

Introdução a Sistemas de Recomendação e Numpy

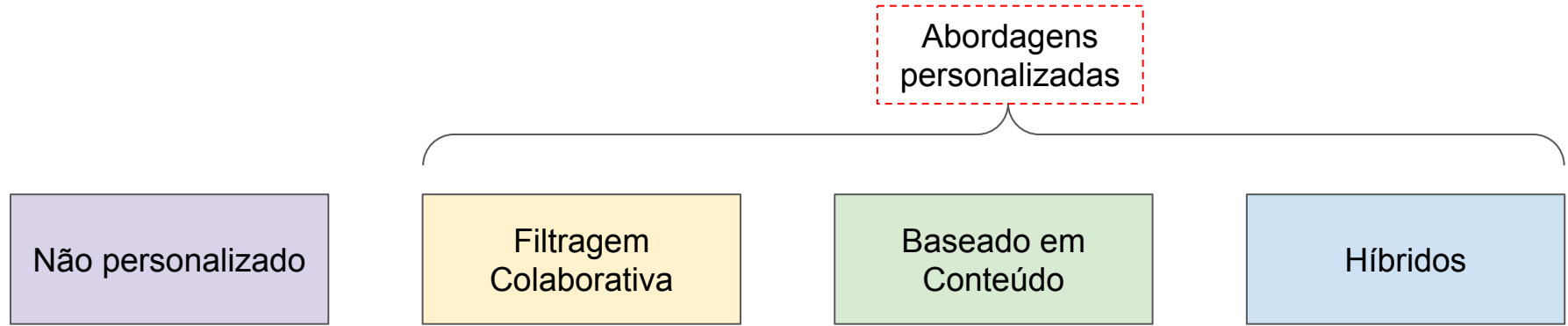
Lucas Félix



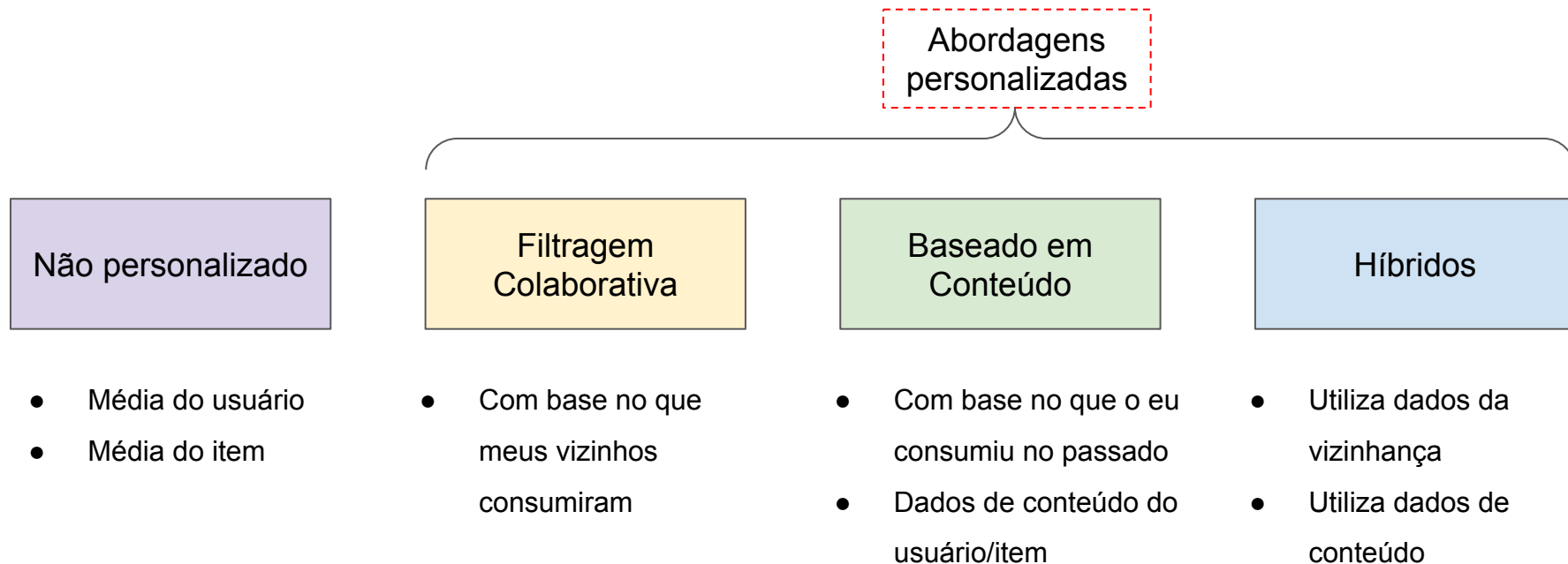
Sistemas de Recomendação:



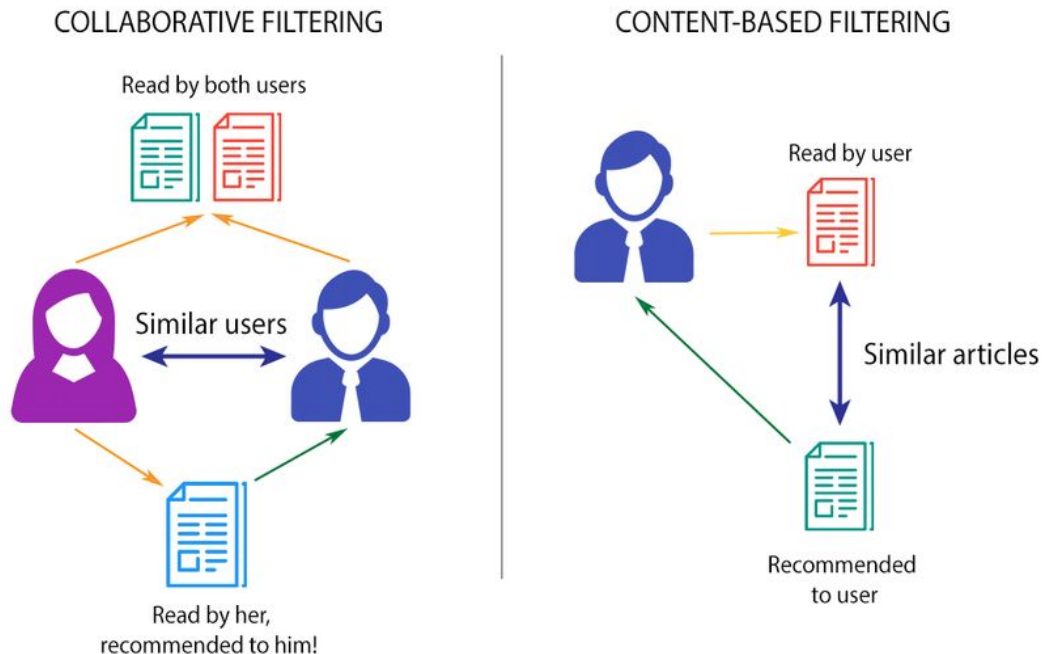
Sistemas de Recomendação: Principais Metodologias



