



UNIVERSIDADE DO MINHO
DEPARTAMENTO DE INFORMÁTICA
DESCOBERTA DE CONHECIMENTO

Ficha 2

22 de Fevereiro de 2019

Francisco Oliveira (A78416)



Conteúdo

1	Exercícios	2
1.1	Exercício 1	2
1.2	Exercício 2	2
1.3	Exercício 3	2
1.4	Exercício 4	3
1.5	Exercício 5	5
1.6	Exercício 6	7

1 Exercícios

1.1 Exercício 1

- A - 24 instâncias;
- B - 5 atributos;
- C - 3 valores: *young*, *pre-presbyopic* e *presbyopic*;
- D - *soft*, *hard* e *none*;
- E - O atributo *tear-prod-rate*;

1.2 Exercício 2

- A - 150 instâncias;
- B - 5 atributos;
- C - Menores valores de “*sepal.length*”;
- D - Maiores valores de “*petal.width*”;
- E - *Petal.width*;

1.3 Exercício 3

- A - Atributos: *outlook*, *temperature*, *humidity*, *windy* (e *play*);
- B - Prever as condições para decidir se o tempo é propício a jogar ou não, lá fora;

1.4 Exercício 4

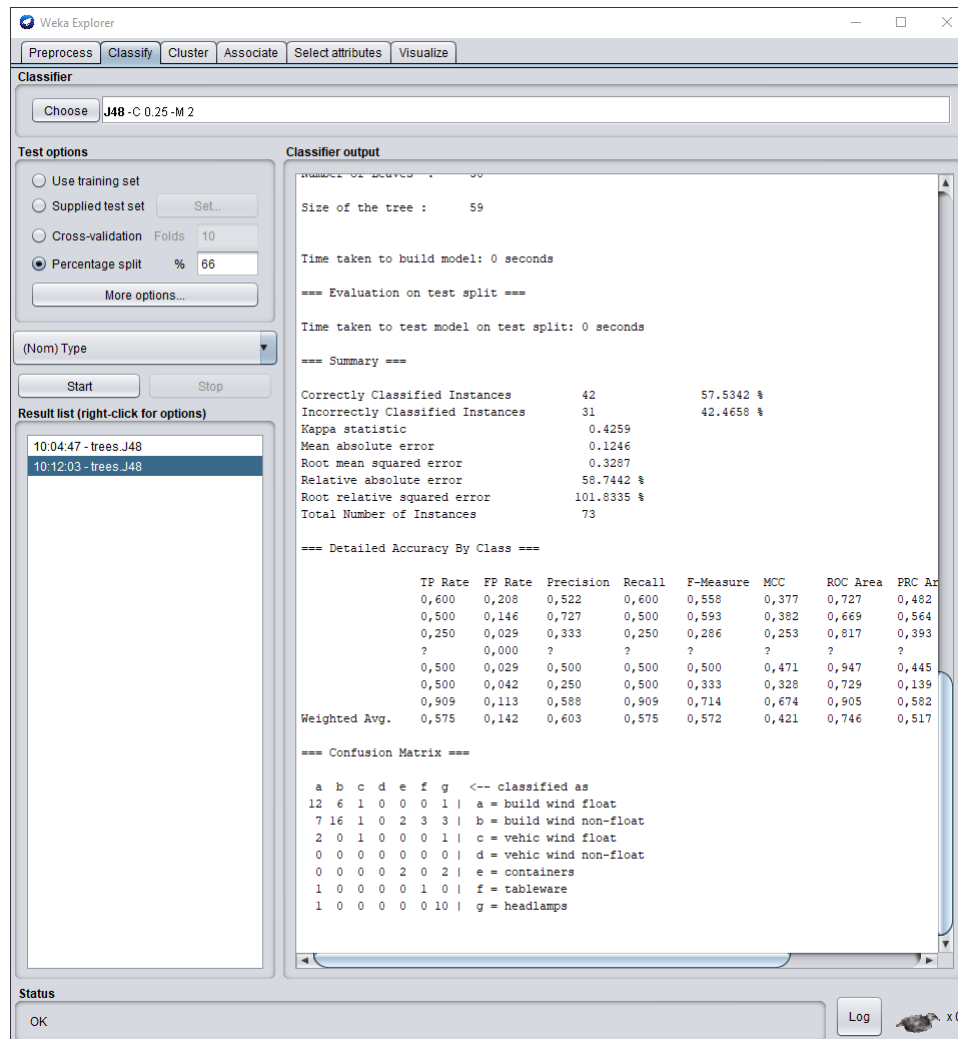


Figura 1: Exercício 4 - Output

