



Universidad
de Huelva



Actividad nº4. Ejercicio inventado para aplicar el algoritmo de aprendizaje simbólico ID3.

Ejercicio de aprendizaje automático simbólico

Inteligencia Artificial - Grado en Ingeniería Informática – 2019/2020

Estudiante: Borja López Pineda

Profesor: Luis Ignacio López Gómez

Índice

Resumen	2
1.-Introducción	2
2.- Ejemplo de prueba	3
2.1.- Datos	3
2.2.- Cálculo manual	3
2.3.- Resultado	4
3.- Ejemplo exploración espacial	5
3.1.- Datos	6
3.3.- Resultado	7
4.- Ejemplo ¿Quién es Quién?	8
4.1.- Datos	11
4.3.- Resultado	12
Anexo I. Detalles de la implementación	14

Resumen

En este trabajo se plantearán tres conjuntos de datos a los que se les aplicará el algoritmo ID3.

El primer ejemplo consiste en unos datos de prueba generados para verificar el correcto funcionamiento de la implementación del algoritmo. Se resolverá también realizando el cálculo manualmente a modo de contraste.

El segundo ejemplo trata de asignar un tipo de misión de exploración espacial a un cuerpo celeste dados sus atributos. Los datos están sacados del videojuego Kerbal Space Program (KSP).

El tercer conjunto de datos ha sido extraído de un tablero que imita al clásico juego de mesa ¿Quién es Quién? Se realizarán unas modificaciones para lograr que el algoritmo ID3 genere la combinación de preguntas con la mayor probabilidad de ganar.

Introducción

ID3 es un algoritmo de clasificación que genera un árbol de decisión a partir de unos datos de entrenamiento. Comúnmente se utiliza para en campos que requieren aprendizaje automático como el procesamiento del lenguaje natural.

Cada problema individual es llamado ejemplo. Los ejemplos constan de campos atributos y un campo clase. Los atributos deben ser los mismos en todos los ejemplos, variando solo el valor de éstos.

Al tratarse de un algoritmo de aprendizaje automático, es capaz de entrenarse con el conjunto de datos que representa el problema. El objetivo es dividir el conjunto de datos resuelto en un grupo de entrenamiento y otro de prueba. El algoritmo procesa el conjunto de entrenamiento. Si se le planteara un caso de este conjunto sabría resolverlo, no obstante, lo ideal es que pueda resolver cualquier caso. Para corroborar esto, se usa el conjunto de datos de prueba que fue reservado. Si los conjuntos se han dividido correctamente y los casos son representativos, el algoritmo debería ser capaz de resolver otros casos que se le planteen.

En los ejemplos de este documento se asume que los atributos de un ejemplo determinan correctamente su clase. En el mundo real esta condición no se cumple ya que la compleja realidad no permite observar todos los atributos o estos resultan en una cantidad elevada que aumenta la complejidad de cálculo. Las mejores implementaciones de ID3 asignan probabilidades cuando no pueden determinar un valor concreto para la clase. Este no es el caso de la implementación que se adjunta, aquí se supone que la clase depende los atributos.

En los casos más sencillos el resultado del algoritmo se utiliza para generar una estructura de condicionantes IF óptima. En los más complejos el resultado se almacena como un árbol de decisión que puede ser usado para clasificar automáticamente otros ejemplos. En nuestro caso la salida del programa es un texto plano que puede ser dibujado en un árbol. En ese árbol los nodos intermedios representan los atributos por los que se clasifica y los nodos hojas representan las clases.

Ejemplo de prueba

Los atributos son: color de ojos, color de cabello y altura. La clase puede ser: + o -. Las abreviaturas utilizadas se detallan en la tabla derecha.

Estos ejemplos no modelan ningún aspecto real.

Datos

Ej	Ojos	Cabello	Altura	Clase	Atributo	Valor	Significado
1	A	R	A	+	Ojos	M	Marrones
2	A	M	M	+		A	Azules
3	M	M	M	-		V	Verdes
4	V	M	M	-	Cabello	M	Castaño
5	V	M	A	+		R	Rubio
6	M	M	B	-	Altura	B	Bajo
7	V	R	B	-		M	Medio
8	A	M	M	+		A	Alto

Calculo manual

Ya que el objetivo de este conjunto de datos es poner a prueba el algoritmo implementado, se adjuntan los cálculos manuales antes de pasar a los resultados que arrojó el programa.

$$\text{Información}(E, C) = -\frac{4}{8} \cdot \log\left(\frac{4}{8}\right) - \frac{4}{8} \cdot \log\left(\frac{4}{8}\right) = 1$$

$$\begin{aligned}\text{Entropía}(E, \text{ojos}) &= \frac{3}{8} \cdot \left(-\frac{3}{3} \cdot \log\left(\frac{3}{3}\right) - \frac{0}{3} \cdot \log\left(\frac{0}{3}\right)\right) + \frac{2}{8} \cdot \left(-\frac{0}{2} \cdot \log\left(\frac{0}{2}\right) - \frac{2}{2} \cdot \log\left(\frac{2}{2}\right)\right) + \frac{3}{8} \\ &\cdot \left(-\frac{1}{3} \cdot \log\left(\frac{1}{3}\right) - \frac{2}{3} \cdot \log\left(\frac{2}{3}\right)\right) = 0.3444\end{aligned}$$

