

Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2025/2026

Guião prático nº 1

Fevereiro, 2026

Tema

Introdução à plataforma KNIME

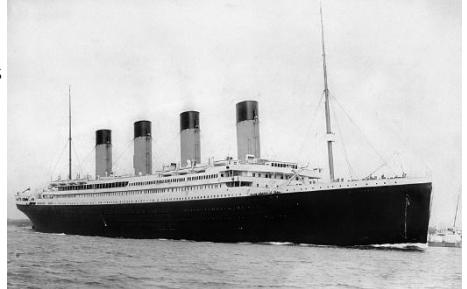
Objetivos de aprendizagem

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

- 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), assim como o *dataset* disponível na plataforma de *e-learning* da Universidade do Minho, com dados dos passageiros do desastre do navio cruzeiro Titanic.



Realize as tarefas seguintes:

1. Use um nodo READER para carregar o ficheiro de dados para o ambiente de trabalho do KNIME;
 - a. CSV READER (carregar dados através de ficheiros separados por vírgula ou ponto e vírgula);
 - b. EXCEL READER (carregar dados em folhas de cálculo MS EXCEL);
2. Procure os nodos necessários para:
 - a. Filtrar as colunas “AGE”, “TICKET” e “CABIN”;
 - b. Transformar a coluna “SURVIVED” para *string*;
 - c. Particionar os dados aleatoriamente, usando 75% para treino (e 25% para teste);
3. Utilize um nodo DECISION TREE LEARNER para treinar uma árvore de decisão:
 - a. Experimentar diferentes atributos em ROOT SPLIT;
4. Utilize um nodo DECISION TREE PREDICTOR para calcular as previsões do modelo obtido;
5. Avalie o desempenho do modelo, usando um nodo SCORER e analisando a matriz de confusão resultante:
 - a. *Accuracy*,
 - b. *Correct/wrong classified*,
6. Experimente diversas parametrizações do nodo DECISION TREE LEARNER, analise e documente os desempenhos alcançados:
 - a. *Quality measure*;
 - b. *Pruning method*;
 - c. *Min. number of records per node*;
7. Desenvolva experiências com diferentes formas de preparação dos dados e analise as consequências:
 - a. As árvores de decisão resultante;
 - b. O desempenho do modelo.

Descrição do dataset TITANIC

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

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