

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

## Actividades resueltas

### Weka: J48 - Árbol de decisión

#### Descripción de la actividad

Se ha obtenido la siguiente salida al aplicar Weka a un conjunto de datos.

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

```

plas <= 127
| mass <= 26.4: tested_negative (132.0/3.0)
| mass > 26.4
| | age <= 28: tested_negative (180.0/22.0)
| | age > 28
| | | plas <= 99: tested_negative (55.0/10.0)
| | | plas > 99
| | | | pedi <= 0.561: tested_negative (84.0/34.0)
| | | | pedi > 0.561
| | | | | preg <= 6
| | | | | age <= 30: tested_positive (4.0)
| | | | | age > 30
| | | | | age <= 34: tested_negative (7.0/1.0)
| | | | | age > 34
| | | | | mass <= 33.1: tested_positive (6.0)
| | | | | mass > 33.1: tested_negative (4.0/1.0)
| | | | | preg > 6: tested_positive (13.0)
plas > 127
| mass <= 29.9
| | plas <= 145: tested_negative (41.0/6.0)
| | plas > 145
| | | age <= 25: tested_negative (4.0)
| | | age > 25
| | | | age <= 61
| | | | mass <= 27.1: tested_positive (12.0/1.0)
| | | | mass > 27.1
| | | | | pres <= 82
| | | | | pedi <= 0.396: tested_positive (8.0/1.0)
| | | | | pedi > 0.396: tested_negative (3.0)
| | | | | pres > 82: tested_negative (4.0)
| | | | age > 61: tested_negative (4.0)
| mass > 29.9
| | plas <= 157

```

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

```

| | | pres <= 61: tested_positive (15.0/1.0)
| | | pres > 61
| | | | age <= 30: tested_negative (40.0/13.0)
| | | | age > 30: tested_positive (60.0/17.0)
| | plas > 157: tested_positive (92.0/12.0)
Number of Leaves :    20

```

Size of the tree : 39

Time taken to build model: 0.08 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	567	73.8281 %
Incorrectly Classified Instances	201	26.1719 %
Kappa statistic	0.4164	
Mean absolute error	0.3158	
Root mean squared error	0.4463	
Relative absolute error	69.4841 %	
Root relative squared error	93.6293 %	
Total Number of Instances	768	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,814	0,403	0,790	0,814	0,802	0,417	0,751	0,811	tested_negative
	0,597	0,186	0,632	0,597	0,614	0,417	0,751	0,572	tested_positive
Weighted Avg.	0,738	0,327	0,735	0,738	0,736	0,417	0,751	0,727	

=== Confusion Matrix ===

```

a   b  <-- classified as
407 93 | a = tested_negative
108 160 | b = tested_positive

```

Contesta a las siguientes cuestiones:

- ¿Qué algoritmo se ha aplicado? ¿Qué tipo de modelo clasificador aprende este algoritmo? ¿A qué categoría de aprendizaje pertenece este algoritmo?

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

- Una nueva instancia con valores de los atributos:

preg = 5  
 plas = 99  
 pres = 74  
 skin = 27  
 insu = 0  
 mass = 29  
 pedi = 0.203  
 age = 32

¿En qué clase se clasificaría? ¿Por qué?

- ¿Cuántas instancias del conjunto de datos de entrenamiento alcanzan el nodo hoja en el que se ha clasificado la instancia? ¿Cuántas de esas instancias pertenecen a la clase indicada en la hoja? ¿y cuantas no pertenecen a dicha clase?

### Resolución de la actividad

- ¿Qué algoritmo se ha aplicado?  
C4.5 llamado J48 en Weka.

*Explicación de la respuesta:*

ID3 solo trabaja con atributos nominales por lo tanto podemos asegurar que no puede ser ID3 porque los atributos son numéricos lo observamos ya en el primer nodo donde se comprueba que el valor del atributo llamado plas toma valores menores o iguales a 127 ( $plas \leq 127$ ).

¿Qué tipo de modelo clasificador aprende este algoritmo?

