruturados e não estruturados
- Arquitetura de dados

# Preparação dos Dados

---

- Campos com formatos incompatíveis
- Dados nulos
- Valores absurdos



# Análise Exploratória

---

- Análises exploratórias
- Comportamento dos dados
- Outliers
- Respostas para questionamentos
- Novas perguntas

# Feature Engineering

---

- O que são as features?
- Quais as mais importantes?

# Modelagem

---

- Aplicação do modelo
- Ajuste de parâmetros

# Avaliação e Apresentação

---

- Métricas de avaliação
- Storytelling
- Data Viz
- Dashboards
- Pílulas

# Deploy

---

- Disponibilizar para uso
- Produto

“Garbage in, garbage out”



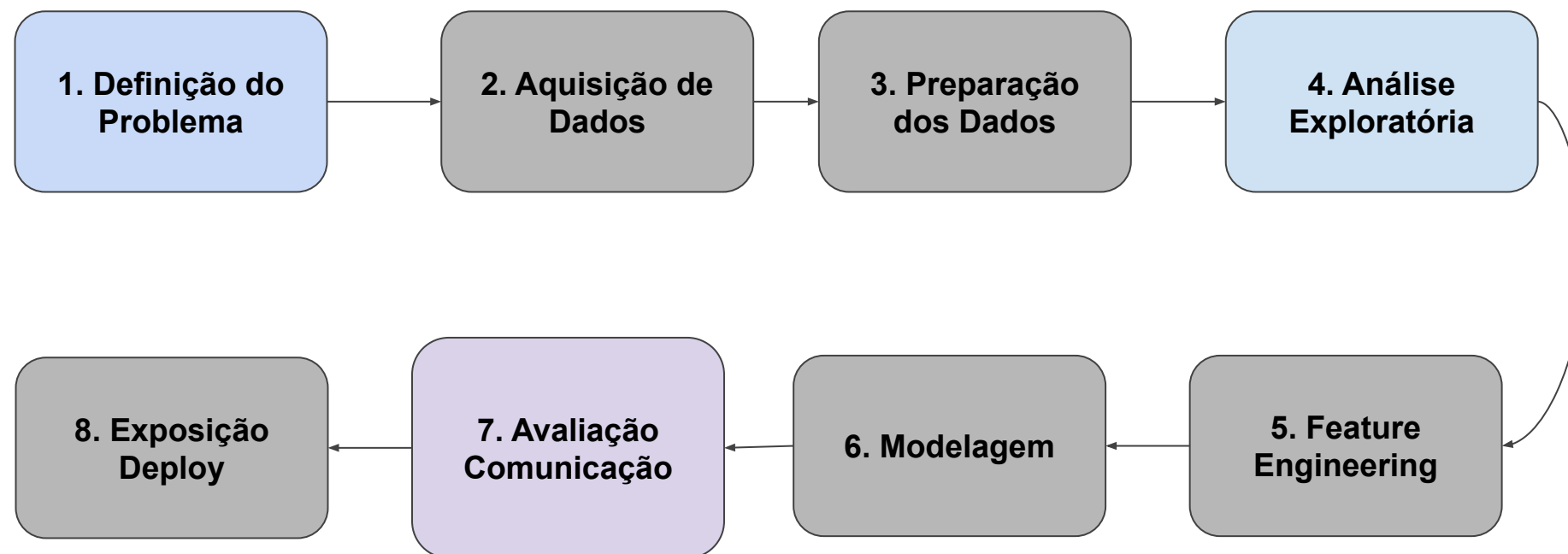
Your analysis is as good as your data.

# Carreiras

---

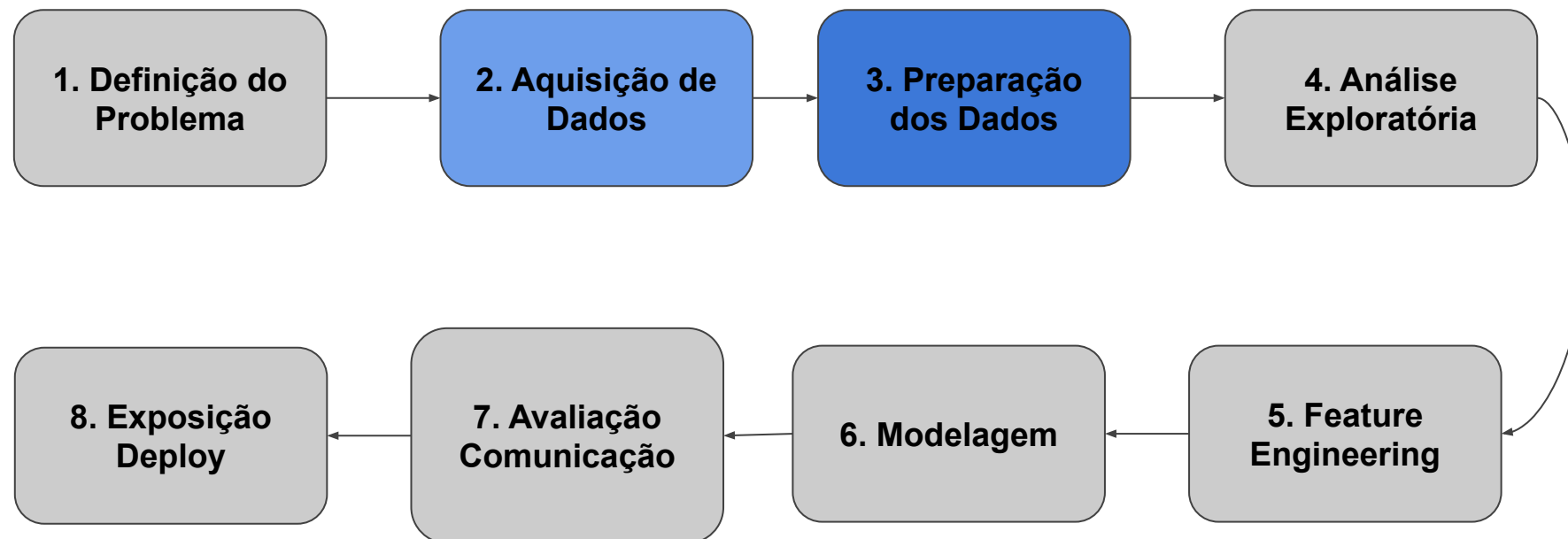
- Engenharia de Dados
- Análise de Dados
- Ciência de Dados

