



Curso: Mestrado Integrado em Engenharia Biomédica
U.C.: Sistemas de Aprendizagem e Extração do Conhecimento

Ficha de Exercícios 02	
Docente:	Hugo Peixoto José Machado
Tema:	Introdução ao Weka
Turma:	PL
Ano Letivo:	2020-2021 – 2º Semestre
Duração da aula:	2 horas

1. Enunciado

Para a execução da presente ficha deverá proceder o *download* e instalação da ferramenta Weka. De seguida, e para facilitar a resolução dos exercícios seguintes, deverá extrair/copiar a pasta “Data” para um local de fácil acesso.

[1] Abrir o Weka / Explorer e carregar o data set “*contact-lens.arff*”. Com este dataset carregado responda às seguintes questões:

- [a] Quantas instâncias (registos) tem este dataset?
- [b] Quantos atributos (colunas) tem este dataset?
- [c] Quantos e quais os valores possíveis para o atributo “*age*”?
- [d] Quais os valores possíveis para o atributo “*contact-lens*”?
- [e] Qual o atributo que tem “*reduced*” como um dos valores?

[2] Abrir o Weka/Explorer e carregar o data set “*iris.arff*”. Com este dataset carregado responda às seguintes questões:

- [a] Quantas instâncias registos tem este data set?
- [b] Quantos atributos (colunas) tem este data set?
- [c] A classe “*iris-setosa*” tende a ter maiores ou menores valores de “*sepal.length*”?
- [d] A classe “*iris-virginica*” tende a ter maiores ou menores valores de “*petal.width*”?
- [e] Qual destes atributos, sozinho, parece dar uma melhor indicação da “*class*”?

[3] Abrir o Weka/Explorer e carregar o data set “*weather.nominal.arff*”. Com este dataset carregado responda às seguintes questões:

- [a] Identificar quais os atributos deste data set?
- [b] A utilização de um algoritmo de classificação poderá trazer conhecimento específico através dos dados apresentados.

Indique um objetivo que possa ser atingido com a aplicação de algoritmos de classificação, quando o mesmo for executado em dados semelhantes mas previamente desconhecidos.

[4] Abrir o Weka / Explorer e carregar o data set “*glass.arff*”. Com este dataset carregado responda às seguintes questões:

- [a] Abrir o separador “Classify” e escolher o algoritmo J48 (“trees”)
- [b] Observar a “*Confusion Matrix*” e indicar quais as maiores falhas no processo de classificação.
- [c] Qual o número de “*headlamps*” que foram classificadas como “*build wind float*”?
- [d] Qual o número de instâncias classificadas corretamente como “*vehic wind non-float*”?
- [e] Qual o número de instâncias classificadas corretamente como “*vehic wind float*”?

[f] Na lista de resultados obtidos clique com o botão direito e selecione “*Visualize tree*”. Copiar os resultados para a ficha de solução e descrever sucintamente o processo de classificação do algoritmo.



[5] Abrir o Weka / Explorer e carregar o data set "*labor.arff*". Com este dataset carregado responda às seguintes questões:

[a] Correr o algoritmo de classificação J48 com os parâmetros por defeito. Indicar a percentagem de instâncias corretamente classificadas.

[b] Utilizando somente 2 casas decimais, abra a configuração do algoritmo J48 e coloque a opção "unpruned" a "True". Corra novamente a classificação e indique a percentagem de instâncias corretamente classificadas.

[6] Abrir o Weka / Explorer e carregar novamente o data set "*glass.arff*". Com este dataset carregado responda às seguintes questões:

[a] Retirar o atributo "Fe". Qual o resultado da classificação?

[b] Retirar todos excepto "Ri", "Mg". Qual o resultado da classificação?