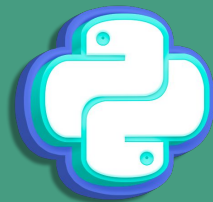


Machine Learning do início ao fim: uma breve visão de um projeto de ML



Python Brasil 2021



Grupo de Estudos em Data Science (GEDS)



Angelica Custodio
Estatista



Deborah Foroni
Atuaria



Laryssa Kato
Física



Mariana Guilardi
Bióloga



Priscila Agostinho
Eng. Telecom.



Material do grupo

O'REILLY®

Hands-On Machine Learning with Scikit-Learn & TensorFlow

CONCEPTS, TOOLS, AND TECHNIQUES
TO BUILD INTELLIGENT SYSTEMS



Aurélien Geron

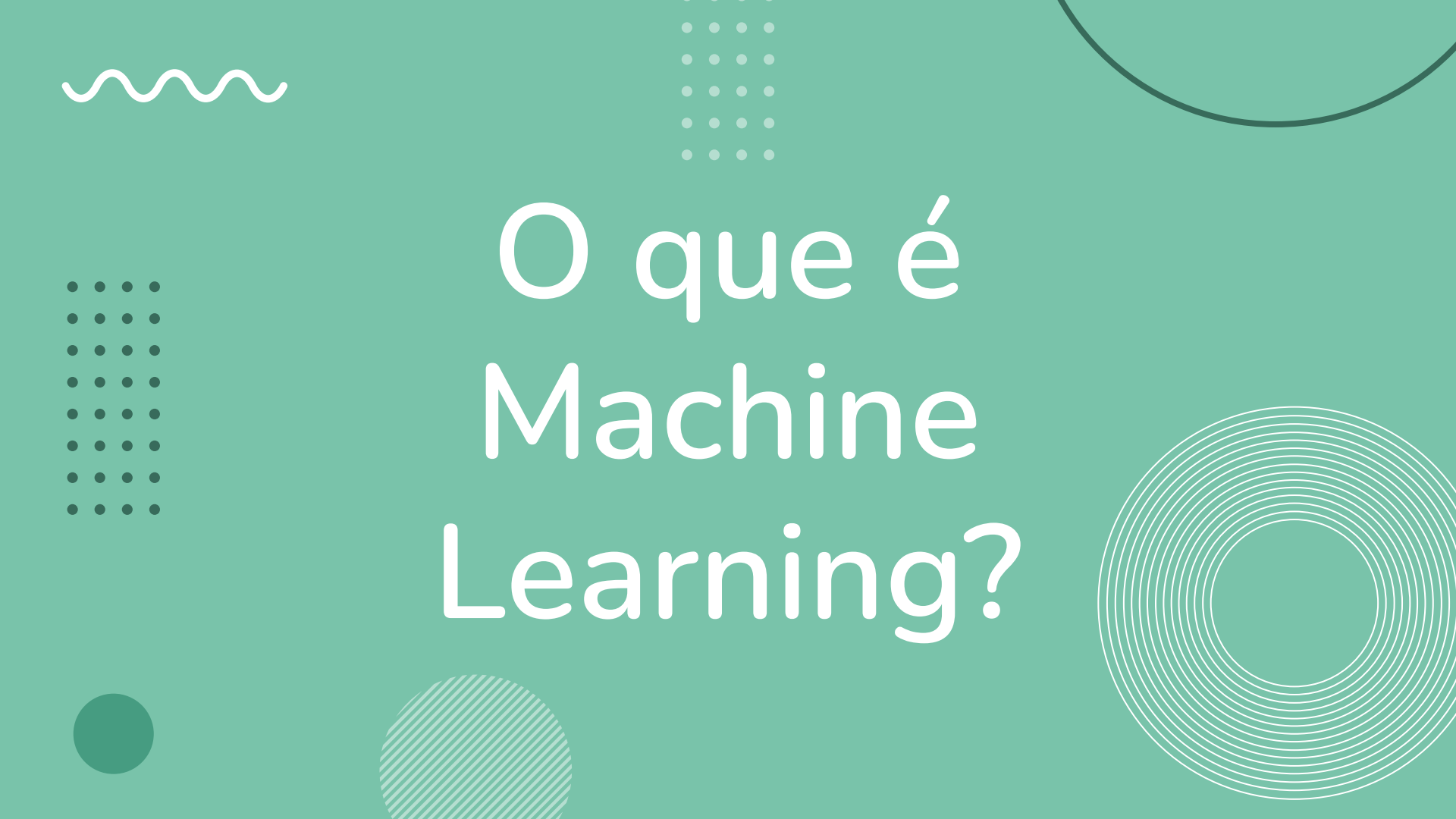
“Hands on Machine Learning
with Scikit-Learn &
TensorFlow”

