

Descoberta de Conhecimento - FE02

Vasco Ramos, PG42852

1.
 - a. 24
 - b. 5 (contudo 4 são atributos e 1 é a classe)
 - c. Existem 3 valores possíveis: young, pre-presbyopic, presbyopic
 - d. soft, hard, none
 - e. Tear-prod-rate
2.
 - a. 150
 - b. 5
 - c. Tende a ter menores valores
 - d. Tende a ter maiores valores
 - e. petal.width
3.
 - a. outlook, temperature, humidity, windy, play
 - b. O objetivo da aplicação de algoritmos de classificação é, essencialmente, criar um modelo que seja capaz de, ao ser fornecido com novos dados sem o mapeamento para a classe que lhes pertence, classificar com relativo (preferivelmente, elevado) sucesso para a respetiva classe. Parte-se do pressuposto que esses novos dados que são passados ao modelo não foram usados na fase de teste.
4.
 - a. Done :)
 - b.

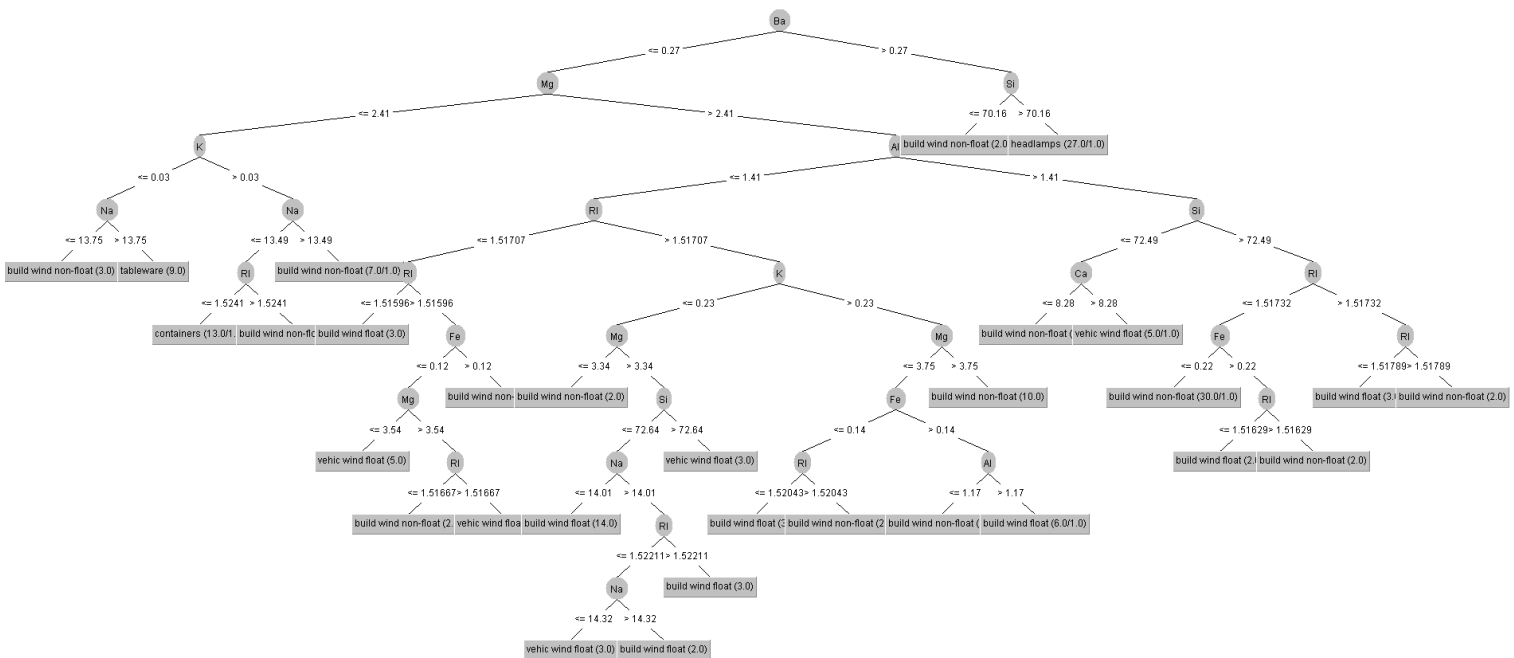
```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  <-- classified as
50 15  3  0  0  1  1 |  a = build wind float
16 47  6  0  2  3  2 |  b = build wind non-float
 5  4  7  0  0  1  0 |  c = vehic wind float
 0  0  0  0  0  0  0 |  d = vehic wind non-float
 0  2  0  0 10  0  1 |  e = containers
 1  1  0  0  0  7  0 |  f = tableware
 3  2  0  0  0  1 23 |  g = headlamps
```

A falha mais notória foi na inconsistência da classificação, tendo em conta que na identificação da classe c houve bastantes erros, provavelmente por haver um baixo número de exemplos, contudo a classe e, por exemplo, também tem um baixo número de exemplos (menos que a classe c) e foi identificada com muito maior sucesso.