Sistemas de Recomendação: Principais Metodologias



Sistemas de Recomendação: Principais Metodologias



Sistemas de Recomendação: Principais Metodologias

Não personalizado

- Média do usuário
- Média do item

Filtragem Colaborativa

- Com base no que meus vizinhos consumiram

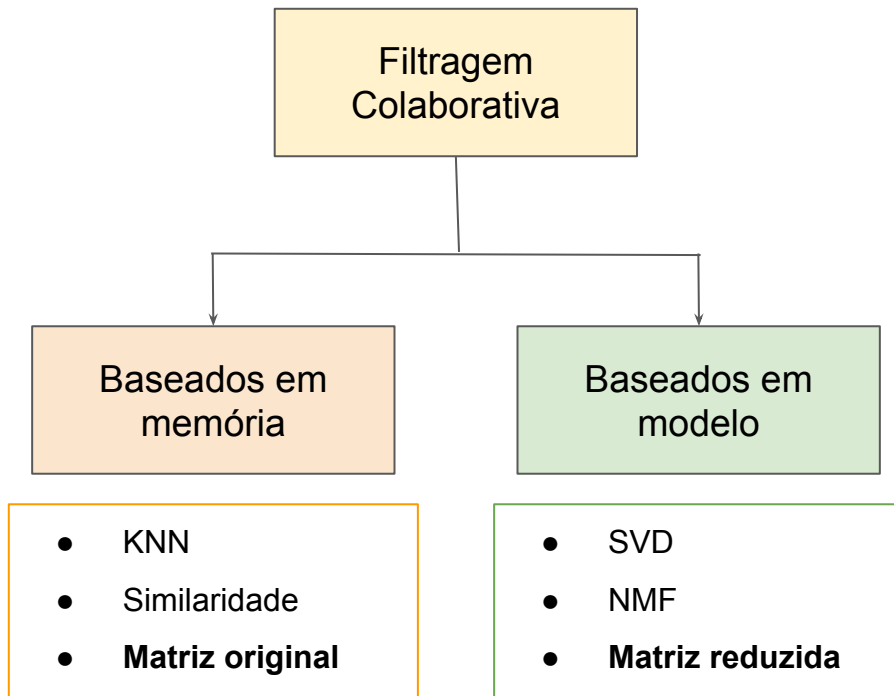
