

**EA072 — EFC 2 - Questão 06****Professor:** Fernando José Von Zuben

Leonardo Rodrigues Marques 178610

3. Existem 5 atributos numéricos e 2 atributos categóricos.

Atributos Numéricos	Estatísticas	Valor
temperatura_media(numérico)	minimum	12.82
	maximum	21.98
	mean	18.448
	StdDev	2.965
umidade_media(numérico)	minimum	68.38
	maximum	97.83
	mean	90.306
	StdDev	7.247
altura_chuva_mensal(numérico)	minimum	9
	maximum	273.7
	mean	107.779
	StdDev	71.769
precip_21_dias(numérico)	minimum	0
	maximum	216.9
	mean	85.179
	StdDev	62.745
numero_dias_chuva(numérico)	minimum	3
	maximum	15
	mean	8.857
	StdDev	3.823

Atributos Categóricos	Rótulo	Contagem	Peso
local	ID	14	14
	IM	14	14
	PM	14	14
	CHI	14	14
	GAL	14	14
class	0	11	11
	1	59	59

7. Matriz de Confusão, Porcentagem de Corretude, Árvore de Decisão.

Matriz de Confusão é uma matriz que apresenta a quantidade de amostras classificadas erroneamente e corretamente para cada uma das classes (acuidade).

```

=== Confusion Matrix ===

  a  b  <-- classified as
 8  3  |  a = 0
 3 56  |  b = 1

```

Figura 1: Matriz de Confusão.

Correctly Classified Instances	64	91.4286 %
Incorrectly Classified Instances	6	8.5714 %
Kappa statistic	0.6764	
Mean absolute error	0.0929	
Root mean squared error	0.2958	
Relative absolute error	34.0726 %	
Root relative squared error	81.1177 %	
Total Number of Instances	70	

Figura 2: Porcentagem de Classificação Correta.

