

## TRABALHO PRÁTICO: Descrição dos Datasets

### Dataset 1 – HEPATITE

Colunas: 14 atributos (números e texto)

Linhas do ficheiro Start: 10

Linhas do ficheiro Train: 600 – possui valores em falta (NA)

Linhas do ficheiro Test: 10

TARGET: classificação 0, 0s, 1, 2, 3 – coluna *Category*

Nos ficheiros *Start* e *Test* os atributos do tipo string foram convertidos da seguinte forma:

Atributo Category	
0=Blood Donor	0
0s=suspect Blood Donor	1
1=Hepatitis	2
2=Fibrosis	3
3=Cirrhosis	4

Atributo Sex	
m	0
f	1

### Dataset 2 – STROKE

Colunas: 12 atributos (números e texto)

Linhas do ficheiro Start: 10

Linhas do ficheiro Train: 620 – possui valores em falta (NA)

Linhas do ficheiro Test: 10

TARGET: classificação 1, 0 – coluna *Stroke*

Nos ficheiros *Start* e *Test* os atributos do tipo string foram convertidos da seguinte forma:

Atributo Smoking_status	
Never smoked	0
formerly smoked	1
Smokes	2
Unknown	3

Atributo Residence_type	
Rural	0
Urban	1

Atributo Gender	
Male	0
Female	1

Attribute ever_married	
No	0
Yes	1

### Dataset 3 – CIRROSE

Colunas: 18 atributos (números e texto)

Linhas do ficheiro Start: 10

Linhas do ficheiro Train: 408 – possui valores em falta (NA)

Linhas do ficheiro Test: 10

TARGET: classificação 1, 2, 3, 4 – coluna *Stage*

Nos ficheiros *Start* e *Test* os atributos do tipo string foram convertidos da seguinte forma:

Atributo Status	
C	0
CL	1
D	2

Atributo Drug	
Placebo	0
D-penicillamine	1

Atributo Sex	
M	0
F	1

Atributos Ascites e Edema	
N	0
Y	1