

# Introdução ao aprendizado de máquina



INFORMAÇÃO,  
TECNOLOGIA  
& INOVAÇÃO

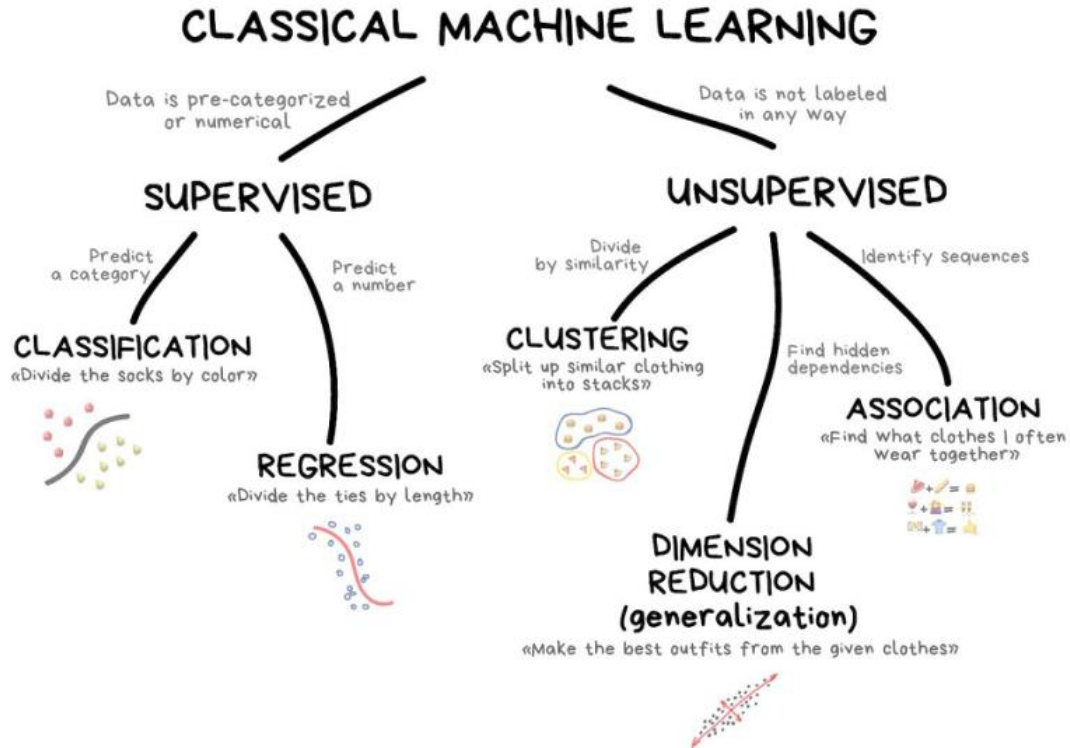
# INTRODUÇÃO



[illegible]

[illegible]

