

## *Machine Learning*



## *Decision Tree, Random Forest e Método Ensemble*

**Bloco 1**

Lucas Claudino



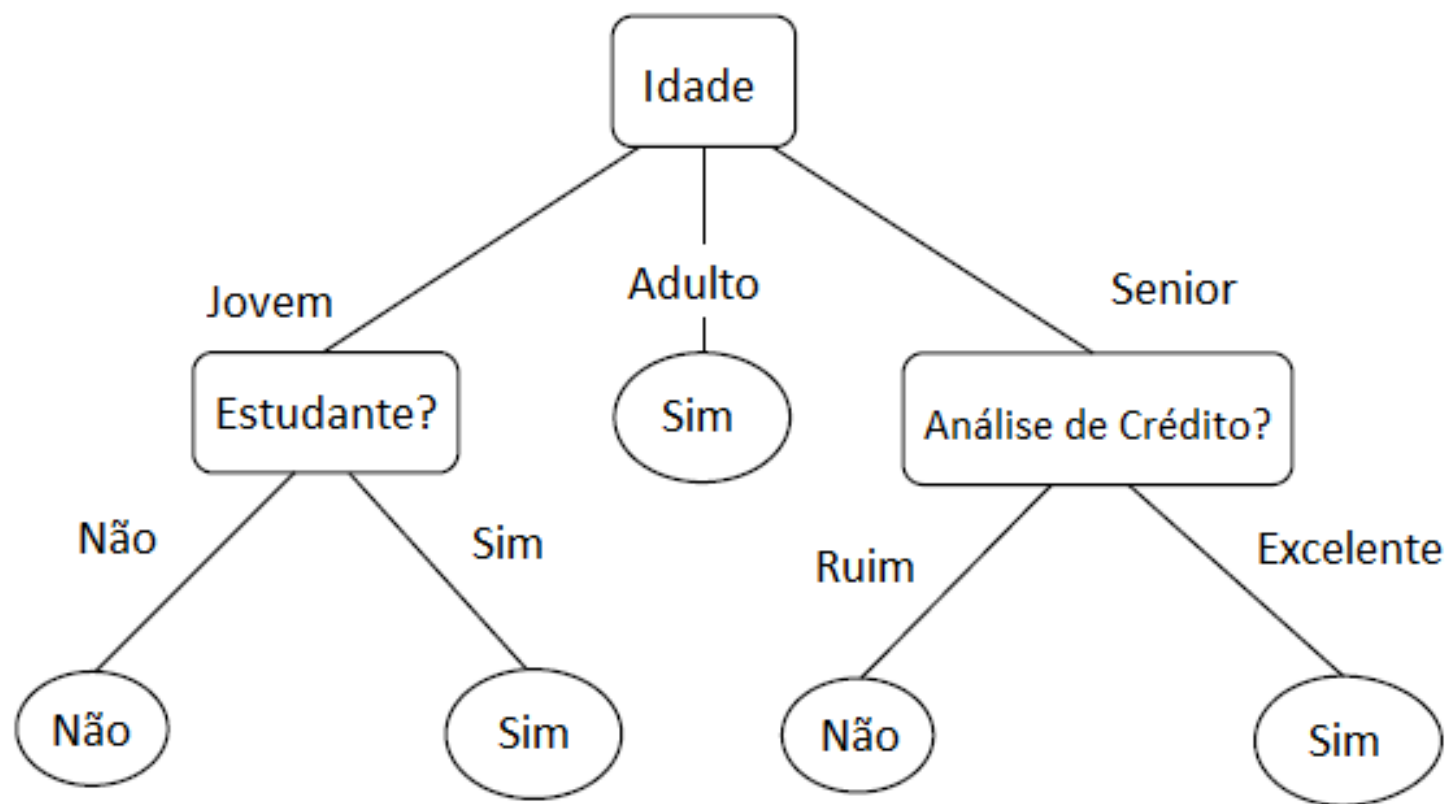


## ► *Decision Tree*

- Inferência a partir dos dados.
- Árvore criada a partir de perguntas e respostas sobre os dados.
- Começar pela raiz e então dividir os dados.
- Divisão para obter o maior ganho de informação.
- Árvore muito grande -> *overfitting* -> métodos de poda.

## ► Árvore de decisão

Figura 1 – Exemplo de árvore de decisão



Fonte: Han, Kamber e Pei (2011).

