



**Universidade do Minho**

Licenciatura em Engenharia Informática

Mestrado Integrado em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º/4º ano, 2º Semestre

Ano letivo 2022/2023

Enunciado Prático n.º 9

Abril, 2023

<b>Tema</b>	Aplicação de técnicas de aprendizagem com KNIME: Segmentação/ <i>Clustering</i>
<b>Objetivos de aprendizagem</b>	Com a realização desta ficha prática pretende-se que os alunos: <ul style="list-style-type: none"><li>• Apliquem nodos de aprendizagem não supervisionada e previsão;</li><li>• Usem nodos de avaliação de modelos;</li></ul>
<b>Enunciado</b>	<p>Os dados incluídos no <i>dataset</i> [iris_data.csv] descrevem características sobre as sépalas e as pétalas de uma flor denominada «iris» (<a href="https://archive.ics.uci.edu/ml/datasets/iris">archive.ics.uci.edu/ml/datasets/iris</a>).</p> <p>Os dados dizem respeito ao comprimento e à largura tanto da sépala como das pétalas da flor. Existem 3 tipos de flores iris: «iris setosa», «iris versicolor» e «iris virginica».</p> <p>O problema incide na construção de modelos suportados por paradigmas de aprendizagem sem supervisão, usando técnicas de segmentação (<i>clustering</i>) com vista à aplicação dos algoritmos k-means e k-medoids para identificar o tipo de flor iris.</p>
<b>Tarefas</b>	<p>Descarregar o ficheiro disponível na plataforma de <i>e-learning</i> da UMinho, secção [Conteúdo] e realizar as tarefas seguintes:</p> <p>T1. Carregar o <i>dataset</i> [iris_data_csv] e aplicar nodos de exploração, preparação e tratamento de dados;</p> <p>T2. Decidir sobre o conhecimento representado nas colunas «id» e «class» e agir em conformidade;</p> <p>T2.1. Quais destas colunas devem ser removidas? Porquê?</p> <p>T3. Aplicar o nodo k-means para treinar um modelo de aprendizagem não supervisionada, para classificar cada caso de estudo como «iris-setosa», «iris-versicolor» ou «iris-virginica» (<i>number of clusters</i> = 3);</p> <p>T3.1. O que acontece se criar modelos com 2 <i>clusters</i>? E com 4? E com 5?</p> <p>T4. Aplicar nodos de visualização (Color Manager e Scatter Plot) para representar graficamente os diferentes casos de estudo e respetivos <i>clusters</i> associados;</p> <p>T5. Aplicar o nodo Cluster Assigner para inferir sobre os dados de teste utilizando o modelo treinado no nodo k-means.</p> <p>T6. Aplicar o nodo Rule Engine para adequar o nome dos <i>clusters</i> atribuídos ("cluster_X") ao respetivo nome da espécie da flor (coluna "class");</p> <p>T6.1. Qual a necessidade de realizar esta tarefa?</p> <p>T7. Avalie o desempenho dos modelos de aprendizagem obtidos com k-means treinados em T3 usando matrizes de confusão e métricas de desempenho.</p> <p>T8. Aplicar o nodo k-medoids para realizar estudo semelhante ao anterior e comparar os resultados.</p>