

CURSO SISTEMAS DE INFORMAÇÃO

MINERAÇÃO DE DADOS

CLUSTERIZAÇÃO UTILIZANDO O PYCARET

Essa atividade tem por objetivo aplicar os conhecimentos de machine learning supervisionado a biblioteca PyCaret para resolver uma tarefa de **agrupamento (clusterização)** com um conjunto **de dados reais**. Dessa forma, é preciso que você explique todos os atributos utilizados do dataset.

Siga os seguintes passos:

- Importe e explore os dados escolhidos.
- Configure o ambiente do PyCaret para a tarefa de clusterização. Obs.: **Não há variável alvo em clusterização**, pois o objetivo é descobrir grupos ocultos nos dados.
- Compare os modelos disponíveis e selecione o melhor desempenho.
- Realize o ajuste de hiperparâmetros com `tune_model()`.
- Avalie e visualize o modelo final. Gere visualizações com `plot_model()` (como elbow, silhouette e cluster plot) para avaliar o número ótimo de clusters e a distribuição dos dados em cada grupo.
- Interprete os resultados e discuta as variáveis mais relevantes
- Justifique a possibilidade desse modelo treinado ser utilizado ou não em um ambiente de produção. Considere aspectos como estabilidade dos clusters, necessidade de atualização e custo de manutenção do modelo.

Entrega:

Código Python comentado (.ipynb). Especifique dentro do código, a origem dos dados e a informação contida em cada atributo.



TOCANTINS
GOVERNO DO ESTADO



Na próxima aula, cada aluno deverá explicar o seu trabalho, inclusive todos os processos realizados pelo PyCaret.

A relação do dataset para cada aluno está em um arquivo separado.