

Projeto Final

Mineração de Dados

Base de Dados: **Employee Analysis | Attrition Report**

<https://www.kaggle.com/datasets/whenamancodes/hr-employee-attrition>

- Base de dados fictício criado pela IBM.
- Fatores que levam ao desgaste (**Attrition**) dos funcionários.

1) Análise Exploratória

- a) Mostrar no formato de **violin plot** a distribuição dos dados do atributo tipo de cargo (**JobRole**)
- b) Apresentar a renda média mensal (**MonthlyIncome**) por escolaridade (**EducationField**) e desgaste (**Attrition**). Utilizar **Pandas Groupby**.

2) Pré-processamento

- a) Transformação dados categóricos
- b) Normalização
- c) Feature Selection Filter x Wrapper
- d) Validar as técnicas de Feature Selection com pelo 1 classificador
- e) Diagrama venn com pelo menos 3 técnicas distintas de Feature Selection

3) Modelos Preditivos e validação

- a) Rodar com o melhor subconjunto de características investigados na etapa 2 (Filter ou Wrapper)
- a) KNN, Decision Tree e Random Forest
- b) Comparar os resultados entre GridSearch e Randon Search
- c) Apresentar o resumo com as métricas Precision, Recall, F1-Score e Accuracy

d) Automatizar a execução de todas as etapas com uso do pipeline

4) Forma de envio:

a) Arquivos: ipynb e pdf