$$\begin{aligned}\text{Entropía}(E, \text{cabello}) &= \frac{2}{8} \cdot \left(-\frac{1}{2} \cdot \log\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log\left(\frac{1}{2}\right)\right) + \frac{6}{8} \cdot \left(-\frac{3}{6} \cdot \log\left(\frac{3}{6}\right) - \frac{3}{6} \cdot \log\left(\frac{3}{6}\right)\right) \\ &= 1\end{aligned}$$

$$\begin{aligned}\text{Entropía}(E, \text{altura}) &= \frac{2}{8} \cdot \left(-\frac{2}{2} \cdot \log\left(\frac{2}{2}\right) - \frac{0}{2} \cdot \log\left(\frac{0}{2}\right)\right) + \frac{4}{8} \cdot \left(-\frac{2}{4} \cdot \log\left(\frac{2}{4}\right) - \frac{2}{4} \cdot \log\left(\frac{2}{4}\right)\right) + \frac{2}{8} \\ &\cdot \left(-\frac{0}{2} \cdot \log\left(\frac{0}{2}\right) - \frac{2}{2} \cdot \log\left(\frac{2}{2}\right)\right) = 0.5\end{aligned}$$

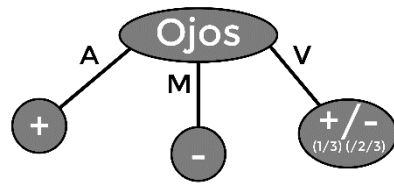
$$\text{Ganancia}(E, \text{ojos}) = 1 - 0.3444 = 0.6556$$

$$\text{Ganancia}(E, \text{cabello}) = 1 - 1 = 0$$

$$\text{Ganancia}(E, \text{altura}) = 1 - 0.5 = 0.5$$

El atributo que minimiza la entropía y, por tanto, maximiza la información es Ojos. Dicho de otro modo, el atributo Ojos es el que aporta más importación sobre la clase.

El primer nodo será Ojos y dos de sus hijas van a ser hojas de una clase determinada, A y M. En el caso del valor V existe una mezcla de clases, se repiten los cálculos para los ejemplos que cumplen la condición.



$$\text{Información}(E, C) = -\frac{1}{3} \cdot \log\left(\frac{1}{3}\right) - \frac{2}{3} \cdot \log\left(\frac{2}{3}\right) = 0.9183$$

$$\text{Entropía}(E, \text{cabello}) = \frac{2}{3} \cdot \left(-\frac{1}{2} \cdot \log\left(\frac{1}{2}\right) - \frac{1}{2} \cdot \log\left(\frac{1}{2}\right)\right) + \frac{1}{3} \cdot \left(-\frac{0}{1} \cdot \log\left(\frac{0}{1}\right) - \frac{1}{1} \cdot \log\left(\frac{1}{1}\right)\right) = 0.6667$$

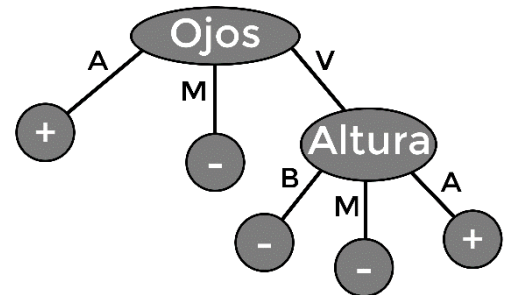
$$\begin{aligned} \text{Entropía}(E, \text{altura}) &= \frac{1}{3} \cdot \left(-\frac{1}{1} \cdot \log\left(\frac{1}{1}\right) - \frac{1}{1} \cdot \log\left(\frac{1}{1}\right)\right) + \frac{1}{3} \cdot \left(-\frac{0}{1} \cdot \log\left(\frac{0}{1}\right) - \frac{1}{1} \cdot \log\left(\frac{1}{1}\right)\right) + \frac{1}{3} \cdot \left(-\frac{0}{1} \cdot \log\left(\frac{0}{1}\right) - \frac{2}{1} \cdot \log\left(\frac{2}{1}\right)\right) = 0 \end{aligned}$$

$$\text{Ganancia}(E, \text{cabello}) = 0.9183 - 0.6667 = 0.2516$$

$$\text{Ganancia}(E, \text{altura}) = 0.9183 - 0 = 0.9183$$

El atributo que más reduce la entropía es Altura. Logra una entropía de 0 que corresponde con un estado completamente ordenado. Así logra maximizar la información.

Este es el árbol finalizado.



Resultado

Al aplicar ID3 obtenemos un árbol que permite deducir la clase de un ejemplo con un máximo de dos comprobaciones. Nos ha servido para saber que la clase no depende del color de pelo, al menos, no es necesario para determinar la clase con estos datos de ejemplo.

La salida del programa adjunta a la derecha, ha sido indentada para facilitar su lectura. Una vez en este estado es fácil dibujar el árbol de decisión. El resultado es exactamente el mismo que al realizar los cálculos a mano.

```

Subarbol Ojos valor Azules
  Hoja de clase: +
Fin de Ojos valor Azules
Subarbol Ojos valor Marrones
  Hoja de clase: -
Fin de Ojos valor Marrones
Subarbol Ojos valor Verdes
  Subarbol Altura valor Medio
    Hoja de clase: -
  Fin de Altura valor Medio
  Subarbol Altura valor Alto
    Hoja de clase: +
  Fin de Altura valor Alto
  Subarbol Altura valor Bajo
    Hoja de clase: -
  Fin de Altura valor Bajo
Fin de Ojos valor Verdes
  
```

Ejemplo exploración espacial

Kerbal Space Program es un simulador de misiones espaciales con un modo carrera. En este modo deberás gestionar tu agencia espacial para obtener puntos de ciencia necesarios para mejorar la tecnología. Estos puntos se ganan al recolectar datos en el sistema solar.