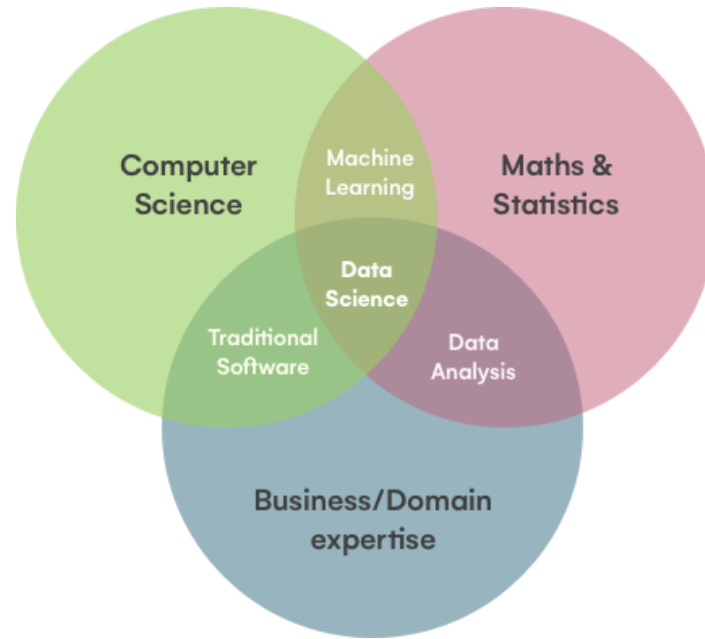
# TIPOS DE APRENDIZADO



# PROJETOS DE DADOS

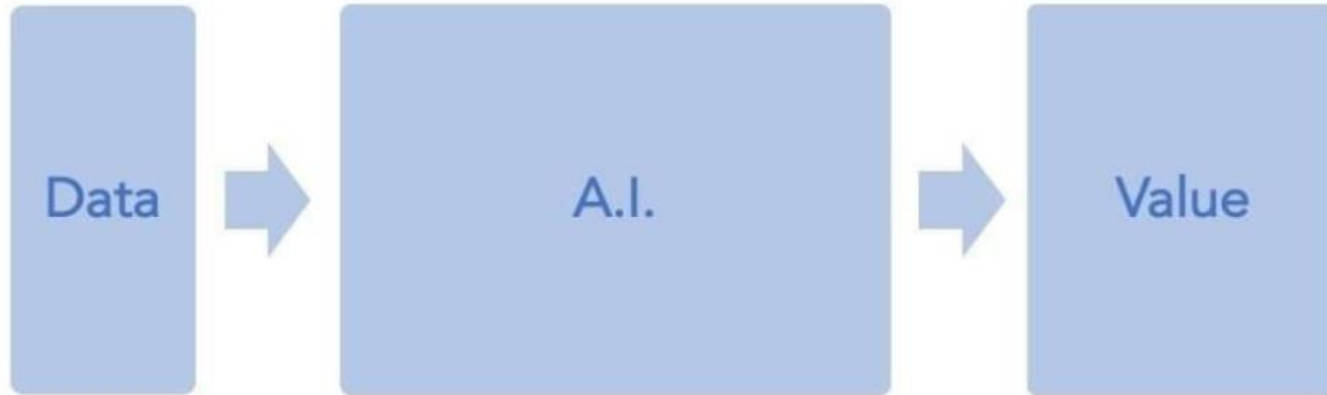


# CIÊNCIA DE DADOS



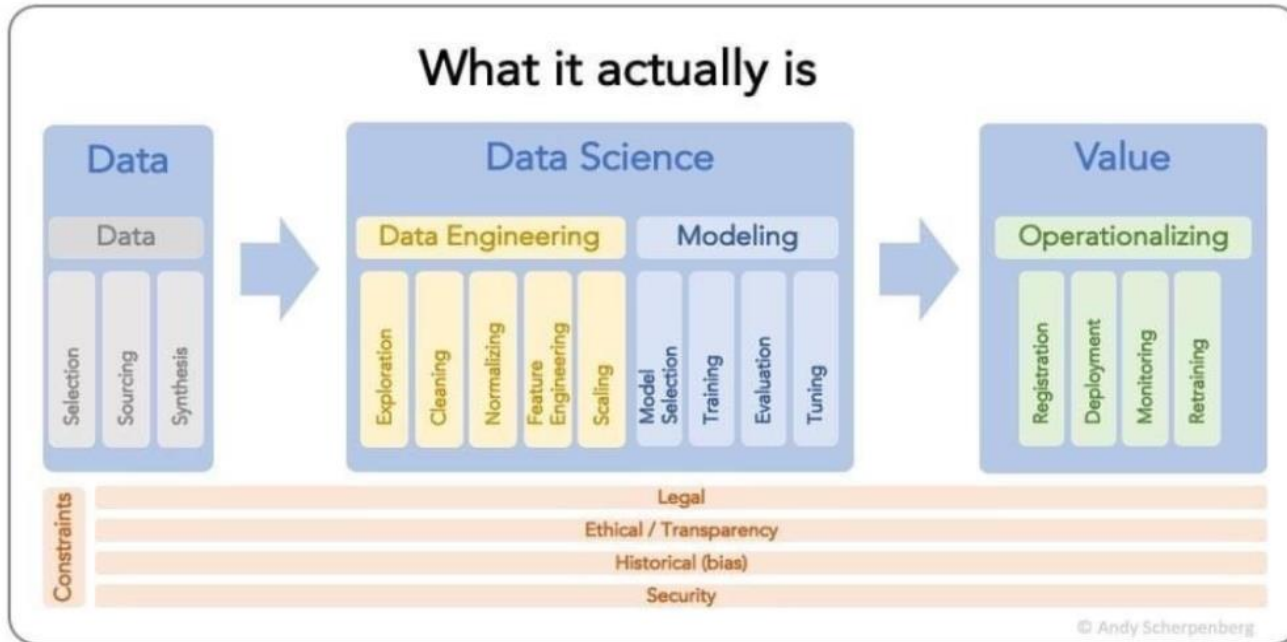
