



# Inteligência Artificial

Obtendo resultados, ontem, hoje e sempre

# Professor



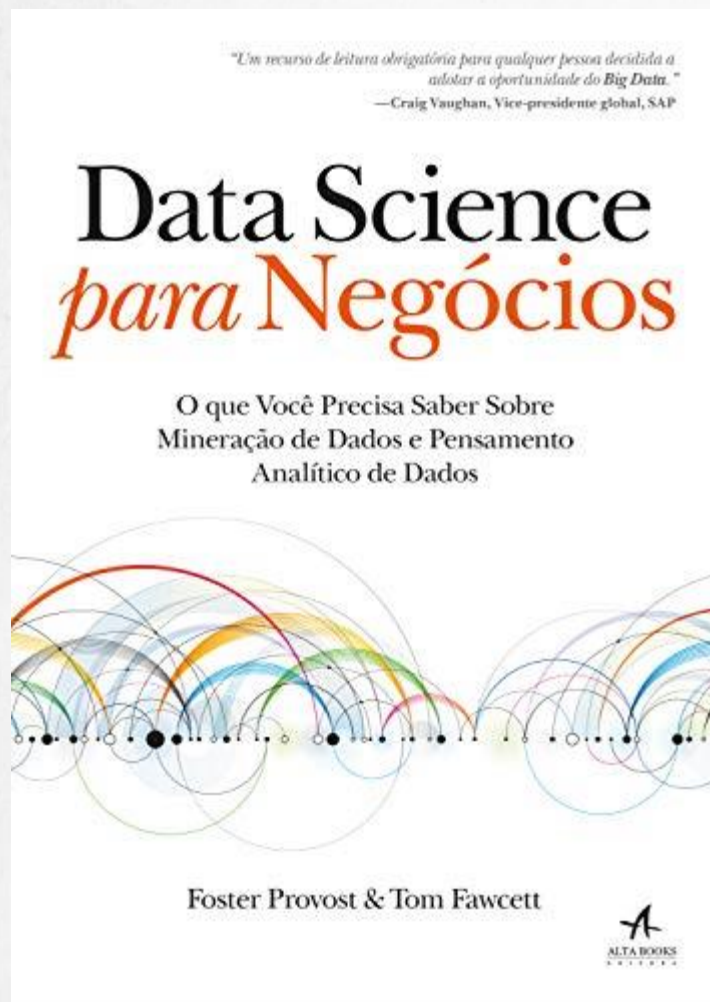
Gustavo Correa Mirapalheta

Eng.Eletrico (UFRGS), Doutor em Adm.de Empresas (FGV). Professor de Tecnologia e Ciência de Dados (FGV). Consultor em Aplicações de Inteligência Artificial. Foi gerente e diretor de vendas de software na IBM Brasil e Sun Microsystems. Seus interesses de pesquisa estão nas áreas de Deep Learning, Natural Language Processing e Quantum Computing.

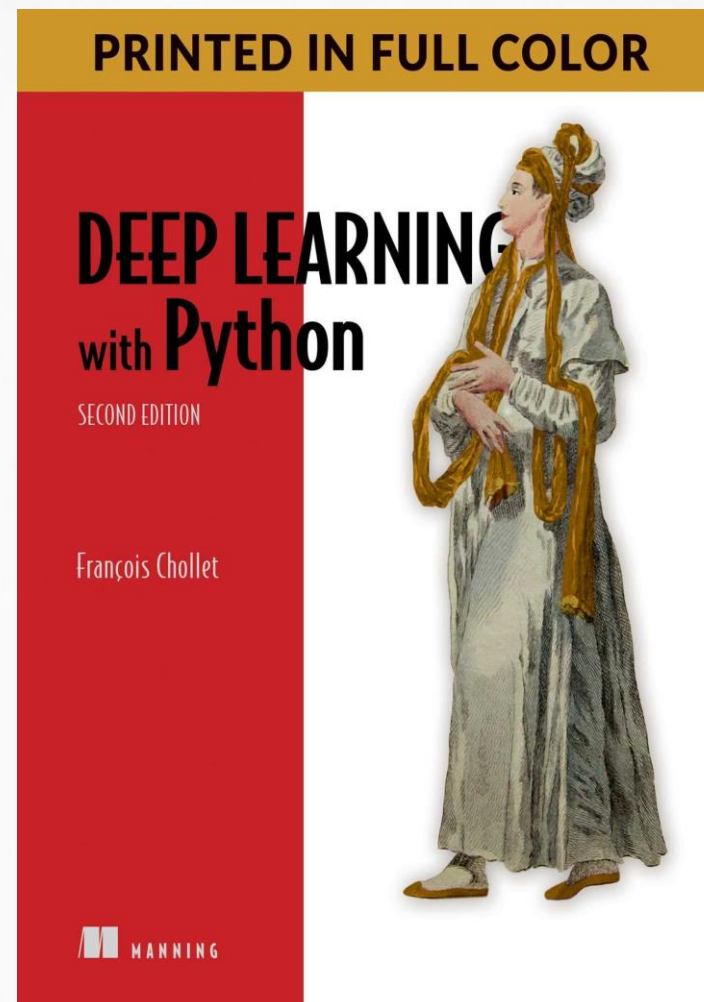
[gustavo.mirapalheta@gmail.com](mailto:gustavo.mirapalheta@gmail.com)

+55 11 9 9204 1201

# Bibliografia – Negócios e Python



PROVOST, Foster & FAWCETT, Tom *Data Science para negócios: O que você precisa saber sobre mineração de dados e pensamento analítico de dados.* Alta Books, 2016, 1ª edição.



CHOLLET, François *Deep Learning with Python.* Manning, 2022, 2nd ed.