En este ejemplo nos centraremos en decidir el tipo de misión más adecuada para explorar un determinado cuerpo celeste. Los datos han sido obtenidos de la Wiki* y se han discretizado. La elección del mejor tipo de misión se ha tomado en base a mi experiencia personal. Por ejemplo: si el planeta tiene una gravedad no excesivamente baja y no demasiado alta, lo mejor es un Rover terrestre, excepto si la atmósfera permite sustentarse con hélices, entonces uno aéreo podría ser mejor opción. Hay otros factores que lo condicionan, si es muy pequeño y la gravedad es baja, no es necesario que tenga ruedas, un simple aterrizador consigue los mismos puntos. Y si el planeta está muy lejos y tiene una gravedad fuerte sin atmósfera, no podemos enviar un Rover porque pesa demasiado como para aterrizarlo tan lejos sin paracaídas. Luego, si no es un planeta rocoso sino gaseoso o se trata de una estrella, no podemos aterrizar, tendremos que enviar un orbitador para que rodee el cuerpo o descienda hasta desintegrarse.

Los atributos que se han seleccionado son:

- Tipo: si se trata de una estrella, planeta o satélite.
- Órbita: en kilómetros. Discretizado.
- Inclinado: si su inclinación es significativa y requiere ajustar planos en la maniobra de inserción.
- Radio: en metros. Discretizado.
- Gravedad: relativa a la gravedad de la tierra, de forma aproximada.
- Atmosfera: relativa a la presión atmosférica terrestre, de forma aproximada.
- Composición: si es un cuerpo rocoso o gaseoso.

Las clases posibles son:

- Orbitador: prueba que se inserta en la órbita y no llega a aterrizar, podría o no llegar a colisionar con el cuerpo.
- Sonda aterrizador: prueba que aterriza suavemente en la superficie.
- Rover aterrizador: prueba que transporta un vehículo capaz de desplazarse por la superficie del cuerpo, debe aterrizar suavemente.
- Rover volador: prueba que transporta un vehículo capaz de desplazarse por la atmósfera del cuerpo, no necesita aterrizar.

Para simplificar, solo se ha tenido en cuenta opciones no tripuladas que no impliquen misiones de retorno.

Órbita		Radio	
<25M	< 25 000 Km	<100M	< 100 000 m
<50M	25 000 - 50 000 Km	<500M	100 000 - 500 000 m
<200M	50 000 - 200 000 Km	<1G	500 000 - 1 000 000 m
<25G	200 000 - 25 000 000 Km	<10G	1 000 000 - 10 000 000 m
<50G	25 000 000 - 50 000 000 Km	>10G	> 10 000 000 m
>50G	> 50 000 000 Km		

Datos

Ejemplo	Tipo	Orbita	Inclinado	Radio	Gravedad	Atmosfera	Composición	Clase
Kerbol	Estrella	-	No	>10G	Infinita	0	Gas	Orbitador
Moho	Planeta	<10G	Si	<500M	.1	0	Roca	Sonda aterrizador
Eve	Planeta	<10G	No	<1G	1.5	5	Roca	Rover volador
Gilly	Satélite	<50M	Si	<100M	.01	0	Roca	Sonda aterrizador
Kerbin	Planeta	<25G	No	<1G	1	1	Roca	Rover volador
Mun	Satélite	<25M	No	<500M	.1	0	Roca	Rover aterrizador
Minmus	Satélite	<50M	No	<100M	.01	0	Roca	Sonda aterrizador
Duna	Planeta	<25G	No	<500M	.5	.05	Roca	Rover aterrizador
Ike	Satélite	<25M	No	<500M	.1	0	Roca	Sonda aterrizador
Dres	Planeta	<50G	Si	<500M	.1	0	Roca	Sonda aterrizador
Jool	Planeta	>50G	No	<10G	1	15	Gas	Orbitador
Laythe	Satélite	<50M	No	<500M	1	.5	Roca	Sonda aterrizador
Vall	Satélite	<50M	No	<500M	.1	0	Roca	Sonda aterrizador
Tylo	Satélite	<200M	No	<1G	1	0	Roca	Rover aterrizador
Bod	Satélite	<200M	No	<100M	.1	0	Roca	Sonda aterrizador
Pol	Satélite	<200M	Si	<100M	.01	0	Roca	Sonda aterrizador
Eeloo	Planeta	>50G	Si	<500M	.1	0	Roca	Rover aterrozdpr