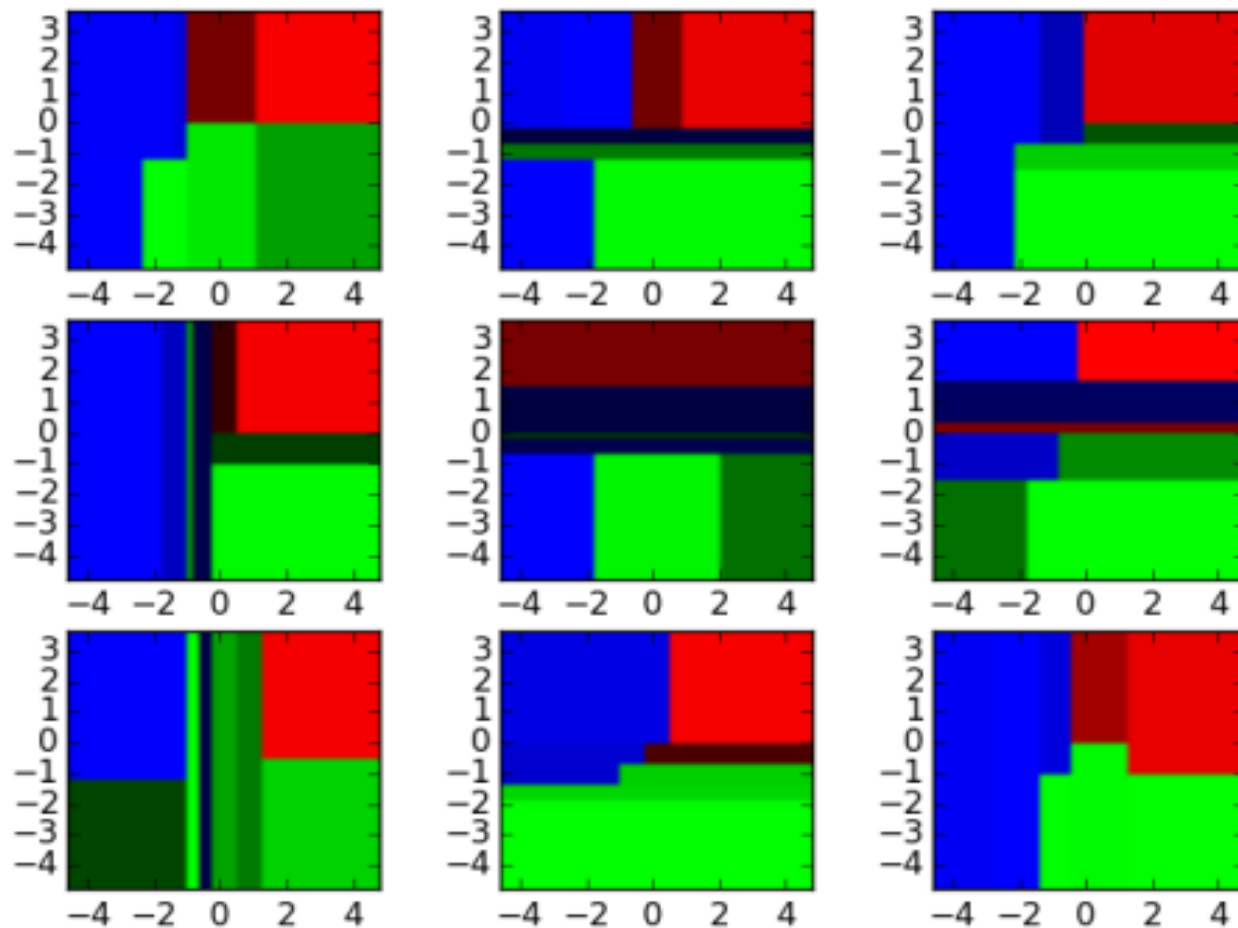


## ► *Random Forest*

- Combinar estimadores fracos para construir um modelo robusto.
- Meta-estimador que agrega diversas árvores de decisão.
  - Quantidade de caract. a serem divididas em cada nó é limitada.
  - Cada árvore extrai uma amostra aleatória.
- Cada árvore criada é um estimado diferente.