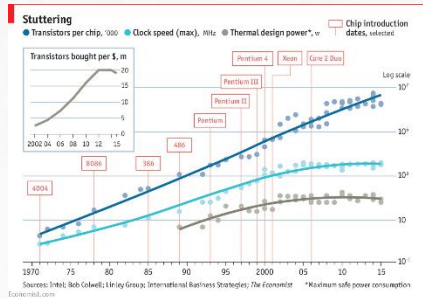
# Agenda

- Evolução Tecnológica
- Aprendizado de Máquina
- Ciclo de vida dos dados
- Modelos Preditivos
- O Dia-a-Dia



# Evolução Tecnológica

## Lei de Moore



Velocidade

Variedade

Inteligência Artificial

Big Data

Volume

Volume

Velocidade

Programação Clássica



# O que nos reserva o futuro?

**Cenário atual: Desafios sociais cuja solução irá requerer a gestão de massas de dados cada vez maiores**

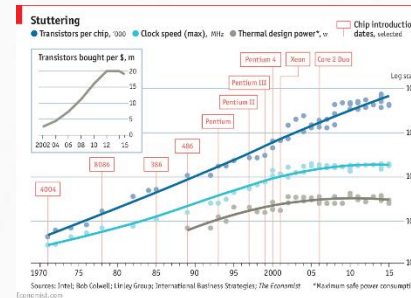


**Big Data**

**Volume**

**Velocidade**

## Lei de Moore



**Cenário atual: O fim da Lei de Moore (tal como a conhecemos)**

**Computação Quântica**



**Velocidade**

**Variedade**

**Volume**

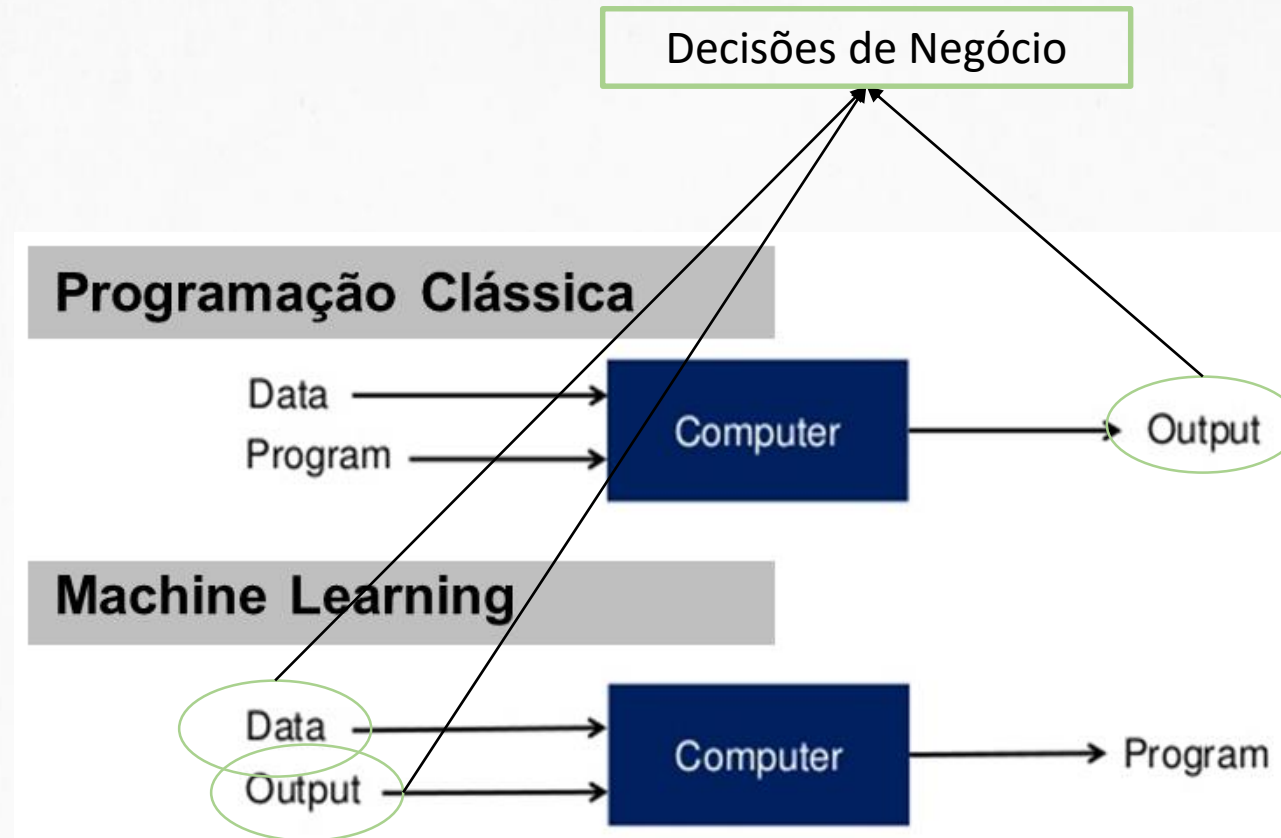
**Programação Clássica**



**Machine Learning**



# Aprendizado de Máquina



# Ciclo de Vida dos Dados





# Exemplos



## Bancos

- Oferta de crédito
- Prevenção de atrito com agência reguladora (Bacen)
- Recuperação de dívidas.



## Marketing

- Análise de sentimento
- Segmentação de perfis



## Seguros

- Precificação de produtos
- Identificação de Fraudes



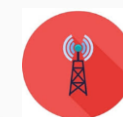
## Saúde

- Atendimentos hospitalares
- Prevenção de doenças



## RH

- Seleção de currículos
- Custos trabalhistas



## Telecom

- *Churning* de contratos
- Compra de serviços
- Minimização de erros em call-center (com text mining)

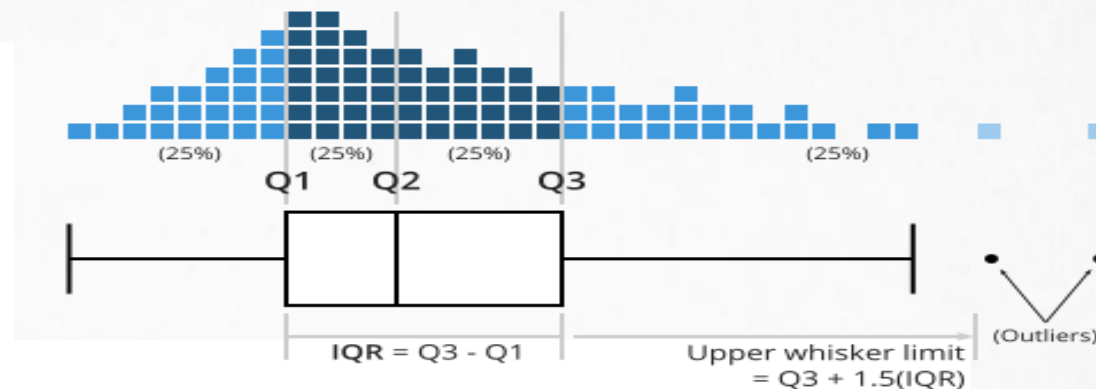
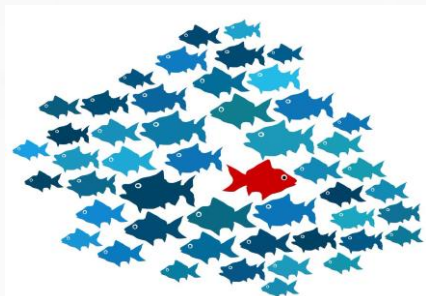
# Ciclo de Vida dos Dados



# Encontrar os Dados Relevantes

## Seleção de Variáveis e Limpeza de Valores

Missing Values



Quais variáveis iremos utilizar?