Pela matriz de confusão é possível verificar que o processo de classificação apresenta maiores falhas quando classifica *build wind float* e *build wind non-float*, já que classificou 6 instâncias *build wind float* como *build wind non-float*, e classificou 7 instâncias *build wind non-float* como *build wind float*.

1 instância real *headlamps* foi classificada como *build wind float*.

Não existe nenhuma instância *vehic wind non-float* detetada corretamente.

Apenas uma instância *vehic wind float* foi classificada corretamente.

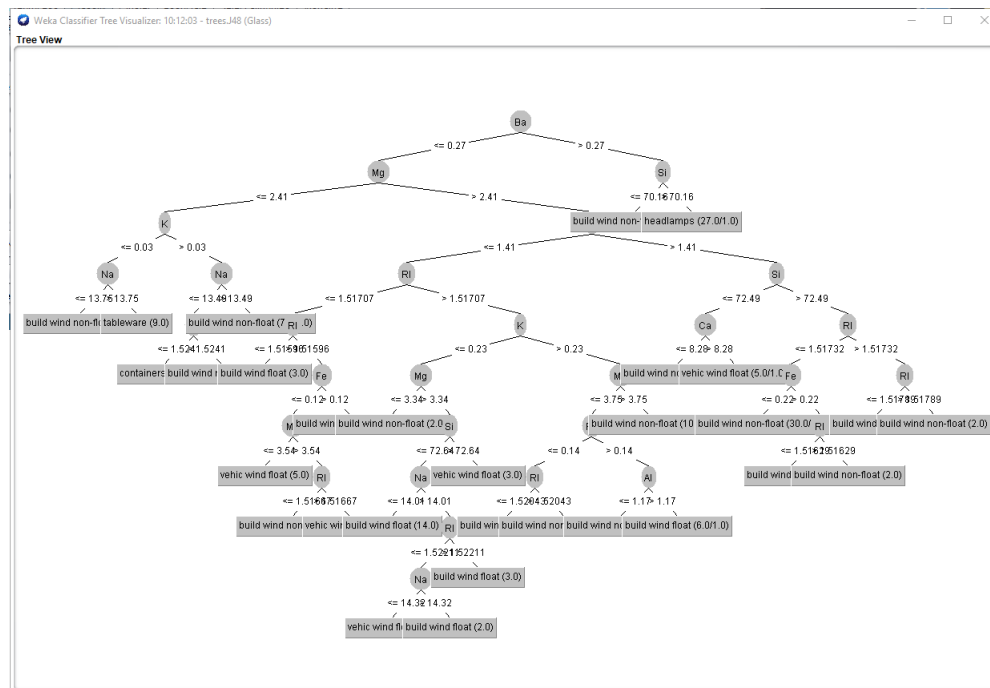


Figura 2: Exercício 4.F - Output

O algoritmo cria uma árvore de decisão, onde cada ramo é uma condição. O processo de classificação compara os valores dos atributos de cada instância com as condições dos ramos e continua até atingir uma folha. As folhas correspondentes às classes de cada instância.

