

# INTRODUÇÃO

Olá,

Este teste consiste em desenvolver um modelo de aprendizado de máquina capaz de classificar as amostras do conjunto de dados [dataset\\_cdjr.parquet.gzip](https://drive.google.com/file/d/1HXq9mczY-5OpFaXK3kk8zAgFEgEgF3jt/view?usp=sharing) em: <<https://drive.google.com/file/d/1HXq9mczY-5OpFaXK3kk8zAgFEgEgF3jt/view?usp=sharing>>.

É um conjunto de dados pequeno de 466 entradas, composto inicialmente por 16 features (feature0, feature1, ..., feature 15) e uma coluna target, que é a classe. A coluna target é zero indicando a **não** ocorrência do evento que desejamos prever e um quando há ocorrência do evento que desejamos prever. Segue uma listagem das variáveis e respectivos tipos de dados.

#	Column	Non-Null Count	Dtype
0	feature0	466 non-null	float64
1	feature1	466 non-null	int64
2	feature2	466 non-null	float64
3	feature3	466 non-null	float64
4	feature4	466 non-null	float64
5	feature5	466 non-null	float64
6	feature6	466 non-null	int64
7	feature7	466 non-null	float64
8	feature8	466 non-null	float64
9	feature9	466 non-null	int64
10	feature10	466 non-null	float64
11	feature11	466 non-null	float64
12	feature12	466 non-null	float64
13	feature13	466 non-null	float64
14	feature14	466 non-null	float64
15	feature15	466 non-null	int64
16	target	466 non-null	int64

## ATIVIDADES

Nossa expectativa é a de observar a sua atuação nas seguintes tarefas:

- ☐ Análise exploratória dos dados
- ☐ Preparação dos dados
- ☐ Modelagem
- ☐ Avaliação da performance do modelo
- ☐ Entrega do modelo

## ENTREGA

- 1) Crie um repositório em <https://github.com/> e compartilhe conosco.
- 2) Documente as atividades mencionadas na seção anterior, através do README do repositório criado no item anterior.
- 3) Informe, também no README do repositório, como podemos executar/"rodar" a solução que você construiu.

**Nota:** Sinta-se à vontade para incrementar a solução; por exemplo, criar um ambiente docker para a execução do ambiente.