

# Data Science em Projetos

Live 02

Sistema de Recomendação



# Data Science em Projetos

## Ao final dessa live você vai ser capaz de:

- ✧ Descrever a importância dos sistemas de recomendações (RecSys) na solução de problemas reais
- ✧ Detalhar as diferenças entre filtro baseado em conteúdo e baseado em colaboração (content-based vs collaborative filtering)
- ✧ Descrever o que é o algoritmo de KNN (K Nearest Neighbors)
- ✧ Construir uma implementação de RecSys utilizando KNN
- ✧ Experimentar recomendações de filmes com o modelo construído

# Data Science em Projetos

## Motivação

- ✧ Primeira live nos assustou (juro que só soube que havia 70+ pessoas só depois, tomei um susto, OBRIGADO!)
- ✧ Muitos perguntaram sobre outliers, vamos fazer conteúdo disso com certeza!
- ✧ Alguns acharam um pouco corrido: infelizmente na live não conseguimos detalhar muito sem dar 3+ horas e deixar vocês mortos, mas podem deixar as dúvidas que respondemos tudo 😊
- ✧ A maioria gostou do Streamlit! Vamos voltar a vê-lo nesse evento, na sexta provavelmente

# Data Science em Projetos

## Sistemas de Recomendação





# Data Science em Projetos

## Sistemas de Recomendação

Livros que você pode gostar



# Data Science em Projetos

## Sistemas de Recomendação

Sugestões para Leon



# Data Science em Projetos

## Sistemas de Recomendação

- ✧ Uma das aplicações mais requisitadas no mercado
- ✧ Aplicável a qualquer tipo de e-commerce (essa palavra é cringe?)
- ✧ Qualquer cientista de dados deve conhecer o básico
- ✧ Pode começar com aplicações simples, até implementações bem complexas
- ✧ Área por si só: RecSys

# Data Science em Projetos

## Sistemas de Recomendação

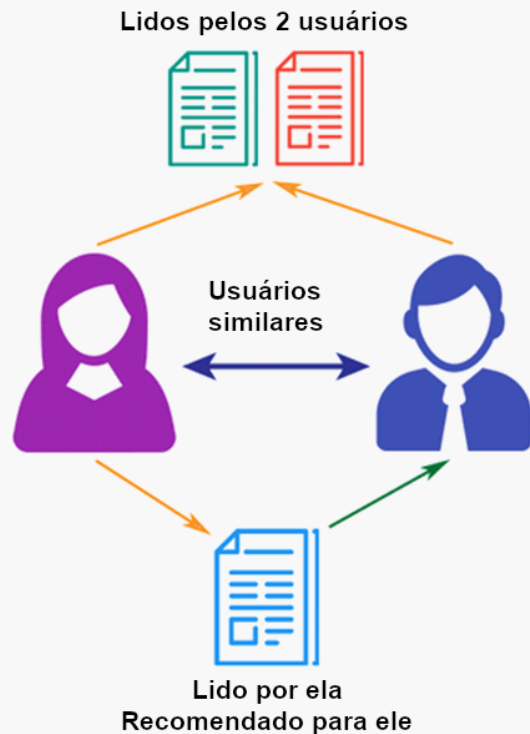
- ✧ Duas formas conhecidas de solucionar o problema de recomendação
- ✧ Filtragem baseada em conteúdo
- ✧ Filtragem colaborativa