Resultado

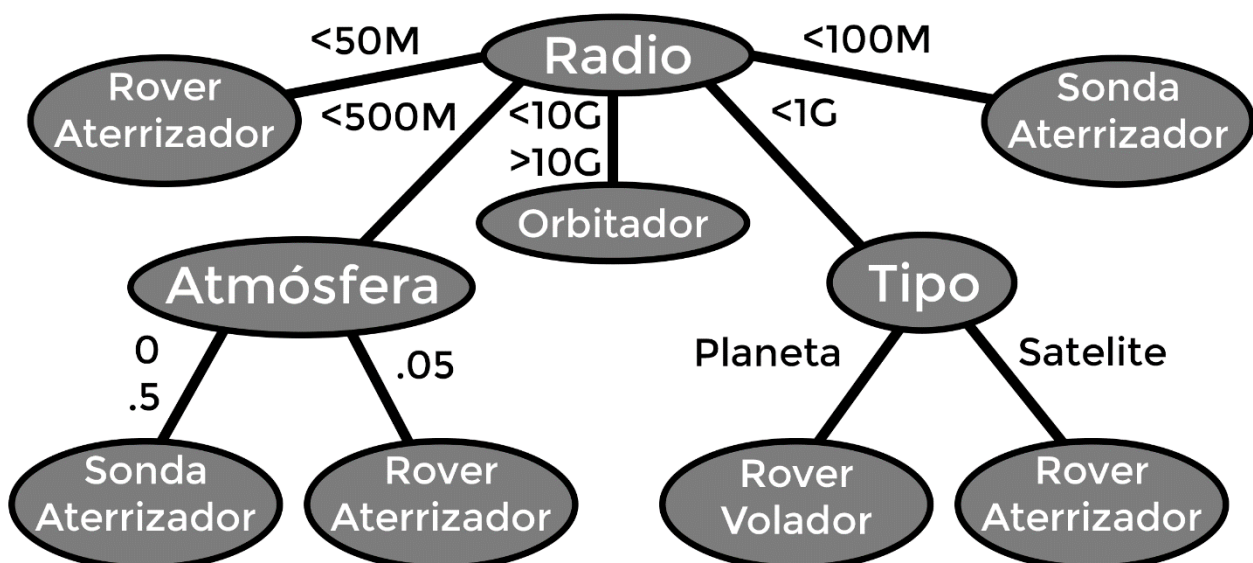
La salida del programa refleja un hecho que parece cumplirse en todos los casos y que está limitando bastante la belleza de este ejemplo.

El tamaño de un cuerpo determina su tipo. Un cuerpo pequeño suele ser un satélite, un poco más grande suele ser un planeta rocoso. Si el tamaño aumenta, ya no existen planetas rocosos, solo gaseosos. Y si es mucho más grande, se tratará de una estrella.

Básicamente el atributo radio resume los atributos tipo y composición. Incluso podría englobar parte del campo órbita porque los planetas gaseosos están más alejados de la estrella.

Por lo demás se nota una falta de casos de ejemplo. Hay muchas decisiones que se toman sin tener en cuenta todos los factores involucrados porque no hay datos de ejemplo que lo obliguen a ello.

Subarbol Radio valor >10G
Hoja de clase: Orbitador
Fin de Radio valor >10G
Subarbol Radio valor <500M
Subarbol Atmosfera valor 0
Hoja de clase: Sonda aterrizador
Fin de Atmosfera valor 0
Subarbol Atmosfera valor .05
Hoja de clase: Rover aterrizador
Fin de Atmosfera valor .05
Subarbol Atmosfera valor .5
Hoja de clase: Sonda aterrizador
Fin de Atmosfera valor .5
Fin de Radio valor <500M
Subarbol Radio valor <1G
Subarbol Tipo valor Planeta
Hoja de clase: Rover volador
Fin de Tipo valor Planeta
Subarbol Tipo valor Satellite
Hoja de clase: Rover aterrizador
Fin de Tipo valor Satellite
Fin de Radio valor <1G
Subarbol Radio valor <100M
Hoja de clase: Sonda aterrizador
Fin de Radio valor <100M
Subarbol Radio valor <50M
Hoja de clase: Rover aterrizador
Fin de Radio valor <50M
Subarbol Radio valor <10G
Hoja de clase: Orbitador
Fin de Radio valor <10G



Ejemplo ¿Quién es Quién?

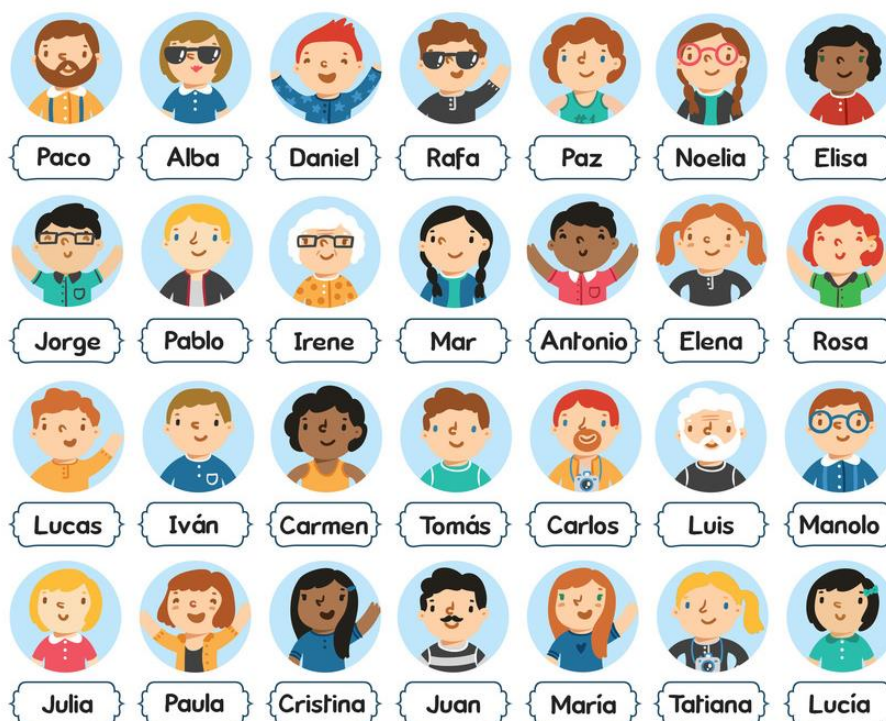
De los creadores de Monopoly y My Little Pony, ¿Quién es Quién? es un juego de mesa que trata de adivinar el personaje del oponente realizando preguntas sobre su apariencia.

