



## Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2024/2025

Guião prático nº 10

Março, 2025

### Tema

Aplicação de técnicas de aprendizagem com KNIME: Regressão.

### 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* «wine quality» é composto por 13 *features*, dos quais «quality1\_10» e «quality\_bin» são os atributos *target*, representando, ambos, uma apreciação da qualidade do vinho.

Os dados descrevem dois problemas idênticos («winequality-red», qualidade de vinhos tintos; «winequality-white», qualidade de vinhos brancos), no sentido em que são caracterizados por atributos com o mesmo significado, sendo as avaliações sobre a sua qualidade resultado de uma apreciação individual e subjetiva de especialistas.

Os dados são disponibilizados em 3 ficheiros:

- Os ficheiros «\*train» devem ser utilizados para criar modelos;
- Os ficheiros «\*test» podem ser usados para testar os modelos desenvolvidos;
- Os ficheiros «\*new» contêm problemas novos sem informação da qualidade dos vinhos.

Realizar as tarefas seguintes:

1. Carregar os *datasets* «winequality-red» e fazer a exploração de dados;
2. Fazer a preparação e o tratamento de dados;
3. Utilizar o nodo LINEAR REGRESSION LEARNER para criar modelos de regressão e o nodo REGRESSION PREDICTOR para calcular as previsões dos modelos;
  - a. Avaliar o desempenho dos modelos através de diferentes métricas (MAE, MSE, RMSE, entre outras);
4. Utilizar o nodo LOGISTIC REGRESSION LEARNER para criar modelos de classificação e o nodo LOGISTIC REGRESSION PREDICTOR para calcular as previsões dos modelos;
  - a. Avaliar o desempenho dos modelos através de diferentes métricas (ACCURARY, PRECISION, entre outras);
5. Carregar o *dataset* «winequality-white» e desenvolver abordagem idêntica ao *dataset* «winequality-red»;
  - a. É **possível** usar os modelos do vinho tinto para prever a qualidade do vinho branco?
  - b. É **admissível** usar os modelos do vinho tinto para prever a qualidade do vinho branco?
6. Avaliar a possibilidade de fazer JOIN dos dois *datasets*:
  - a. É **admissível** proceder deste modo para realizar uma análise coerente do problema da qualidade dos vinhos, independentemente do tipo de vinho (branco/tinto)?
  - b. É **possível** tratar os dois problemas deste modo? Se sim, como? Se não, porquê?



## Descrição do *dataset* WINE-QUALITY

ATRIBUTO	DESCRIÇÃO
<b>fixed acidity</b>	Acidez fixa
<b>volatile acidity</b>	Acidez volátil
<b>citric acid</b>	Ácido cítrico
<b>residual sugar</b>	Açúcar residual
<b>chlorides</b>	Cloretos
<b>free sulfur dioxide</b>	Dióxido de enxofre livre
<b>total sulfur dioxide</b>	Dióxido de enxofre total
<b>density</b>	Densidade
<b>pH</b>	pH
<b>sulfates</b>	Sulfatos
<b>alcohol</b>	Álcool
<b>quality1_10</b>	Qualidade do vinho [1..10]
<b>quality_bin</b>	Qualidade do vinho (normal = normal, good = bom)

Mais detalhes sobre estes dados podem ser encontrados neste *link*: [archive.ics.uci.edu/dataset/186/wine+quality](https://archive.ics.uci.edu/dataset/186/wine+quality)