Source 1

Source 2

⋮

Source n

Select and  
merge

Clean and  
transform

Raw  
Data

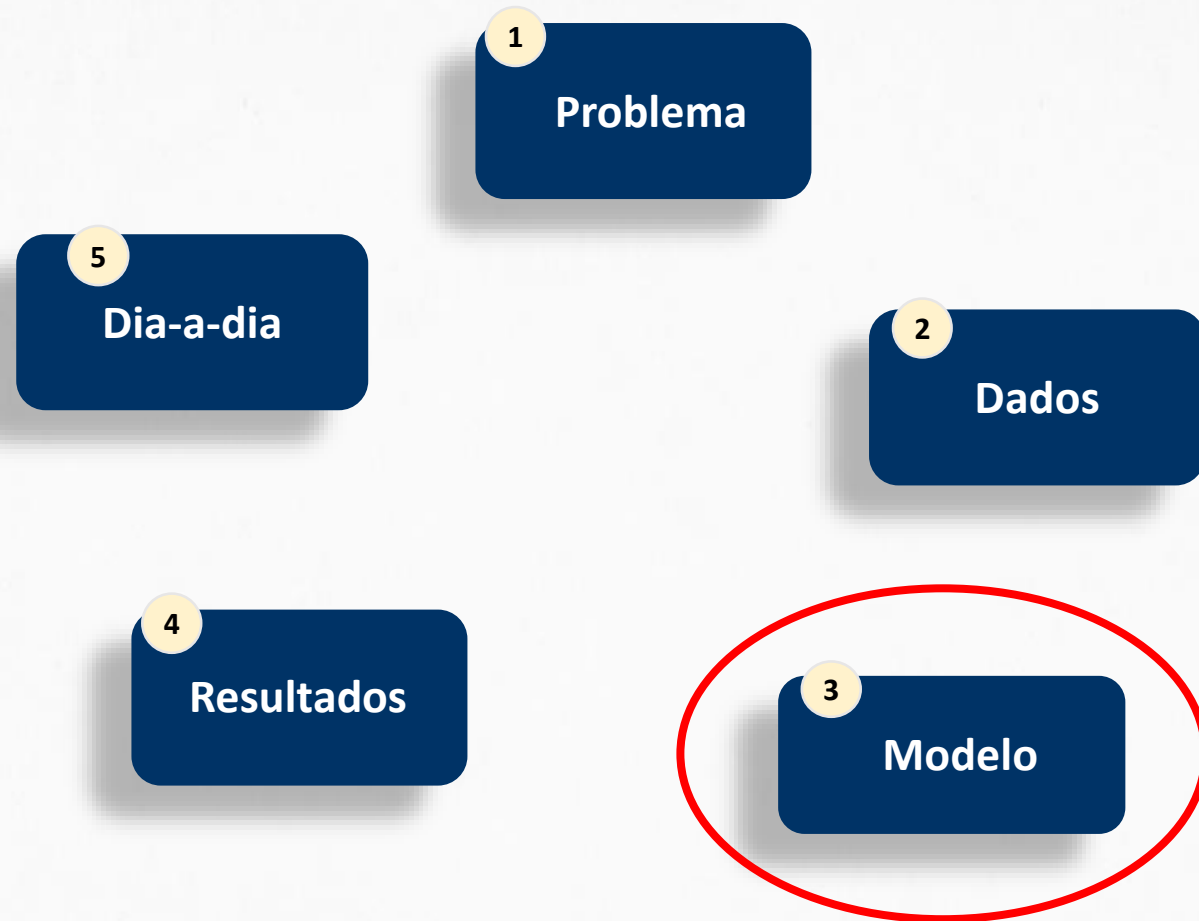
Features

Modeling

Insights

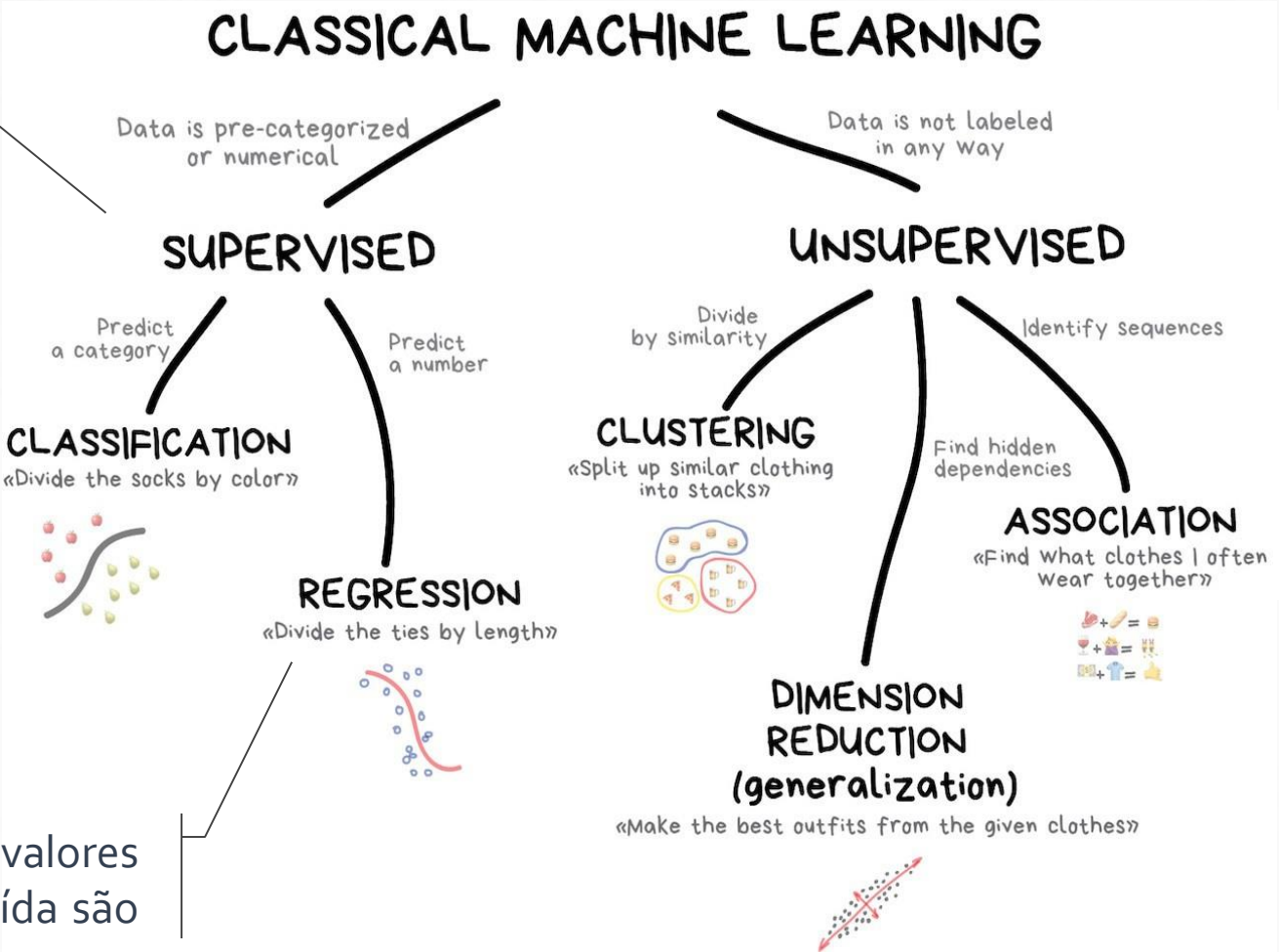
Ciclo de Vida dos Dados

# Ciclo de Vida dos Dados



# Modelos Supervisionados

Temos exemplos dos resultados esperados



X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>p</sub>	Y