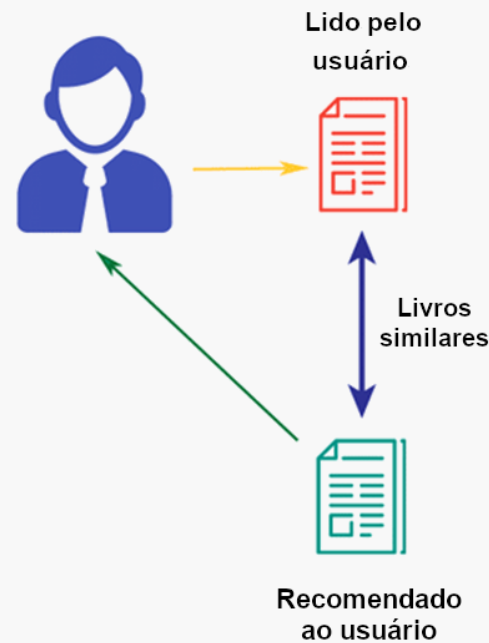
# Data Science em Projetos

## Sistemas de Recomendação

FILTRAGEM COLABORATIVA



FILTRAGEM BASEADA EM CONTEÚDO



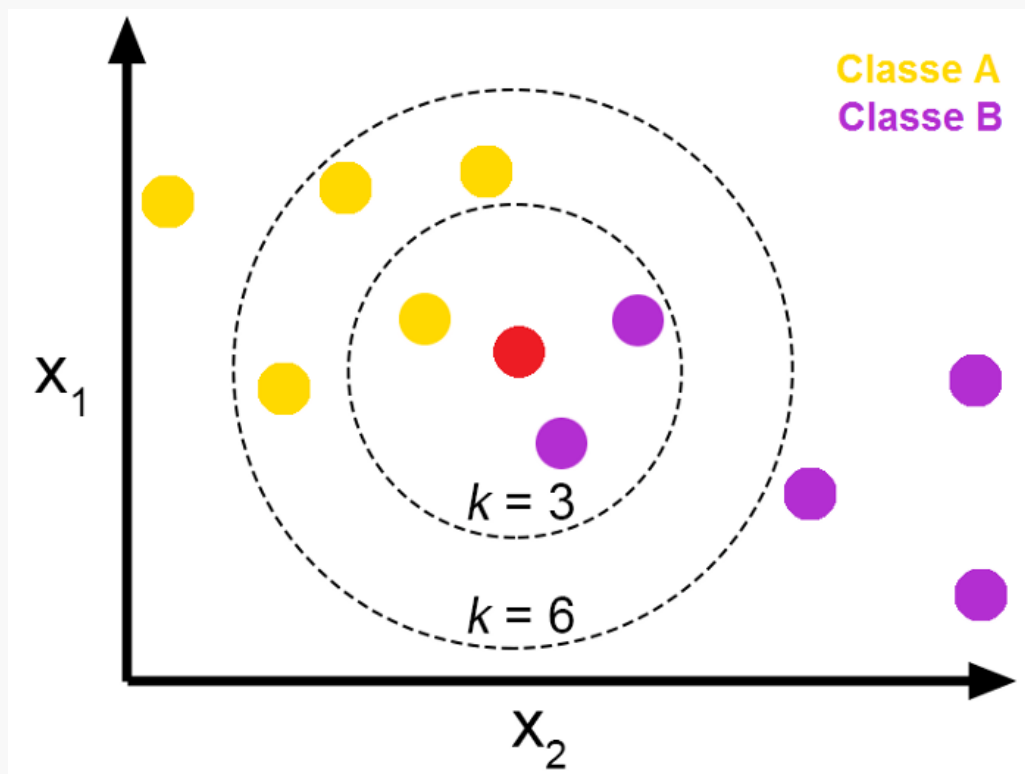
# Data Science em Projetos

## Sistemas de Recomendação

- ✧ Identificar similaridades: KNN
- ✧ K Nearest Neighbors
- ✧ Para classificação e regressão: procura as diferenças entre uma amostra e os seus “vizinhos” para determinar qual a sua classe (ou valor para regressão)
- ✧ As variáveis independentes (características das observações) são utilizadas no cálculo dessa distância para determinar quem se parece mais com a amostra que não conhecemos

# Data Science em Projetos

## Sistemas de Recomendação



# Data Science em Projetos

## Sistemas de Recomendação

- ✧ Nesse sistema de recomendações, vamos buscar os conteúdos similares a partir das distâncias definidas nos filmes
- ✧ Quando houver um filme que foi visto por João, vamos buscar outros similares para recomendá-lo
- ✧ A distância entre os filmes quem vai nos fornecer é o modelo KNN