## ► *Random Forest*

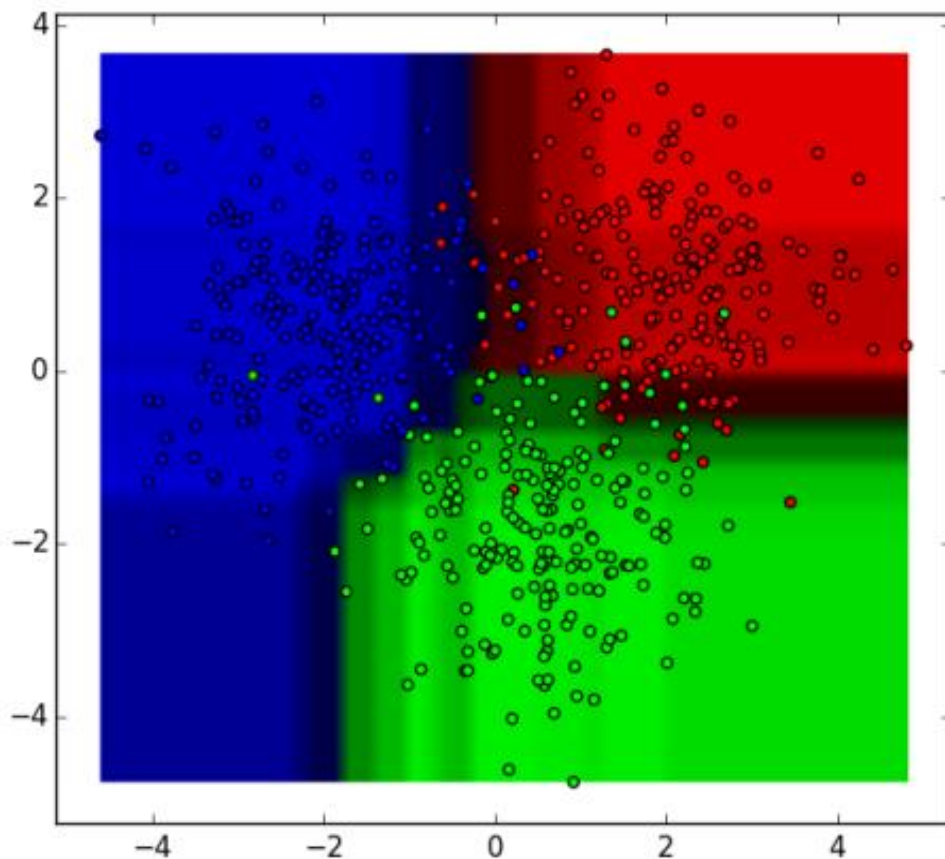
Figura 2 – Árvores de decisão



Fonte: Maini e Sabri (2017, p. 53).

## ► *Random Forest*

Figura 3 – Resultado da combinação das  
nove árvores de decisão



Fonte: Maini e Sabri (2017, p. 53).