# Análise de Dados

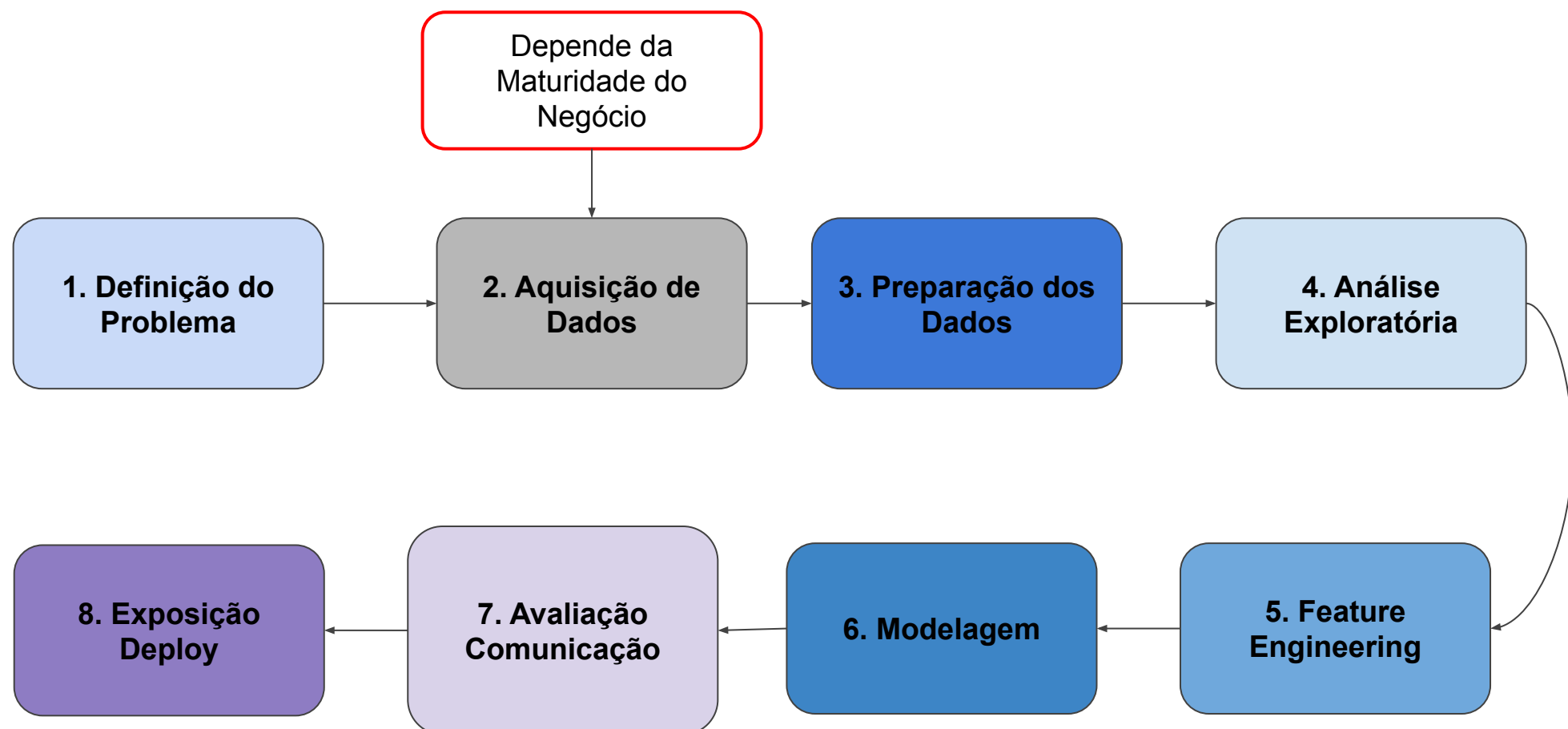




# Engenharia de Dados

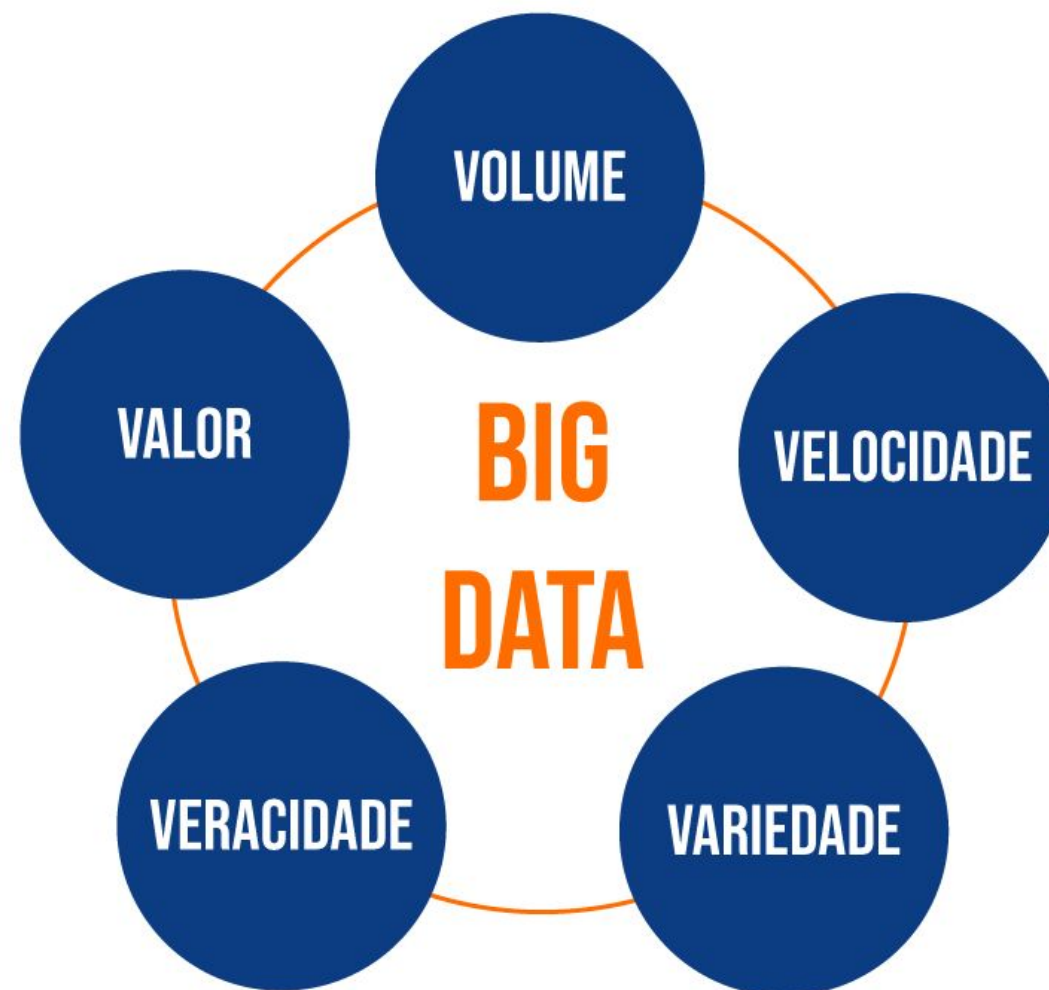


# Ciência de Dados



# E Big Data?

---



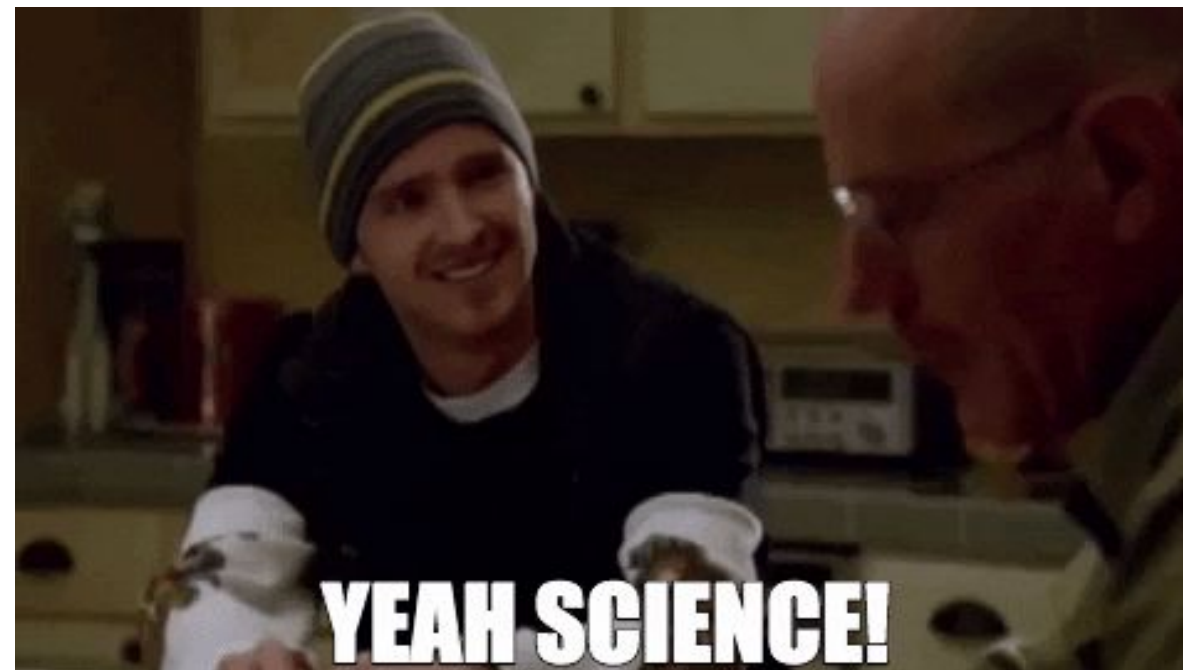
# Nunca esqueça

---

O que faz um  
**cientista** de  
dados?

# Ciência.

---



## Nunca esqueça

---

“Sem **dados** você é apenas mais uma pessoa com uma **opinião**”

William Deming

# Este slide vale 1 milhão de dólares

---

## **Antes de começar qualquer coisa:**

- Conheça o problema
- Pesquise o que já foi feito
- Encontre o estado da arte
- Descubra os desafios e oportunidades

# Fluxo de Trabalho Inteligente





# Compreenderam?

---

Uma empresa de e-commerce fechou seu balanço do ano e deseja estimar qual será o faturamento do ano seguinte se baseado no crescimento executado somado ao esperado.

**Preditiva**

**Descritiva**

**Prescritiva**

**Diagnóstica**

# Compreenderam?

---

Uma empresa de marketing digital realizou no mês passado, uma campanha de lançamento de um novo produto e precisa entender como foi a performance dessa campanha.

**Preditiva**

**Descritiva**

**Prescritiva**

**Diagnóstica**

# Compreenderam?

---

Uma empresa que atua no mercado de ações deseja maximizar os retornos de aplicações, identificando quais ações têm maior probabilidade de ser rentáveis.

**Preditiva**

**Descritiva**

**Prescritiva**

**Diagnóstica**

# Compreenderam?

---

Em uma reunião semanal, o gestor da área, utilizando o dashboard de controle percebe que o número de vendas da semana foi muito superior das anteriores. Ao analisar melhor o dashboard, percebeu que foi realizada um grande evento de promoção gerente novos clientes que tinham um ticket médio elevado.