Árbol de decisión.

*Explicación de la respuesta:*

El algoritmo J48 genera un árbol de decisión. Podemos ver que no se han aprendido reglas de clasificación porque aunque haya comprobaciones de los valores de las variables en los nodos del árbol y éstas se pudieran mapear a reglas, en el modelo generado no aparece la estructura de las reglas:

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

If atributo = valor then clase.

¿A qué categoría de aprendizaje pertenece este algoritmo?

Aprendizaje supervisado.

*Explicación de la respuesta:*

Todos los árboles de decisión, también J48, pertenecen al aprendizaje supervisado porque a partir de los valores asignados a los atributos de las instancias nos permiten identificar la clase en la que se clasificaría una nueva instancia.

- Una nueva instancia con valores de los atributos:

preg = 5  
 plas = 99  
 pres = 74  
 skin = 27  
 insu = 0  
 mass = 29  
 pedi = 0.203  
 age = 32

¿En qué clase se clasificaría?

tested\_negative

¿Por qué?

Para determinar en qué clase se clasificaría la instancia tenemos que recorrer el árbol desde la raíz a las hojas e ir evaluando cada una de las condiciones. Empezamos comprobando el valor del atributo plas porque es el que aparece en el nodo raíz del árbol. Se observa que la división en dos ramas del árbol se da para los valores plas  $\leq 127$  (primera rama) y plas  $> 127$  (segunda rama). Miramos el valor que tiene el atributo plas en la nueva instancia y éste toma el valor plas = 99. Por lo tanto nos movemos a través de la primera rama del árbol porque se cumple la condición plas  $\leq 127$ . En el siguiente nivel del árbol tenemos que comprobar el valor del atributo mass que en este caso toma el valor mass = 29 y por tanto nos movemos por la rama mass  $> 26.4$  al cumplirse esta condición. Entonces bajamos un nivel más en el árbol y comprobamos el valor

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

del atributo age que toma el valor age = 32 en la instancia a clasificar. Por lo tanto nos movemos por la rama age > 28 bajando un nivel más en el árbol y nos encontramos que debemos comprobar otra vez el valor del atributo plas. Como hemos dicho antes este atributo toma el valor plas = 99 por lo que se cumple la condición plas <= 99. Al movernos a esta rama llegamos a un nodo hoja que nos permite saber cómo se ha clasificado la instancia. El nodo hoja es plas <= 99: tested\_negative (55.0/10.0) por lo tanto observamos que la instancia queda clasificada en la clase tested\_negative.

*Nota: En la respuesta del examen era no necesaria una explicación tan detallada.*

- ¿Cuántas instancias del conjunto de datos de entrenamiento alcanzan el nodo hoja en el que se ha clasificado la instancia?  
55 instancias

*Explicación de la respuesta:*

En la generación del modelo clasificador 55 instancias del conjunto de datos de entrenamiento han alcanzado la hoja tested\_negative (55.0/10.0). El primer número que se muestra entre paréntesis junto a la clase indica el número total de instancias que alcanzan la hoja. Es decir el número de instancias de los datos de entrenamiento que alcanzan la hoja y quedan clasificadas como de la clase indicada en la hoja, ya sea correcta o incorrectamente.

¿Cuántas de esas instancias pertenecen a la clase indicada en la hoja?  
45 instancias

*Explicación de la respuesta:*

De las 55 instancias del conjunto de datos de entrenamiento que han alcanzado la hoja tested\_negative (55.0/10.0), 10 no pertenecen a dicha clase y por tanto quedarían mal clasificadas (nos lo indica el segundo número que se muestra entre paréntesis), entonces 45 sí pertenecen a la clase tested\_negative.

Asignatura	
Técnicas de Inteligencia Artificial	Claudia Villalonga Palliser

¿y cuantas no pertenecen a dicha clase?

10 instancias

*Explicación de la respuesta:*

Como ya se ha comentado nos lo indica el segundo número que se muestra entre paréntesis junto al nombre de la clase.