# CIÊNCIA DE DADOS NA PRÁTICA

What companies think A.I. looks like

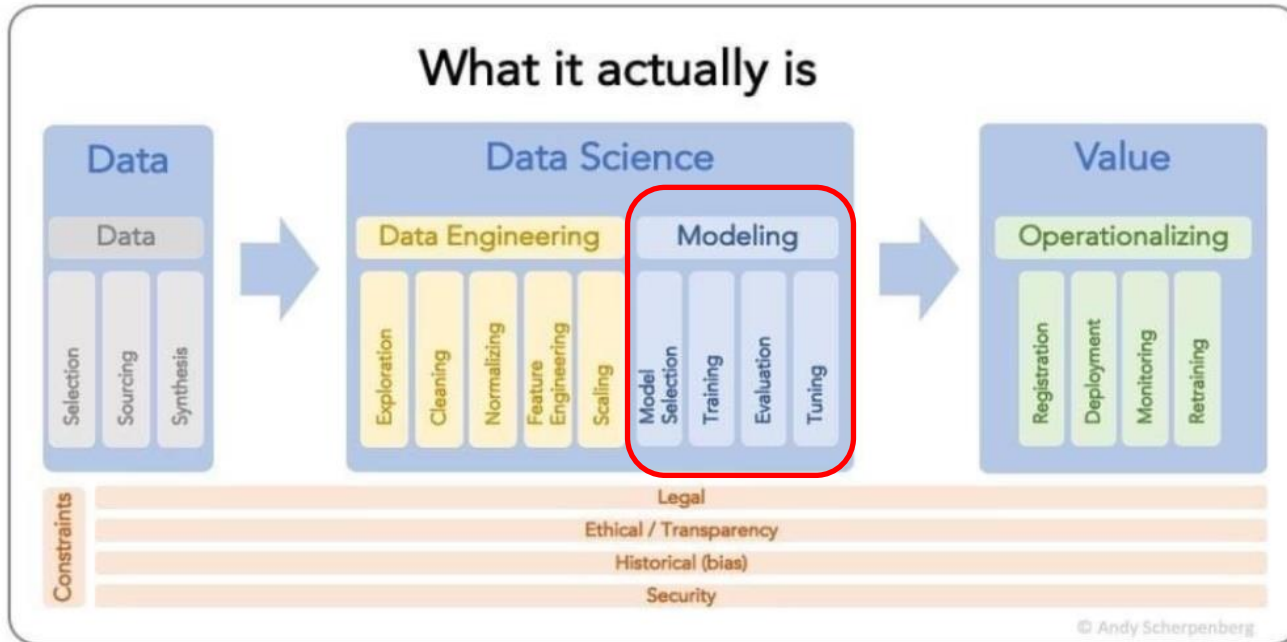




# CIÊNCIA DE DADOS NA PRÁTICA



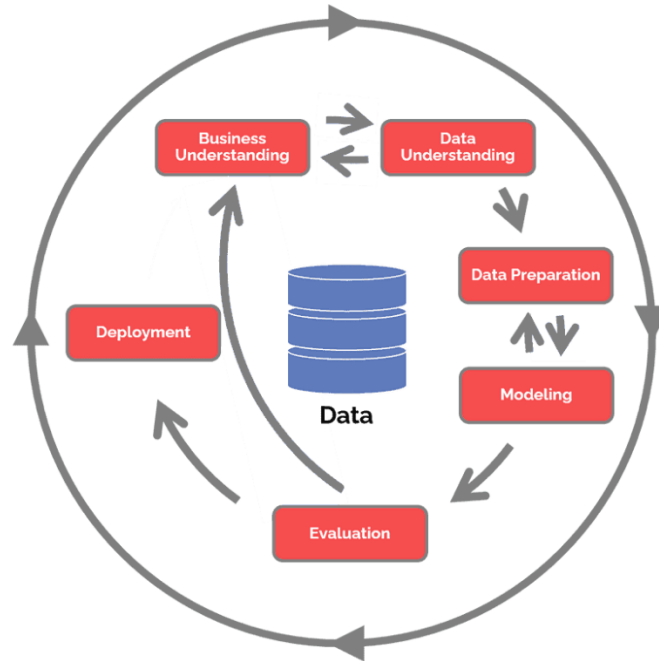
# CIÊNCIA DE DADOS NA PRÁTICA



Nosso foco será na modelagem dos dados!