**Preditiva**

**Descritiva**

**Prescritiva**

**Diagnóstica**

# Imprima essa lista

---

- Leia
- Questione
- Produza
- Comunique
- Documente
- Conecte-se



# Portfólio

---

- LinkedIn
- Medium
- FastPages

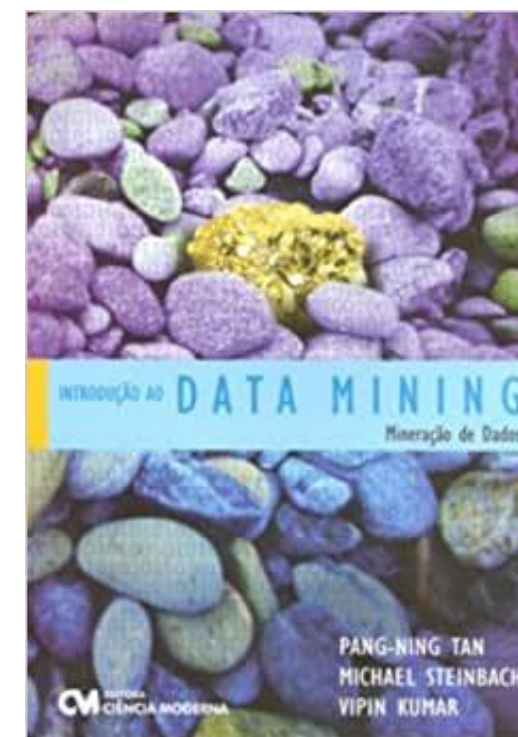
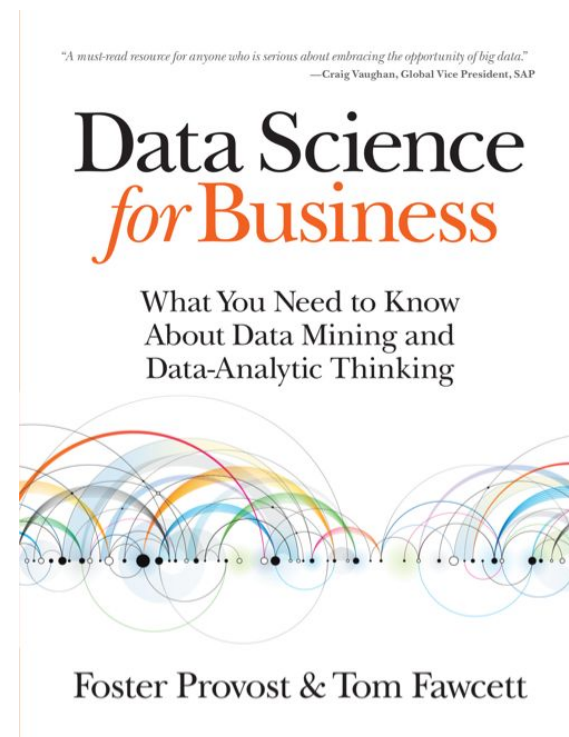
# Repositórios