Estrutura da Apresentação

**Machine
Learning**

**Seis passos
para um
projeto
de ML**

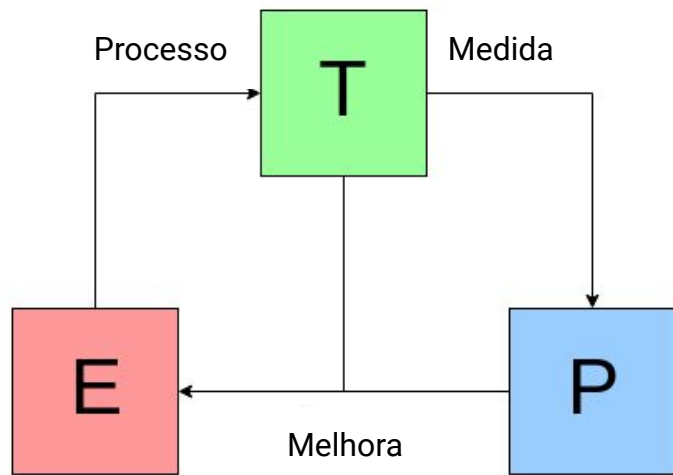
**Conclusões
e para
saber mais**



O que é Machine Learning?

Machine Learning

Machine Learning é a ciência (e arte) de programar computadores para que eles possam aprender com os dados.



T = tarefa

P = performance

E = experiência

(Mitchell, 1997)