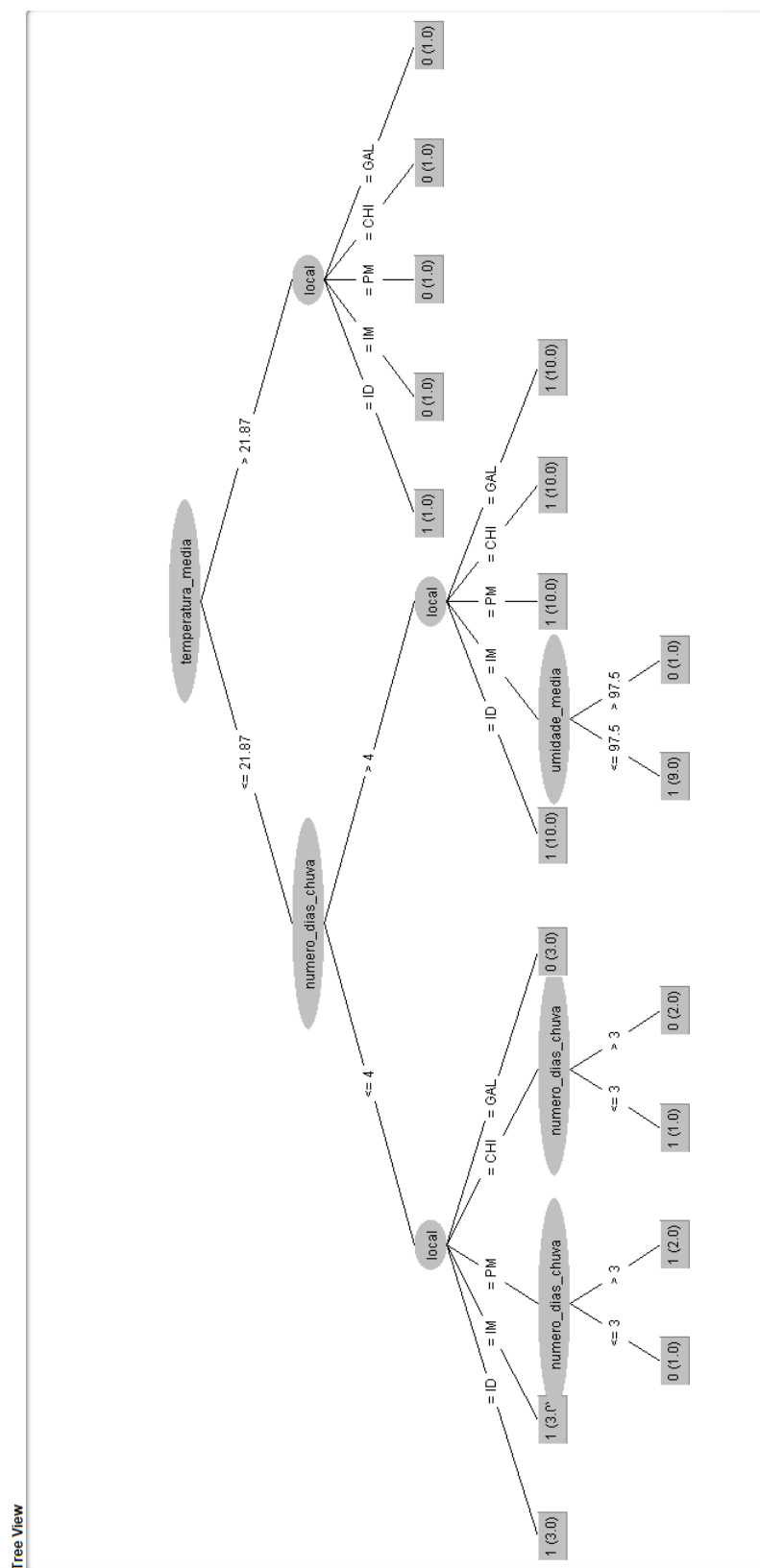


Figura 3: Árvore de Decisão em modo View.

```
1 J48 unpruned tree
2 -----
3
4 temperatura_media <= 21.87
5 |   numero_dias_chuva <= 4
6 | |   local = ID: 1 (3.0)
7 | |   local = IM: 1 (3.0)
8 | |   local = PM
9 | | |   numero_dias_chuva <= 3: 0 (1.0)
10 | | |   numero_dias_chuva > 3: 1 (2.0)
11 | |   local = CHI
12 | | |   numero_dias_chuva <= 3: 1 (1.0)
13 | | |   numero_dias_chuva > 3: 0 (2.0)
14 | |   local = GAL: 0 (3.0)
15 |   numero_dias_chuva > 4
16 | |   local = ID: 1 (10.0)
17 | |   local = IM
18 | | |   umidade_media <= 97.5: 1 (9.0)
19 | | |   umidade_media > 97.5: 0 (1.0)
20 | |   local = PM: 1 (10.0)
21 | |   local = CHI: 1 (10.0)
22 | |   local = GAL: 1 (10.0)
23 temperatura_media > 21.87
24 |   local = ID: 1 (1.0)
25 |   local = IM: 0 (1.0)
26 |   local = PM: 0 (1.0)
27 |   local = CHI: 0 (1.0)
28 |   local = GAL: 0 (1.0)
```

8.

```
1 R1: IF temperatura_media <= 21.87 AND numero_dias_chuva <= 4
2 AND local = ID then 1(3.0)
3
4 R2: IF temperatura_media <= 21.87 AND numero_dias_chuva <= 4
5 AND local = IM then 1(3.0)
6
7 R3: IF temperatura_media <= 21.87 AND numero_dias_chuva <= 4
8 AND local = PM AND numero_dias_chuva <= 3 then 0(1.0)
9
10 R4: IF temperatura_media <= 21.87 AND numero_dias_chuva <= 4
11 AND local = PM AND numero_dias_chuva > 3 then 1(2.0)
12
13 R5: IF temperatura_media <= 21.87 AND numero_dias_chuva <= 4
14 AND local = CHI AND numero_dias_chuva <= 3 then 1(1.0)
15
16 R6: IF temperatura_media <= 21.87 AND numero_dias_chuva <= 4
17 AND local = CHI AND numero_dias_chuva > 3 then 0(2.0)
18
19 R7: IF temperatura_media <= 21.87 AND numero_dias_chuva <= 4
20 AND local = GAL then 0(3.0)
21
22 R8: IF temperatura_media <= 21.87 AND numero_dias_chuva > 4
23 AND local = ID then 1(10.0)
24
25 R9: IF temperatura_media <= 21.87 AND numero_dias_chuva > 4
26 AND local = IM AND umidade_media <= 97.5 then 1(9.0)
27
28 R10: IF temperatura_media <= 21.87 AND numero_dias_chuva > 4
29 AND local = IM AND umidade_media > 97.5 then 0(1.0)
30
31 R11: IF temperatura_media <= 21.87 AND numero_dias_chuva > 4
32 AND local = PM then 1(10.0)
33
34 R12: IF temperatura_media <= 21.87 AND numero_dias_chuva > 4
35 AND local = CHI then 1(10.0)
```

```
36
37 R13: IF temperatura_media <= 21.87 AND numero_dias_chuva > 4
38 AND local = GAL then 1(10.0)
39
40 R14: IF temperatura_media > 21.87 AND local = ID then 1(1.0)
41
42 R15: IF temperatura_media > 21.87 AND local = IM then 0(1.0)
43
44 R16: IF temperatura_media > 21.87 AND local = PM then 0(1.0)
45
46 R17: IF temperatura_media > 21.87 AND local = CHI then 0(1.0)
47
48 R18: IF temperatura_media > 21.87 AND local = GAL then 0(1.0)
```

9. Ao alterarmos o parâmetro **numMinObj** de 1 para 2, observamos que houve uma diminuição de 91.4286% para 90% na classificação correta de instâncias, uma diminuição de 18 para 7 no número de folhas, e diminuição de 26 para 10 no tamanho da árvore. Esse parâmetro define o número mínimo de instâncias em uma folha.

Ao alterarmos o parâmetro **unpruned** de *true* para *false*, observamos que houve uma grande diminuição de 91.4286% para 87.1429% na classificação correta das instâncias, uma diminuição de 18 para 2 no número de folhas, e diminuição de 26 para 3 no tamanho da árvore. Esse parâmetro é uma ferramenta muito importante para eliminar informações irrelevantes ou ruidosas da árvore de decisão afim de otimizar o algoritmo. Entretanto em árvores pequenos, como a utilizada, essa poda pode retirar informações necessárias para aumento na classificação correta das instâncias.

10. Utilizamos uma base de dados[1] com 1728 instâncias e 7 atributos categóricos. Ao aplicar a poda, não conseguimos aumentar o desempenho de corretude classificatória. Entretanto ao mudar o parâmetro **numMinObj** de 1 para 2, obtemos uma árvore mais simplifica e com ganho aproximado de 0.7523% de corretude classificatória em relação a árvore original.

numMinObj 1 - Correctly Classified Instances 1614 93.4028

numMinObj 2 - Correctly Classified Instances 1627 94.1551 %

[1] <https://github.com/jaredly/decision-tree>