---

- Github
- Kaggle
- Notebooks
- Projetos Open

# Bibliografia

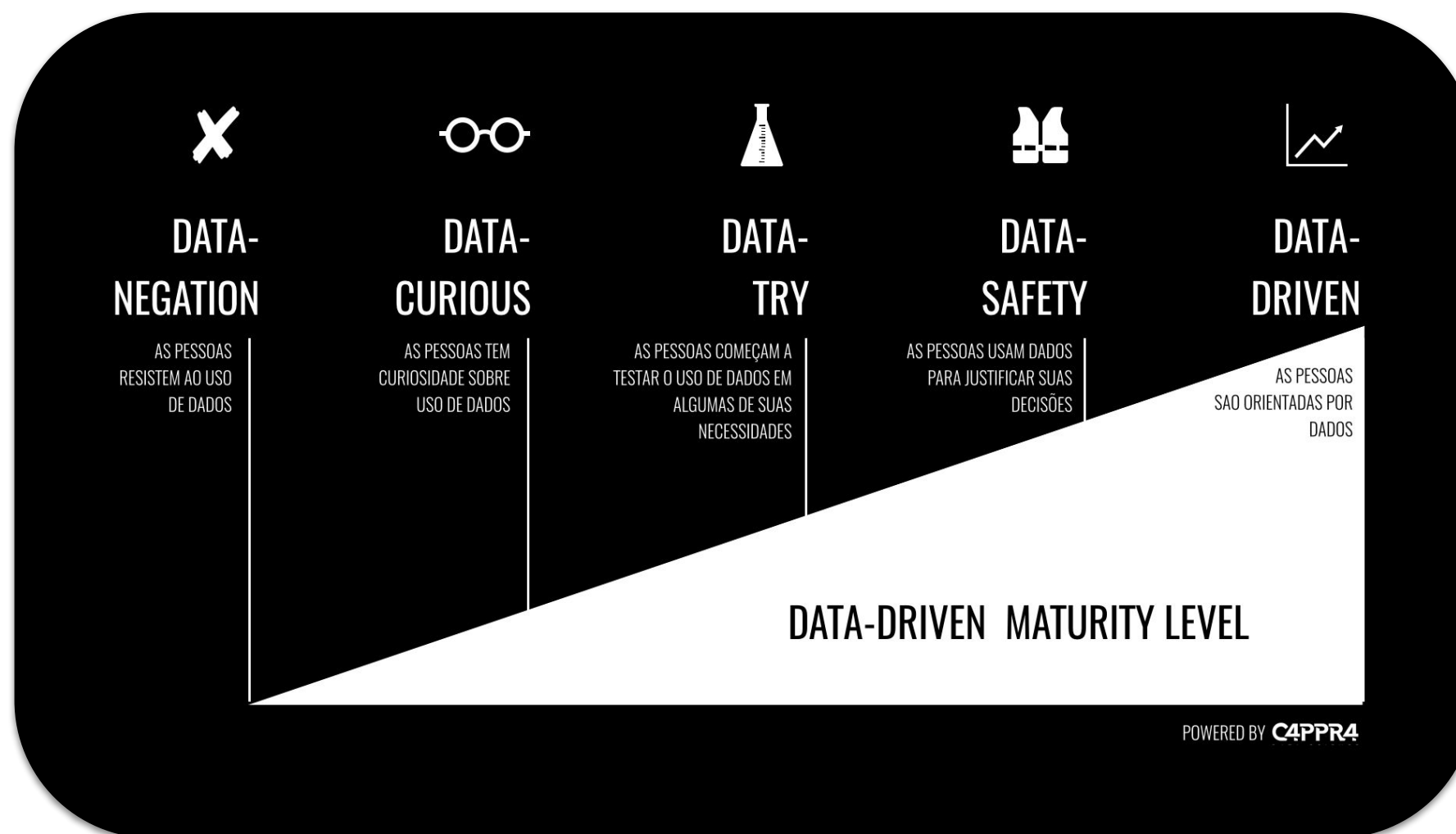
---





# Uma rápida visão do mercado

# Mercado



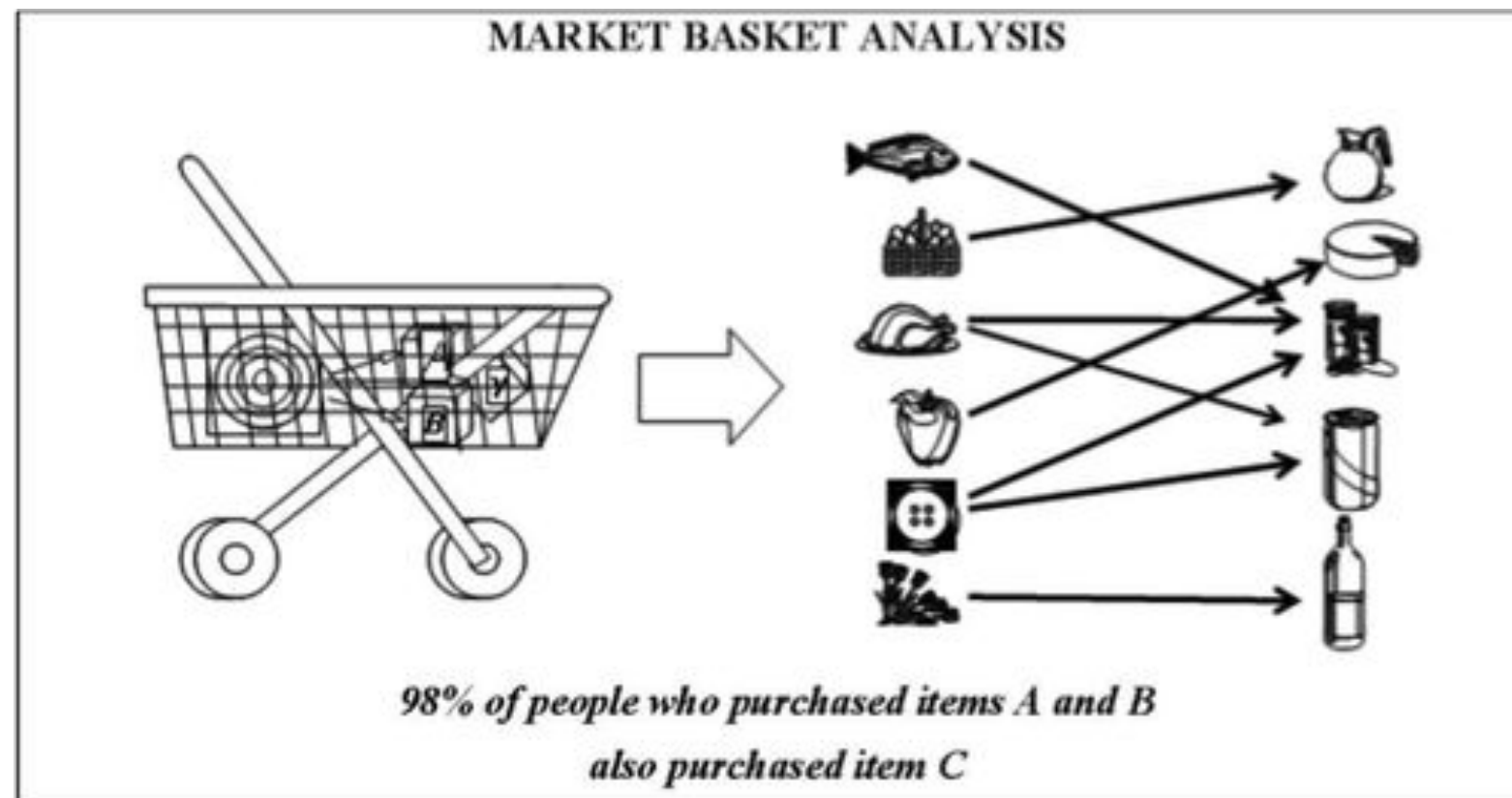
Fonte: Cappra Institute  
Data-driven precisa ser cultura e não um projeto.

# Questionamentos

---

- Tenho dados suficientes?
- Dados confidenciais
- Capacidade de processamento
- Armazenamento