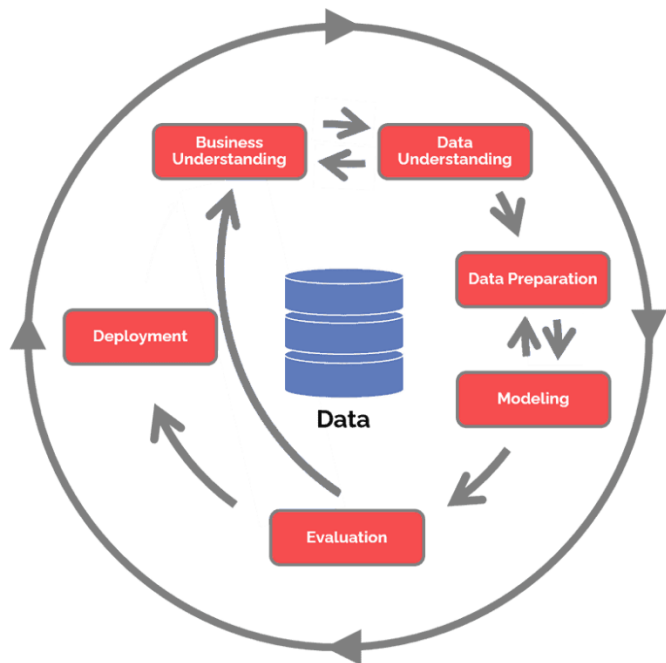
# O PROCESSO CRISP - DM

CRISP – DM (*Cross Industry Standard Process for Data Mining*)



# O PROCESSO CRISP-DM

CRISP – DM (*Cross Industry Standard Process for Data Mining*)

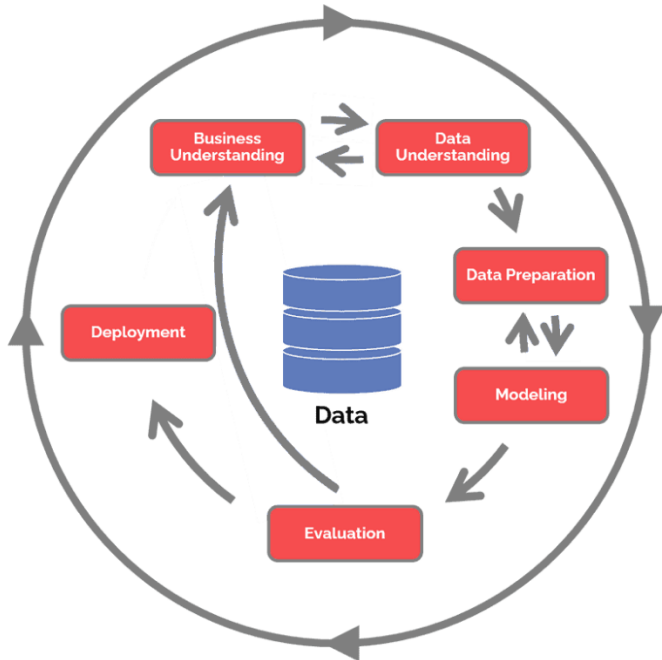


## Entendimento do negócio

- Determinar os objetivos de negócio
- Avaliar a situação atual e recursos
- Determinar os objetivos dos modelos de dados
- Elaborar um plano de projeto

# O PROCESSO CRISP-DM

CRISP – DM (*Cross Industry Standard Process for Data Mining*)

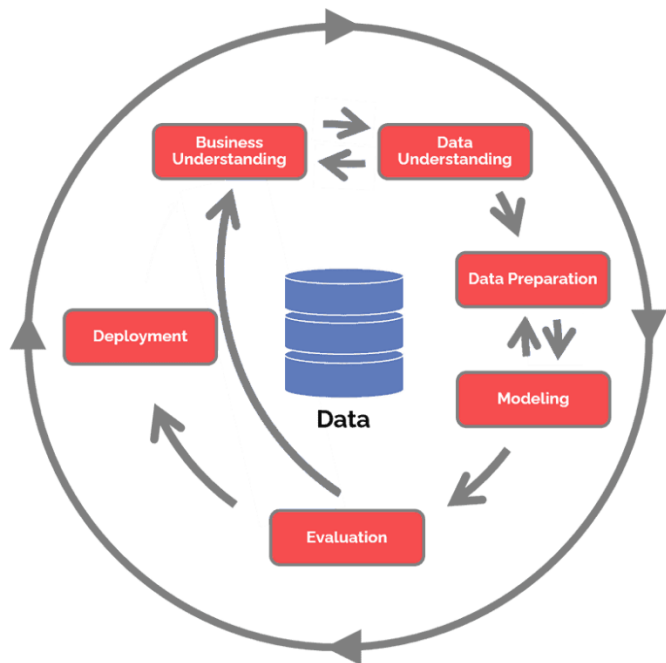


## Entendimento dos dados

- Coletar dados iniciais
- Descrever os dados
- Explorar os dados
- Verificar a qualidade dos dados

