

- 1) Para um determinado dataset obteve-se as seguintes contagens de termos frequentes:  $\text{sup}(A) = 125$ ,  $\text{sup}(B) = 100$ ,  $\text{sup}(C) = 99$ ,  $\text{sup}(AC) = 45$ ,  $\text{sup}(BC) = 84$ ,  $\text{sup}(AB) = 54$ ,  $\text{sup}(ABC) = 45$ . Assumindo que queremos eliminar redundâncias, que regras devem ser derivadas para o consequente C (apresente todas as medidas de cada regra)? Justifique.
  
- 2) Considere a figura 2 onde estão representadas duas árvores de decisão sobre dataset "*diabetes*".
  - a. Explique sucintamente o processo que permite obter a segunda árvore tendo por base a primeira árvore.
  
  - b. Apresente as previsões nos dois modelos para o novo caso:  
 $\text{plas}=120$ ,  $\text{mass}=0$ ,  $\text{pres} = 50$ ,  $\text{age}=25$ ,  $\text{pedi}=1.0$ ,  $\text{preg}=8$ .
  
  - c. Apresente a expectativa de acerto (score) para a previsão anterior nos dois modelos. Quantos casos suportam esta decisão?

```

plas <= 127
| mass <= 26.4
| | preg <= 7: tested_negative (117.0/1.0)
| | preg > 7
| | | mass <= 0: tested_positive (2.0)
| | | mass > 0: tested_negative (13.0)
| mass > 26.4
| | age <= 28: tested_negative (180.0/22.0)
| | age > 28
| | | plas <= 99: tested_negative (55.0/10.0)
| | | plas > 99
| | | | pedi <= 0.561: tested_negative (84.0/34.0)
| | | | pedi > 0.561
| | | | | preg <= 6
| | | | | age <= 30: tested_positive (4.0)
| | | | | age > 30
| | | | | age <= 34: tested_negative (7.0/1.0)
| | | | | age > 34
| | | | | mass <= 33.1: tested_positive (6.0)
| | | | | mass > 33.1: tested_negative (4.0/1.0)
| | | | | preg > 6: tested_positive (13.0)
plas > 127
| mass <= 29.9
| | plas <= 145: tested_negative (41.0/6.0)
| | plas > 145
| | | age <= 25: tested_negative (4.0)
| | | age > 25
| | | | age <= 61
| | | | | mass <= 27.1: tested_positive (12.0/1.0)
| | | | | mass > 27.1
| | | | | pres <= 82
| | | | | | pedi <= 0.396: tested_positive (8.0/1.0)
| | | | | | pedi > 0.396: tested_negative (3.0)
| | | | | | pres > 82: tested_negative (4.0)
| | | | | age > 61: tested_negative (4.0)
| mass > 29.9
| | plas <= 157
| | | pres <= 61: tested_positive (15.0/1.0)
| | | pres > 61
| | | | age <= 30: tested_negative (40.0/13.0)
| | | | age > 30: tested_positive (60.0/17.0)
| | plas > 157: tested_positive (92.0/12.0)

```

Size of the tree : 43

Incorrectly Classified Instances 65 24.9042 %

Figura 2 (árvore 1)

```

plas <= 127
| mass <= 26.4: tested_negative (132.0/3.0)
| mass > 26.4
| | age <= 28: tested_negative (180.0/22.0)
| | age > 28
| | | plas <= 99: tested_negative (55.0/10.0)
| | | plas > 99
| | | | pedi <= 0.561: tested_negative (84.0/34.0)
| | | | pedi > 0.561
| | | | | preg <= 6
| | | | | age <= 30: tested_positive (4.0)
| | | | | age > 30
| | | | | age <= 34: tested_negative (7.0/1.0)
| | | | | age > 34
| | | | | mass <= 33.1: tested_positive (6.0)
| | | | | mass > 33.1: tested_negative (4.0/1.0)
| | | | | preg > 6: tested_positive (13.0)
plas > 127
| mass <= 29.9
| | plas <= 145: tested_negative (41.0/6.0)
| | plas > 145
| | | age <= 25: tested_negative (4.0)
| | | age > 25
| | | | age <= 61
| | | | mass <= 27.1: tested_positive (12.0/1.0)
| | | | mass > 27.1
| | | | | pres <= 82
| | | | | pedi <= 0.396: tested_positive (8.0/1.0)
| | | | | pedi > 0.396: tested_negative (3.0)
| | | | | pres > 82: tested_negative (4.0)
| | | | age > 61: tested_negative (4.0)
| mass > 29.9
| | plas <= 157
| | | pres <= 61: tested_positive (15.0/1.0)
| | | pres > 61
| | | | age <= 30: tested_negative (40.0/13.0)
| | | | age > 30: tested_positive (60.0/17.0)
| | plas > 157: tested_positive (92.0/12.0)

```

Size of the tree : 39

Incorrectly Classified Instances 62 23.7548 %

Figura 2 (árvore 2)