

Trabalho 1: Clustering (Agrupamento)

Mineração de Dados - Pós-Dev. IA - IFES - TRE

Nome do Aluno/a: _____

Data de Entrega: **12 de Abril de 2024**

Professor: Sérgio N. Simões

O objetivo deste segundo trabalho é aprender conceitos de análise agrupamento (*clustering*) e revisar os conceitos de pré-processamento de dados utilizando a biblioteca `Pandas + Scikit-Learn`! Entretanto, conforme mencionado nas aulas, o mais importante é que você aprenda os conceitos e consiga implementá-los na linguagem que lhe for mais conveniente.

Exercício 1 - Interpretando os resultados das análises de *clustering*:

O objetivo deste exercício é treinar o estudante para interpretar análises de *clustering* e Segmentação de dados. Imagine que você trabalha em uma empresa que realiza market direcionado aos seus clientes. A sua tarefa é, através das análises de *clustering* e Segmentação de dados, auxiliar na realização de marketing direcionado. Inicialmente, é fornecido um Notebook com uma análise de *clustering* e a sua tarefa será buscar entender e interpretar os resultados.

<https://colab.research.google.com/drive/1vFRy6g0JPIm8pItFqXyplU41ul-Bn?usp=sharing>

TAREFA: Primeiramente, leia e execute o notebook acima tentando entender o que faz cada célula. Em seguida, faça as atividades abaixo:

1-1) Análise Exploratória: comente sobre as distribuições dos dados de cada atributo (gerais) destacando algumas (3-5) observações que você julgar importante.

1-2) Análise de *Clustering*: gere um relatório sobre os perfis encontrados em cada *cluster*.

1-3) Nas células da seção `Multivariate Conclusions` explique respectivamente a partir de quais gráficos é possível extrair a conclusão de cada item. Justifique sua resposta.

1-4) Explicar quais gráficos embasam a conclusão do cluster 0, 1 e 2.

1-5) Descreva quais seriam as estratégias de marketing que você sugeriria para realizar uma propaganda direcionada ao cluster 2.

1-6) Faça uma pesquisa e explique com suas palavras como utilizar árvores de decisão para interpretar os clusters.

Lembre-se de documentar cada etapa do trabalho e fornecer explicações claras para cada exercício.

Exercício 2 - Alguns Conceitos importantes sobre *clustering*:

Tarefa: responda as questões abaixo.

2-1) Explique com suas palavras o que é *clustering* e quais suas aplicações.

2-2) Porque clustering é considerado aprendizado **não-supervisionado**?

2-3) Qual a diferença entre clustering hierárquico e particional?

2-4) Explique com suas palavras como funciona o K-means. Cite 3 outros métodos de clustering que você acha interessante.

2-5) Quais as métricas de distâncias mais usadas em *clustering*?

2-6) Quais são os critérios de *linkage* e para que servem?

Algumas Referências

- <https://www.ic.unicamp.br/~reltech/2005/05-03.pdf>
- <https://dcm.ffclrp.usp.br/~augusto/teaching/ami/AM-I-Clustering.pdf>
- <https://medium.com/analytics-vidhya/partitional-clustering-181d42049670>
- <https://lzpdatascience.medium.com/what-is-the-difference-between-hierarchical-and-partitional-clustering-edc0d488c7c4>
- <https://machinelearningmastery.com/clustering-algorithms-with-python/>
- <https://towardsdatascience.com/k-means-clustering-with-scikit-learn-6b47a369a83c>
- <https://stackabuse.com/hierarchical-clustering-with-python-and-scikit-learn/>
- <https://www.datacamp.com/community/tutorials/k-means-clustering-python>
- <https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68>
- <https://towardsdatascience.com/hierarchical-clustering-explained-e58d2f936323>
- <https://medium.com/leukemiaairesearch/clustering-techniques-with-gene-expression-data-4b35a04f87d5>
- <https://towardsdatascience.com/which-countries-are-affected-the-most-by-covid-19-4d4570852e31>

Bom trabalho!