Baseado em Conteúdo

- Com base no que o eu consumiu no passado
- Dados de conteúdo do usuário/item

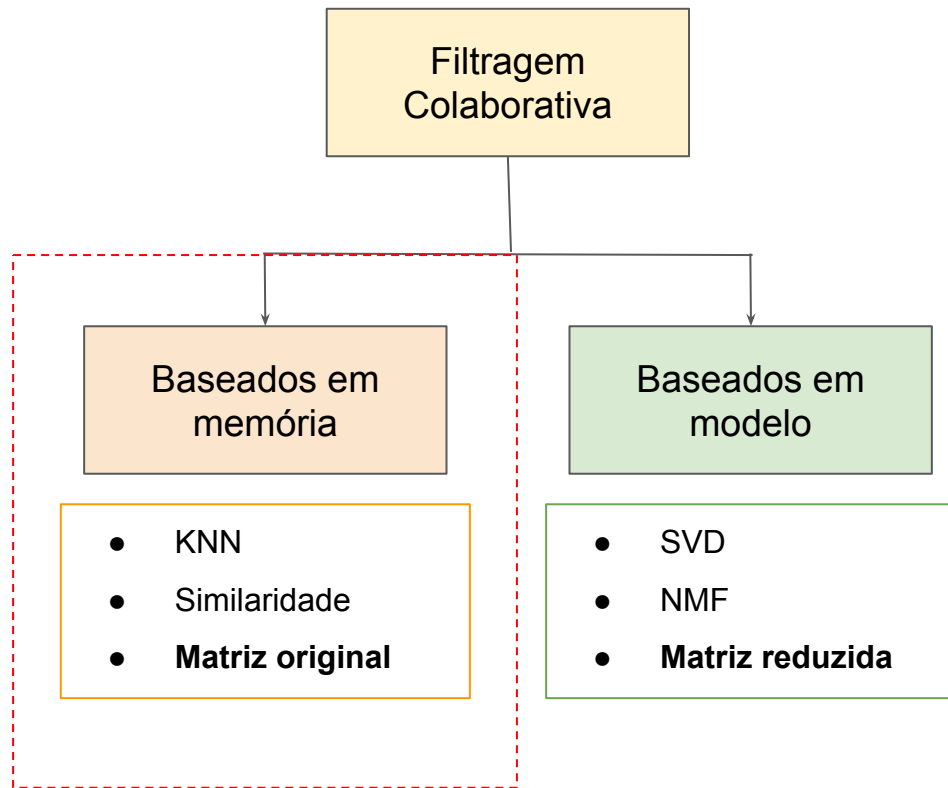
Híbridos

- Utiliza dados da vizinhança
- Utiliza dados de conteúdo

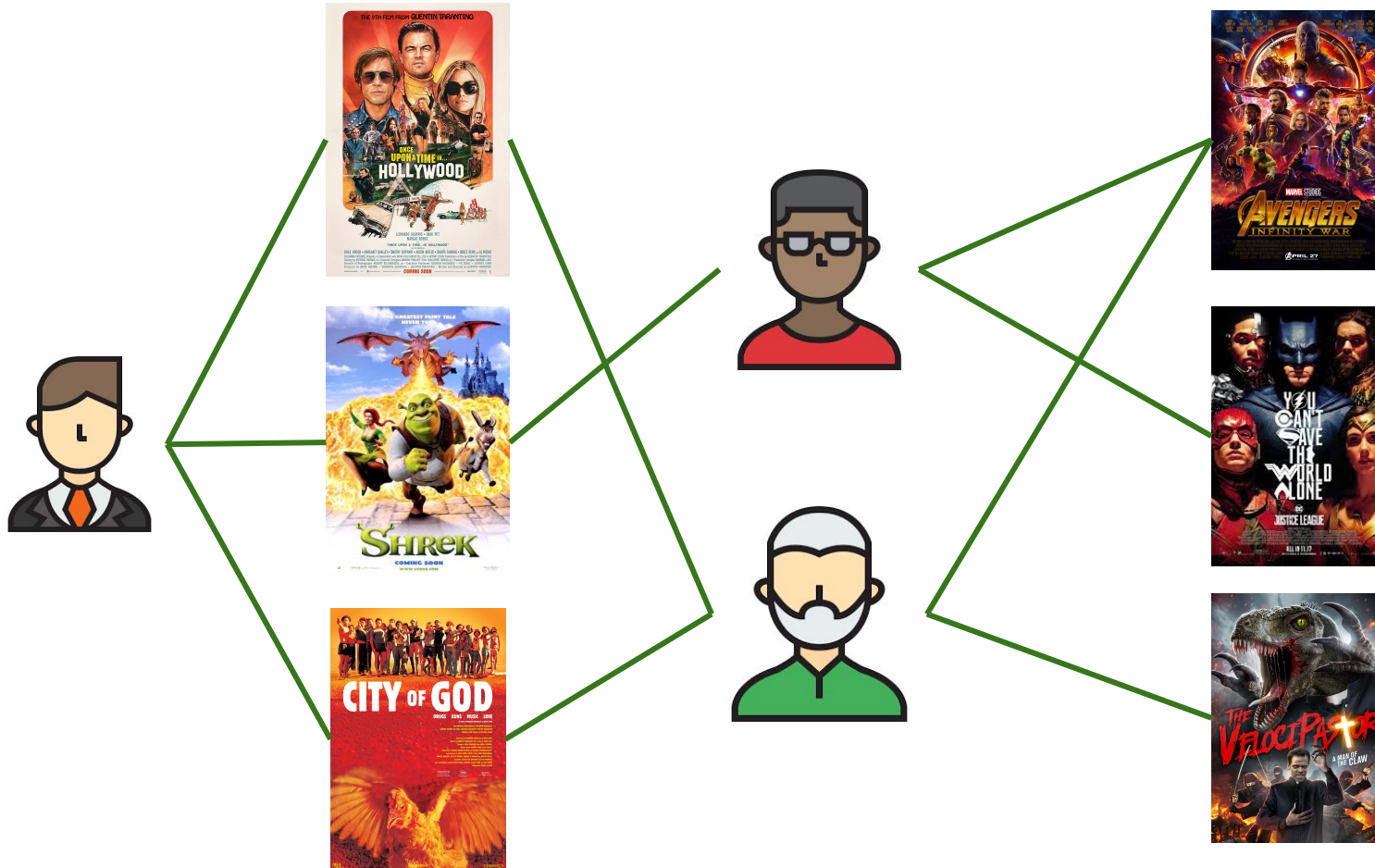
Sistemas de Recomendação: Filtragem Colaborativa



