



**Universidade do Minho**

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2023/2024

Ficha prática nº 1

Fevereiro, 2024

**Tema**

Introdução à plataforma KNIME.

**Objetivos de aprendizagem**

Com a realização desta ficha prática pretende-se que os alunos:

- Conheçam a plataforma de análise de dados KNIME;
- Experimentem o KNIME na aplicação de técnicas de *machine learning*;

**Enunciado**

Descarregue e instale a plataforma KNIME ([knime.com/downloads](https://www.knime.com/downloads)), assim como o *dataset* disponível na plataforma de *e-learning* da Universidade do Minho, com dados sobre os passageiros do desastre Titanic.

Realize as tarefas seguintes:

- Use um nodo READER para carregar o ficheiro de dados para o ambiente de trabalho do KNIME;
- Procure os nodos necessários para:
  - Filtrar as colunas AGE, TICKET e CABIN
  - Transformar a coluna SURVIVED para *string*
  - Particionar os dados aleatoriamente, usando 75% para treino (e 25% para teste)
- Utilize um nodo DECISION TREE LEARNER para treinar uma árvore de decisão;
- Utilize um nodo DECISION TREE PREDICTOR para calcular as previsões do modelo obtido;
- Avalie o desempenho do modelo pela ACCURACY, usando um nodo SCORER e analisando a matriz de confusão resultante;
- Experimente diversas parametrizações do nodo DECISION TREE LEARNER, analise e documente os desempenhos alcançados;
- Desenvolva experiências com diferentes formas de preparação dos dados e analise as consequências.

Descrição do *dataset*  
TITANIC

ATRIBUTO	DESCRIÇÃO
<b>PassengerID</b>	Número de identificação do passageiro no <i>dataset</i>
<b>Survived</b>	Valor binário sobre se o passageiro sobreviveu (1 = SIM) ou não sobreviveu (0 = NÃO)
<b>Pclass</b>	Classe do passageiro (1 = 1ª classe; 2 = 2ª classe; 3 = 3ª classe)
<b>Name</b>	Nome do passageiro
<b>Sex</b>	Sexo do passageiro (male = masculino; female = feminino)
<b>Age</b>	Idade do passageiro
<b>SibSp</b>	Número de irmãos/cônjuges ( <i>siblings/spouses</i> ) a bordo
<b>Parch</b>	Número de pais/filhos ( <i>parents/children</i> ) a bordo
<b>Ticket</b>	Número do bilhete
<b>Fare</b>	Preço do bilhete (libra esterlina - £)
<b>Cabin</b>	Cabine
<b>Embarked</b>	Porto de embarque (S = Shouthampton; Q = Queenstown; C = Cherbourg)

Mais detalhes sobre os dados do problema podem ser encontrados nesta ligação: [kaggle.com/titanic](https://www.kaggle.com/titanic)