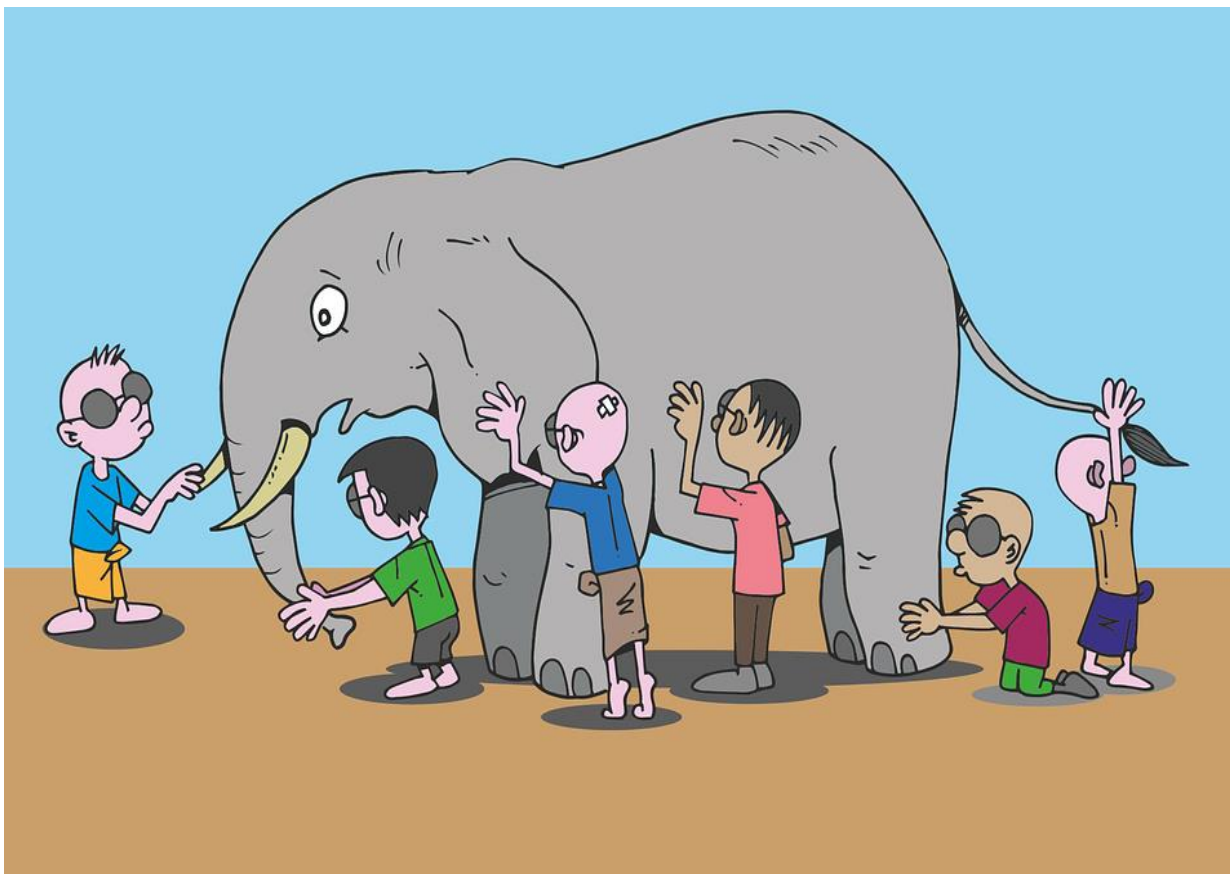
## ► Método *Ensemble*

- Método de agrupamento.
- Tenta mitigar os três maiores problemas: ruído, viés e variância.
- Combina resultados vindos de vários modelos.
- Predição mais precisa que todos os resultados.



## ► Naive Bayes

Figura 4 – Observações individuais e enviesadas



Fonte: <https://towardsdatascience.com/simple-guide-for-ensemble-learning-methods-d87cc68705a2> . Acesso em 27 ago. 2019.

# PÓS-GRADUAÇÃO

## Implementações e peculiaridades

### Bloco 2

Lucas Claudino





## ► *Decision Tree*

- Critérios pra divisão: ganho de informação; entropia.
- Como evitar *overfitting*?
  - Limitar o número de interações.
  - Técnicas de poda.

## ► Poda em árvore de decisão

Figura 5 – Pseudo-código de poda de árvore de decisão

---

### Algorithm 2 Poda em Árvore de decisão

---

**Data:** função  $f(T, m)$  (estimativa para o erro de generalização, baseado no tamanho  $m$  da amostra)

**Result:** Substituição ou não do nó

```
for nó  $j$  em um caminho de baixo para cima (das folhas à raiz) do
|  encontre  $T'$  que minimize  $f(T, m)$ , aonde  $T'$  é qualquer um dos:
|  a árvore atual após substituir o nó  $j$  por uma folha 1
|  a árvore atual após substituir o nó  $j$  por uma folha 0
|  a árvore atual após substituir o nó  $j$  por sua sub-árvore esquerda
|  a árvore atual após substituir o nó  $j$  por sua sub-árvore direita
end
```

