

# APRIORI-ID3-ARBOL.pdf



patrivc



Apuntes Variados



4º Grado en Ingeniería Informática



Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación  
Universidad de Granada



MÁSTER EN

## Inteligencia Artificial & Data Management

MADRID

Formamos  
**talento** para un futuro  
**Sostenible**

saber más



Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato  
→ Planes pro: más coins

perdo  
espacio



Necesito  
concentración

ali ali ooh  
esto con 1 coin me  
lo quito yo...

WUOLAH

APRIL 21

TID	Artículos
L1	Dátiles, Peras, Plátanos, Uvas
L2	Naranjas, Uvas, Peras, Kiwis
L3	Peras, Plátanos
L4	Peras, Kiwis, Uvas
L5	Plátanos, Naranjas, Uvas, Kiwis, Peras

Sopros y confidencia → 50%, 75%.

C1

~~Dátiles →  $\frac{1}{5} = 20\%$~~   
Peras →  $\frac{5}{5} = 100\%$   
Plátanos →  $\frac{3}{5} = 60\%$   
Uvas →  $\frac{4}{5} = 80\%$   
~~Naranjas →  $\frac{2}{5} = 40\%$~~   
Kiwis →  $\frac{3}{5} = 60\%$

L1

Peras  
Plátanos  
Uvas  
Kiwis

Tachamos los que no  
negan el 50%.

C2

Peras → Plátanos  $\frac{3}{5} = 60\%$   
Peras → Uvas  $\frac{4}{5} = 80\%$   
Peras → Kiwis 60%  
~~Plátanos → Uvas 40%~~  
~~Plátanos → Kiwis 20%~~  
Uvas → Kiwis 60%.

L2

Peras Plátanos  
Peras Uvas  
Peras Kiwis  
Uvas Kiwis

C3

~~Peras Plátanos Uvas 40%~~  
~~Peras Plátanos Kiwis 20%~~  
Peras Uvas Kiwis 60%.

L3

Peras Uvas Kiwis

• Reglas de asociación

C2

<del>Peras → Plátanos [60%, 60%]</del>	Plátanos → Peras [60%, 100%]
Peras → Uvas [80%, 80%]	Uvas → Peras [60%, 100%]
<del>Peras → Kiwis [60%, 60%]</del>	Kiwis → Peras [60%, 100%]
Uvas → Kiwis [60%, 75%]	Kiwis → Uvas [60%, 100%]

→  
si se aparece uvas, debe  
aparecer kiwis

C3

Peras Uvas → Kiwi [60%, 75%]  
Peras Kiwi → Uvas [60%, 100%]  
Uvas Kiwi → Peras [60%, 100%]

• Otra zona

Kiwi → Peras Uvas [60%, 100%]  
Uvas → Peras Kiwi [60%, 75%]  
~~Peras → Uvas Kiwi [60%, 60%]~~

WUOLAH

## ALGORITMO ID3 con INDICE GINI + ARBOL DE DECISION

Un árbol de decisión es un clasificador que en función de un conjunto de atributos permite determinar a que clase pertenece el caso objeto de estudio. Su estructura es la siguiente:

- **Hoja**: es una categoría (clase) de la variable objeto de la clasificación
- **Nodo**: es un nodo de decisión que especifica una prueba simple a realizar
- **Descendientes**: cada nodo son los posibles resultados de la prueba del nodo.

El proceso de generación de un árbol de decisión consiste en 2 fases:

1. **Construcción del árbol**: al principio todos los ejemplos de entrenamiento están en el nodo raíz. Se van dividiendo recursivamente en base a los atributos seleccionados.
2. **Poda del árbol**: identificar e quitar ramas que describen ruidos anómalos.

**ID3** selecciona el atributo con la mayor ganancia de información. Mide la capacidad predictiva de una clase. Está basado en la Entropía. Su esquema es el siguiente:

1. Selecciona el atributo A que maximice la ganancia  $G(S,A)$
2. Crea un nodo para ese atributo con tantos sucesos como valores tenga
3. Introduce los ejemplos en los sucesos según el valor que tenga el atributo A
4. Por cada sucesor:

