



Universidade do Minho

Licenciatura em Engenharia Informática

Aprendizagem e Decisão Inteligentes

3º Ano, 2º Semestre

Ano letivo 2023/2024

Ficha prática nº 7

Março, 2024

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 treinar/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.

Realize as tarefas seguintes:

T1. Carregar os *datasets* «winequality-red» e fazer a exploração de dados;

T2. Fazer a preparação e o tratamento de dados;

T3. Utilizar o nodo LINEAR REGRESSION LEARNER para criar modelos de regressão e o nodo REGRESSION PREDICTOR para calcular as previsões dos modelos;

T3.1. Avaliar o desempenho dos modelos criados através de diferentes métricas (MAE, MSE, RMSE, entre outras);

T4. Utilizar o nodo LOGISTIC REGRESSION LEARNER para criar modelos de classificação e o nodo LOGISTIC REGRESSION PREDICTOR para calcular as previsões dos modelos;

T4.1. Avaliar o desempenho dos modelos criados através de diferentes métricas (ACCURARY, PRECISION, entre outras);

T5. Carregar o *dataset* «winequality-white» e desenvolver abordagem idêntica ao *dataset* «winequality-red»;

T5.1. É **possível** usar os modelos do vinho tinto para prever a qualidade do vinho branco?

T5.2. É **admissível** usar os modelos do vinho tinto para prever a qualidade do vinho branco?

T6. Avaliar a possibilidade de fazer JOIN dos dois *datasets*:

- É admissível proceder deste modo para realizar uma análise coerente do problema da qualidade dos vinhos, independentemente do tipo de vinho (branco/tinto)?
- É possível tratar os dois problemas deste modo? Se sim, como? Se não, porquê?

Descrição dos
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 nesta ligação: ics.uci.edu/dataset/186/wine+quality