

## **Projeto A3 - UCs de Inteligência Artificial (IA)**

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)
- “JoiningYear”: 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)
- “PaymentTier”: coluna referente a categorização dos funcionários em diferentes níveis salariais (Tipo float)
- “Age”: coluna referente a idade (Tipo Int)
- “Gender”: coluna referente a identidade de gênero (Tipo String)
- “EverBenched”: coluna que indica se um funcionário já esteve temporariamente sem trabalho atribuído (Tipo String)
- “ExperienceInCurrentDomain”: coluna referente aos anos de experiência do funcionário na área atual (Tipo String)
- “LeaverOrNot”: coluna referente ao destino (Tipo String)

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

Umas das variáveis que precisam ser modificadas serão as “Education” e “Gender”, elas serão transformadas em uma variáveis do tipo ‘Binária’, removendo características não binárias.

## 1º Método de aprendizagem de máquina

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

```
accuracy          0.40    1862
macro avg    0.11    0.16    0.12    1862
weighted avg    0.26    0.40    0.29    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.

## 2º Método de aprendizagem de máquina

Como segundo método de aprendizagem de máquina utilizamos KNN (K - *Nearest Neighbors*) que deram os seguintes resultados:

```
[[1036 15]
 [113    0]]
0.8900343642611683
```

Com esse método obtivemos 89% de acerto em relação a pessoas que estão empregadas e desempregadas. De 1036 que estão empregadas tivemos apenas 15 previsões erradas e de 113 que estão desempregadas tivemos 100% de acerto nas previsões. Esse método se mostrou eficaz nas previsões podendo passar por melhorias.

## 3º Método de aprendizagem de máquina

Como terceiro método de aprendizagem de máquina utilizamos Naive Bayes que retornaram os seguintes resultados:

```
A acurácia do modelo 3 foi 90.64%
[[1055    0]
 [ 109    0]]
```

O Naive Bayes é um classificador probabilístico simples, mas eficaz. O fato de alcançar uma taxa de acerto de 90.64% sugere que o modelo está realizando bem a tarefa em relação à distinção entre pessoas empregadas e desempregadas.

## **Conclusão**

Os três testes de análise de dados revelaram insights sobre a predição do status de emprego. O modelo de regressão mostrou uma acurácia de 40%, ou seja, não é possível identificar se as pessoas já ficaram desempregadas ou não. Em contrapartida, o KNN obteve 89% de acurácia, indicando melhor desempenho na identificação de padrões. O Naive Bayes alcançou a maior acurácia, com 90.64%, destacando sua eficácia e assertividade na análise de dados relacionados ao desemprego. Os resultados enfatizam a importância de algoritmos mais robustos, como KNN e Naive Bayes, para uma predição mais precisa e significativa sobre os padrões comportamentais relacionados ao afastamento de funcionários ou à falta de atribuições de tarefas.