1.5 Exercício 5

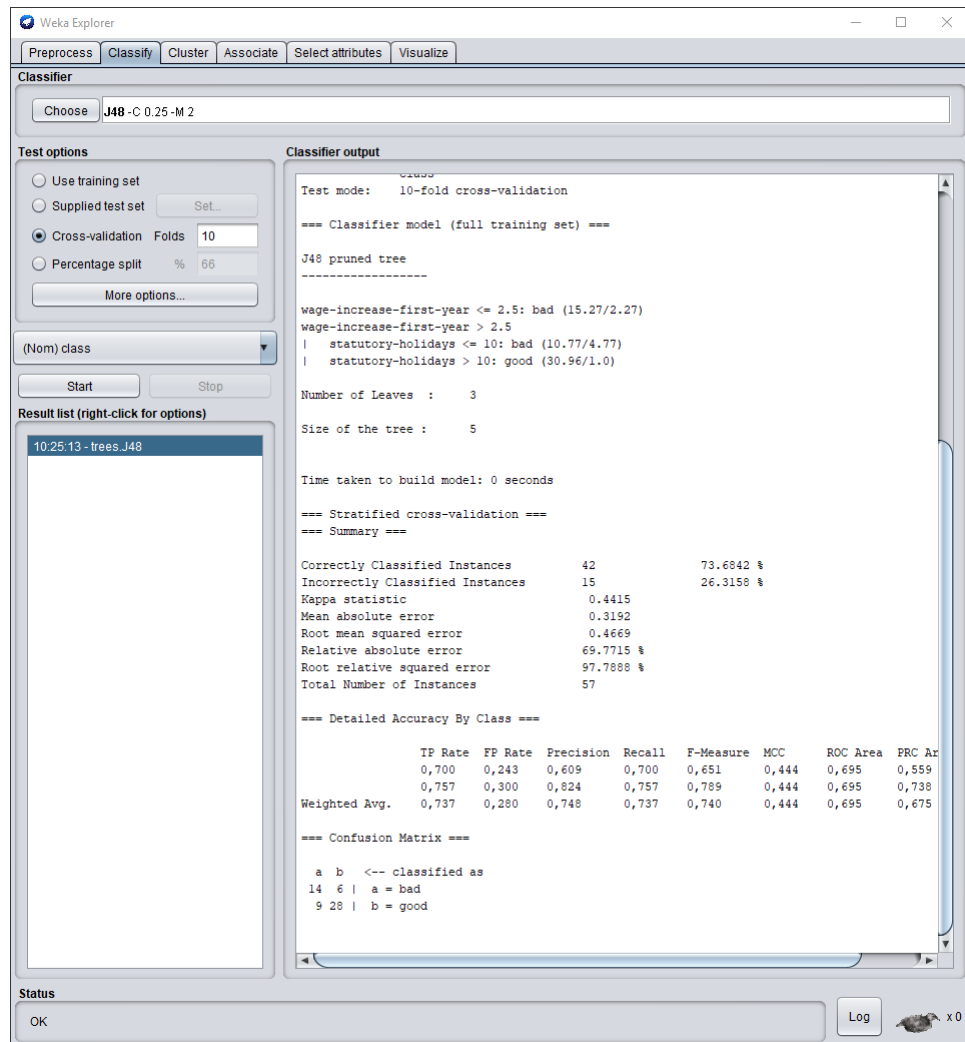


Figura 3: Exercício 5.A - Output

J48 Tree pruned: Cerca de 73.6842% das instâncias (42 de 57) foram classificadas corretamente.

