



Data Science Academy

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

## Machine Learning

### Projeto 2

Construindo um Sistema de Recomendação  
de Filmes (Boosted Collaborative Filtering)

## Especificação

Como estudamos no capítulo 13, construir um sistema de recomendação de filmes não é uma tarefa complexa, mas automatizar este processo pode ser desafiador. Um ponto importante a ser considerado é a medida de similaridade que será usada, já que isso vai influenciar no resultado e precisão das recomendações.

Neste projeto construímos um sistema de recomendação de filmes totalmente automatizado e com cada função desenvolvida em linguagem Python. Você encontra ainda o código necessário para calcular a similaridade pelo coeficiente de Pearson, a similaridade cosine e a similaridade Jaccard. A medida de similaridade será um dos parâmetros a ser usado pelo sistema de recomendação. Além disso, unimos as duas técnicas de filtro colaborativo, item based e user based para trabalhar com a técnica Boosted Collaborative Filtering, formando na verdade um sistema de recomendação híbrido.

O dataset de entrada possui os ratings (avaliações) de usuários a uma série de filmes. Nosso sistema vai processar esses dados e gerar como saída a lista de usuários e os filmes que devem ser recomendados. Esse não é um sistema online e seu processamento leva algumas horas (esse projeto foi processado em 5 horas nos equipamentos da DSA). Conforme conversamos, muitos sistemas de recomendação são processados durante à noite e no dia seguinte as recomendações podem ser feitas por e-mail, por exemplo (Amazon e Netflix possuem essa abordagem).

Para executar o sistema, abra um terminal ou prompt de comando, navegue até o diretório onde estão os arquivos que você baixou e execute:

```
python recommender.py ratings.csv Pearson
```

python – nome do interpretador

recommender.py – nome do seu aplicativo (script)

ratings.csv – nome do arquivo com os ratings dos usuários

Pearson – medida de similaridade (pode ser ainda cosine ou Jaccard e o script está preparado para executar com uma das 3 medidas)