

APRENDENDO COM OS DADOS

UMA ABORDAGEM DE CIÊNCIA DE DADOS E APRENDIZADO DE MÁQUINA UTILIZANDO R (PARTE 2)

Prof. Rafael G. Mantovani

04/09/2019



Dois Vizinhos - PR, Brasil
Setembro, 2019

Universidade Tecnológica Federal do Paraná (UTFPR)
IV Semana Acadêmica do curso de Engenharia de Software

Roteiro

- 1 Revisão
- 2 Classificação holdout
- 3 Classificação CV
- 4 Benchmark
- 5 Seleção de Features / Tuning
- 6 Titanic
- 7 Referências

Material

Link: https://github.com/rgmantovani/saes2019_dataScience

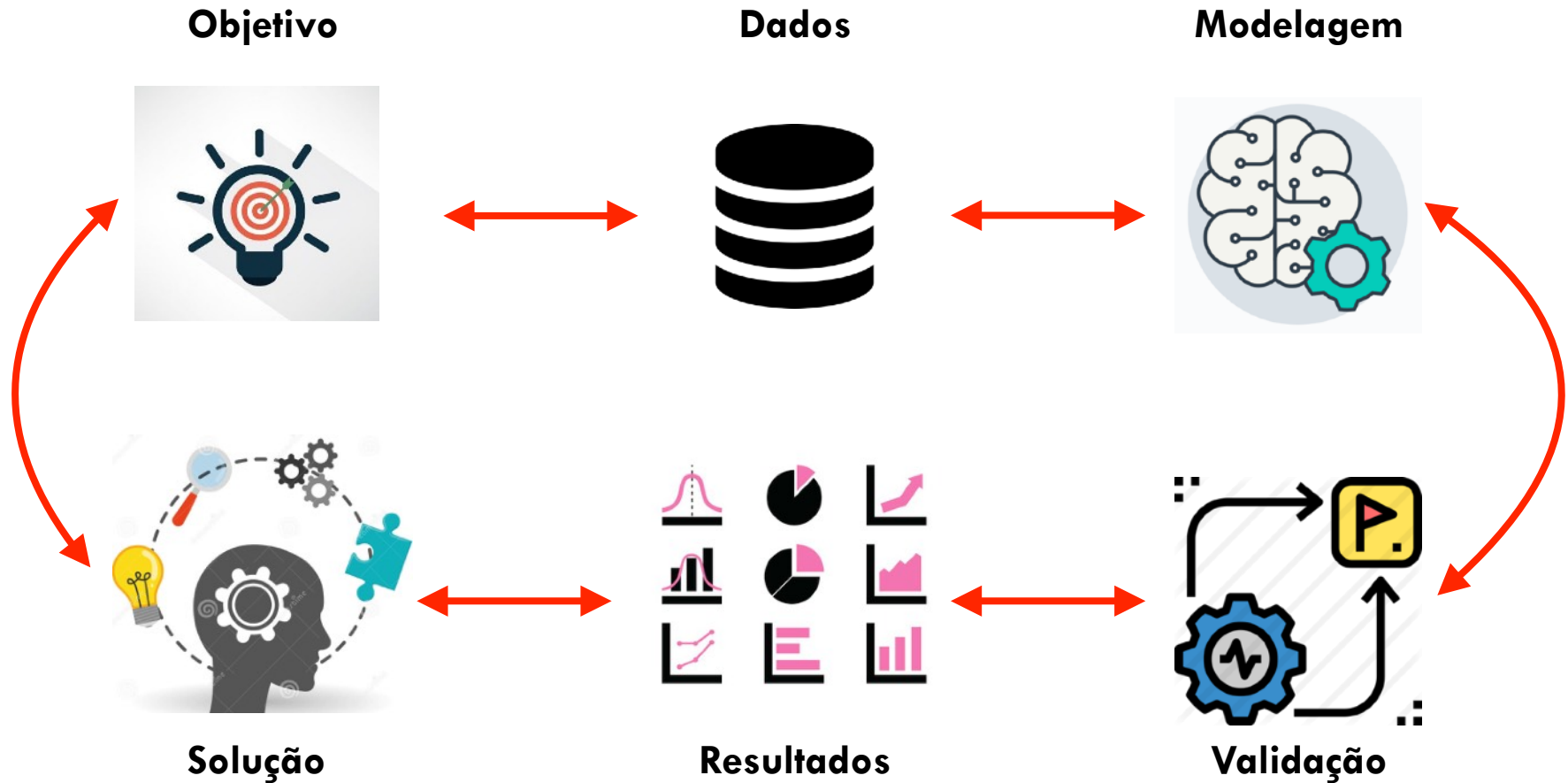
The screenshot shows the GitHub repository page for `rgmantovani / saes2019_dataScience`. The repository has 4 commits, 1 branch, and 0 releases. The main branch is `master`. The repository contains the following files and folders:

File/Folder	Description
<code>datasets</code>	adding config scripts
<code>scripts</code>	adding examples
<code>ggplot2-cheatsheet.pdf</code>	adding cheat sheets
<code>r-cheat-sheet-3.pdf</code>	adding cheat sheets

Roteiro

- 1 Revisão**
- 2 Classificação holdout**
- 3 Classificação CV**
- 4 Benchmark**
- 5 Seleção de Features / Tuning**
- 6 Titanic**
- 7 Referências**

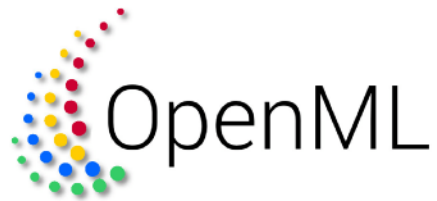
Fluxo de Ciência de Dados



Fluxo de Ciência de Dados



Ferramentas



Roteiro

- 1 Revisão
- 2 Classificação holdout
- 3 Classificação CV
- 4 Benchmark
- 5 Seleção de Features / Tuning
- 6 Titanic
- 7 Referências

Exercício 01

- Dataset sussa
 - ler, rodar os dados
 - rodar um algoritmo c holdout

Roteiro

- 1 Revisão
- 2 Classificação holdout
- 3 Classificação CV
- 4 Benchmark
- 5 Seleção de Features / Tuning
- 6 Titanic
- 7 Referências

Exercício 02

- Dataset suasa
 - ler, rodar os dados
 - rodar um algoritmo c CV

Roteiro

- 1 Revisão
- 2 Classificação holdout
- 3 Classificação CV
- 4 Benchmark
- 5 Seleção de Features / Tuning
- 6 Titanic
- 7 Referências

Exercício 03

- Dataset suasa
 - ler, rodar os dados
 - usar CV
 - varios algoritmos
 - varias medidas de desempenho

Exercício 04

- Dataset sussa
 - ver os resultados do exercício 3
 - graficos
 - teste de hipotese (friedman, nemenyi)

Roteiro

- 1 Revisão
- 2 Classificação holdout
- 3 Classificação CV
- 4 Benchmark
- 5 Seleção de Features / Tuning
- 6 Titanic
- 7 Referências

Exercício 05

- Dataset medio
 - NAs
 - features correlacionadas
 - normalizacao

Roteiro

- 1 Revisão
- 2 Classificação holdout
- 3 Classificação CV
- 4 Benchmark
- 5 Seleção de Features / Tuning
- 6 Titanic
- 7 Referências

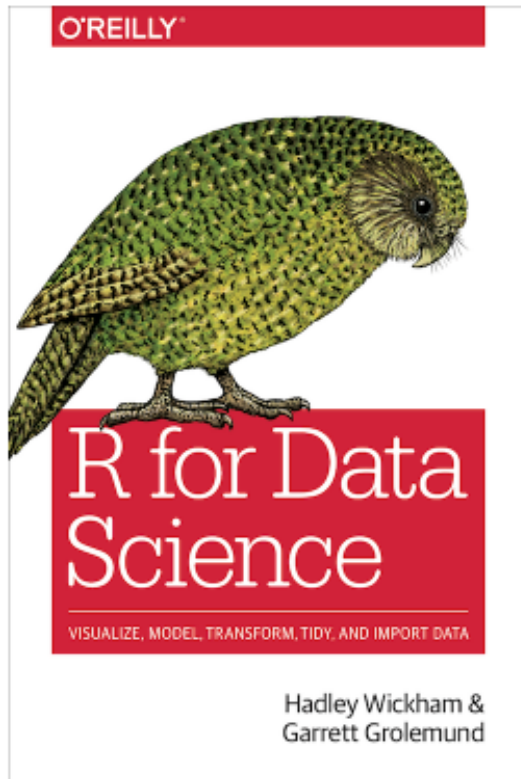
Exercício 06

- Titanic (OpenML data id = 40945)
- descobrir quem sobreviveu, quem morreu
 - Classificação