Seis passos para um projeto de ML

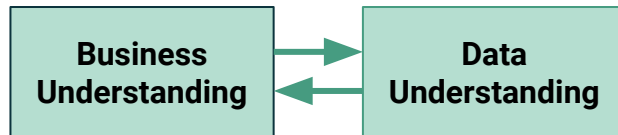


1. Olhar para o problema

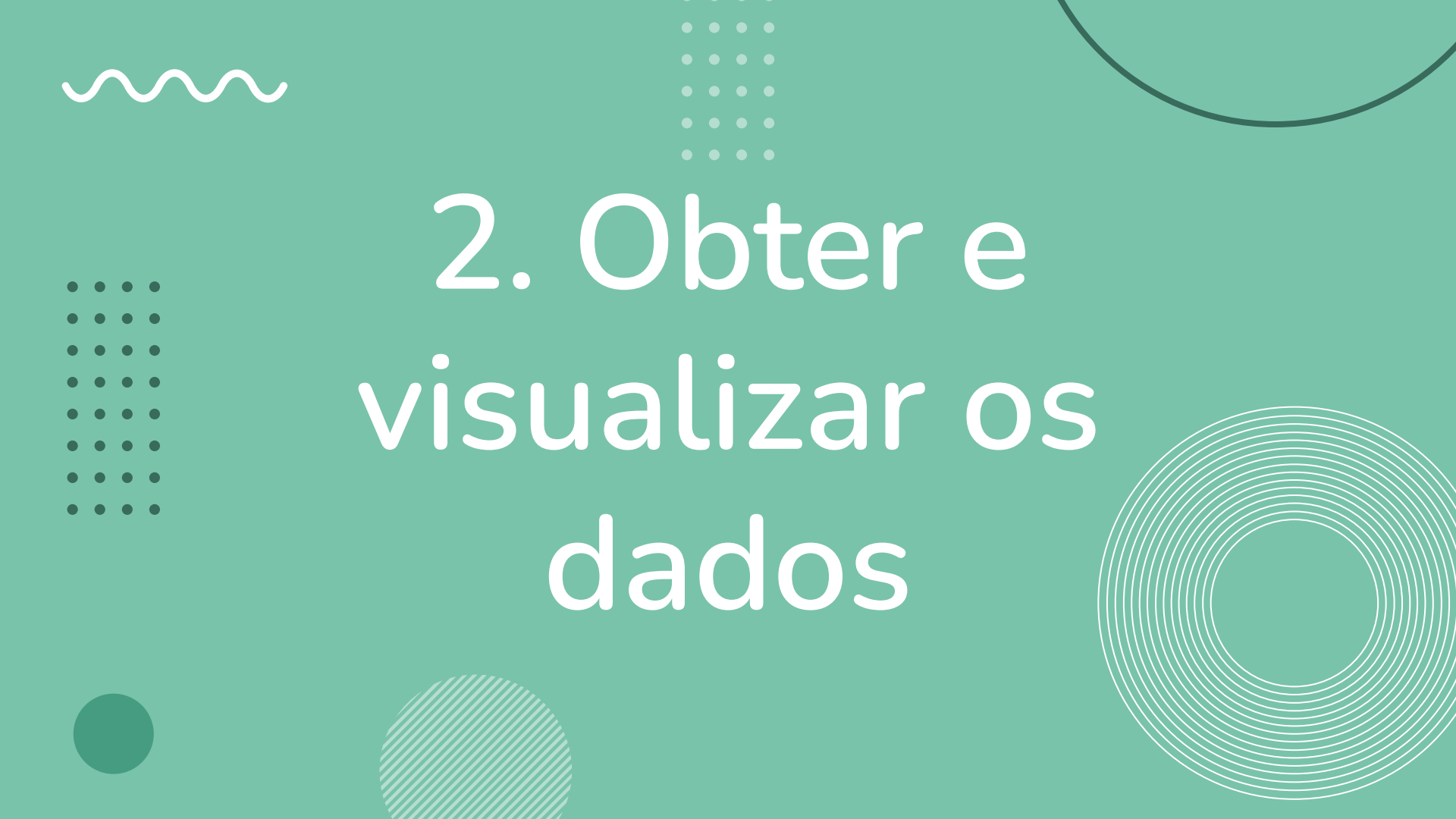
Olhar para o problema

Qual o problema que deseja resolver?

- Como você espera usar e se beneficiar deste modelo?
- Qual a solução no momento (se houver)?
- Qual o modelo você usará?
- Quais métricas você usará para medir o desempenho do modelo?
- Quais hipóteses foram feitas até agora (por você ou por outros)?