Si nos fijamos, la estrategia ganadora trata de descartar la mayor cantidad de posibilidades con cada pregunta. Podemos hacer una pregunta muy concreta que separe el tablero en un grupo de 1 personaje y un grupo con el resto. Si tenemos suerte, ganaremos con una sola pregunta, pero el resto de las veces habremos prácticamente desaprovechado una pregunta. A largo plazo, la mejor estrategia es no arriesgar y realizar siempre la pregunta que divide de forma más equitativa las opciones a descartar y a conservar.

Precisamente cuando más equilibrada sea la división, más probable es ganar y menos entropía tiene el estado siguiente. Es por esto que ID3 puede usarse para generar un árbol de decisión que indique la jugada perfecta al ¿Quién es quién?

Para lograrlo tenemos que hacer algunas modificaciones. Cada ejemplo es un personaje del juego, la clase del ejemplo es el nombre del personaje. Los atributos son las diferentes preguntas que se pueden realizar.

Por otro lado, a fin de jugar justamente, he incluido en la siguiente tabla la definición y reglas de cada pregunta. Todas las preguntas se responden con sí o no, no se permite otro tipo de pregunta. No se pueden combinar atributos físicos en una misma pregunta. Ej: “¿Tiene los ojos negros?” es una pregunta válida, pero “¿De qué color son sus ojos?” y “¿Tiene los ojos negros o verdes?” no son preguntas válidas.



Este es el tablero que usará para jugar al ¿Quién es quién?

Los atributos se han sacado a partir de las preguntas posibles. En la siguiente tabla se detallan todos los atributos y se especifican los matices de cada uno.

Grupo de preguntas	Pregunta / Atributo	Observaciones
Color de ojos	Ojos negros	Si sus ojos son no son visibles (criterio de Ojos visibles), todos estos atributos valen No.
	Ojos castaños	
	Ojos azules	
	Ojos verdes	
Ojos	Ojos visibles	Sus ojos serán no visibles si están cerrados o están cubiertos por gafas teñidas.
Gafas	Gafas de sol	Gafas de ver engloba las gafas cuadradas y las no cuadradas
	Gafas de ver	
	Gafas de ver cuadradas	
Cámara	Cámara	
Color de Pelo	Pelo pelirrojo	
	Pelo castaño	
	Pelo negro	
	Pelo rubio	
	Pelo blanco	
Longitud de pelo	Pelo corto	
	Pelo largo	
Adornos de pelo	Pelo con coletas/s	
	Pelo con trenzas	
	Horquilla en el pelo	
Tipos de pelo	Pelo lacio	
	Pelo rizado	
Vello facial	Barba	La pregunta Barba y la pregunta Bigote son las dos verdaderas si se cumple Barba y Bigote.
	Bigote	
	Barba y bigote	
Mangas	Sin mangas	Se considera sin mangas si la camiseta es de tirantes. Se considera manga corta si sus brazos son visibles. Si solo pueden verse sus manos, no es manga corta. Se considera manga larga si sus brazos no son visibles, para ello deben estar en la imagen, aunque en el dibujo no se alcancen a ver y la ropa pueda dar lugar a pensar que lleva manga corta, si no se ve pueden ver los brazos, se asume manga larga.
	Manga corta	
	Manga larga	
Tipo de prenda	Camisa	

	Polo	Las camisas tienen botones en todo el largo de la prenda. Los polos solo tienen botones en la parte superior. Las chaquetas deben ser abiertas.
	Chaqueta	
	Tirantes	
Color de prenda	Prenda azul	En caso de tener prendas de más de un color, se tomará el color distinto a negro. Si ninguna de las prendas es negra, se tomará el color de la prenda más interior.
	Prenda amarilla	
	Prenda roja	
	Prenda verde	
	Prenda negra	
Boca	Boca abierta	
Brazos	Brazos bajados	Dos brazos levantados no engloban un brazo levantado.
	Brazo levantado	
	Brazos levantados	
Tono de piel	Color	
Género	Macho	

Algunas preguntas han sido descartadas por tener el mismo efecto que otra pregunta ya existente. Por ejemplo “¿Es una persona mayor?” es equivalente a “¿Tiene el pelo blanco?” ya que todos los señores mayores, y solo los señores mayores, tienen el pelo blanco. O la pregunta “¿Lleva camiseta de tirantes?” es equivalente a “¿Lleva camiseta sin mangas?”, las personas con camiseta de tirantes es la única prenda sin mangas.

Otras preguntas han sido eliminadas por su subjetividad. Por ejemplo “¿Es un niño/a?” o “¿Es joven?”