# O PROCESSO CRISP-DM

CRISP – DM (*Cross Industry Standard Process for Data Mining*)

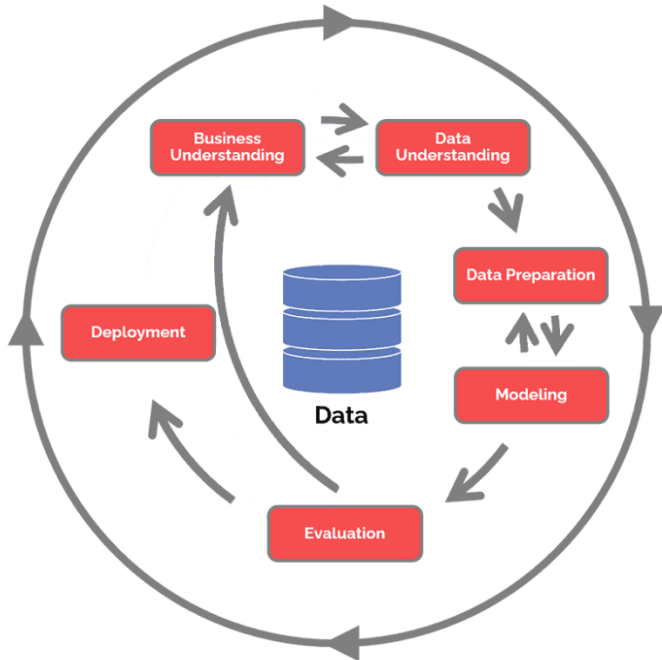


## Preparação dos dados

- Selecionar os dados
- Senear os dados
- “Construir” novos dados
- Integrar dados
- Formatar os dados

# O PROCESSO CRISP-DM

CRISP – DM (*Cross Industry Standard Process for Data Mining*)

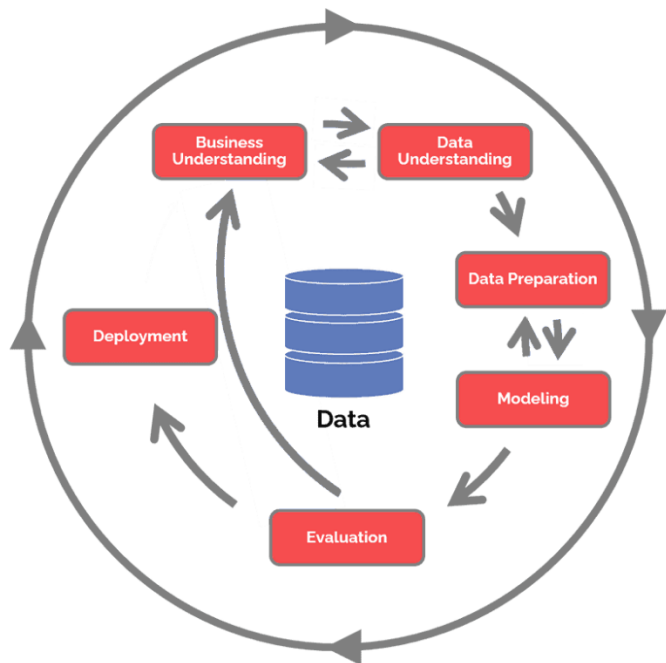


## Modelagem

- Selecionar técnicas
- Elaborar um cenário de teste
- Treinar os modelos
- Avaliar os modelos

# O PROCESSO CRISP-DM

CRISP – DM (*Cross Industry Standard Process for Data Mining*)