Sistemas de Recomendação: Filtragem Colaborativa



Sistemas de Recomendação: Baseado em Memória



Sistemas de Recomendação: Baseado em Memória

						
	5	3	5			
		4		5	5	
	4		3	2		5

Sistemas de Recomendação: Baseado em Memória

Qual é o problema dessa representação ?

						
	5	3	5			
		4		5	5	
	4		3	2		5

Similaridade:

$$\text{similarity}(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n A_i^2} \times \sqrt{\sum_{i=1}^n B_i^2}}$$



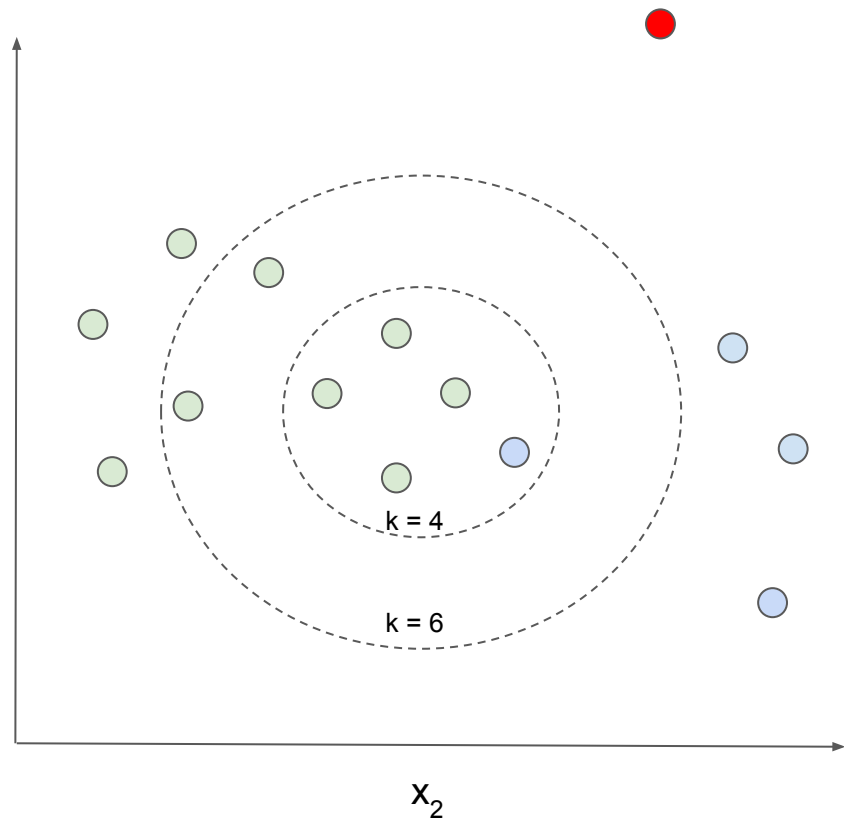
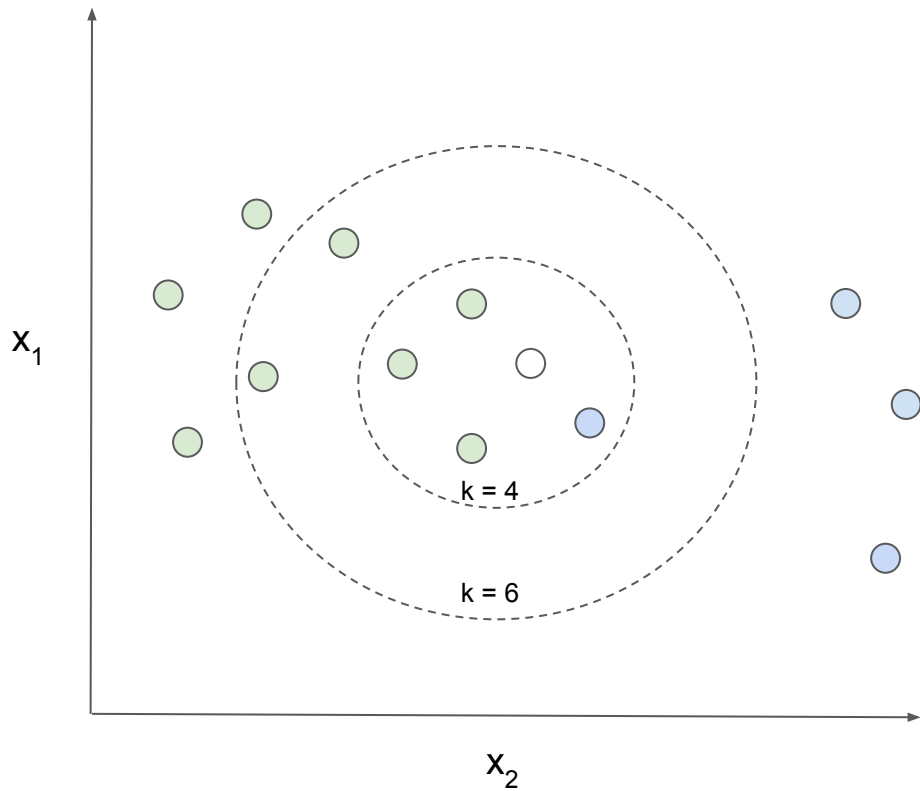
5	3	5			
	4		5	5	