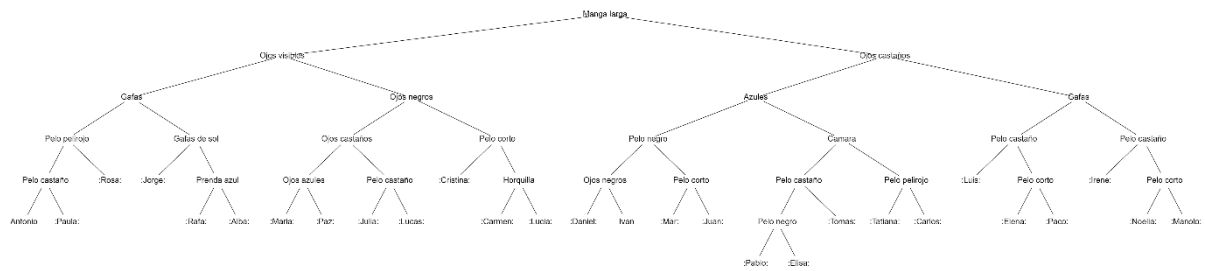
En la siguiente tabla se ha marcado con una cruz todas las combinaciones de preguntas cuya respuesta es un sí.

Datos

Personaje	Color de ojos					Gafas						Color de pelo					Largo de pelo		Adornos de pelo			Tipo de pelo		Vello facial			Mangas			Tipo de ropa				Color de ropa						Brazos					
	Ojos negros	Ojos castaños	Ojos azules	Ojos verdes	Ojos visibles	Gafas	Gafas de sol	Gafas de ver	Gafas de ver cuadradas	Cámara	Pelo pelirrojo	Pelo castaño	Pelo negro	Pelo rubio	Pelo blanco	Pelo corto	Pelo largo	Pelo con coleta/s	Pelo con trenzas	Horquilla en el pelo	Pelo lacio	Pelo rizado	Barba	Bigote	Barba y Bigote	Sin mangas	Manga corta	Manga larga	Camisa	Polo	Chaqueta	Tirantes	Prenda azul	Prenda amarilla	Prenda roja	Prenda verde	Prenda negra	Boca abierta	Brazos bajados	Un brazo levantado	Dos brazos levantados	Color	Macho		
Paco		X			X							X				X						X	X	X			X	X			X		X					X					X		
Alba						X	X					X				X					X		X	X	X		X		X			X					X								
Daniel										X						X											X					X					X			X					
Rafa						X	X					X				X					X						X			X							X		X					X	
Paz			X		X							X				X						X				X										X			X						
Noelia		X			X	X		X				X					X		X							X		X			X					X			X						
Elisa			X		X								X			X						X						X							X				X				X		
Jorge						X		X	X				X			X						X					X		X								X				X			X	
Pablo			X		X									X		X					X							X			X							X						X	
Irene		X			X	X		X	X						X	X						X						X							X				X						
Mar	X				X								X				X		X								X			X						X				X					
Antonio													X			X						X					X		X							X				X	X	X			
Elena		X			X							X					X	X									X		X								X			X					
Rosa										X						X						X					X		X								X				X				
Lucas		X			X							X				X						X					X			X						X				X				X	
Iván	X				X							X				X					X						X			X								X						X	
Carmen	X				X								X			X						X				X													X				X		
Tomas			X		X							X				X						X					X												X					X	
Carlos			X		X					X	X					X					X		X	X	X			X			X							X						X	
Luis		X			X										X	X						X	X	X			X										X			X				X	
Manolo		X			X	X		X				X				X					X						X	X						X				X						X	
Julia		X			X									X		X					X						X			X								X							
Paula												X				X					X						X		X									X			X				
Cristina	X				X								X				X			X	X						X			X				X				X			X				X
Juan	X				X								X			X						X		X				X									X		X					X	
María				X	X							X					X				X						X							X				X		X					
Tatiana			X		X					X					X			X									X		X									X		X					
Lucia	X				X							X				X				X	X						X			X							X			X					

Resultado

Este es el árbol de decisión que se obtiene del texto plano que proporcionó el algoritmo.



Ha sido dibujado usando una herramienta para representar árboles binarios modificada.

La altura del árbol nos da una idea del número de preguntas necesarias para determinar un personaje. Algunos personajes pueden averiguarse con 4 preguntas, otros con 6, la mayoría con 5.

Este resultado también da pie a escoger un personaje difícil de adivinar con fundamento. Objetivamente, si solo podemos realizar las preguntas aquí recogidas y suponemos que el adversario va a jugar de la mejor forma posible (siguiendo este árbol) los personajes Pablo y Elisa son más difíciles de adivinar puesto que requieren más preguntas que la media.

Este mismo árbol ha sido puesto a prueba con 4 oponentes distintos. En todas las partidas, es el humano el que comienza preguntando. Los resultados fueron: dos victorias y dos derrotas. Las dos derrotas se debieron a que el humano comenzó primero y los dos jugadores necesitaron el mismo número de preguntas. Una de las derrotas fue por suerte y en la otra el jugador eligió su movimiento con fundamentos.

Se adjunta una imagen en alta definición del árbol así como los programas compilados necesarios para generar las salidas.