- Si solo hay ejemplos de una clase,  $C_k$ , lo etiquetamos con  $C_k$
- Si no, llamar a ID3 con un conjunto de ejemplos formado por los ejemplos de ese nodo, eliminando la columna de A

$$GAIN(S,A) = Entropia(S) - \sum_{i=1}^{|S|} \frac{|S_i|}{|S|} Entropia(S_i)$$

### Ejercicio examen

Ganancia Información

$$\Delta = G(\text{padre}) - \sum_{\text{hijo}} \frac{N_i}{N} G(\text{Hijo}_i)$$

Gini

$$G = 1 - \sum_{i=1}^{\text{clases}} p(i|t)^2$$

Entropía

$$H = - \sum_{i=1}^{\text{clases}} p(i|t) \log_2(p(i|t))$$

Empezamos con el árbol. Vamos que atributo es el mejor para dividir el árbol:

$$G_{\text{base}} = 1 - \left(\frac{5}{10}\right)^2 - \left(\frac{5}{10}\right)^2 = 1 \quad (G_{\text{padre}})$$

Color

$$G_{\text{azul}} = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0.5$$

$$G_{\text{rojo}} = 1 - \left(\frac{0}{2}\right)^2 - \left(\frac{2}{2}\right)^2 = 0$$

$$G_{\text{naranja}} = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0.375$$

$$\Delta_{\text{color}} = G_{\text{padre}} - \frac{4}{10} \cdot G_{\text{azul}} - \frac{4}{10} \cdot G_{\text{rojo}} - \frac{4}{10} \cdot G_{\text{naranja}} = 1 - \left(\frac{4}{10} \cdot 0.5\right) - \left(\frac{4}{10} \cdot 0\right) - \left(\frac{4}{10} \cdot 0.375\right) = 0.15$$



# Imagínate aprobando el examen

## Necesitas tiempo y concentración

Planes	 PLAN TURBO	 PLAN PRO	 PLAN PRO+
 Descargas sin publi al mes	10 	40 	80 
 Elimina el video entre descargas			
 Descarga carpetas			
 Descarga archivos grandes			
 Visualiza apuntes online sin publi			
 Elimina toda la publi web			
 Precios <span>Anual <input type="checkbox"/></span>	0,99 € / mes	3,99 € / mes	7,99 € / mes

Ahora que puedes conseguirlo,  
¿Qué nota vas a sacar?



# WUOLAH

## Apuntes Variados



Banco de apuntes de la

WUOLAH



**Comparte estos flyers en tu clase y consigue más dinero y recompensas**

- 1** Imprime esta hoja
- 2** Recorta por la mitad
- 3** Coloca en un lugar visible para que tus compis puedan escanar y acceder a apuntes
- 4** Llévate dinero por cada descarga de los documentos descargados a través de tu QR



Altura

$$G_{BASO} = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0.5$$

$$G_{ALTO} = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0.5$$

$$\Delta_{ALTURA} = 0.5 - \left(\frac{4}{10} \cdot 0.5\right) - \left(\frac{6}{10} \cdot 0.5\right) = 0$$

RAYAS

$$G_{SI} = 1 - \left(\frac{4}{6}\right)^2 - \left(\frac{2}{6}\right)^2 = 0.44$$

$$G_{NO} = 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{3}{4}\right)^2 = 0.375$$

$$\Delta_{RAYAS} = 0.5 - \left(\frac{6}{10} \cdot 0.44\right) - \left(\frac{4}{10} \cdot 0.375\right) = 0.086$$

Textura

$$G_{SUAVE} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0$$

$$G_{VENOZA} = 1 - \left(\frac{3}{6}\right)^2 - \left(\frac{3}{6}\right)^2 = 0.5$$

$$G_{PELUDO} = 1 - \left(\frac{1}{8}\right)^2 - \left(\frac{7}{8}\right)^2 = 0.44$$

$$\Delta_{TEXTURA} = 0.5 - \left(\frac{1}{10} \cdot 0\right) - \left(\frac{9}{10} \cdot 0.44\right) - \left(\frac{3}{10} \cdot 0.5\right) = 0.066$$

Entonces:

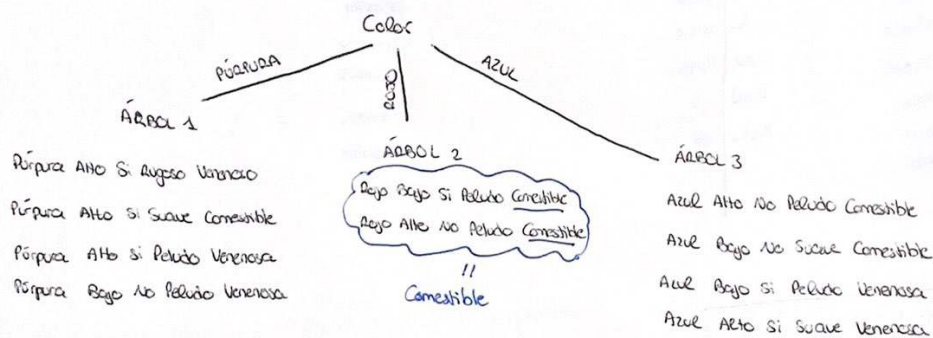
$$\Delta_{COLOR} = 0.15$$

← Empezamos dividiendo por color que es el más grande

$$\Delta_{ALTURA} = 0$$

$$\Delta_{RAYAS} = 0.086$$

$$\Delta_{TEXTURA} = 0.066$$



• Vamos de nuevo cuál es el mejor atributo para seguir dividiendo

• Para árbol 1 (reglas mirando árbol 1)

$$G_{PÚRPURA} = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0.375 \text{ (púrpura)}$$