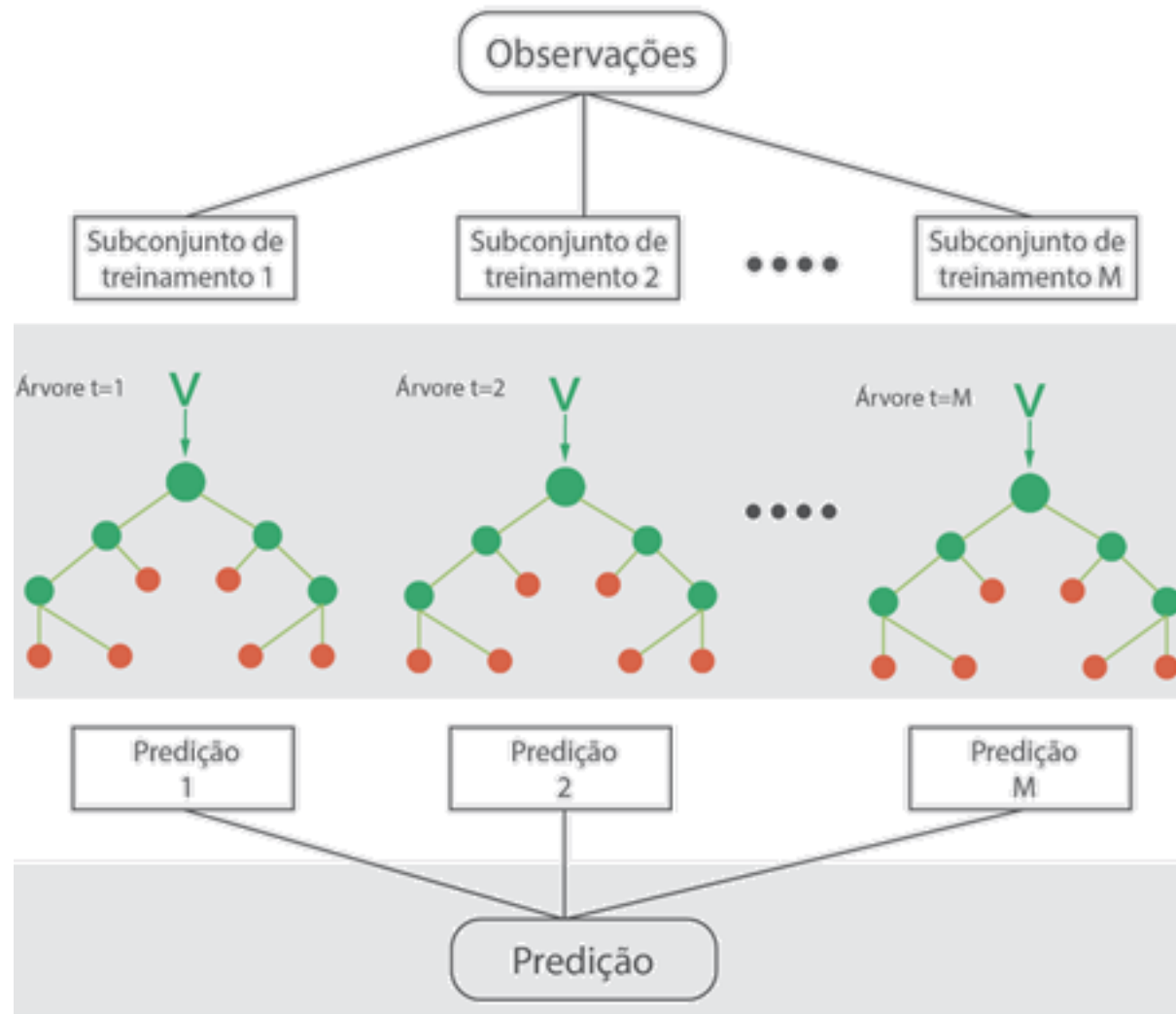
Fonte: elaborada pelo autor.



## ► Tipos de métodos *ensemble*

- Soluções simples: moda e média.
- Soluções avançadas: *Bagging*, *Boosting* e *Random Forest*.

Figura 6 – Estrutura de um modelo *bagging*



Fonte: elaborada pelo autor.

# PÓS-GRADUAÇÃO

## Teoria em prática

### Bloco 3

Lucas Claudino







## ► Teoria em prática

- Criar um *ensemble* para relacionar ozônio e velocidade do vento.
- Comando “require(data.table)” e “bagging\_data=data.table(airquality)”.

Ozone: média de ozônio em partes por bilhão.

Solar.R: radiação solar, em Langleys.

