

ÁRBOLES DE DECISIÓN ID3

Guillermo Siles Bonilla

Fernando Sánchez Muñoz

Francisco Molina Sánchez

Karen Andrea Pérez Velásquez

Modelo de clasificación también conocido como ID3 que significa "inducción mediante árboles de decisión", capaz de tomar decisiones con gran precisión.

Sistema de aprendizaje supervisado que aplica la estrategia "divide y vencerás" para hacer la clasificación, implementando métodos y técnicas para la realización de procesos inteligentes, representando así el conocimiento y el aprendizaje, con el propósito de automatizar tareas.

Fuente: Wikipedia

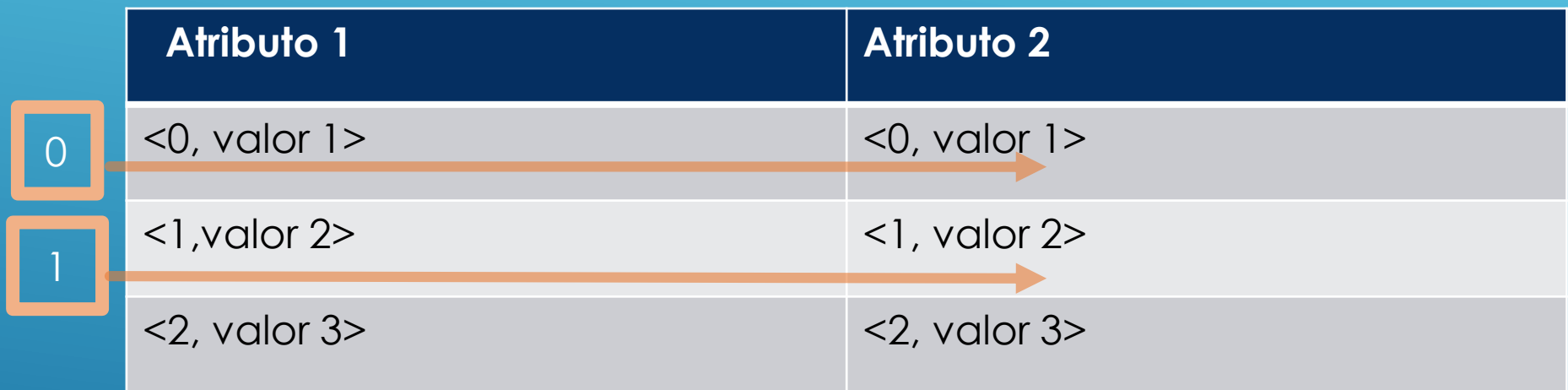
DEFINICIÓN

Como hemos visto en clase, es necesario contar con un fichero de datos para la evaluación de los mismo. Existen distintos formatos de entrada que nos permitirían analizar estos datos. Por ejemplo *xml*, *json*, *csv*.

En nuestro caso, hemos optado por utilizar archivos en formato CSV.

TOMA DE DATOS

La estructura seguida para los datos es un TreeMap, de manera que para cada conjunto <clave, valor>, donde la clave es el número de fila a la que pertenece y el valor es una estructura List con los valores de dicha fila.



Atributo 1	Atributo 2
<0, valor 1>	<0, valor 1>
<1, valor 2>	<1, valor 2>
<2, valor 3>	<2, valor 3>

ESTRUCTURA DE DATOS

Con las fórmulas vistas en clase, en esta parte del algoritmo se halla la entropía general del conjunto y la ganancia para cada uno de los atributos, pues se empezarán a tratar los datos a partir del atributo que más ganancia tenga.

Luego, seguimos tratando los datos, generando para cada una de las iteraciones un nuevo conjunto de elementos que consta de todos los atributos menos los que ya han sido tratados.

De esta manera se genera el árbol de decisión ID3.

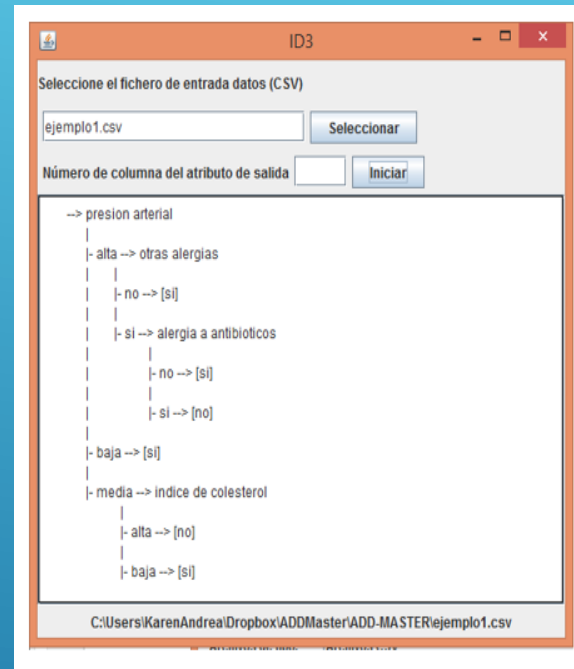
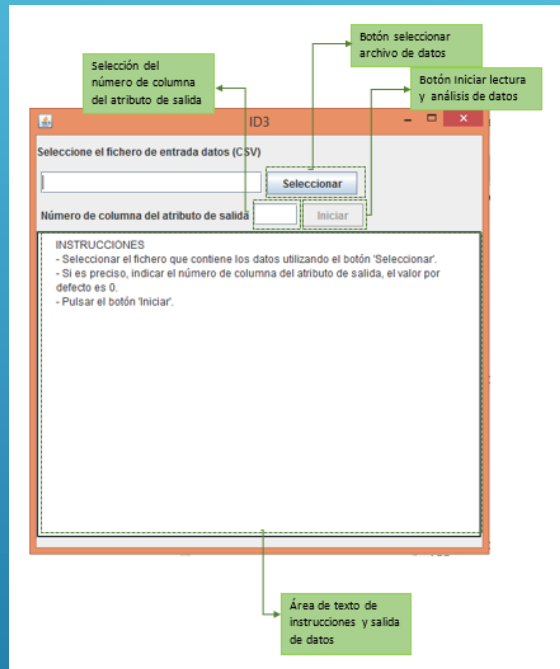
ANÁLISIS DE DATOS

Los datos generados por el programa se muestran en una estructura de árbol, como se puede ver en el ejemplo.

```
--> presion arterial
  | - alta --> otras alergias
  |   | - no --> [si]
  |   | - si --> alergia a antibioticos
  |     | - no --> [si]
  |     | - si --> [no]
  | - baja --> [si]
  | - media --> indice de colesterol
    | - alta --> [no]
    | - baja --> [si]
```

RESULTADOS

Para facilitar el uso y la comprensión de los resultados, hemos creado una interfaz gráfica sencilla.



INTERFAZ GRÁFICA

GRACIAS POR SU ATENCIÓN