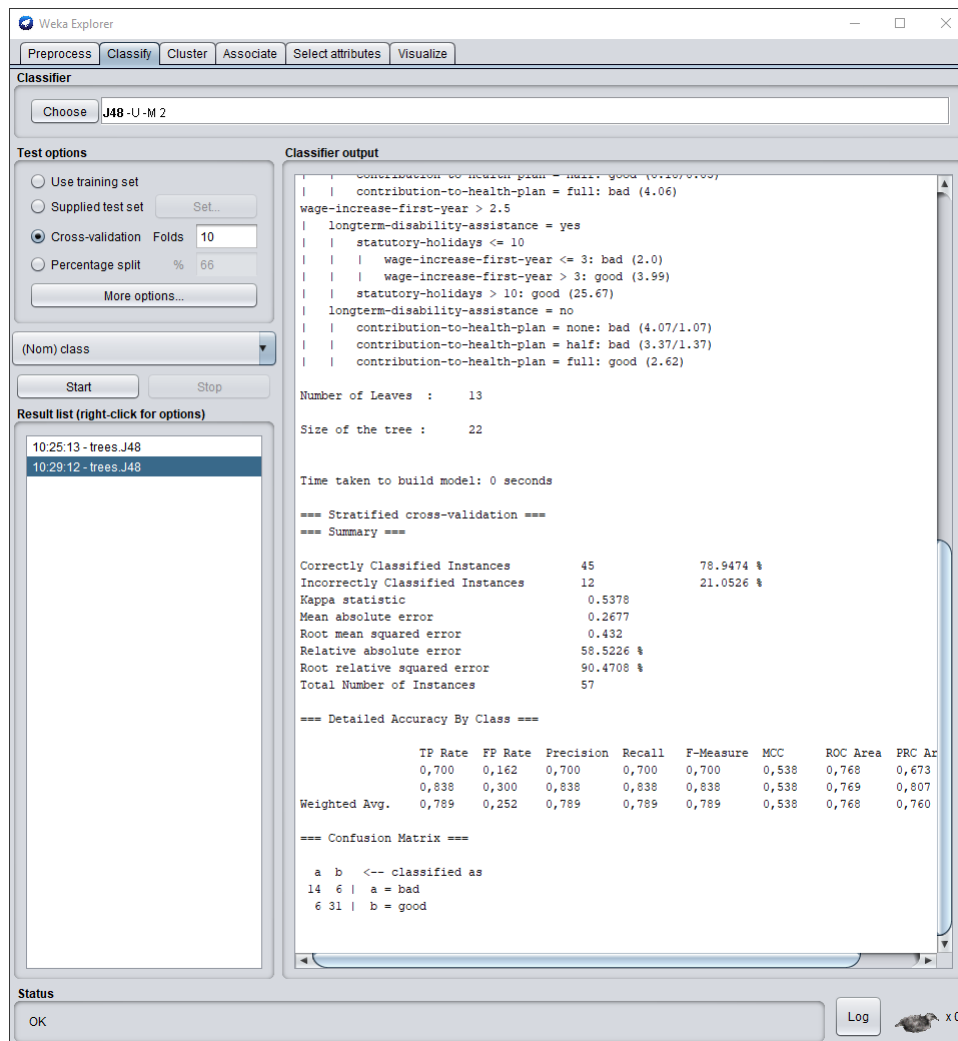


Figura 4: Exercício 5.B - Output

J48 Tree unpruned: Cerca de 78.9474% das instâncias (45 de 57) foram classificadas corretamente (**overfitting**).

1.6 Exercício 6

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) Type

Start Stop

Result list (right-click for options)

10:36:35 - trees.J48

Classifier output

```

J48 <= 70.16: build wind non float (27.0/1.0)
| Si > 70.16: headlamps (27.0/1.0)

Number of Leaves :    26
Size of the tree :    51

Time taken to build model: 0 seconds

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

Correctly Classified Instances      144          67.2897 %
Incorrectly Classified Instances    70           32.7103 %
Kappa statistic                    0.5519
Mean absolute error                 0.1029
Root mean squared error             0.285
Relative absolute error             48.5797 %
Root relative squared error         87.8206 %
Total Number of Instances          214

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Ar
0,657   0,160   0,667   0,657   0,662   0,499   0,790   0,678
0,697   0,217   0,639   0,697   0,667   0,471   0,786   0,617
0,294   0,030   0,455   0,294   0,357   0,323   0,799   0,344
?        0,000   ?         ?         ?         ?         ?         ?
0,769   0,005   0,909   0,769   0,833   0,827   0,876   0,656
0,778   0,029   0,538   0,778   0,636   0,629   0,931   0,565
0,793   0,022   0,852   0,793   0,821   0,795   0,866   0,738
Weighted Avg.   0,673   0,136   0,674   0,673   0,670   0,541   0,811   0,632

=== 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  6  5  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