Here, the target variable Y can "supervise" how the algorithm builds the model

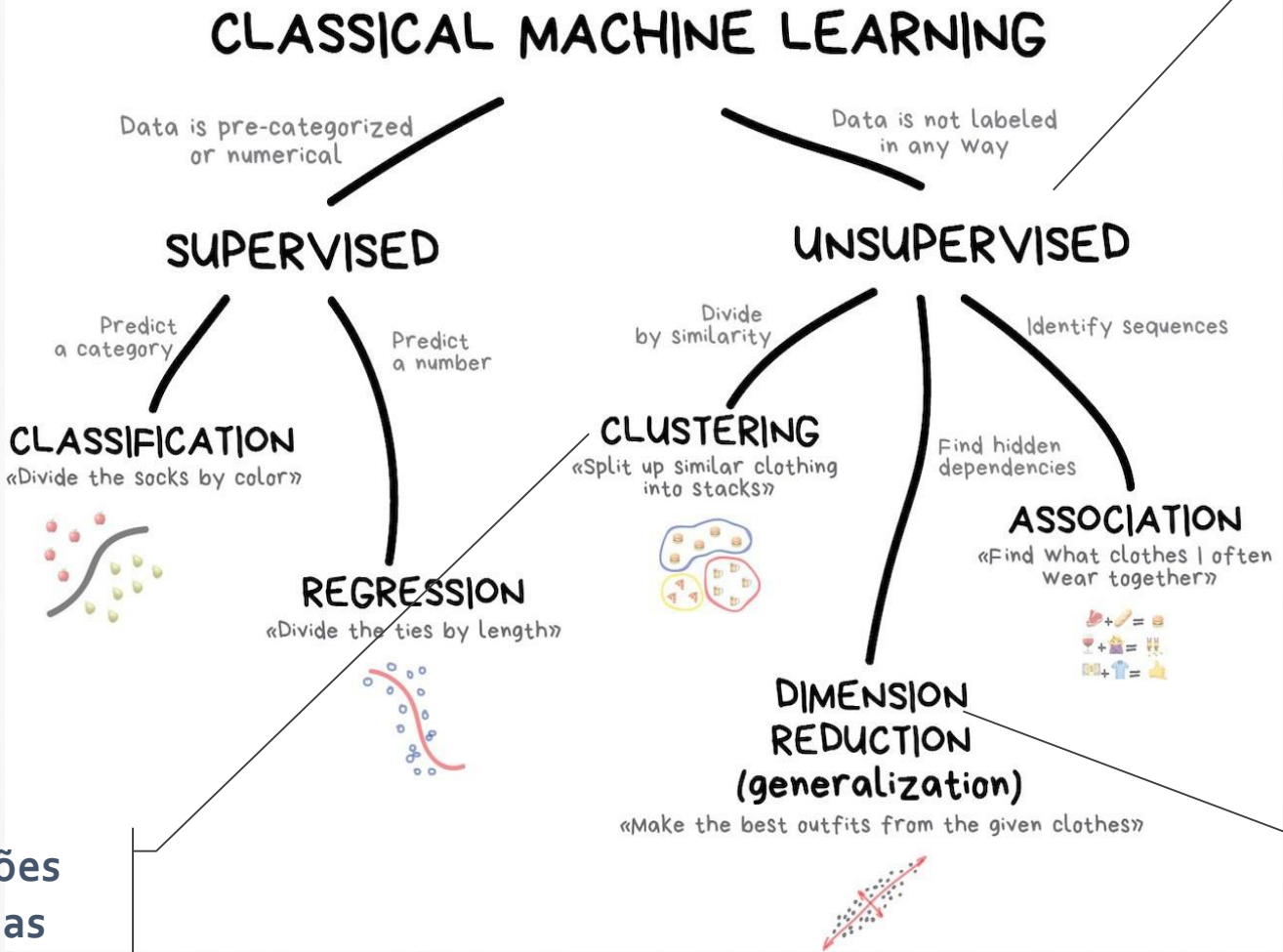
Os valores de saída são categóricos

Os valores de saída são contínuos



# Modelos Não Supervisionados

A máquina  
irá gerar os  
valores de  
saída para  
nós



In unsupervised learning, there is a set of input variables

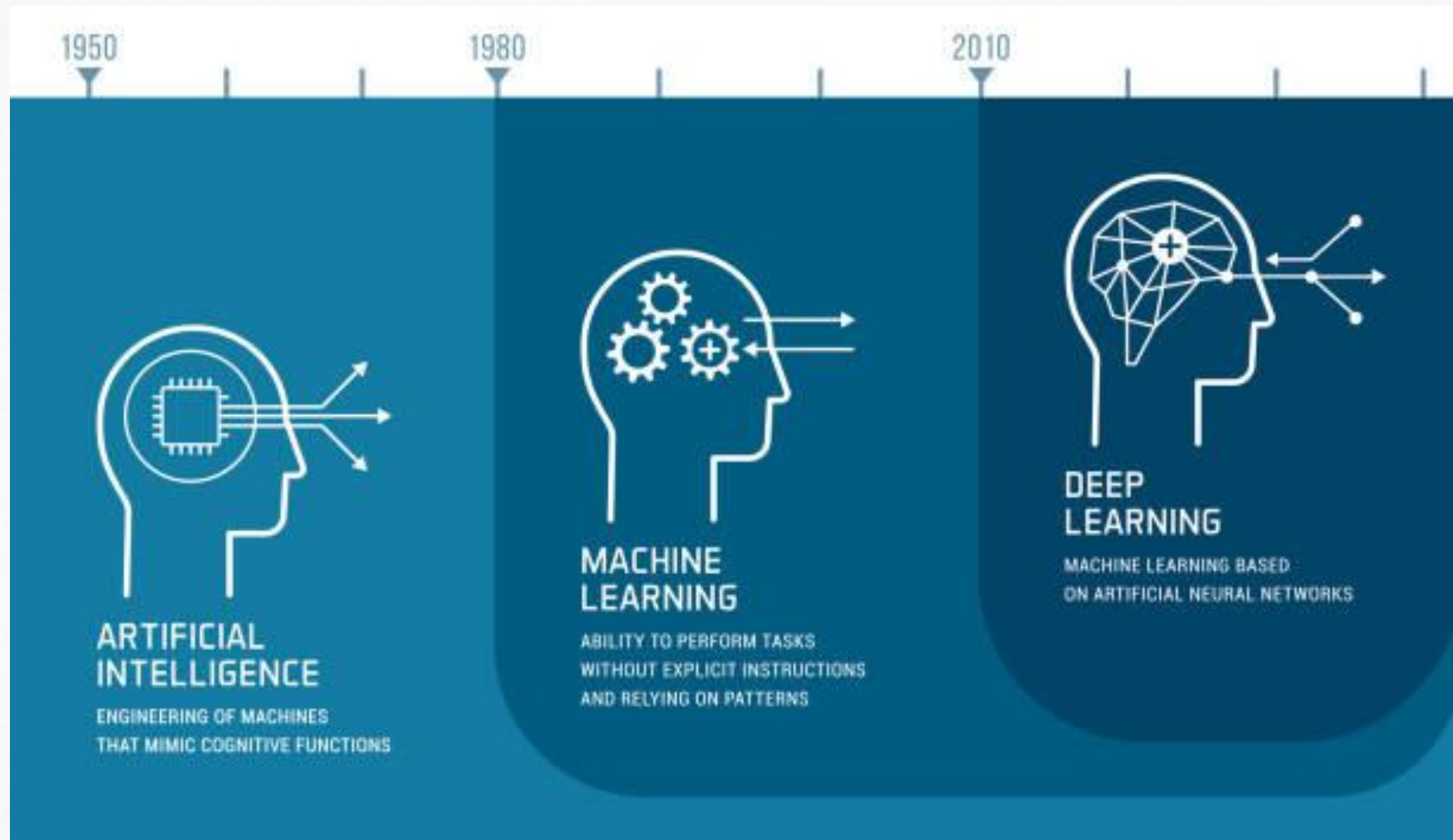
X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>p</sub>	Y