CRISP-DM Methodology



2. Obter e visualizar os dados

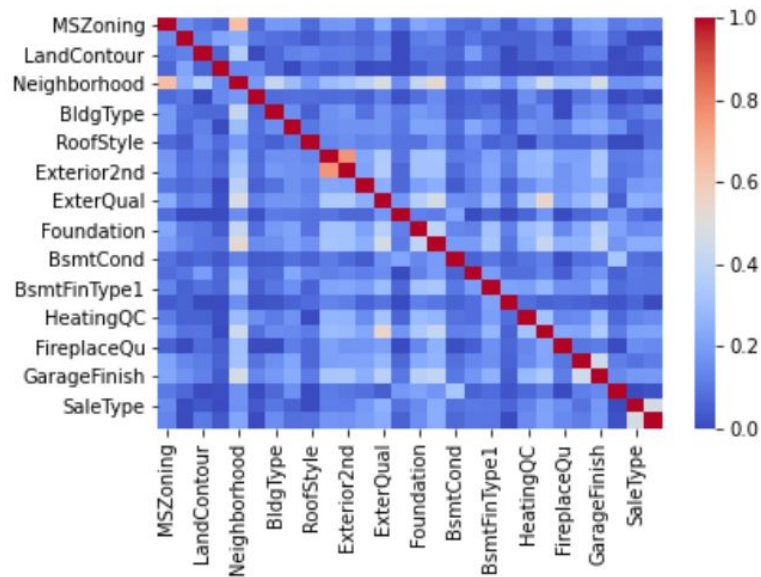


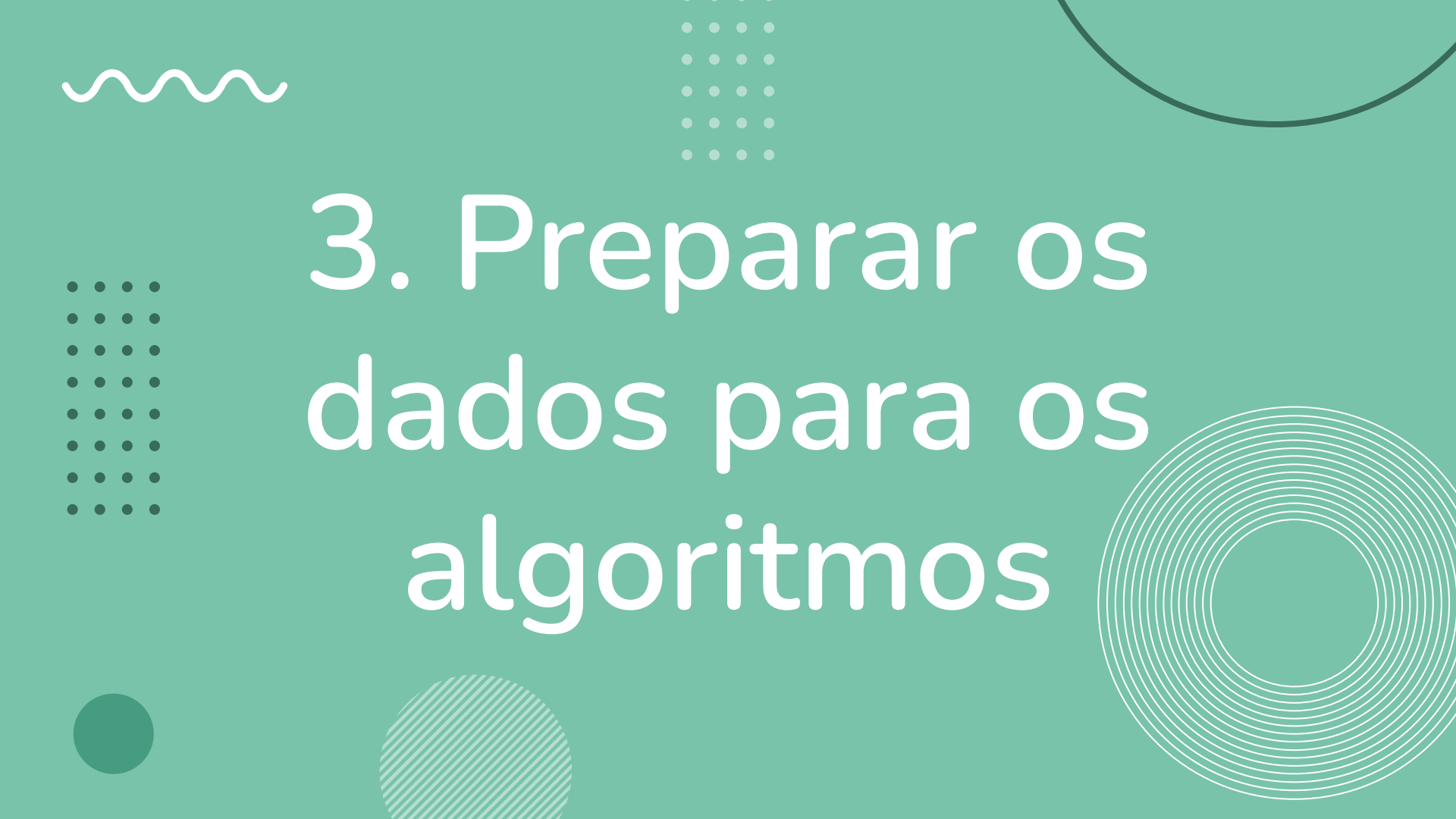
Obter e visualizar os dados

- Download dos dados
- Verificar a estrutura dos dados
- Criar um conjunto de teste para evitar viéses pré-análises

Visualizar os dados

- Visualizar os dados graficamente para identificar padrões
- Buscar correlações entre cada par de atributos
- Experimentando combinações de atributos

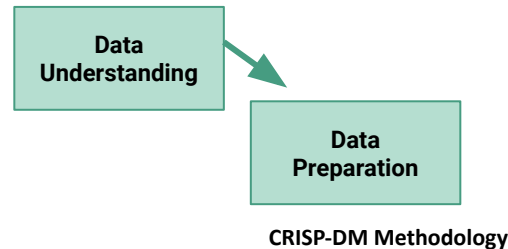




3. Preparar os dados para os algoritmos

Preparar os dados para os algoritmos

- Criar funções para limpar os dados para otimização das tarefas.
- Manipular textos e transformar textos em atributos categóricos.
- Customizar transformadores: combinação de variáveis.
- Escalonar características (normalizar os dados).