$$\text{sim}(A, B) = \frac{5 \times 0 + 3 \times 4 + 5 \times 0 + 0 \times 5 + 0 \times 5 + 0 \times 5}{\sqrt{5^2 + 3^2 + 5^2} \times \sqrt{4^2 + 5^2 + 5^2}} = 0.19$$

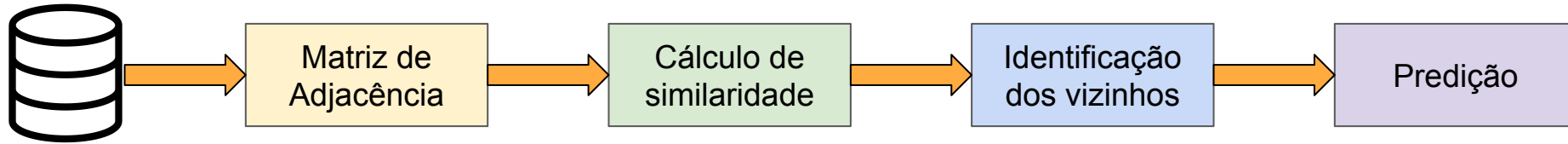
Predição:

$$y = \frac{\sum_{j \in N_u^k(i)} sim(i, j) \times r_{u,j}}{\sum_{j \in N_u^k(i)} sim(i, j)}$$

Variável	Significado
y	Predição
N	Vizinhos do usuário/item u (i)
k	Quantidade total de vizinhos
u	Usuário alvo
i	Item alvo
j	Item vizinho
r	Nota que o usuário u deu ao item j



Resumo da Metodologia:



Problemas da representação:

- Esparsidade
- Complexidade
- Similaridade sensível a amostragem de dados
- **Como contornar isso ?**
 - Boas implementações
 - Estrutura de dados
 - Bibliotecas de alto desempenho
 - Pandas, **Numpy**, Numba

NumPy:



- Biblioteca open-source
- Operações de computação numérica
- Estruturas de dados próprias
- Funções da biblioteca são escritas em C