

### Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes 3° Ano, 2° Semestre Ano letivo 2023/2024

Ficha prática nº 10 Abril, 2024

### Tema

## Objetivos de aprendizagem

Aplicação de técnicas de aprendizagem com KNIME: Segmentação/ Clustering.

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

- Apliquem nodos de aprendizagem não supervisionada, de segmentação;
- Usem nodos de avaliação de modelos;

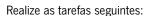
### Enunciado

O problema descrito pelos dados do *dataset* «iris» respeita a um conjunto de informações obtidas de flores "iris" que se distinguem em 3 espécies: "setosa", "versicolor" e "virginica".

Os dados registados no *dataset* incluem a identificação de cada instância de uma flor através do comprimento e largura das pétalas e das sépalas, para as 3 espécies identificadas.

O problema incide na construção de modelos suportados por paradigmas de aprendizagem sem supervisão, usando técnicas de segmentação (*clustering*) com vista à aplicação dos algoritmos k-means e k-medoids para identificar o tipo de flor iris.

Neste contexto, a aplicação de técnicas de segmentação deve descartar a utilização do atributo classificador.



- T1. Carregar o dataset «iris» e aplicar nodos de exploração, preparação e tratamento de dados;
- T2. Decidir sobre o conhecimento representado nas colunas «id» e «class» e agir em conformidade;
  - T2.1. Quais destas colunas devem ser removidas? Porquê?
- T3. Aplicar o nodo K-MEANS para construir um modelo de aprendizagem não supervisionada, para classificar cada caso de estudo como «iris-setosa», «iris-versicolor» ou «iris-virginica» (*number of clusters* = 3);
  - T3.1. O que acontece se criar modelos com 2 *clusters*? E com 4? E com 5?
- T4. Aplicar nodos de visualização (COLOR MANAGER e SCATTER PLOT) para representar graficamente os diferentes casos de estudo e respetivos *clusters* associados;
- T5. Aplicar o nodo CLUSTER ASSIGNER para inferir sobre os dados de teste utilizando o modelo treinado no nodo K-MEANS.
- T6. Aplicar o nodo RULE ENGINE para adequar o nome dos *clusters* atribuídos ("cluster\_X") ao respetivo nome da espécie da flor (coluna "class");
  - T6.1. Qual a necessidade de realizar esta tarefa?
- T7. Avalie o desempenho dos modelos de aprendizagem obtidos com K-MEANS treinados em T3 usando matrizes de confusão e métricas de desempenho.
- T8. Aplicar o nodo K-MEDOIDS para realizar estudo semelhante ao anterior e comparar os resultados.
- T9. Como se comparam os modelos criados nesta ficha prática (segmentação) com os desenvolvidos na ficha 6 (árvores de decisão)?



# Descrição do *dataset* IRIS

ATRIBUTO	DESCRIÇÃO
id	Identificador do registo de dados
sepal_length	Comprimento da sépala, em cm
sepal_width	Largura da sépala, em cm
petal_length	Comprimento da pétala, em cm
petal_width	Largura da pétala, em cm
Class	Classificação da flor
	Iris Setosa
	Iris Versicolr
	Iris Virginica

Mais detalhes sobre estes dados podem ser encontrados nesta ligação: ics.uci.edu/dataset/53/iris