Preparar os dados para os algoritmos


- Pipelines de transformação de dados?

Há muitas etapas de transformação de dados que precisam ser executadas na ordem certa, mas por sorte, o `sckit-learn` tem a classe `Pipeline` que ajuda com todas as transformações que precisamos fazer.

Conseguimos fazer transformações de:

- Dados Nulos
- Dados Categoricals
- Customizar a transformação
- Padronização e Normalização dos dados





4. Treinar o modelo

Treinar o modelo

Identificar qual o tipo de problema e avaliar algoritmos no conjunto de treino

1a Etapa

Conjuntos de
Treino do Modelo

Problema é de
classificação,
regressão ou não
supervisionado

2a Etapa

Técnicas avaliadas:

- Árvore de decisão
- Regressão Logística
- Random Forest
- Regressão Logística

3a Etapa

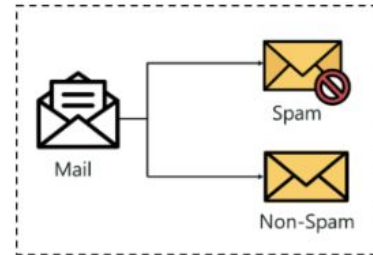
Definição das métricas
de qualidade de ajuste
do seu modelo

4a Etapa

Tabela resumo com as
medidas de acurácia de
cada modelo

Treinar o modelo

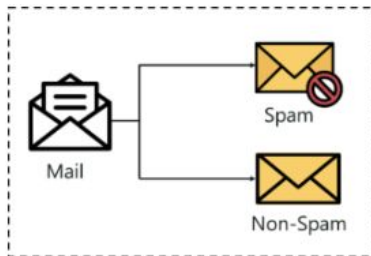
Exemplo: Classificação de e-mails Spam ou não Spam



Treinar o modelo

Exemplo: Classificação de e-mails Spam ou não Spam

1a Etapa



2a Etapa

- Árvore de decisão
- Regressão Logística
- Random Forest

3a Etapa

Não queremos que um e-mail importante seja classificado como Spam

Treinar o modelo

3a Etapa: Não queremos que um e-mail importante seja classificado como Spam

Observado	Predito	
	Não Spam	Spam
	Não Spam	Spam
	Verdadeiro Negativo	Falso Positivo
	Falso Negativo	Verdadeiro Positivo

Medidas de Avaliação do Modelo de Classificação

Acurácia: conta verdadeiros positivos e negativos/total

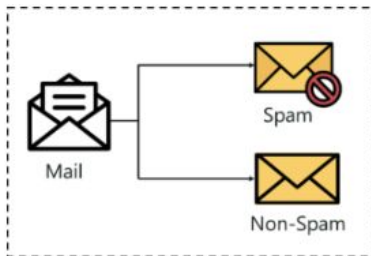
Recall: conta falsos negativos e verdadeiros positivos

Precisão: conta falsos positivos e verdadeiros positivos

Treinar o modelo

Exemplo: Classificação de e-mails Spam ou não Spam

1a Etapa



2a Etapa

- Árvore de decisão
 - Regressão Logística
 - Random Forest
- Não queremos que um e-mail importante seja classificado como Spam