```

Status

OK Log x 0

Figura 5: Exercício 6.A - Classificação sem atributo Fe

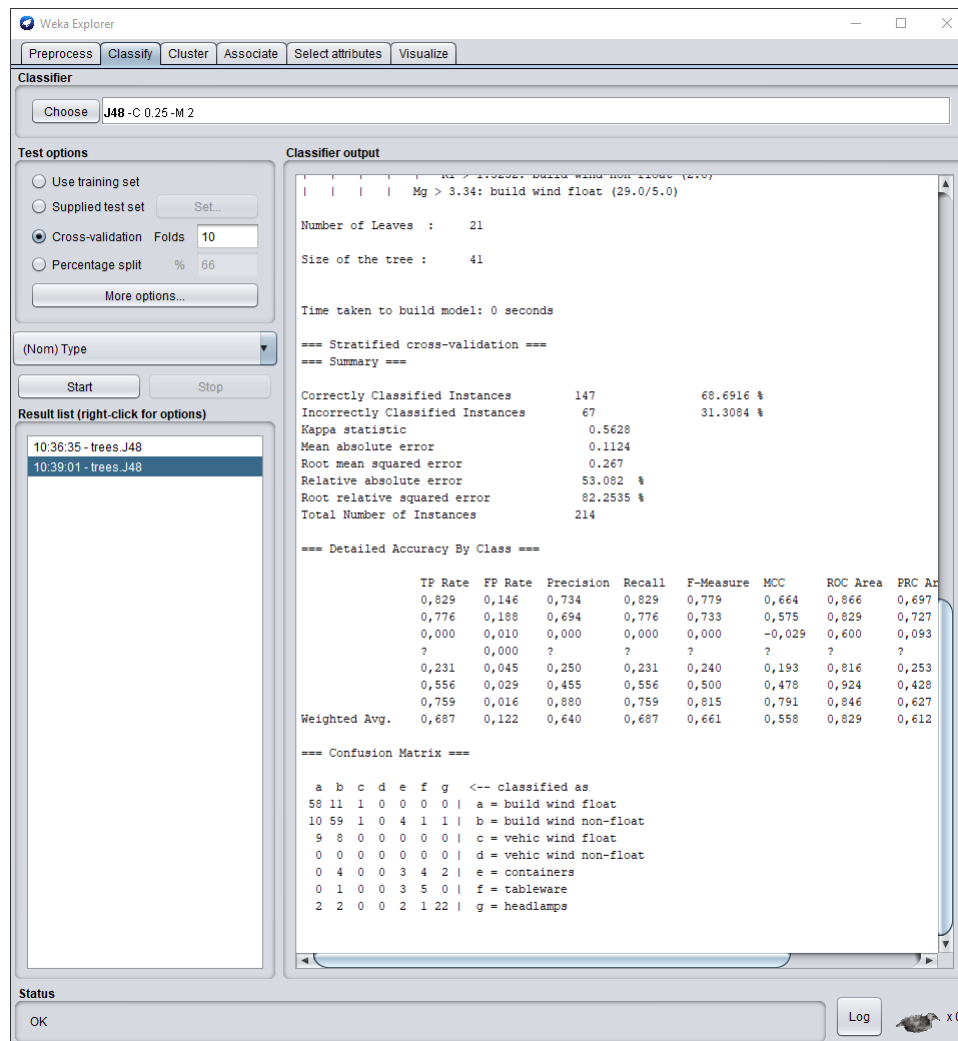


Figura 6: Exercício 6.B - Classificação só com atributos Ri e Mg

A classificação foi melhor apenas com os atributos Ri e Mg, pois para as instâncias usadas como teste (que são as mesmas que as de treino) obteve-se resultados melhores apenas com estes atributos. Isto, pois ocorre o processo de *overfitting*. Já o processo de classificação com os outros atributos gera um modelo mais generalizado para outros casos para além dos de treino, sendo possível então obter melhores resultados para outros testes sem ser os próprios casos de treino.