But, there's no target variable, Y.

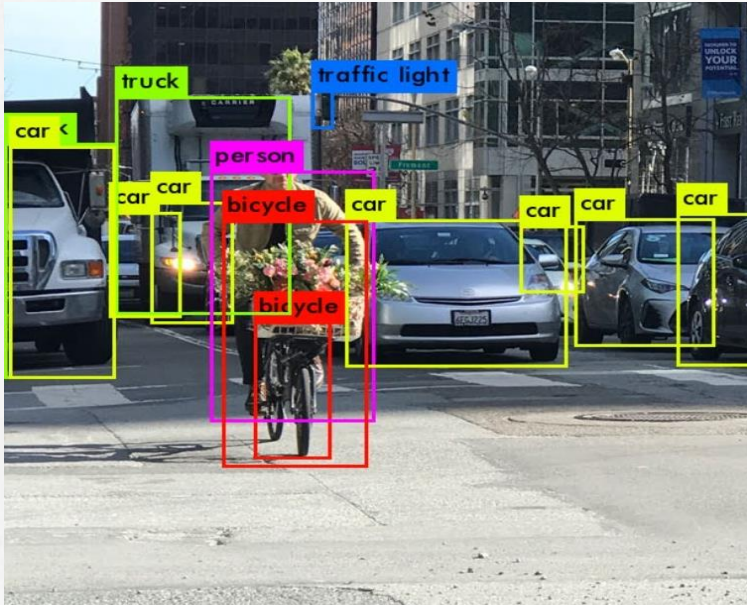
Marcações  
nas linhas

Marcações  
nas colunas

# Machine Learning *on Esteroide*: Redes Neurais (a.k.a. Deep Learning)



# Aplicações Recentes e Futuras (nem tanto...)



Reconhecimento de Imagens

Criação de Imagens

Veículos Autônomos

Logística: automação do *last mile* de entregas



RH: Recrutamento e Seleção.

