A base de dados escolhida para o projeto A3 foi **“Employee dataset”**, está disponível no Kaggle: <https://www.kaggle.com/datasets/tawfikelmetwally/employee-dataset>

As colunas da base de dados estão divididas da seguinte forma:

* “education”: coluna referente a qualificação educacional dos funcionários (Tipo String)
* “joining\_year”: coluna referente ao ano em que o funcionário entrou na empresa (Tipo Int)
* “city”: coluna referente ao local ou cidade onde cada funcionário está (Tipo String)
* “payment\_tier”: coluna referente a categorização dos funcionários em diferentes níveis salarias (Tipo Float)
* “age”: coluna referende a idade (Tipo Int)
* “gender”: coluna referente a identidade de gênero (Tipo String)
* “ever\_benched”: coluna que indica se um funcionário já esteve temporariamente sem trabalho atribuído (Tipo String)
* “experience”: coluna referente aos anos de experiencia do funcionário na área atual (Tipo String)
* “leaver\_or\_not”: coluna referente ao destino (Tipo String)

A variável target será Ever Benched, tem objetivo de prever se existem padrões de comportamento de afastamento dos funcionários ou falta de atividades a serem atribuidas. Escolhemos essa variável para entender e facilitar a tomada de decisões e prevenir problemas futuros com funcionários.

Uma variável que precisa ser modificada será a “PaymentTier”, ela será transformada em uma variável string, removendo características não numéricas.

As colunas “joining\_year” e “age” são variáveis numéricas que podem estar em diferentes escalas. Com a padronização, ela pode ser aplicada para garantir que todas as variáveis numéricas tenham uma escala comparável.

**1º Método de aprendizagem de máquina**

Como primeiro método de aprendizagem de máquina utilizamos a Regressão Logística que deram os seguintes resultados:

accuracy 0.14 1862

macro avg 0.02 0.05 0.03 1862

weighted avg 0.06 0.14 0.08 1862

Ou seja, os resultados indicam que o modelo pode precisar de melhorias, como em termos de precisão e recall para várias classes. Pode ser útil explorar técnicas de ajuste de modelo ou considerar estratégias de tratamento de desequilíbrios nas classes.