- c. 3
- d. 0
- e. 6

f.



O algoritmo aplica uma árvore de decisão em que são avaliadas as propriedades pela sua ordem de importância na classificação. No exemplo em concreto, o atributo BA é o primeiro a ser analisado, consoante o seu valor vai analisar o atributo MG (se BA ≤ 0.27) ou o atributo SI (se BA > 0.27). Se for para o MG, a próxima decisão vai ser entre K (se MG ≤ 2.41) ou Al (se MG > 2.41). Se, por outro lado, de BA foi para SI, a próxima decisão vai ser entre build wind non-float (se SI ≤ 70.16) ou headlamps (se SI > 70.16). O restante processo de decisão toma um comportamento semelhante e vai prosseguindo este algoritmo de decisão até que percorra a árvore de forma a obter uma classificação para os valores introduzidos.

5.

- 73.6842 %
- 78.9474 %

6.

a.

```

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      145          67.757 %
Incorrectly Classified Instances    69          32.243 %
Kappa statistic                    0.5591
Mean absolute error                0.0996
Root mean squared error            0.2806
Relative absolute error            47.0303 %
Root relative squared error        86.4631 %
Total Number of Instances         214

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,657    0,160    0,667    0,657    0,662    0,499    0,797    0,695    build wind float
0,697    0,210    0,646    0,697    0,671    0,480    0,793    0,620    build wind non-float
0,353    0,030    0,500    0,353    0,414    0,379    0,813    0,375    vehic wind float
?         0,000    ?         ?         ?         ?         ?         ?         vehic wind non-float
0,769    0,005    0,909    0,769    0,833    0,827    0,876    0,660    containers
0,778    0,029    0,538    0,778    0,636    0,629    0,931    0,565    tableware
0,793    0,022    0,852    0,793    0,821    0,795    0,869    0,738    headlamps
Weighted Avg.    0,678    0,134    0,681    0,678    0,676    0,548    0,817    0,641

=== Confusion Matrix ===

 a b c d e f g <-- classified as
46 19 3 0 0 1 1 | a = build wind float
14 53 3 0 1 3 2 | b = build wind non-float
 5  5 6 0 0 1 0 | c = vehic wind float
 0  0 0 0 0 0 0 | d = vehic wind non-float
 0  2 0 0 10 0 1 | e = containers
 1  1 0 0 0 7 0 | f = tableware
 3  2 0 0 0 1 23 | g = headlamps

```

b.

```

Correctly Classified Instances      144          67.2897 %
Incorrectly Classified Instances    70          32.7103 %
Kappa statistic                    0.5428
Mean absolute error                0.1128
Root mean squared error            0.2755
Relative absolute error            53.2717 %
Root relative squared error        84.893 %
Total Number of Instances         214

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
0,800    0,146    0,727    0,800    0,762    0,639    0,860    0,704    build wind float
0,776    0,203    0,678    0,776    0,724    0,559    0,832    0,730    build wind non-float
0,000    0,020    0,000    0,000    0,000    -0,041    0,563    0,085    vehic wind float
?         0,000    ?         ?         ?         ?         ?         ?         vehic wind non-float
0,231    0,045    0,250    0,231    0,240    0,193    0,780    0,230    containers
0,444    0,024    0,444    0,444    0,444    0,420    0,924    0,458    tableware
0,759    0,016    0,880    0,759    0,815    0,791    0,860    0,629    headlamps
Weighted Avg.    0,673    0,127    0,632    0,673    0,650    0,541    0,824    0,615

=== Confusion Matrix ===

 a b c d e f g <-- classified as
56 12 2 0 0 0 0 | a = build wind float
11 59 1 0 3 1 1 | b = build wind non-float
 9  8 0 0 0 0 0 | c = vehic wind float
 0  0 0 0 0 0 0 | d = vehic wind non-float
 0  4 0 0 3 4 2 | e = containers
 0  1 0 0 4 4 0 | f = tableware
 1  3 1 0 2 0 22 | g = headlamps

```