Esta es la salida del algoritmo con tabulaciones para facilitar su lectura.

```
Subarbol Manga larga valor Si
  Subarbol Ojos castaños valor Si
    Subarbol Gafas valor No
      Subarbol Pelo castano valor Si
        Subarbol Pelo corto valor Si
          Hoja de clase: Paco
          Fin de Pelo corto valor Si
          Subarbol Pelo corto valor No
            Hoja de clase: Elena
            Fin de Pelo corto valor No
          Fin de Pelo castano valor Si
          Subarbol Pelo castano valor No
            Hoja de clase: Luis
            Fin de Pelo castano valor No
          Fin de Gafas valor No
          Subarbol Gafas valor Si
            Subarbol Pelo castano valor Si
              Subarbol Pelo corto valor No
                Hoja de clase: Noelia
                Fin de Pelo corto valor No
              Subarbol Pelo corto valor Si
                Hoja de clase: Manolo
                Fin de Pelo corto valor Si
              Fin de Pelo castano valor Si
              Subarbol Pelo castano valor No
                Hoja de clase: Irene
                Fin de Pelo castano valor No
            Fin de Gafas valor Si
            Fin de Ojos castaños valor Si
            Subarbol Ojos castaños valor No
              Subarbol Ojos azules valor No
                Subarbol Pelo negro valor No
                  Subarbol Ojos negros valor No
                    Hoja de clase: Daniel
                    Fin de Ojos negros valor No
                  Subarbol Ojos negros valor Si
                    Hoja de clase: Ivan
                    Fin de Ojos negros valor Si
                  Fin de Pelo negro valor No
                  Subarbol Pelo negro valor Si
                    Subarbol Pelo corto valor No
                      Hoja de clase: Mar
                      Fin de Pelo corto valor No
                    Subarbol Pelo corto valor Si
                      Hoja de clase: Juan
                      Fin de Pelo corto valor Si
                    Fin de Pelo negro valor Si
                  Fin de Ojos azules valor No
                  Subarbol Ojos azules valor Si
                    Subarbol Camara valor No
                      Subarbol Pelo castano valor No
                        Subarbol Pelo negro valor Si
                          Hoja de clase: Elisa
                          Fin de Pelo negro valor Si
                        Subarbol Pelo negro valor No
                          Hoja de clase: Pablo
                          Fin de Pelo negro valor No
                        Fin de Pelo castano valor No
                        Subarbol Pelo castano valor Si
                          Hoja de clase: Tomas
                          Fin de Pelo castano valor Si
                        Fin de Camara valor No
                        Subarbol Camara valor Si
                          Subarbol Pelo pelirrojo valor Si
                            Hoja de clase: Carlos
                            Fin de Pelo pelirrojo valor Si
                            Subarbol Pelo pelirrojo valor No
                              Hoja de clase: Tatiana
                              Fin de Pelo pelirrojo valor No
                            Fin de Camara valor Si
                          Fin de Ojos azules valor Si
                          Fin de Ojos castaños valor No
                          Fin de Manga larga valor Si
                          Subarbol Manga larga valor No
                            Subarbol Ojos visibles valor No
                              Subarbol Gafas valor Si
                                Subarbol Gafas de sol valor Si
                                  Subarbol Prenda azul valor Si
                                    Hoja de clase: Alba
                                    Fin de Prenda azul valor Si
                                    Subarbol Prenda azul valor No
                                      Hoja de clase: Rafa
                                      Fin de Prenda azul valor No
                                  Fin de Gafas de sol valor Si
                                  Subarbol Gafas de sol valor No
                                    Hoja de clase: Jorge
                                    Fin de Gafas de sol valor No
                                Fin de Gafas valor Si
                                Subarbol Gafas valor No
                                  Subarbol Pelo pelirrojo valor No
                                    Subarbol Pelo castano valor No
                                      Hoja de clase: Antonio
                                      Fin de Pelo castano valor No
                                      Subarbol Pelo castano valor Si
                                        Hoja de clase: Paula
                                        Fin de Pelo castano valor Si
                                      Fin de Pelo pelirrojo valor No
                                      Subarbol Pelo pelirrojo valor Si
                                        Hoja de clase: Rosa
                                        Fin de Pelo pelirrojo valor Si
                                  Fin de Gafas valor No
                                  Fin de Ojos visibles valor No
                                  Subarbol Ojos visibles valor Si
                                    Subarbol Ojos negros valor No
                                      Subarbol Ojos castaños valor No
                                        Subarbol Ojos azules valor Si
                                          Hoja de clase: Paz
                                          Fin de Ojos azules valor Si
                                          Subarbol Ojos azules valor No
                                            Hoja de clase: Maria
                                            Fin de Ojos azules valor No
                                          Fin de Ojos castaños valor No
                                          Subarbol Ojos castaños valor Si
                                            Hoja de clase: Lucas
                                            Fin de Pelo castano valor Si
                                            Subarbol Pelo castano valor No
                                              Hoja de clase: Julia
                                              Fin de Pelo castano valor No
                                            Fin de Ojos castaños valor Si
                                          Fin de Ojos negros valor No
                                          Subarbol Ojos negros valor Si
                                            Subarbol Pelo corto valor Si
                                              Subarbol Orquilla en el pelo valor No
                                                Hoja de clase: Carmen
                                                Fin de Orquilla en el pelo valor No
                                              Subarbol Orquilla en el pelo valor Si
                                                Hoja de clase: Lucia
                                                Fin de Orquilla en el pelo valor Si
                                              Fin de Pelo corto valor Si
                                              Subarbol Pelo corto valor No
                                                Hoja de clase: Cristina
                                                Fin de Pelo corto valor No
                                              Fin de Ojos negros valor Si
                                              Fin de Ojos visibles valor Si
                                              Fin de Manga larga valor No
                                            Fin de Ojos negros valor Si
                                            Fin de Ojos visibles valor Si
                                            Fin de Manga larga valor No
                                          Fin de Ojos negros valor Si
                                          Fin de Ojos visibles valor Si
                                          Fin de Manga larga valor No
                                        Fin de Ojos negros valor Si
                                        Fin de Ojos visibles valor Si
                                        Fin de Manga larga valor No
                                      Fin de Ojos negros valor Si
                                      Fin de Ojos visibles valor Si
                                      Fin de Manga larga valor No
                                    Fin de Ojos negros valor Si
                                    Fin de Ojos visibles valor Si
                                    Fin de Manga larga valor No
                                  Fin de Ojos negros valor Si
                                  Fin de Ojos visibles valor Si
                                  Fin de Manga larga valor No
                                Fin de Ojos negros valor Si
                                Fin de Ojos visibles valor Si
                                Fin de Manga larga valor No
                              Fin de Ojos negros valor Si
                              Fin de Ojos visibles valor Si
                              Fin de Manga larga valor No
                            Fin de Ojos negros valor Si
                            Fin de Ojos visibles valor Si
                            Fin de Manga larga valor No
                          Fin de Ojos negros valor Si
                          Fin de Ojos visibles valor Si
                          Fin de Manga larga valor No
                        Fin de Ojos negros valor Si
                        Fin de Ojos visibles valor Si
                        Fin de Manga larga valor No
                      Fin de Ojos negros valor Si
                      Fin de Ojos visibles valor Si
                      Fin de Manga larga valor No
                    Fin de Ojos negros valor Si
                    Fin de Ojos visibles valor Si
                    Fin de Manga larga valor No
                  Fin de Ojos negros valor Si
                  Fin de Ojos visibles valor Si
                  Fin de Manga larga valor No
                Fin de Ojos negros valor Si
                Fin de Ojos visibles valor Si
                Fin de Manga larga valor No
              Fin de Ojos negros valor Si
              Fin de Ojos visibles valor Si
              Fin de Manga larga valor No
            Fin de Ojos negros valor Si
            Fin de Ojos visibles valor Si
            Fin de Manga larga valor No
          Fin de Ojos negros valor Si
          Fin de Ojos visibles valor Si
          Fin de Manga larga valor No
        Fin de Ojos negros valor Si
        Fin de Ojos visibles valor Si
        Fin de Manga larga valor No
      Fin de Ojos negros valor Si
      Fin de Ojos visibles valor Si
      Fin de Manga larga valor No
    Fin de Ojos negros valor Si
    Fin de Ojos visibles valor Si
    Fin de Manga larga valor No
  Fin de Ojos negros valor Si
  Fin de Ojos visibles valor Si
  Fin de Manga larga valor No
Fin de Ojos negros valor Si
Fin de Ojos visibles valor Si
Fin de Manga larga valor No
```

