

# CPS833 - Datamining

[Página inicial](#) ► [Meus cursos](#) ► [CPS833](#) ► [16 abril - 22 abril](#) ► [Tarefa de Clusterização](#)

[Descrição](#) [Enviar](#) [Editar](#) [Visualizar envios](#)

## Tarefa de Clusterização

**Disponível até:** terça, 30 abr 2019, 21:00

**Número máximo de arquivos:** 1

**Tipo de trabalho:** Trabalho individual

### Tarefa de Clusterização

Encontre clusters de filmes semelhantes nas bases do MovieLens de 100k e de 1M.

Para isso, **utilize apenas as notas dos usuários como atributos** para os filmes. Faça dois tipos de abordagens:

1. Você deve considerar o vetor de notas de um filme, **preenchendo com a média do filme as notas 0**, e calcular a **distância usando, por exemplo, cosseno**. Esse agrupamento obtido não deverá ser muito bom, e deverá ser mais lento. Rode mais de uma vez para verificar o que acontece com os grupos. **Faça um gráfico em função do número de grupos**. Use os algoritmos **K-means e DBScan**. Talvez haja dificuldades para rodar no dataset de 1M.
2. Faça um préprocessamento: **reduza a dimensionalidade** do conjunto de dados usando o **PCA**. Escolha um número Q de dimensões (experimente! plote os valores singulares, por exemplo). **Depois aplique a métrica baseada em distância euclidiana**. Rode mais de uma vez para verificar o que acontece com os grupos. **Faça um gráfico em função do número de grupos**. Use os algoritmos **K-means e DBScan**.

No seu relatório liste os **5 grupos mais coesos com mais de 5 filmes**. Se um grupo listado tiver mais de 30 filmes, não é necessário listar o nome de todos. Compare os grupos gerados pelo k-means com os gerados pelo DBScan, e verifique se tem algum grupo parecido. Coloque uma tabela com os tempos médios de execução de cada algoritmo.

Comente sobre dois agrupamentos que contenham filmes muito vistos (com muitas avaliações) e filmes pouco vistos nos dois métodos (item 1 e 2, K-means e DBScan)

VPL 3.2.4

### NAVEGAÇÃO



[Página inicial](#)


■ [Painel](#)

[Páginas do site](#)

[Meus cursos](#)

[CPS833](#)

[Participantes](#)

 [Emblemas](#)