# Mercado



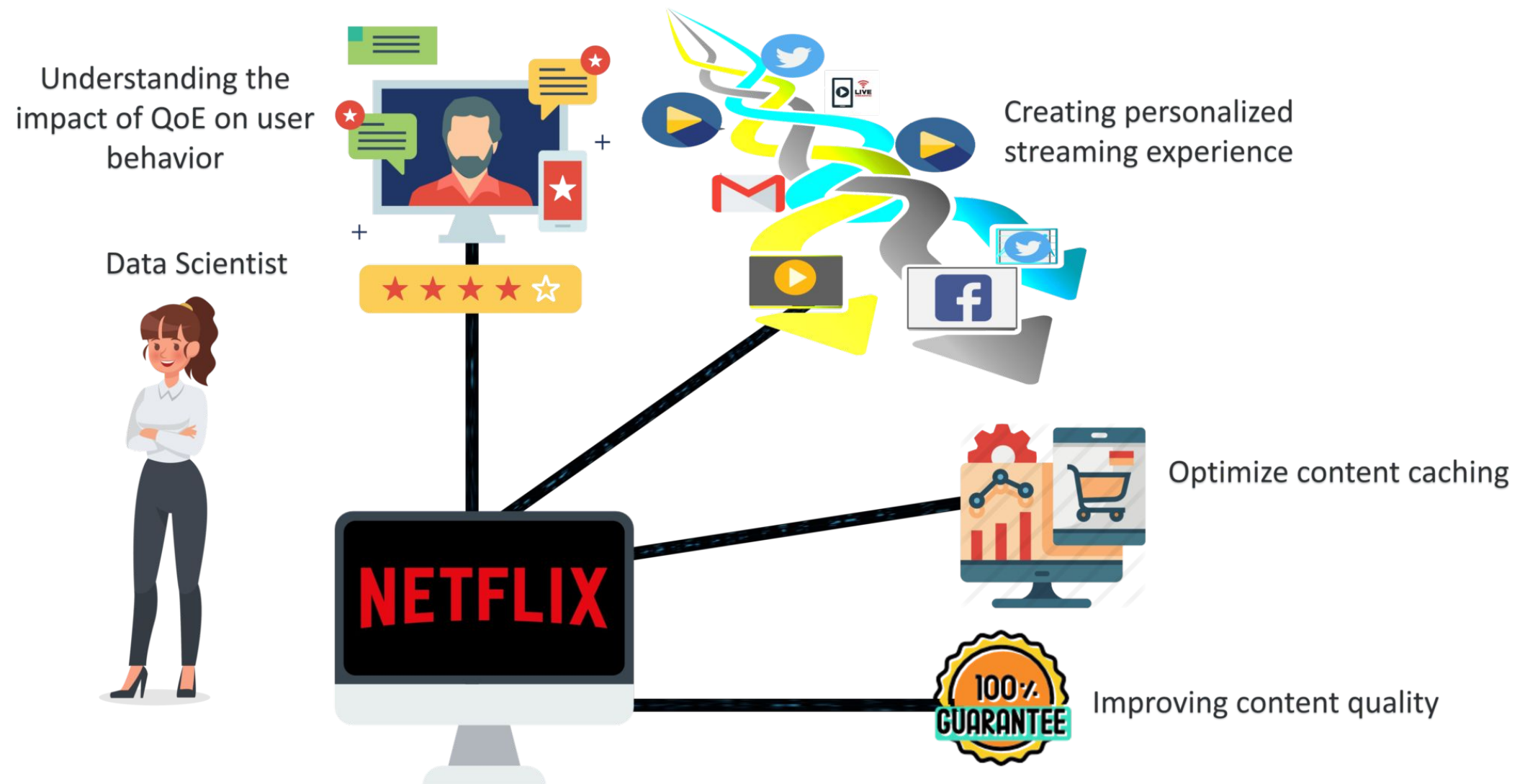
# Mercado

---

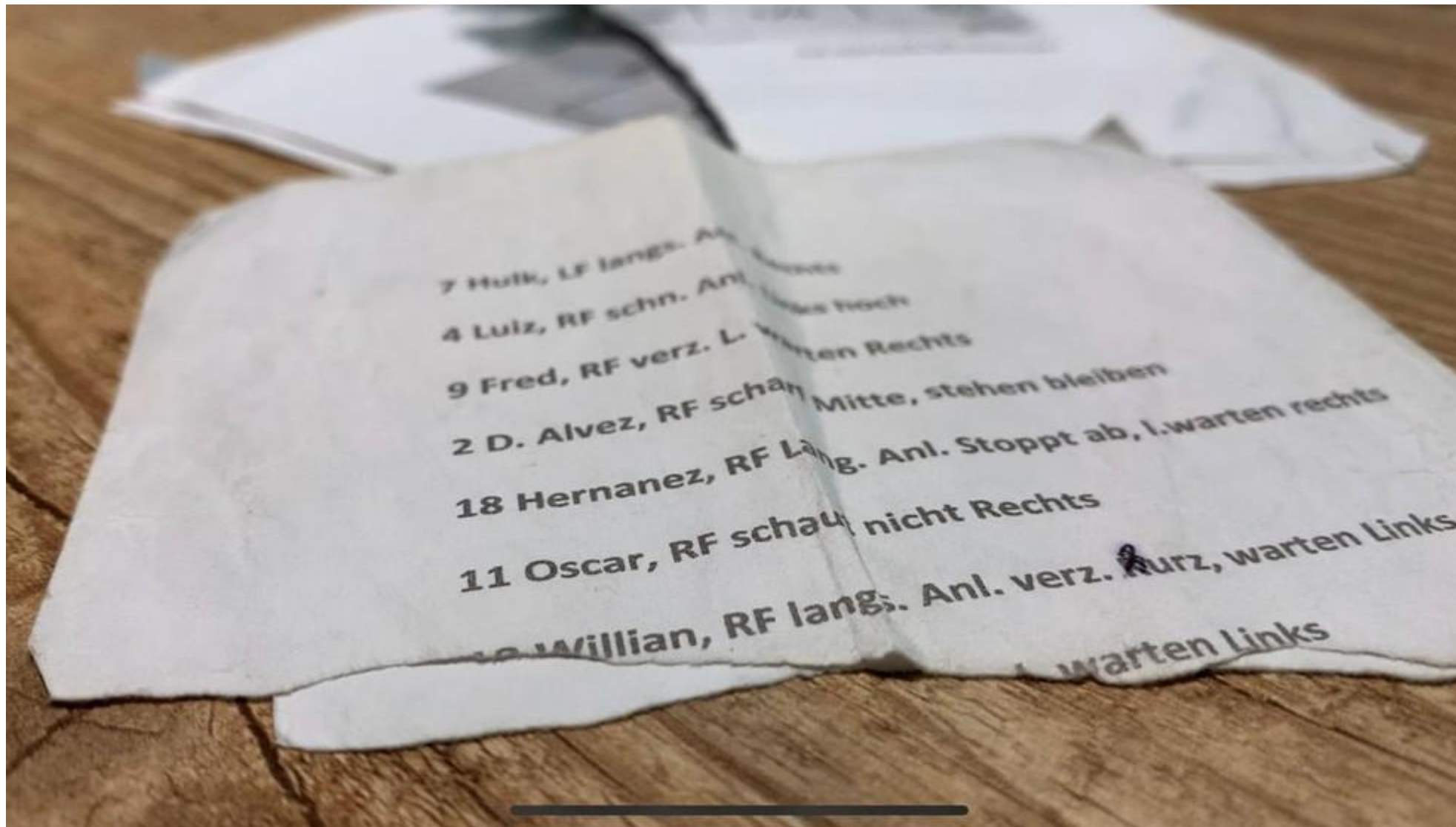


Walmart 

# Mercado

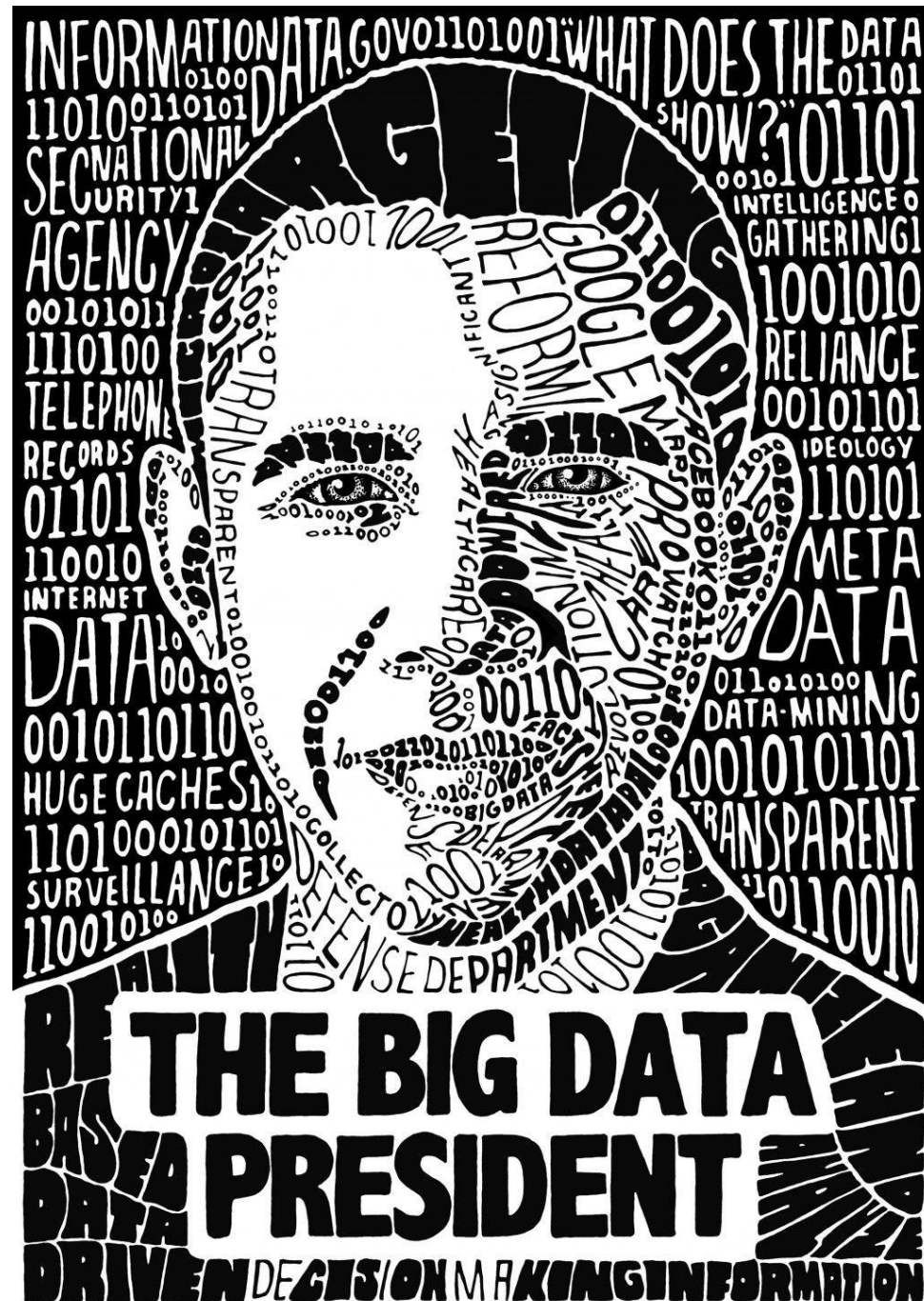


# Mercado



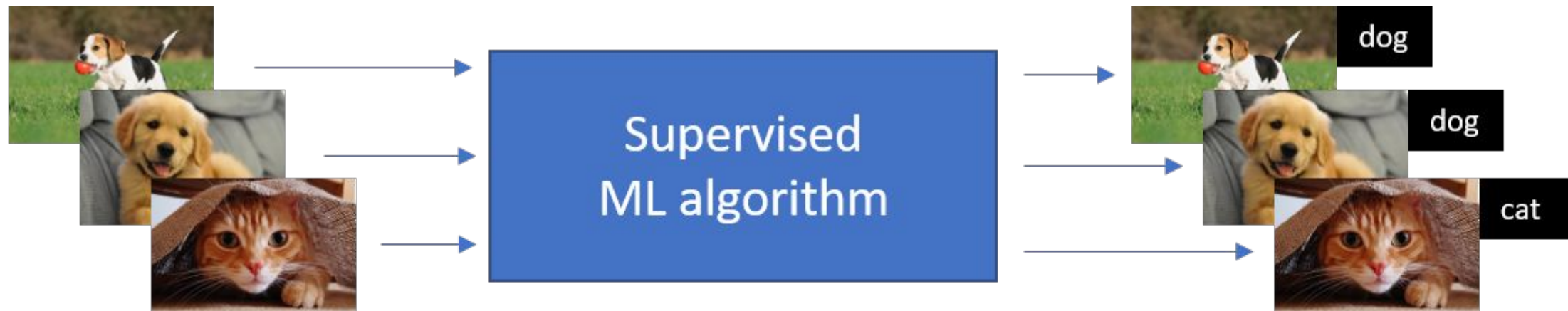