Altura

$$G_{ALTO} = 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2 = 0.44$$

$$G_{BAJO} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0$$

$$\Delta_{ALTURA} = 0.375 - \left(\frac{3}{4} \cdot 0.44\right) - \left(\frac{1}{4} \cdot 0\right) = 0.04$$

• Para árbol 2 (mirando reglas árbol 2)

$$G_{COMESTIBLE} = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0.5$$

Altura

$$G_{ALTO} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0.5$$

$$G_{BAJO} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0.5$$

$$\Delta_{ALTURA} = 0.5 - \left(\frac{2}{4} \cdot 0.5\right) - \left(\frac{2}{4} \cdot 0.5\right) = 0$$

RAYAS

$$G_{SI} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0$$

$$G_{NO} = 1 - \left(\frac{3}{4}\right)^2 - \left(\frac{1}{4}\right)^2 = 0$$

$$\Delta_{RAYAS} = 0.5 - \left(\frac{1}{2} \cdot 0\right) - \left(\frac{3}{4} \cdot 0\right) = 0.5$$

Textura

$$G_{SUAVE} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0$$

$$G_{VENOZA} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0.5$$

$$G_{PELUDO} = 1 - 0^2 - 1^2 = 0$$

$$\Delta_{TEXTURA} = 0.5 - \left(\frac{1}{4} \cdot 0\right) - \left(\frac{3}{4} \cdot 0.5\right) - \left(\frac{1}{4} \cdot 0\right) = 0$$

Textura

$$G_{SUAVE} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0$$

$$G_{VENOZA} = 1 - \left(\frac{2}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0.5$$

$$G_{PELUDO} = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0$$

$$\Delta_{TEXTURA} = 0.375$$



Importante

Puedo eliminar la publi de este documento con 1 coin

¿Cómo consigo coins? → Plan Turbo: barato  
→ Planes pro: más coins

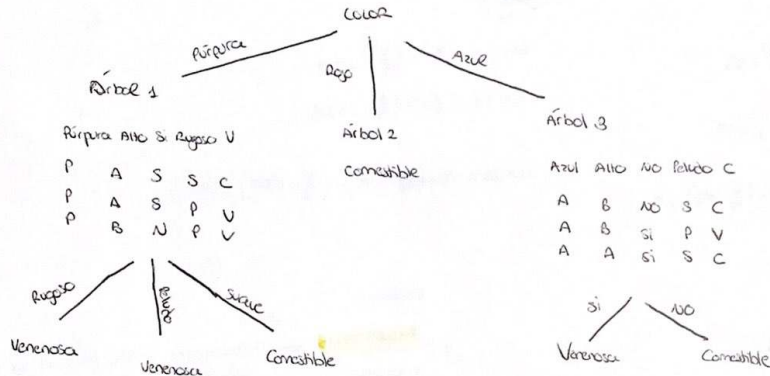
perdo  
espacio



Necesito  
concentración

ali ali ooh  
esto con 1 coin me  
lo quito yo...

WUOLAH



Reglas a partir del árbol de decisión

Antecedente 1	Antecedente 2	Clase		
Var	Valor	Var	Valor	Valor
color	Púrpura	textura	Peludo	Venenoza
color	Púrpura	textura	Suave	Comestible
color	Púrpura	textura	Rugoso	Venenoza
color	Azul	Rayos	Si	Venenoza
color	Azul	Rayos	no	Comestible

Matriz Confusión

Real \ Predicho	Venenosa	Comestible
Venenosa	TP 2	FN 0
Comestible	FP 1	TN 3

Tomemos que mejor de conjunto de prueba y ver si se cumple o no

Prueba

11	A	B	SI	R	V	→ TP
12	R	B	N	S	C	→ TN
13	A	B	S	S	V	→ TP
14	R	A	S	P	C	→ TN
15	P	A	N	R	C	→ FP
16	P	A	N	S	C	→ TN

WUOLAH