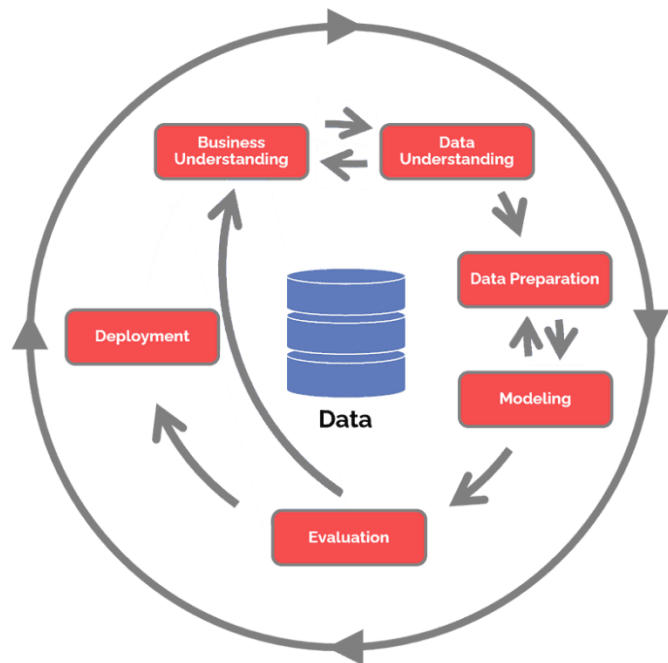
## Avaliação

- Avaliar os resultados
- Rever todo o processo
- Determinar os próximos passos



# O PROCESSO CRISP-DM

CRISP – DM (*Cross Industry Standard Process for Data Mining*)



## Implantação

- Planejar a implantação
- Planejar monitoramento e manutenção do modelo
- Elaborar relatório final
- Rever o projeto

# Alguma dicas

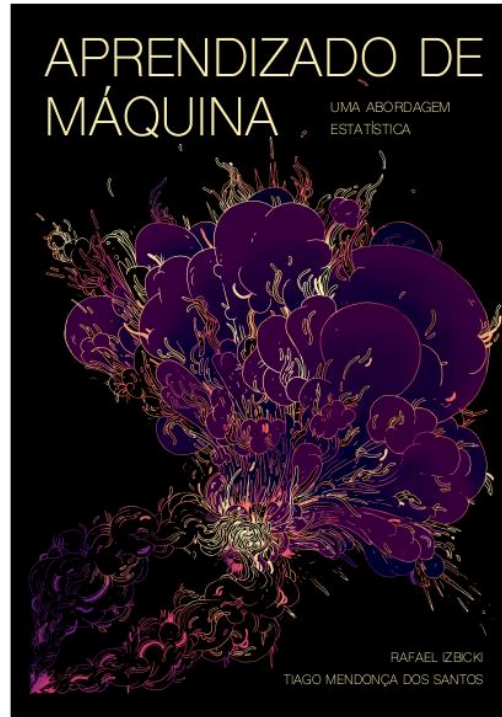
- Defina muito bem os requisitos / objetivos do negócio / aplicação
- Entenda a diferença entre KPIs de negócio / projeto e as métricas de performance do modelo
- Conheça bem os *stakeholders* e tenha diferentes estratégias de apresentação do projeto
- Documente todo o conhecimento e lições aprendidas durante os ciclos realizados no projeto
- Documente e tenha controle de versão de todo o código desenvolvido
- Catalogue todos os dados; eles poderão ser (e serão) úteis no futuro
- **Trabalhe verticalmente:** apresentar valor de maneira ágil é um fator de sucesso em projetos de dados



