

---

## **Atividade com k-NN, Decision Tree e Random Forest para Classificação de Sobreviventes do Titanic**

**Objetivo:** O objetivo desta atividade é criar e comparar três modelos de classificação, k-NN, Decision Tree e Random Forest, utilizando a técnica de hold-out com o conjunto de dados do Titanic para prever se um passageiro sobreviveu ou não ao naufrágio.

**Conjunto de Dados:** O conjunto de dados do Titanic está disponível em diversos repositórios de dados, como Kaggle (<https://www.kaggle.com/c/titanic>). Ele contém informações sobre passageiros do navio, incluindo características como idade, sexo, classe da cabine, entre outras, além da variável de saída que indica se o passageiro sobreviveu ou não.

### **Tarefas:**

- Baixe o conjunto de dados do Titanic e carregue-o em seu ambiente de programação.
- Explore o conjunto de dados para entender suas características, como dimensões, tipos de dados, estatísticas descritivas e a distribuição das classes (sobreviventes e não sobreviventes).
- Realize o pré-processamento dos dados, tratando dados ausentes ou inconsistentes e selecionando as variáveis relevantes para a classificação.
- Utilize a técnica de divisão em treino (70%), validação (10%) e teste (20%)
- Implemente os modelos de k-NN, Decision Tree e Random Forest utilizando a biblioteca Scikit-learn em Python.
- Treine os modelos com conjunto de treino.
- Avalie o desempenho dos modelos utilizando métricas apropriadas, como acurácia, precisão, recall, F1-score e matriz de confusão, com base nos resultados do conjunto de teste.
- Compare o desempenho dos modelos, utilizando as métricas de desempenho calculadas com o conjunto de teste.
- Discuta as vantagens e desvantagens de cada modelo e faça conclusões sobre a capacidade de ambos em prever se um passageiro sobreviveu ou não ao naufrágio do Titanic com base nos resultados.

### **Bônus:**

- Realize ajustes nos hiperparâmetros dos modelos para melhorar seu desempenho e compare novamente os resultados.

- Visualize a árvore de decisão gerada pelo modelo de Decision Tree para entender como as características são utilizadas nas decisões de classificação.