Anexo I. Detalles de la implementación

La función principal que hace todo el trabajo se llama dividir y recibe un vector de ejemplos y un vector de atributos.

Los ejemplos se guardan como una estructura que almacena el nombre del ejemplo, la clase a la que pertenece y un mapa que asocia nombres de atributos con su valor.

El vector de atributos podría omitirse, pero se perdería tiempo obteniéndolos de los ejemplos, cuando van a tener que ser extraídos múltiples veces. Las clases se detectan en cada llamada porque van a variar a medida que se reduzcan los ejemplos y es sencillo de manipular. Igual ocurre con los valores por atributo, se podían pasar en lugar de extraerlos, pero llevar la cuenta de los valores que van desapareciendo resultaría complicado.

La función continúa mapeando los valores de cada atributo en una lista asociada con el nombre del atributo. Para todos los atributos, se busca sus posibles valores.

La entropía se calcula con una suma ponderada por valor del atributo. El procedimiento se asemeja al conteo que podría realizarse manualmente por cada clase por cada valor del atributo. Una vez terminado, se comprueba si la entropía es menor que la del atributo con la entropía más baja encontrado. No es necesario calcular la información puesto que la elección por menor entropía será la misma que si se decidiera por mayor información.

A continuación, nos preparamos para subdividir los conjuntos creados. Para ello eliminamos el atributo seleccionado de la lista de atributos, al dividir los ejemplos por valor de ese atributo todos los conjuntos tienen un mismo valor. Itera por todos los valores del atributo creando un subconjunto con los ejemplos filtrados, se llama a sí mismo con el nuevo conjunto filtrado y la lista de atributos modificada.

El caso base será que el conjunto de ejemplos tenga una sola clase o esté vacío.

La salida del algoritmo se realiza por pantalla, la traza que muestra permite reconstruir el árbol que han seguido las llamadas y las decisiones que ha tomado.

Como se ha explicado con anterioridad, este algoritmo no está preparado para considerar probabilidades, es por esto que necesita que todos los conjuntos de valores de los atributos tengan asociado una única clase.

Un algoritmo que considere probabilidades generaría un árbol incluso si los ejemplos incluyen este extracto.

Ej	A	B	C	Clase
E1	1	1	1	+
E2	1	1	1	+
E3	1	1	1	-

El resultado consideraría que A=1, B=1, C=1 debe clasificarse como clase + porque la probabilidad es superior.

El programa que he hecho, con suerte, no crashearía si se introducen estos ejemplos.