



Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2024/2025

Guião prático nº 11

Março, 2025

Tema

Utilização do KNIME para resolução de problemas.

Objetivos de aprendizagem

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

- Utilizem modelos de aprendizagem na tomada de decisão;
- Calculuem soluções para problemas novos;

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.

Utilize os (melhores) modelos criados no Guião 10 para fazer a previsão da qualidade do novo conjunto de vinhos (colheita nova) descrito nos *datasets* «wine quality new».

Tratando-se de *datasets* que representam conjuntos de 100 problemas novos (“red wine” e “white wine”), não há informação sobre o *output* respetivo.

1. Para cada modelo criado para resolver cada um dos problemas (classificação, regressão, “red wine”, “white wine”, diferentes técnicas, variadas parametrizações), criar um ficheiro .CSV com o resultado da previsão contendo uma única coluna com a designação “prediction red” ou “prediction white” e as 100 soluções propostas para cada um dos casos dos *datasets*.
2. Para o problema de classificação com os *datasets* “red wine” e “white wine”, utilizar o [link bit.ly/4bM9WGP](https://bit.ly/4bM9WGP) para fazer o *upload* do ficheiro de resultados:
 - a. O ficheiro deverá ser submetido em formato .CSV;
 - b. O nome do ficheiro deverá ser o número do estudante, p.ex., “a123456.csv”;
 - c. Após a submissão, a equipa docente informará sobre as métricas de desempenho calculadas para os resultados submetidos;
 - d. Os estudantes que pretendam entrar na competição por um bónus na classificação final, deverão enviar por correio eletrónico para o docente do turno PL o *workflow* KNIME do qual resultou o ficheiro de resultados submetido durante a respetiva aula PL.



Descrição do *dataset* WINE-QUALITY

ATRIBUTO	DESCRIÇÃO
fixed acidity	Acidez fixa
volatile acidity	Acidez volátil
citric acid	Ácido cítrico
residual sugar	Açúcar residual
chlorides	Cloretos
free sulfur dioxide	Dióxido de enxofre livre
total sulfur dioxide	Dióxido de enxofre total
density	Densidade
pH	pH
sulfates	Sulfatos
alcohol	Álcool
quality1_10	Qualidade do vinho [1..10]
quality_bin	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