



Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2023/2024

Ficha prática nº 7

Março, 2024

**Tema**

Aplicação de técnicas de aprendizagem com KNIME: Árvores de Decisão.

**Objetivos de aprendizagem**

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

- Apliquem nodos de aprendizagem e previsão;
- Usem nodos de avaliação de modelos;

**Enunciado**

O problema descrito pelos dados do *dataset* «bcw» (*breast cancer wisconsin*) é composto por 11 atributos (*features*), dos quais «class» é o *target*.

Em síntese, os dados referem-se a diversas características que oncologistas especialistas em cancro de mama avaliaram em termos da sua relevância.

Realize as tarefas seguintes::

T1. Carregar o *dataset* «bcw» e usar nodos de exploração de dados, analisando:

- a) Tendências centrais;
- b) Dispersão estatística;
- c) Correlação entre *features*.

T2. Crie gráficos para visualizar as características dos dados;

T3. Tratar os dados em falta (*missing values*);

T4. Utilize um nodo DECISION TREE LEARNER para treinar uma árvore de decisão e um nodo DECISION TREE PREDICTOR para calcular as previsões do modelo obtido;

T5. Avalie o desempenho dos modelos criados através de diferentes métricas (ACCURARY, PRECISION, entre outras);

T6. Experimente diversas parametrizações do nodo DECISION TREE LEARNER, analise e documente os desempenhos alcançados;

T7. Experimente outras formas de preparação dos dados e analise as consequências:

- a) Selecionar *features* mais relevantes;
- b) Fazer *pruning*.

Descrição do *dataset*  
BCW

ATRIBUTO	DESCRIÇÃO
<b>idCode</b>	Número de código de amostra
<b>Clump_Thickness</b>	Espessura do aglomerado [1..10]
<b>Uniformity_of_Cell_Size</b>	Uniformidade do tamanho da célula [1..10]
<b>Uniformity_of_Cell_Shape</b>	Uniformidade da forma da célula [1..10]
<b>Marginal_Adhesion</b>	Adesão marginal [1..10]
<b>Single_Epithelial_Cell_Size</b>	Tamanho de célula epitelial [1..10]
<b>Bare_Nuclei</b>	Núcleos [1..10]
<b>Bland_Chromatin</b>	Cromatina [1..10]
<b>Normal_Nucleoli</b>	Nucléolos [1..10]
<b>Mitoses</b>	Mitoses [1..10]
<b>Class</b>	Classe (2 = benigno, 4 = maligno)

Mais detalhes sobre estes dados podem ser encontrados nesta ligação: [ics.uci.edu/dataset/15/breast+cancer](https://ics.uci.edu/dataset/15/breast+cancer)