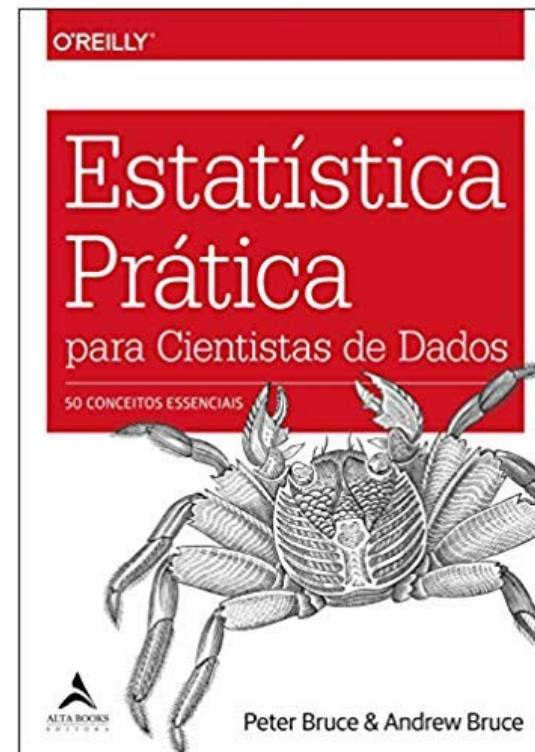
Roteiro

- 1 Revisão
- 2 Classificação holdout
- 3 Classificação CV
- 4 Benchmark
- 5 Seleção de Features / Tuning
- 6 Titanic
- 7 Referências

Referências

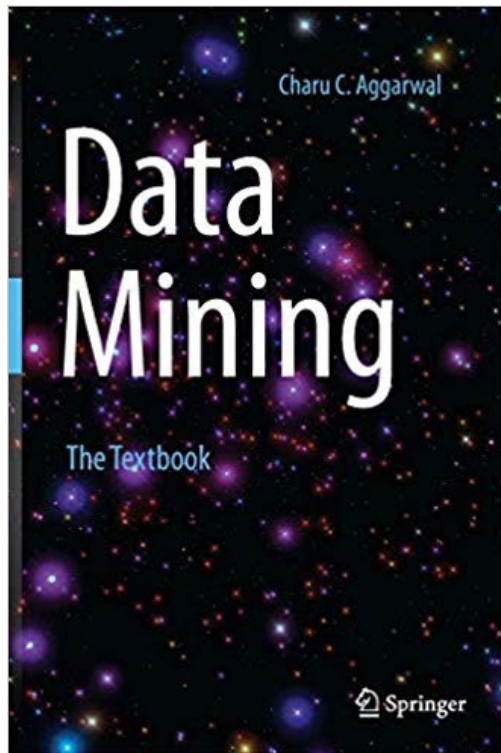


[Wickham & Grolemund, 2018]

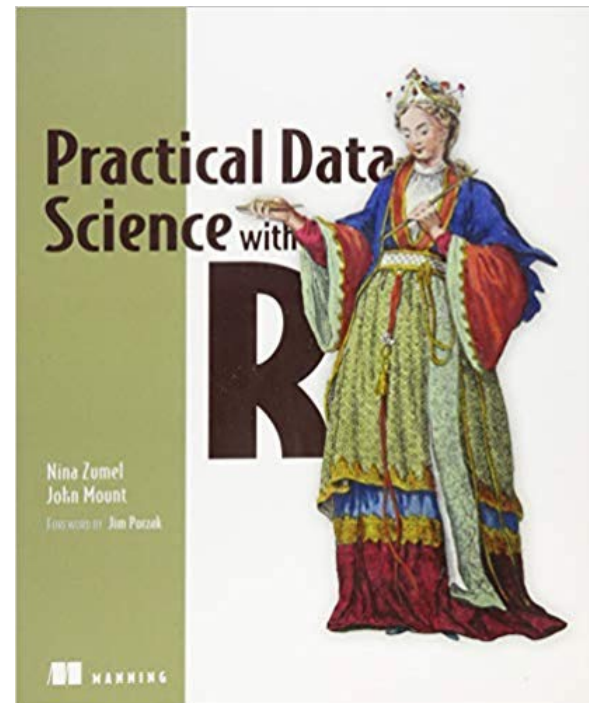


[Bruce & Bruce, 2019]

Referências



[Aggarwal, 2015]



[Zumel and Mount, 2014]

Perguntas?

Prof. Rafael G. **Mantovani**

rafaelmantovani@utfpr.edu.br