# MATERIAL



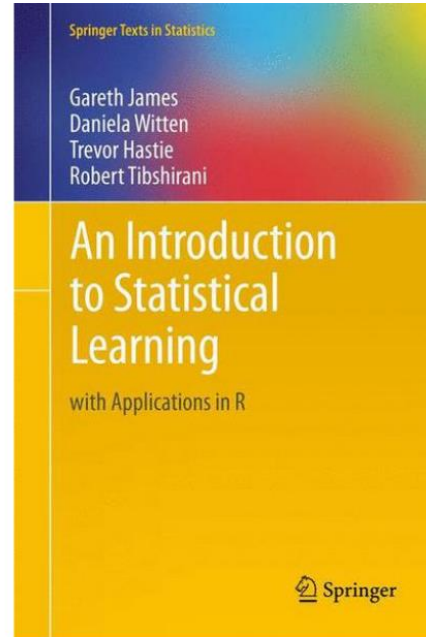
# MATERIAL



<http://www.rizbicki.ufscar.br/ame>



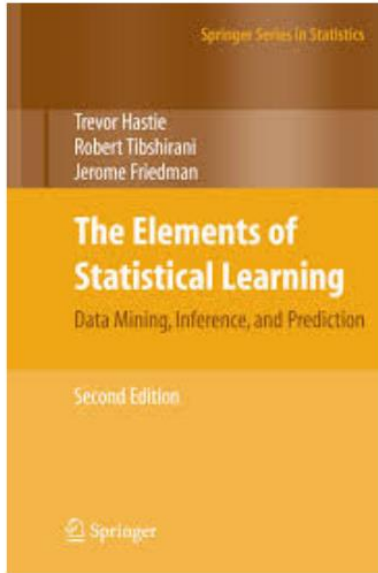
# MATERIAL



<http://faculty.marshall.usc.edu/gareth-james/ISL/>



# MATERIAL



<https://web.stanford.edu/~hastie/ElemStatLearn/>



# Aprendizado Supervisionado



INFORMAÇÃO,  
TECNOLOGIA  
& INOVAÇÃO

# APRENDIZADO SUPERVISIONADO

Aprendizado **supervisionado**: Dadas medições  $(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_n, Y_n)$ , aprender um modelo para **prever**  $Y_i$  baseado em  $\mathbf{X}_i$



