

PRÁCTICA 2. INTRODUCCIÓN A WEKA Y PYTHON

Descripción del conjunto de datos:

El conjunto de datos viene descrito por la siguiente función lógica:

$$(A \wedge B \vee (\neg C \wedge \neg D)) \text{ xor } E$$

El archivo de datos es generado a partir de una hoja de cálculo, y contiene cinco atributos binarios que son las entradas de la función (A, B, C, D, E) y la salida de la función (CLASE). Para generar los datos se introducen todas las combinaciones posibles de valores de los atributos y como aplicación de la función, obtenemos el valor binario de la clase que será 0 o 1.

Experimento en Weka:

Para realizar el experimento en Weka tenemos que aplicar una técnica de preprocesamiento. Realizamos la discretización de los atributos ya que el clasificador ID3 solo trabaja con atributos nominales y nuestro dataset está formado por valores numéricos.

Utilizaremos el conjunto de datos entero para entrenar y probar el clasificador ID3.

```
=== Classifier model (full training set) ===
```

```
Id3
```

```
E = False
```

```
| A = False
```

```
| | C = False
```

```
| | | D = False: True
```

```
| | | D = True: False
```

```
| | C = True: False
```

```
| A = True
```

```
| | B = False
```

```
| | | C = False
```

```
| | | | D = False: True
```

```
| | | | D = True: False
```

```
| | | C = True: False
```

```
| | B = True: True
```

```
E = True
```

```
| A = False
```

```
| | C = False
```

```
| | | D = False: False
```

```
| | | D = True: True
```

```
| | C = True: True
```

```
| A = True
```

```
| | B = False
```

```
| | | C = False
```

```
| | | | D = False: False
```

```
| | | | D = True: True
```

```
| | | C = True: True
```

```
| | B = True: False
```

=== Confusion Matrix ===

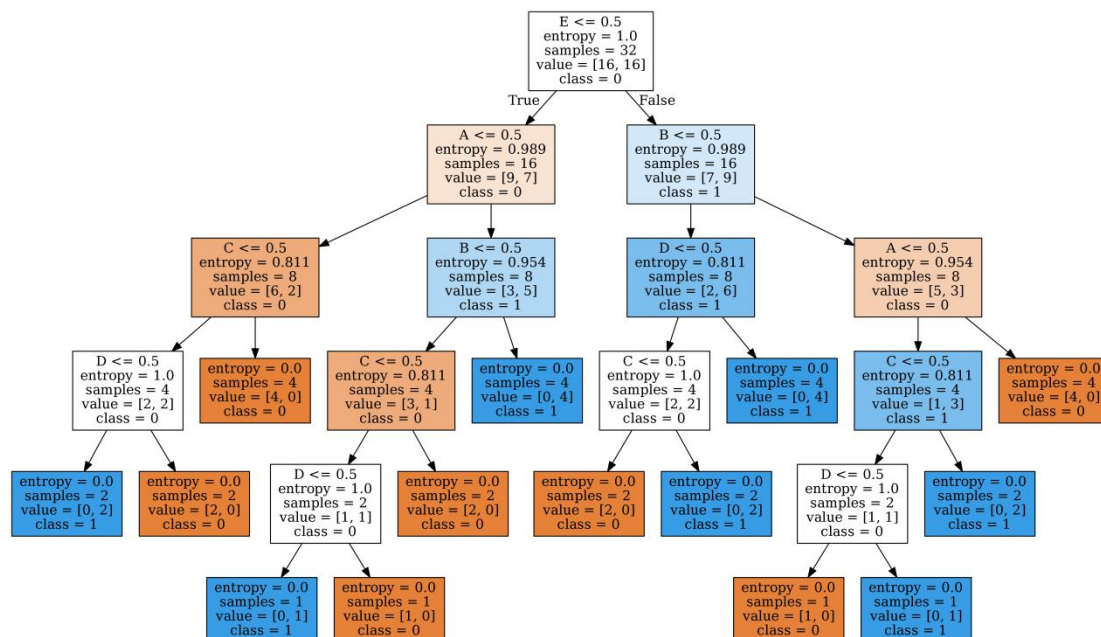
```
a b  <-- classified as
16 0 | a = True
0 16 | b = False
```

Correctly Classified Instances	32	100	%
Incorrectly Classified Instances	0	0	%

Experimento en Python:

No realizamos tratamiento de datos, simplemente cargamos el conjunto de datos generado.

Utilizamos todo el conjunto de datos (32 instancias) para el conjunto de entrenamiento del clasificador. Como queremos que la tasa de acierto sea del 100% y el conjunto de datos es pequeño, utilizaremos este para test.



Matriz de confusión:

```
[16 0]
[ 0 16]
```

Porcentaje de ejemplos mal clasificados: 0.0 %