Educação: Correção automática de provas

Treinamento: Instrução automática e personalizada

Direito: Criação inicial de documentos

Suporte e Atendimento: Automatização do 1º Nível

Geração de Texto



# Novas Soluções vs. Antigas Questões

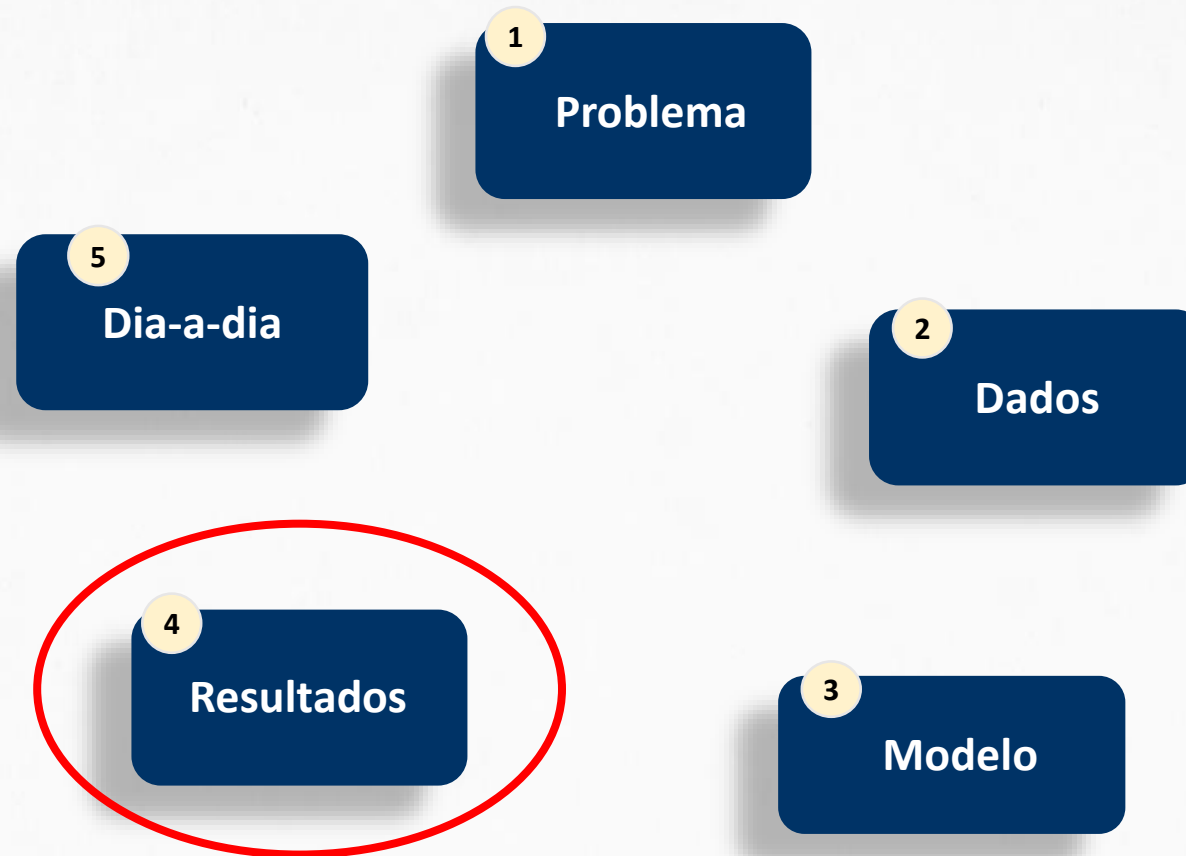
Ele vai acabar passando no *Teste de Turing*.  
Ele estará *consciente*?



Mas como vamos diferenciar uma *máquina* que imita a consciência da *consciência* “real”?

