



Data Science Academy

[www.datascienceacademy.com.br](http://www.datascienceacademy.com.br)

Big Data Real-Time Analytics com  
Python e Spark

Mini-Projeto 3  
Sistema de Recomendação

Um Sistema de Recomendação combina várias técnicas computacionais para selecionar itens personalizados com base nos interesses dos usuários e conforme o contexto no qual estão inseridos. Tais itens podem assumir formas bem variadas como, por exemplo, livros, filmes, notícias, música, vídeos, anúncios, links patrocinados, páginas de internet, produtos de uma loja virtual, etc. Empresas como Amazon, Netflix e Google são reconhecidas pelo uso intensivo de sistemas de recomendação com os quais obtém grande vantagem competitiva.

Sistemas de Recomendação podem ser vistos ao realizar buscas em sites de pesquisa da internet, em compras online, ou até mesmo ao visualizamos nossos e-mails. São o mecanismo por trás da propaganda personalizada que recebemos na web, com indicações de sites para visitarmos ou produtos para compramos.

Com o advento do consumo em dispositivos móveis, a propagação o e-commerce e o Big Data, sistemas de recomendação tornaram-se um tema extremamente atrativo. Através de algoritmos simples e facilmente integráveis a aplicações web, eles agregam valor ao negócio online, promovendo itens de consumo direcionados a um público alvo.

## **Tipos de Sistemas de Recomendação**

Podem ser classificados em basicamente três categorias, a partir de como a recomendação é feita:

### **Baseado em Conteúdo**

Um sistema de recomendação baseado em conteúdo recomenda ao usuário produtos que sejam semelhantes ao que ele preferiu no passado. A recomendação é feita a partir de tags "descritoras" de itens. Itens com características próximas destas tags são recomendados. Em um cenário de recomendação de filmes, por exemplo, um usuário que, assiste e gosta do filme "Star Wars" teria recomendações do gênero ação e ficção científica.

Vantagens deste tipo de sistema é que são simples para dados textuais e não necessitam de muitas informações sobre um usuário para sugerir itens. Todavia, além de serem difíceis de aplicar em contextos multimídia, podem oferecer recomendações repetitivas, recomendando sempre assuntos que o usuário já conhece.

### Filtragem Colaborativa

Filtragem colaborativa consiste na recomendação de itens que pessoas com gosto semelhante preferiram no passado. Analisa-se a vizinhança do usuário a partir da regra: "Se um usuário gostou de A e de B, um outro usuário que gostou de A também pode gostar de B". Esse tipo de recomendação apresenta resultados positivos na prática e evita o problema de recomendações repetitivas. Uma desvantagem é que requer grande número de informações sobre o usuário e sua vizinhança para funcionar precisamente.

### Sistemas Híbridos

Por fim, um sistema híbrido consiste em combinar as duas abordagens mencionadas, tentando fortificá-las e superar suas desvantagens.

Dentro da categoria Filtragem Colaborativa, pode-se ainda dividir os sistemas em mais duas categorias: Item-Based e User-Based.

Neste projeto, veremos como aplicar um algoritmo de Sistema de Recomendação e calcular o affinity score, a fim de descobrir se um produto deve ou não ser recomendado a um usuário.