# Mercado

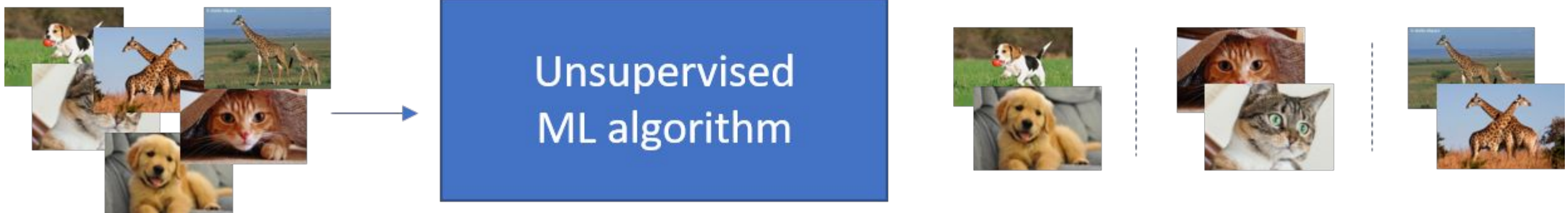




# Algoritmos supervisionados



# Algoritmos Não-supervisionados



# Exemplos

Aprovo ou nego um crédito consignado de R\$2.000,00?

