



Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2024/2025

Guião prático nº 8

Março, 2025

Tema

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

Objetivos de aprendizagem

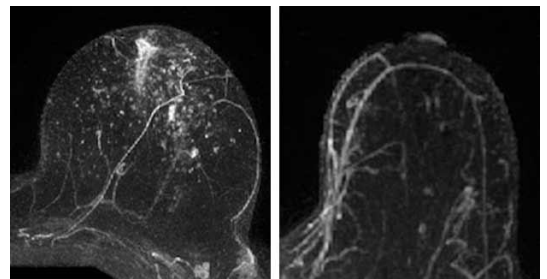
Com a realização desta ficha prática pretende-se que os estudantes:

- 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 imagiológicas que oncologistas especialistas em cancro de mama avaliaram em termos da sua relevância para a identificação de sinais cancerígenos.



Realizar as tarefas seguintes:

1. Carregar o *dataset* «bcw» e usar nodos de exploração de dados, analisando:
 - a. Tendências centrais;
 - b. Dispersão estatística;
 - c. Correlação entre *features*;
2. Criar gráficos para visualizar as características dos dados;
3. Tratar os dados em falta (*missing values*);
4. Utilizar um nodo DECISION TREE LEARNER para treinar uma árvore de decisão e um nodo DECISION TREE PREDICTOR para calcular as previsões do modelo obtido;
5. Avaliar o desempenho dos modelos criados através de diferentes métricas (ACCURARY, PRECISION, entre outras);
6. Experimente diversas parametrizações do nodo DECISION TREE LEARNER, analise e documente os desempenhos alcançados;
7. 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
idCode	Número de código de amostra
Clump_Thickness	Espessura do aglomerado [1..10]
Uniformity_of_Cell_Size	Uniformidade do tamanho da célula [1..10]
Uniformity_of_Cell_Shape	Uniformidade da forma da célula [1..10]
Marginal_Adhesion	Adesão marginal [1..10]
Single_Epithelial_Cell_Size	Tamanho de célula epitelial [1..10]
Bare_Nuclei	Núcleos [1..10]
Bland_Chromatin	Cromatina [1..10]
Normal_Nucleoli	Nucléolos [1..10]
Mitoses	Mitoses [1..10]
Class	Classe (2 = benigno, 4 = maligno)

Mais detalhes sobre estes dados podem ser encontrados neste *link*: archive.ics.uci.edu/dataset/15/breast+cancer