



▼ Tarefa de Aprofundamento 6

CASO: Clustering, Segmentando Países pela Força de Trabalho

Com base na atividade *Tarefa de Aprofundamento 5*, **Pesquise e Discuta apenas uma** das perguntas a seguir.

1. Pesquise outros modelos de clusterização e exemplifique uma situação em
- ▼ que um desses modelos pesquisados apresenta resultados melhores que os modelos de clusterização hierárquica e kmédias.

Talvez você queira aplicar esse outro modelo aos dados da Tarefa 5 e compartilhar com os colegas. Mas isso é opcional.

Sua resposta aqui em 1 (um) parágrafo.

2. A clusterização é bastante empregada para a agrupamento de documentos como emails, páginas de notícias etc. Pesquise qual a métrica de similaridade
- ▼ (distância) normalmente empregada na clusterização de documentos e explique por quê.

A escolha de uma medida de similaridade ou distância entre documentos pode ser bastante subjetiva. A medida de distância mais utilizada é Euclidiana. Normalmente por ser levado em consideração o modelo de representação dos documentos, a natureza das variáveis que podem ser discretas, contínuas ou binárias. Também considerando a escala de medida, se é nominal, ordinal, intervalo

entre outras. É importante também levar em consideração o uso de uma medida padrão caso não

▼ Referências

https://www.maxwell.vrac.puc-rio.br/14103/14103_6.PDF

https://www2.ufjf.br/lates//files/2016/12/Conte%3%bado-5-%e2%80%93-N_An%c3%aallise-de-

<https://www.cin.ufpe.br/~if796/aulas/clustering-2010.pdf>

<http://www.leg.ufpr.br/~walmes/ensino/mintex/slides/07-similar.pdf>

Todos acessados em 4 de dezembro de 2022 as 21:31

Adicione aqui as referências empregadas.

[Colab paid products](#) - [Cancel contracts here](#)