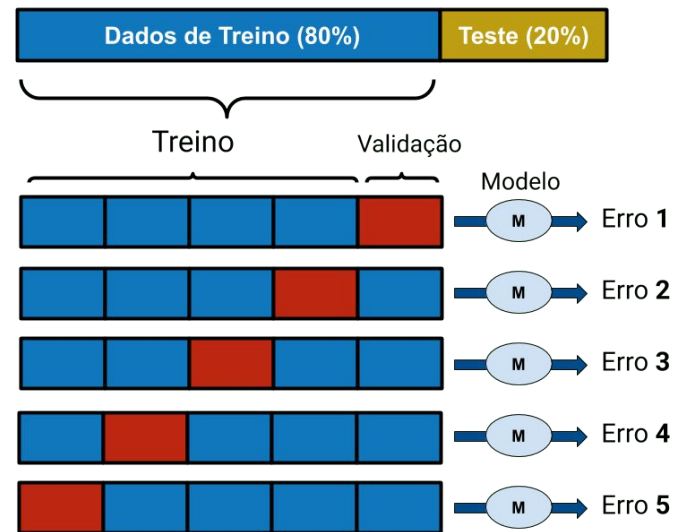
3a Etapa

4a Etapa

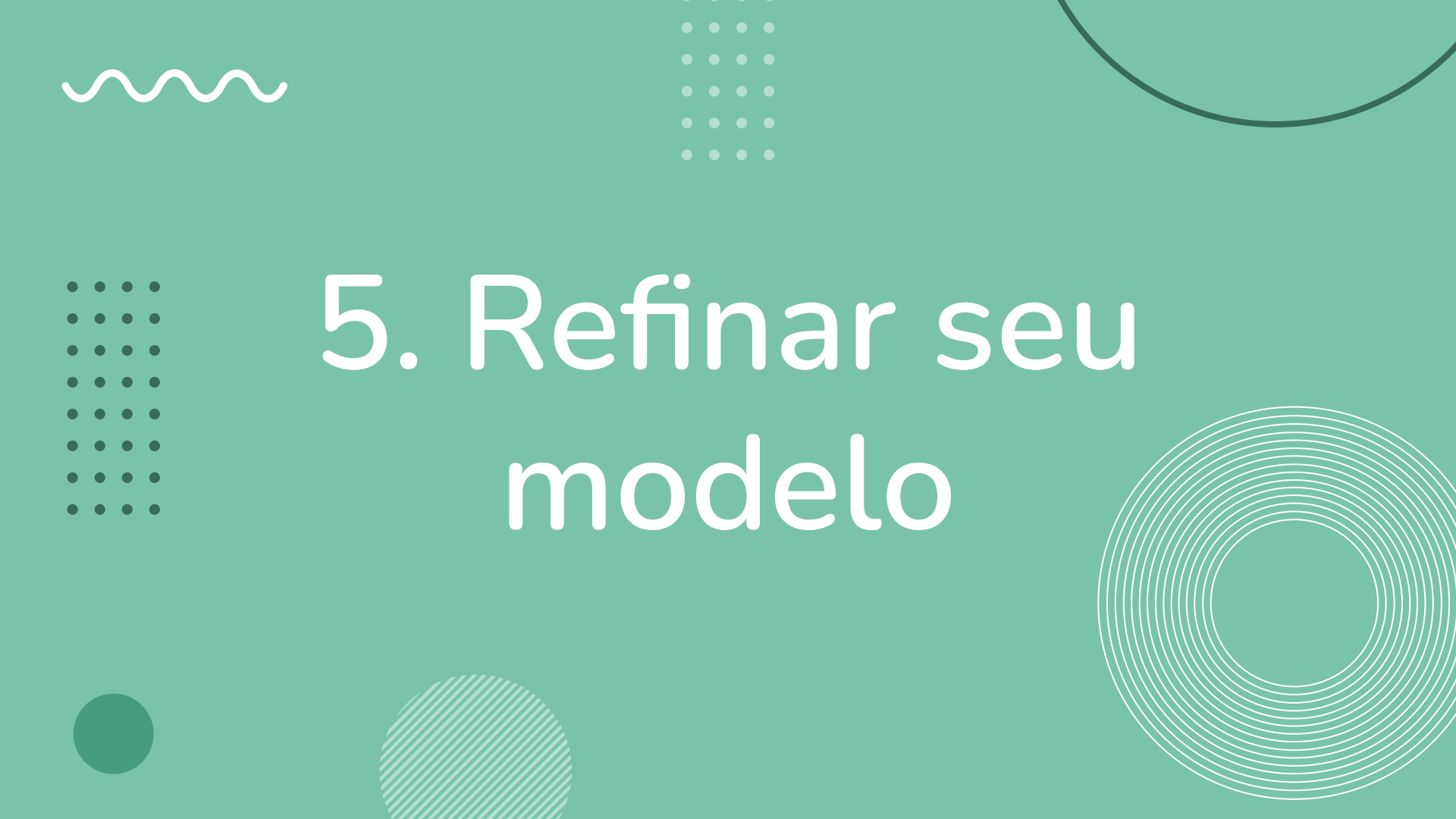
Modelo	Precisão
Modelo 1	0.85
Modelo 2	0.86
Modelo 3	0.91