## Risco de Crédito

Idade	24
Profissão	Engenheiro
Renda Mensal	R\$5.000,00
Gênero	Masculino
Anos no emprego	2
Anos na mesma residência	12
Gastos no cartão de crédito no último mês	R\$1.800,00
Total de investimentos	R\$20.000,00
...	...
Total de crédito anterior	R\$0

# Exemplos

---

Identificar números escritos à mão

7	2	1	0	4	1	4	9	5	9
0	6	9	0	1	5	9	7	3	4
9	6	6	5	4	0	7	4	0	1
3	1	3	4	7	2	7	1	2	1
1	7	4	2	3	5	1	2	4	4

# Exemplos

---

## Classificar e-mails como SPAM

- Sinalizar na caixa de entrada se um e-mail é ou não é spam.
- Reconhecer assuntos comuns de SPAM.
- Blacklist

# Exemplos

---

- A partir dos seus critérios você pode construir regras para cada um desses problemas e então seguir essas regras para cada novo e-mail, novo cliente, etc.
- Você constrói essas regras com base na sua experiência de negócio ou nas experiências passadas por especialistas no negócio.
- Por exemplo:
  - Se assunto “ganhe 1 milhão” então SPAM
  - Se assunto “vitamina muda vida” então SPAM

## Questão

---

Como **fornecer** essa experiência para a máquina?

## Uma boa resposta

---

Apontando **atributos**  
(features) e identificando  
os **padrões**.



# Exemplo

## Diferentes Abordagens

Idade	24
Profissão	Engenheiro
Renda Mensal	R\$5.000,00
Gênero	Masculino
Anos no emprego	2
Anos na mesma residência	12
Gastos no cartão de crédito no último mês	R\$1.800,00
Total de investimentos	R\$20.000,00
...	...
Total de crédito anterior	R\$0

- Concede crédito?
  - Sim ou Não
- Nível de Risco:
  - [0 a 5]

# Exemplo

---

## Características Textuais:

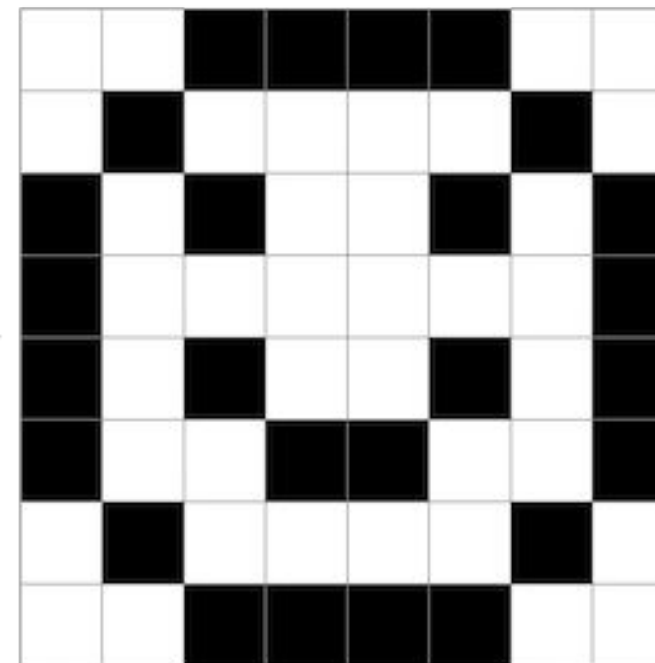
- Frequência de palavras em mensagens classificadas como spam ou não spam.

	george	you	hp	free	!	edu	remove
spam	0.00	2.26	0.02	0.52	0.51	0.01	0.28
email	1.27	1.27	0.90	0.07	0.11	0.29	0.01

# Exemplo

## Imagens e Similaridade

1	1	0	0	0	0	1	1
1	0	1	1	1	1	0	1
0	1	0	1	1	0	1	0
0	1	1	1	1	1	1	0
0	1	0	1	1	0	1	0
0	1	1	0	0	1	1	0
1	0	1	1	1	1	0	1
1	1	0	0	0	0	1	1



# Similar não é igual

---



# Segmentação

---



# Diversas Fontes

---

- Bancos de Dados
- Dispositivos de IOT
- Arquivos
- Apis
- Web
- ...

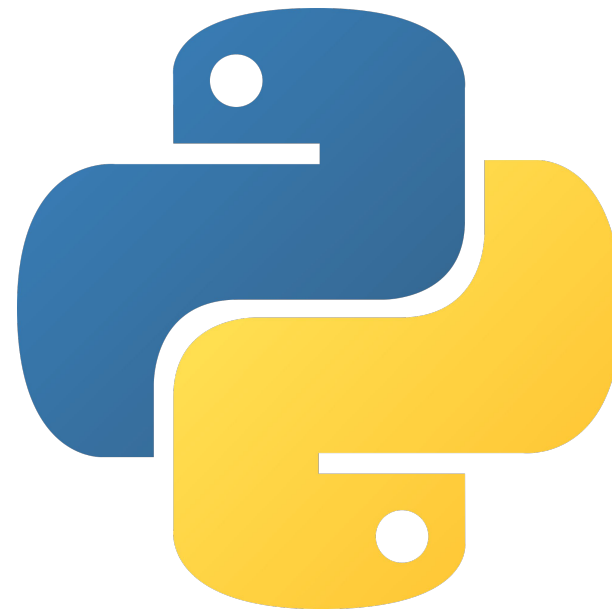
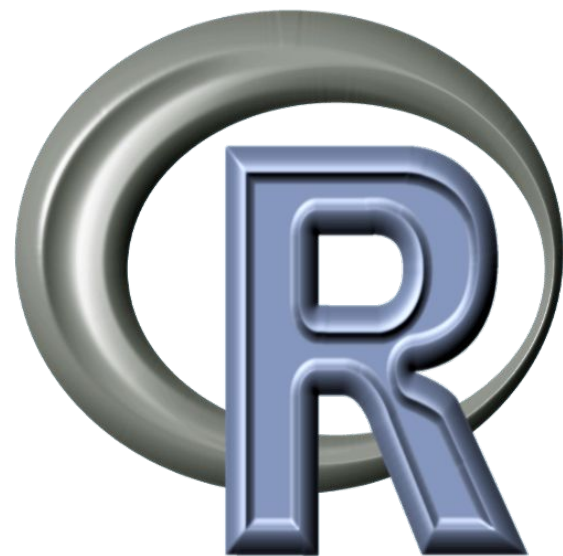


# Algumas Ferramentas



# Linguagens de Programação

---





# De 0 a 5:

---

**O qual confortável você está com:**

- Python
- Estatística

# Mão na Massa

---