# Data Science em Projetos

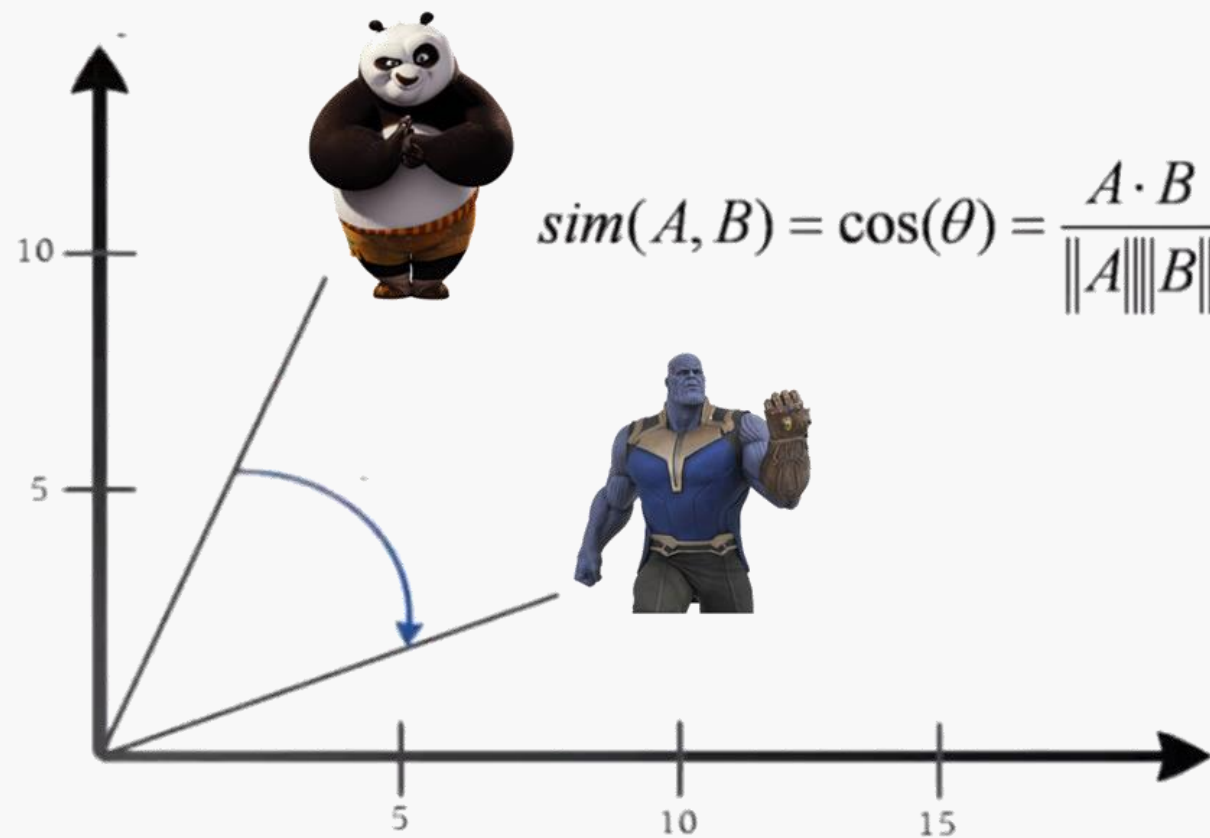
## Sistemas de Recomendação

### SIMILARIDADE POR COSSENO

$\cos 0^\circ = 1$   
 $\cos 90^\circ = 0$

$\cos 0 = \text{mesmo vetor}$

Mesmo vetor =  
mesmas notas dos  
mesmos usuários





# Data Science em Projetos

## Sistemas de Recomendação

- ✧ Vamos utilizar, portanto, o KNN para uma análise não supervisionada
- ✧ KNN usado em classificação e regressão são análises supervisionadas, ou seja, aprender com o histórico para determinar o futuro (modelos preditivos)
- ✧ Nesse caso, KNN vai nos ajudar a calcular as distâncias dos filmes calculando a similaridade de cossenos entre eles!

# Data Science em Projetos

## Sistemas de Recomendação

✧ Vamos ao código!

# Data Science em Projetos

## Obrigado!!

- ✧ Diz pra gente o que achou nos comentários!!
- ✧ O que achou do conteúdo até agora??
- ✧ Até sexta-feira!!!