Treinar o modelo

- Possibilidade de reavaliar modelo com a **Validação-Cruzada** a depender do problema avaliado.



$$\text{Erro médio de validação} = (\text{Erro 1} + \dots + \text{Erro 5}) / 5$$



5. Refinar seu modelo

Refinar seu modelo

Com os algoritmos escolhidos, é hora de refinar seus resultados considerando o conjunto de validação.

Hiperparâmetros

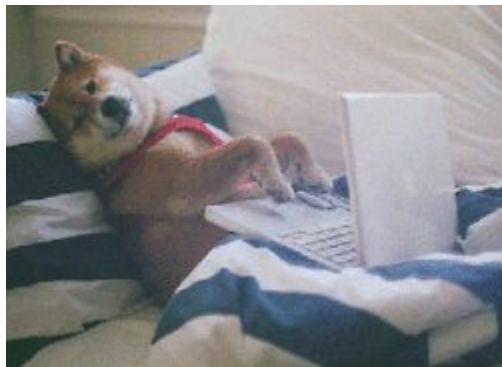
- Ajuste dos hiperparâmetros do modelo: Gridsearch, Randomizedsearch, Bayes search.
- Análise do melhor modelo e os erros no conjunto de validação.

Refinar seu modelo

- Análise dos vieses do seu modelo
 - **Interpretabilidade** de Modelos Opacos
 - Importância de variáveis, PDP, ICE, ALE, LIME, Shapley Values
 - Análise da acurácia do Modelo por **variáveis-chave**

Refinar seu modelo

Avalie seu modelo no **conjunto de teste**.





6. Apresentar solução

Apresentar solução

- Traduzir para o contexto do problema!

Apresentar solução

- Traduzir para o contexto do problema!
- Divida o seu processo:
 - O que você aprendeu?
 - O que funcionou? O que não funcionou?
 - Quais as limitações?

Apresentar solução

- Traduzir para o contexto do problema!
- Divida o seu processo:
 - O que você aprendeu?
 - O que funcionou? O que não funcionou?
 - Quais as limitações?
- Dividir os resultados de forma clara, objetiva e utilizar recursos de visualização!

Apresentar solução

- Utilização do modelo?