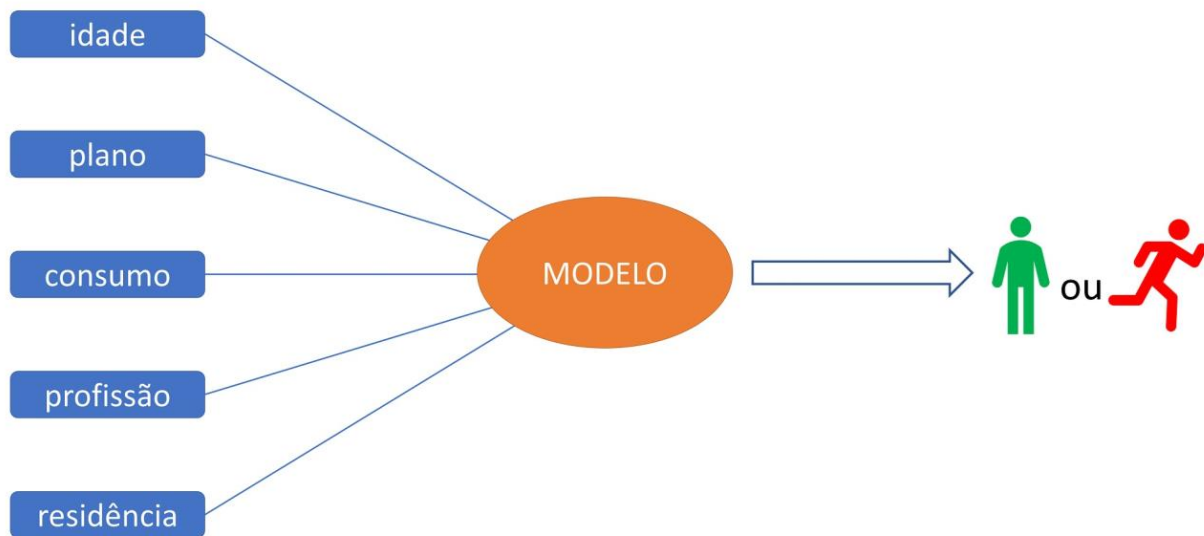


# EXEMPLOS

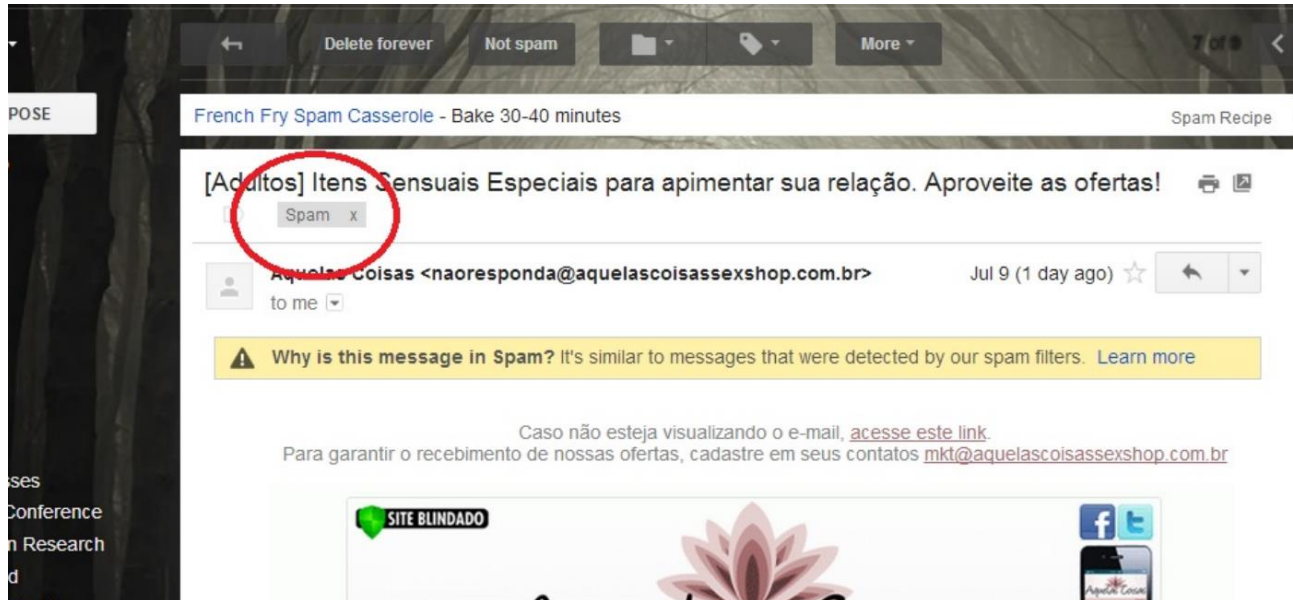


# EXEMPLO - CHURN

cancelamento de um serviço/cliente.



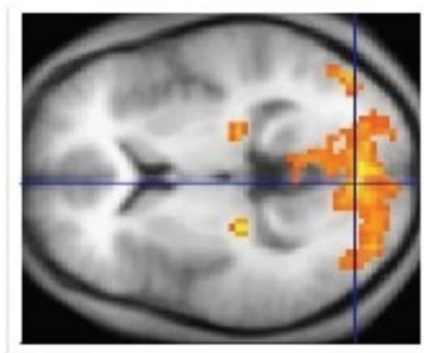
# EXEMPLO - DETECÇÃO DE SPAMS



# EXEMPLO - RECONHECIMENTO DE DÍGITOS

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

# EXEMPLO - LEITURA DE PENSAMENTOS



# EXEMPLO – PREVISÃO DE DEMANDA



# EXEMPLO – PREVISÃO DE VALOR DE ATIVOS

29,89 BRL

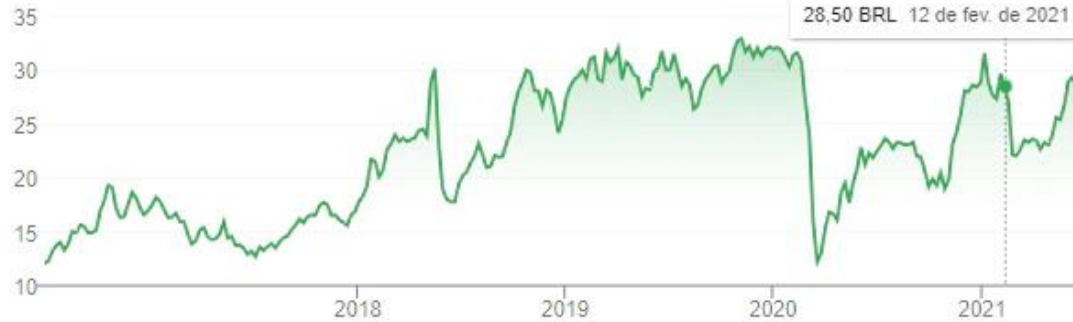
+17,83 (147,84%) ↑ últimos 5 anos

24 de jun. 17:13 BRT · Exoneração de responsabilidade

BVMF: PETR3

+ Seguir

1 D | 5 D | 1 M | 6 M | YTD | 1 A | 5 A | Máx



# EXEMPLO – OUTRAS APLICAÇÕES

- Concessão de crédito
- Detecção de fraude
- Direcionamento de campanhas de *marketing*
- Otimização de produção
- Manutenção preditiva de equipamentos
- Roteirização
- **Infinitas possibilidades!**





# Aprendizado Supervisionado

## Unidade 1: Regressão

- Função de predição
- Risco
- Overfitting
- Seleção de Modelos (Data-Splitting e Validação Cruzada)
- Método de Mínimos Quadrados
- Lasso (Penalização L1)

## Unidade 2: Classificação

- Regressão logística
- Árvores e Random Forests
- Boosting
- KNN
- SVM
- Desbalanceamento de classes
- Explainable ML
- Viés de Seleção e Concept Shift
- Considerações Práticas