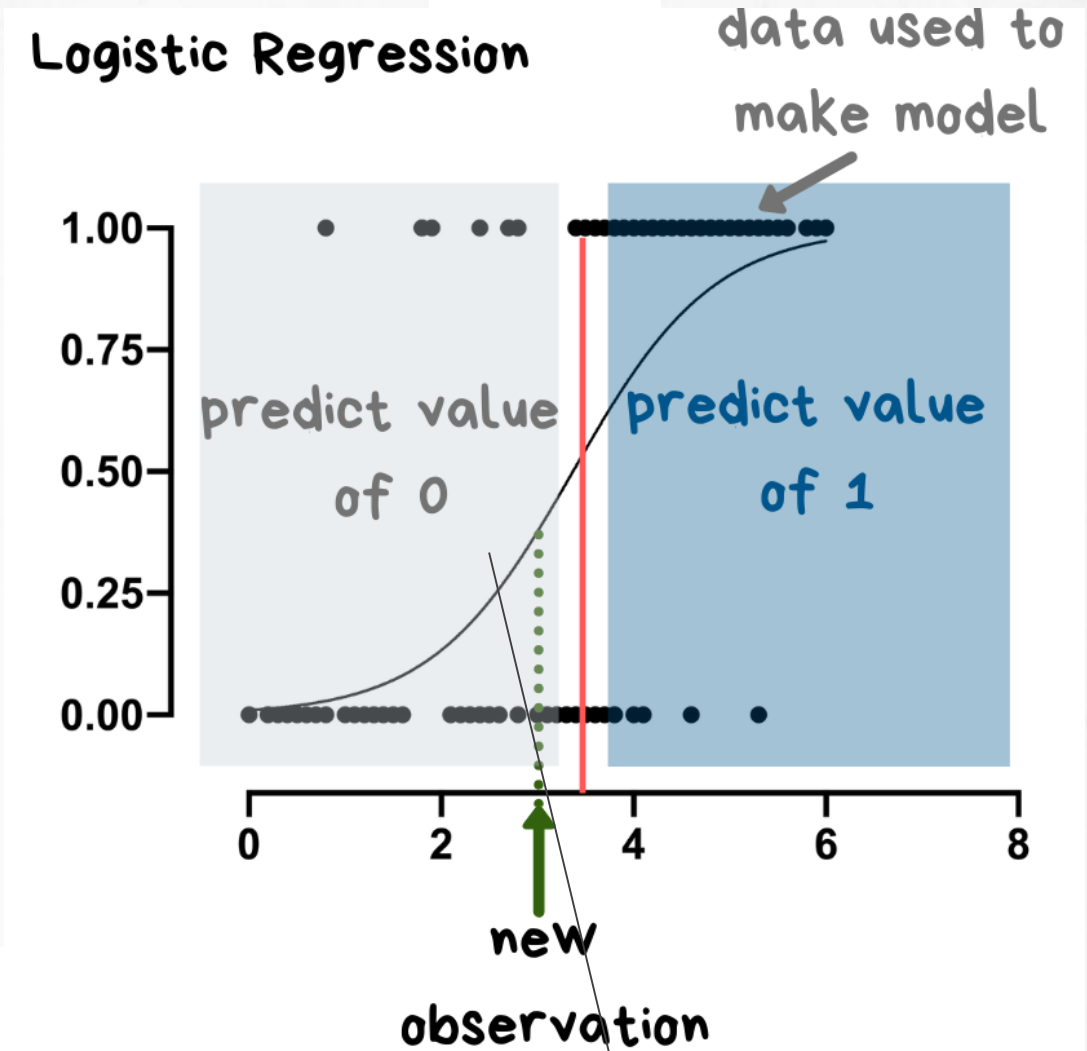
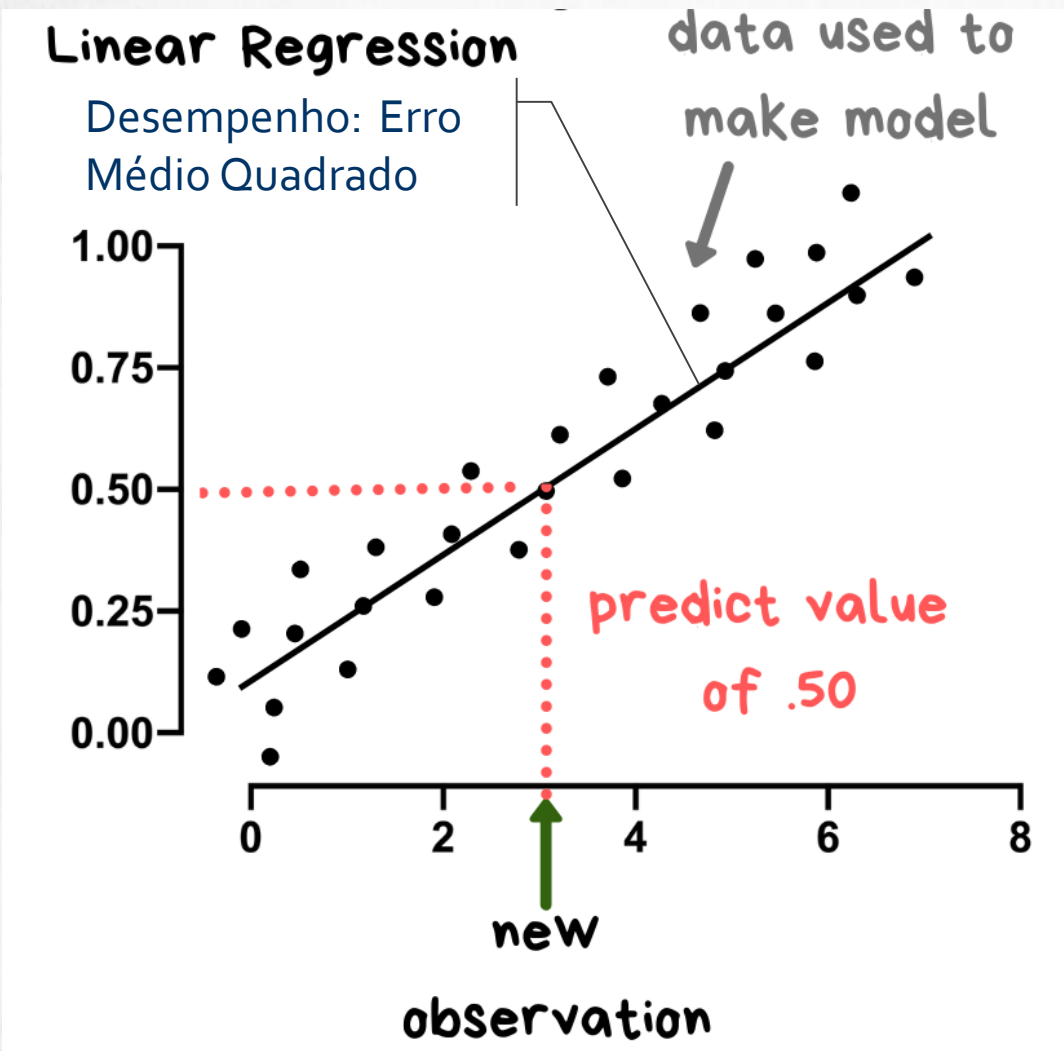


# Ciclo de Vida dos Dados



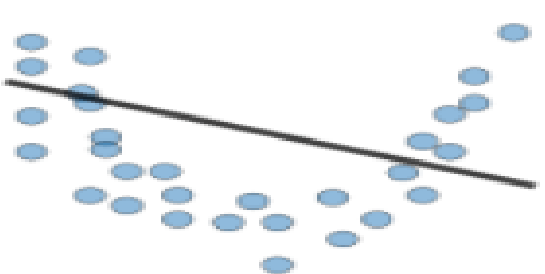
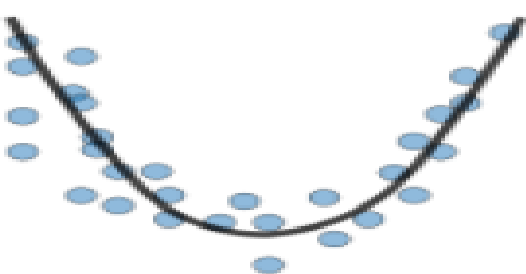

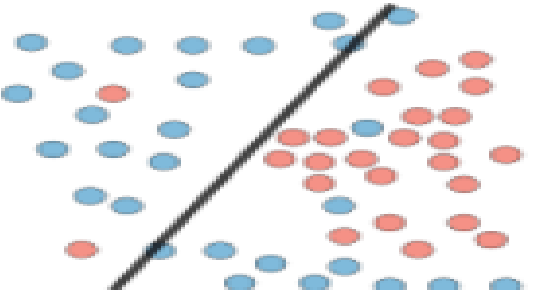
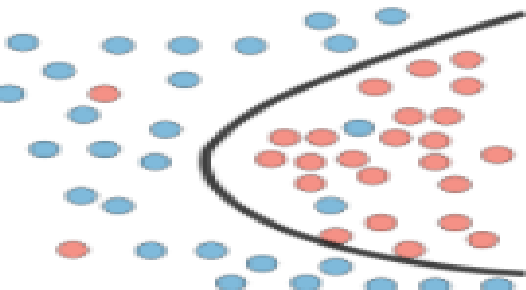
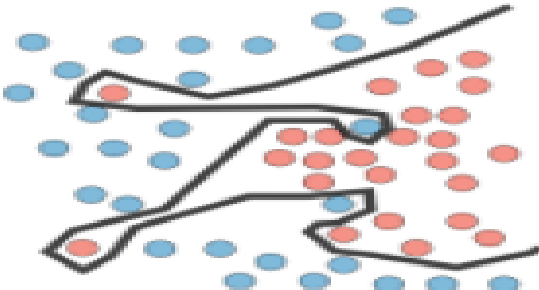