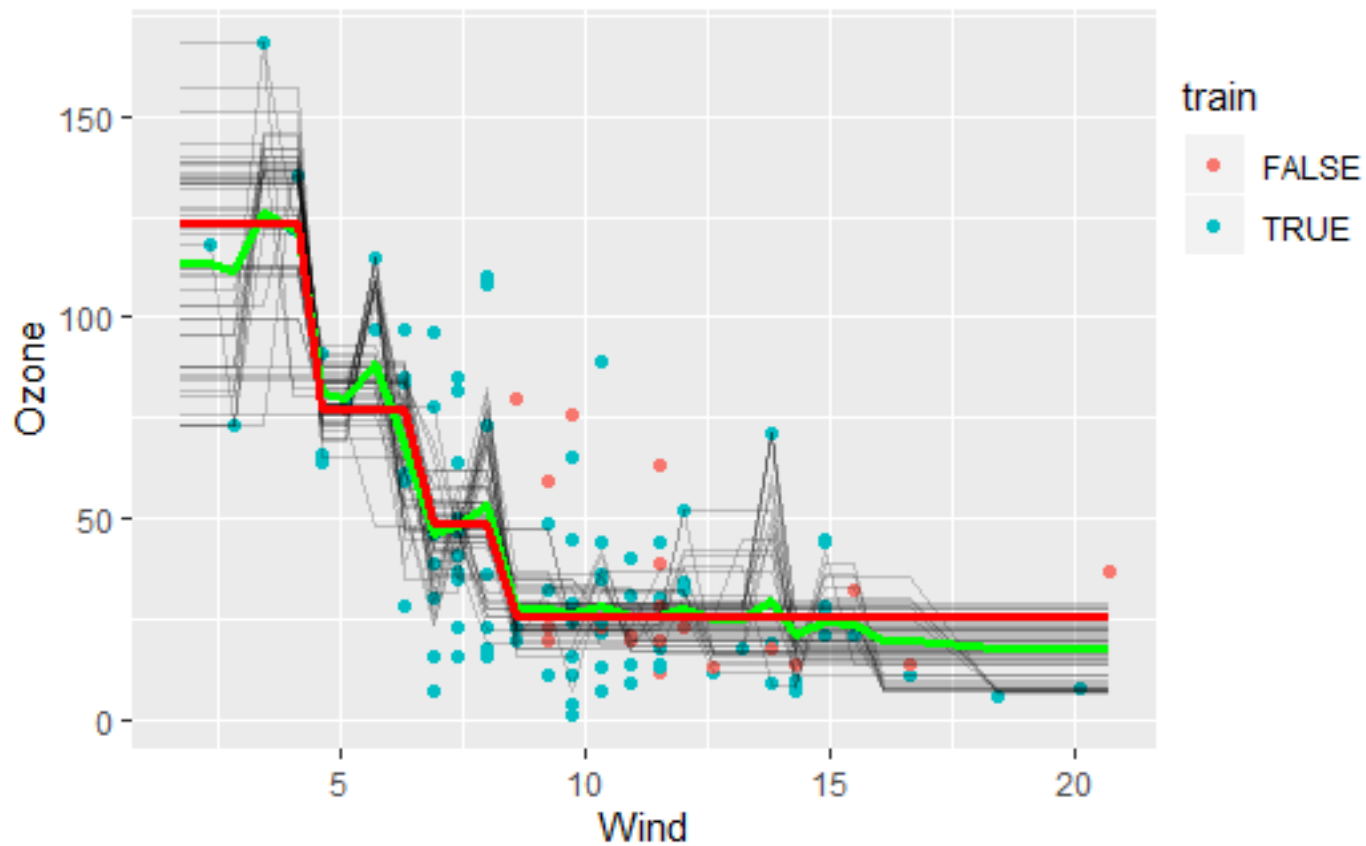
Wind: velocidade média do vento, em milhas por hora.

Temp: temperatura diária máxima, em graus Fahrenheit.



## ► Teoria em prática

Figura 7 – Resolução de problema utilizando árvores aleatórias e método *ensemble*



Fonte: elaborada pelo autor.

## Dica do professor

Bloco 4

Lucas Claudino





## ► Dica do professor

- Fórum com modelos, exemplos e algoritmos prontos para aprendizado de máquina.
- R-bloggers.



## Referências

FACELI, K. *et al.* **Inteligência artificial**: uma abordagem de aprendizado de máquina. São Paulo: LTC Editora, 2011.

HAN, J.; KAMBER, M.; PEI, J. **Data mining**: concepts and techniques. Elsevier, 2011.

MAINI, Vishal; SABRI, Samer. **Machine Learning for Humans**. 2017. Disponível em: <https://everythingcomputerscience.com/books/Machine%20Learning%20for%20Humans.pdf>. Acesso em: 26 ago. 2019.

RAMZAI, Juhi. **Simple guide for ensemble learning methods**. 2019. Disponível em: <https://towardsdatascience.com/simple-guide-for-ensemble-learning-methods-d87cc68705a2>. Acesso em: 27 ago. 2019.

