



Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2024/2025

Guião prático nº 16

Abril, 2025

Tema

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

Objetivos de aprendizagem

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.



Realizar as tarefas seguintes:

1. Carregar o *dataset* «iris» e aplicar nodos de exploração, preparação e tratamento de dados;
2. Decidir sobre o conhecimento representado nas colunas «id» e «class» e agir em conformidade;
 - a. Quais destas colunas devem ser removidas? Porquê?
3. 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);
 - a. O que acontece se criar modelos com 2 *clusters*? E com 4? E com 5?
4. Aplicar nodos de visualização (COLOR MANAGER e SCATTER PLOT) para representar graficamente os diferentes casos de estudo e respetivos *clusters* associados;
5. Aplicar o nodo CLUSTER ASSIGNER para inferir sobre os dados de teste utilizando o modelo treinado no nodo K-MEANS.
6. Aplicar o nodo RULE ENGINE para adequar o nome dos *clusters* atribuídos (“cluster_X”) ao respetivo nome da espécie da flor (coluna “class”);
 - a. Qual a necessidade de realizar esta tarefa?
7. Avalie o desempenho dos modelos de aprendizagem obtidos com K-MEANS usando matrizes de confusão e métricas de desempenho.
8. Aplicar o nodo K-MEDOIDS para realizar estudo semelhante ao anterior e comparar os resultados.
9. Como se comparam os modelos criados neste guião prático (segmentação) com os desenvolvidos no guião 7 (árvores de decisão)?

Descrição do *dataset* IRIS

ATRIBUTO	DESCRIÇÃO
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 neste *link*: ics.uci.edu/dataset/53/iris