# Modelos Supervisionados - Desempenho

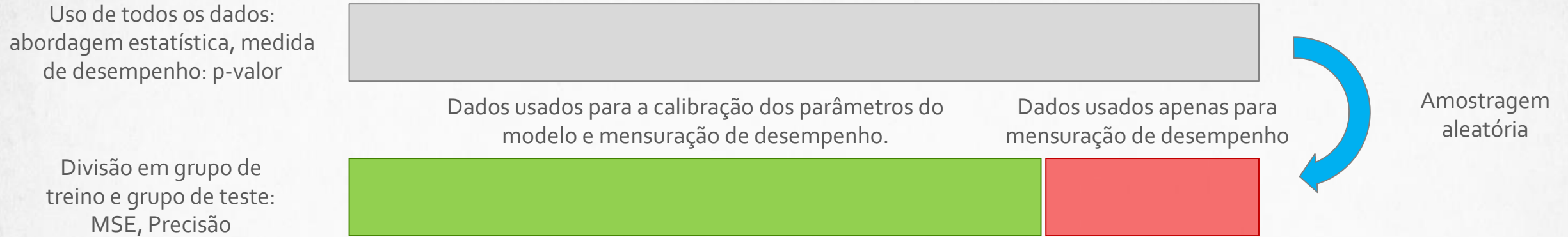


# Modelos Supervisionados – Sobre ajuste

- O ajuste de um modelo aos dados pode ser visualizado de duas formas:
  - Na previsão de valor pela melhor linha que passa pelo meio dos dados
  - Na previsão de categoria pela linha que separa pontos de cores distintas de cada lado da mesma

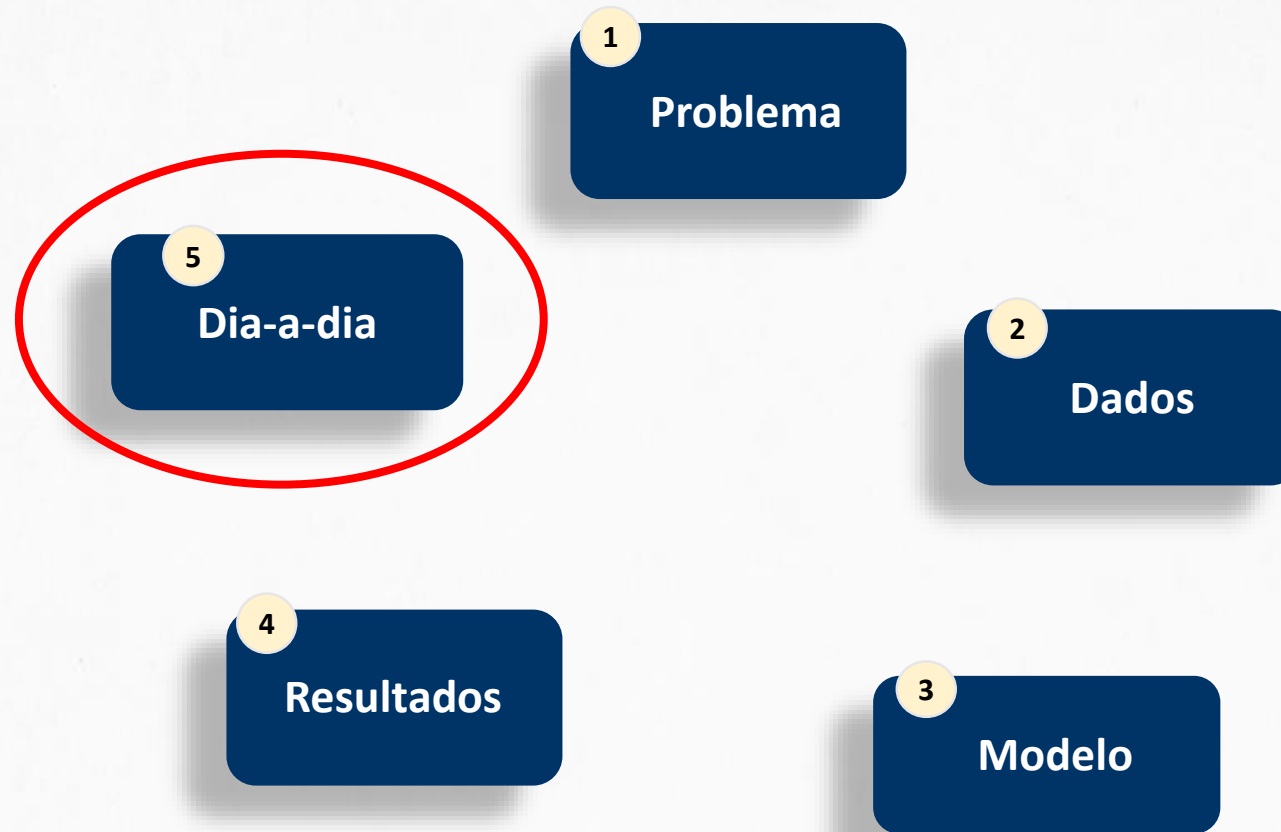
	Underfitting	Just right	Overfitting
Symptoms	Previsor pouco flexível Baixa complexidade Alto viés, Baixa Variância	Previsor muito flexível Alta complexidade Baixo viés, Baixa Variância	Previsor muito flexível Alta complexidade Baixo viés, Alta Variância
Regression illustration			
Classification illustration			

# Modelos Supervisionados – Treino & Teste



- Com o aumento do volume de dados disponíveis para cálculo dos parâmetros de um modelo a partir do início do século XXI passou-se a adotar a técnica de dividir o conjunto em dois grupos treino e teste.
- Os parâmetros do modelo são calculados no grupo de treino. O modelo tem seu desempenho avaliado no conjunto de teste.
- Este tipo de técnica é utilizada para evitar um problema conhecido como sobreajuste (*overfitting*)

# Ciclo de Vida dos Dados



**Levem em conta os possíveis usos da nova ferramenta...**





# Título

## Subtítulo

### Conteúdo