

# Mineração de Dados e Descoberta de Conhecimento

*Profs. Heitor S. Lopes e Thiago H. Silva (UTFPR, 2022) - Exercício #4*

## A) Classificação

1. **Objetivo:** explorar a classificação de dados com árvores de decisão utilizando validação cruzada, Knn e redes neurais
2. **Questões:**
  - 2- Considere ainda uma árvore de decisão para classificar se um indivíduo sobreviveu ou não com base no dataset "titanic.csv". Qual o resultado médio de acurácia e F1-score utilizando a estratégia de validação cruzada 5-fold? Discuta os resultados.
  - 3- Ainda considerando o dataset "titanic.csv", construa um modelo utilizando k-NN para prever se uma pessoa sobreviveu ou não. Considere diferentes valores de k vizinhos:  $k=\{2,3,4,5,6,7\}$ . Use validação cruzada 5-fold na avaliação. Houve variação significativa nos diferentes modelos testados? Algum deles foi melhor do que a estratégia baseada em árvore de decisão?
  - 4- Avalie o resultado de uma rede neural com, pelo menos, duas camadas escondidas e dez neurônios em cada. Use validação cruzada 5-fold na avaliação. Existe tendência de melhora em relação às tentativas anteriores?

## B) Regressão

3. **Dados:** use o dataset 'datasetCarros.csv' em todos os exercícios.
4. **Objetivo:** explorar os conceitos de regressão.
5. **Questões:**
  - 1
    - a. Faça um modelo de regressão linear simples utilizando a variável 'KmRodado' para prever a 'PrecoVenda'.
    - b. Calcule o  $R^2$  para o modelo criado.
  - 2
    - c. Separe o dataset em teste (5%) e treino (95%). Use o método 'train\_test\_split' do sklearn; configure o parâmetro random\_state=10.
    - d. Treine um modelo de regressão linear múltipla no dataset de treino utilizando todas

as variáveis (exceto 'Nome') para prever a 'PrecoVenda' e exiba os coeficientes do modelo.

- e. Avalie o modelo encontrado utilizando o dataset de teste. Calcule o  $R^2$  e MSE.