

FICHA 2

- 1) Abrir o Weka / Explorer e carregar o data set "contact-lens.arff". Com este data set carregado responda às seguintes questões:
 - a) Quantas instâncias (registos) tem este data set?
R.: 24
 - b) Quantos atributos (colunas) tem este data set?
R.: 5
 - c) Quantos e quais os valores possíveis para o atributo "age"?
R.: 3. Young, pre-presbyopic, presbyopic
 - d) Quais os valores possíveis para o atributo "contact-lens"?
R.: soft, hard, none
 - e) Qual o atributo que tem "reduced" como um dos valores?
R.: tear-prod-rate

- 2) Abrir o Weka/Explorer e carregar o data set "iris.arff". Com este data set carregado responda às seguintes questões:
 - a) Quantas instâncias registos tem este data set?
R.: 150
 - b) Quantos atributos (colunas) tem este data set?
R.: 5
 - c) A classe "iris-setosa" tende a ter maiores ou menores valores de "sepal.length"?
R.: menores
 - d) A classe "iris-viginica" tende a ter maiores ou menores valores de "petal.width"?
R.: maiores
 - e) Qual destes atributos, sozinho, parece dar uma melhor indicação da "class"?
R.: petallength

- 3) Abrir o Weka/Explorer e carregar o data set "weather.nominal.arff". Com este data set carregado responda às seguintes questões:
 - a) Identificar quais os atributos deste data set?
R.: Outlook, temperature, humidity, windy, play
 - 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.
R.: Neste caso específico, saber se é possível ir jogar, dadas as condições meteorológicas.

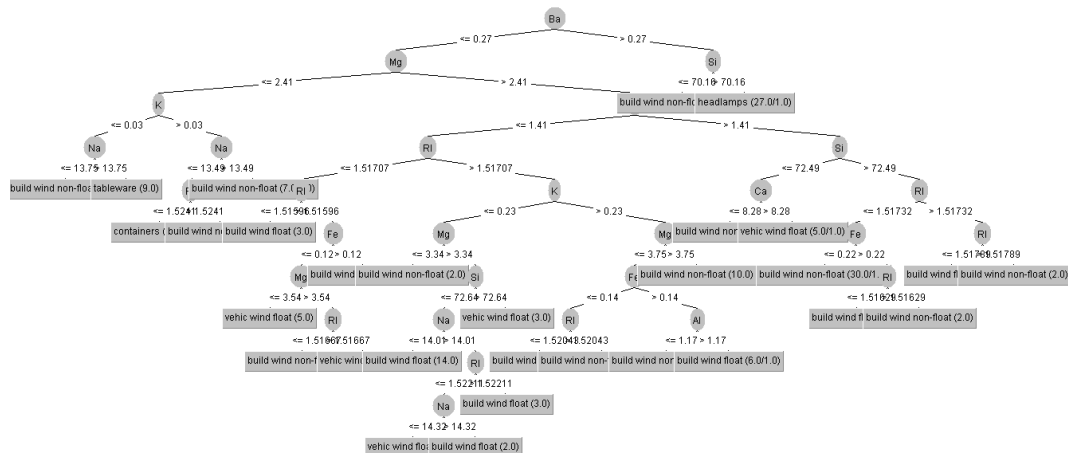
- 4) Abrir o Weka / Explorer e carregar o data set "glass.arff". Com este data set carregado responda às seguintes questões:
 - a) Abrir o separador "Classify" e escolher o algoritmo J48 ("trees")
R.: Algoritmo executado com opções por default e com cross validation a 10 folds

- b) Observar a “Confusion Matrix” e indicar quais as maiores falhas no processo de classificação.
R.: As maiores falhas encontram-se 1) na classificação de 'a' como 'b' e de 'b' como 'a'.

=== Confusion Matrix ===

| | a | b | c | d | e | f | g | |
|---|----|----|---|---|----|---|----|--------------------------|
| a | 50 | 15 | 3 | 0 | 0 | 1 | 1 | a = build wind float |
| b | 16 | 47 | 6 | 0 | 2 | 3 | 2 | b = build wind non-float |
| c | 5 | 5 | 6 | 0 | 0 | 1 | 0 | c = vehic wind float |
| d | 0 | 0 | 0 | 0 | 0 | 0 | 0 | d = vehic wind non-float |
| e | 0 | 2 | 0 | 0 | 10 | 0 | 1 | e = containers |
| f | 1 | 1 | 0 | 0 | 0 | 7 | 0 | f = tableware |
| g | 3 | 2 | 0 | 0 | 0 | 1 | 23 | g = headlamps |

- c) Qual o número de “headlamps” que foram classificadas como “build wind float”?
R.: 3
- d) Qual o número de instâncias classificadas corretamente como “vehic wind non-float”?
R.: 0
- e) Qual o número de instâncias classificadas corretamente como “vehic wind float”?
R.: 6
- f) Na lista de resultados obtidos clicar com o botão direito e selecionar “Visualizetree”. Copiar os resultados para a ficha de solução e descrever sucintamente o processo de classificação do algoritmo.
R.: O algoritmo divide os valores (numéricos, neste caso) dos atributos, organizando-os numa árvore de decisão. Percorrendo os ramos, é possível tomar uma decisão quando um novo caso surgir, através dos valores (atributos) desse novo caso.



- 5) Abrir o Weka / Explorer e carregar o data set “labor.arff”. Com este data set carregado responda às seguintes questões:
- a) Correr o algoritmo de classificação J48 com os parâmetros por defeito. Indicar a percentagem de instâncias corretamente classificadas.
R.: 73.6842%

- 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.
R.: 78.9474%

- 6) Abrir o Weka / Explorer e carregar novamente o data set "glass.arff". Com este data set carregado responda às seguintes questões:

- a) Retirar o atributo "Fe". Qual o resultado da classificação?
R.: 67.2897% de acertos, 32.7103% de instâncias mal classificadas.

- b) Retirar todos excepto "Ri", "Mg". Qual o resultado da classificação?
R.: 68.6916% de acertos, 31.3084% de instâncias mal classificadas.