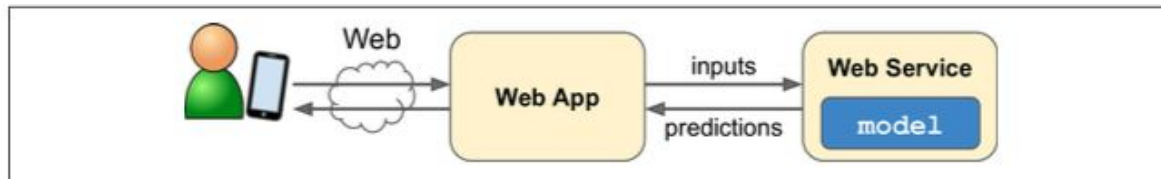
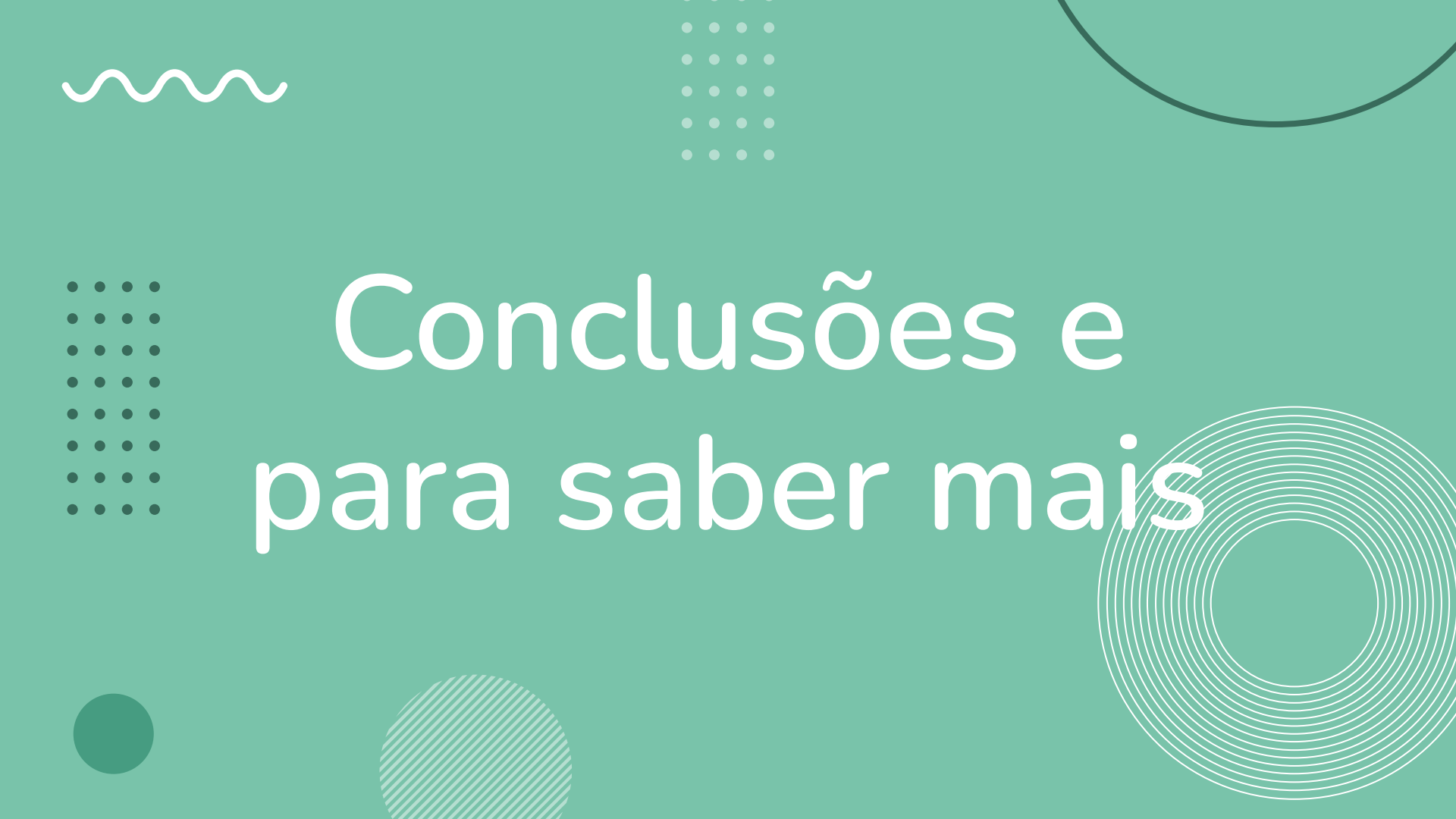


Imagem retirada do livro Hands On Machine Learning - Aurelién Géron

- Deploy, Monitoramento do Modelo e suas variáveis, Manutenção



Conclusões e
para saber mais

Seis passos para um projeto de ML

Olhar para o problema

Entender qual o problema que devemos solucionar em conjunto com a área de negócios.

Obter e visualizar dos dados

Explorar bem os dados obtidos.

Preparar os dados para os algoritmos

Tratar e manipular os dados para a etapa de modelagem

Seis passos para um projeto de ML

Treinar um modelo

Identificar qual o tipo de problema e avaliar algoritmos no conjunto de treino.

Refinar seu modelo

Refinar seu modelo considerando o conjunto de validação. Apurar resultados na base de teste.

Apresentar solução

Dividir os resultados de forma clara, objetiva, trazendo para o contexto do problema.

+ monitoramento do sistema

Para saber mais

- Livro Hands On Machine Learning - Aurelién Géron e seu [notebook de cada capítulo](#)
- Livro [Aprendizado de Máquina](#) (em linguagem R) sobre modelos
- Algoritmos de Destruição em Massa - Cathy O'Neil



Muito obrigada!



Perguntas?

Notebook com Análise House Pricing

Essa apresentação foi criada por Slidesgo, incluindo ícones de por Flaticon e infográficos e imagens por Feepik