

Práctica 6. Introducción al aprendizaje computacional con WEKA

Cristina Díaz García

Abril 2019

Índice

Índice general	1
1. Enunciado	2
2. Pregunta 1	2
2.1. Solución proporcionada	2
3. Pregunta 2	3
3.1. Solución proporcionada	3
4. Pregunta 3	3
4.1. Solución proporcionada	3
5. Pregunta 4	3
5.1. Solución proporcionada	4
6. Pregunta 5	6
6.1. Solución proporcionada	6

1. Enunciado

Tarea: Contesta a las preguntas que figuran a continuación.

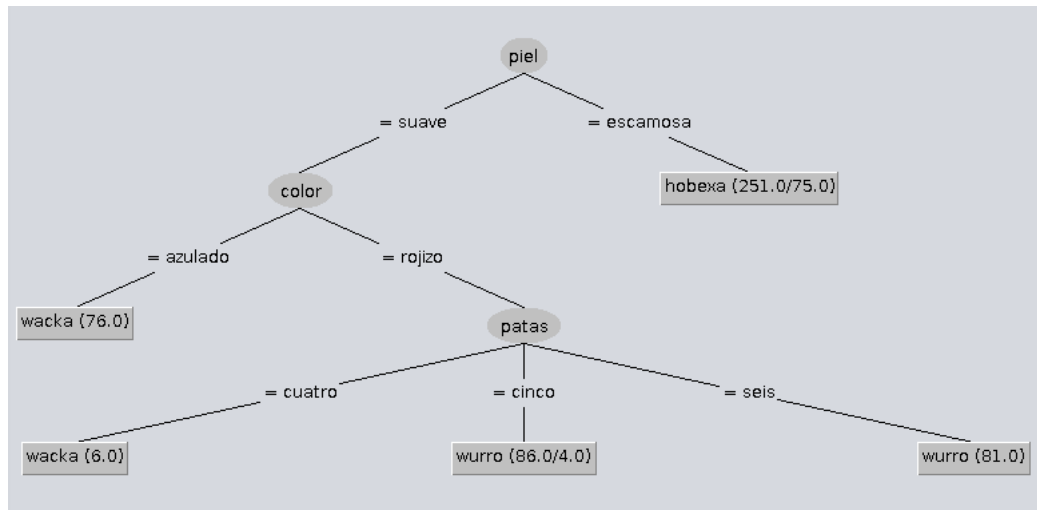
Entrega: Documento pdf con la solución (capturas de pantalla y textos descriptivos).

2. Pregunta 1

¿Cómo clasificaríamos a un bicho rojizo que cojea según este modelo? ¿Y a un bicho de piel escamosa? ¿Y a un bicho de piel suave, rojizo y con cuatro patas?

2.1. Solución proporcionada

Usando el árbol proporcionado:



- **Bicho rojizo que cojea:** En la primera decisión no tenemos información suficiente, así que con este modelo, no se podría clasificar este ejemplo.
- **Bicho de piel escamosa:** En la primera decisión, llegamos a la determinación de que es un hobexa.
- **Bicho de piel suave, rojizo y con cuatro patas:** En la primera decisión, tomaríamos el camino de la piel suave, llegando a la decisión del color de piel, que como es rojizo, necesitamos el número de patas, que al ser 4, el resultado es wacka.

3. Pregunta 2

¿Encuentras alguna explicación razonable a que los errores de clasificación se cometan con las wackas?

3.1. Solución proporcionada

Porque de los wackas es de los que menos información segura tenemos: ninguno de sus atributos son indistinguiblemente suyos.

4. Pregunta 3

¿Cómo clasificaríamos a un bicho rojizo que cojea según este modelo? ¿Y a un bicho de piel escamosa? ¿Y a un bicho de piel suave, rojizo y con cuatro patas?

4.1. Solución proporcionada

Usando las reglas proporcionadas:

```
JRIP rules:
=====

(color = azulado) and (piel = suave) => animal=wacka (76.0/0.0)
(color = azulado) and (patas = cinco) => animal=wacka (64.0/31.0)
(patatas = cuatro) and (piel = suave) => animal=wacka (6.0/0.0)
(patatas = cuatro) and (color = azulado) and (anomalia = false) => animal=wacka (53.0/26.0)
(piel = suave) => animal=wurro (167.0/4.0)
=> animal=hobexa (134.0/15.0)

Number of Rules : 6
```

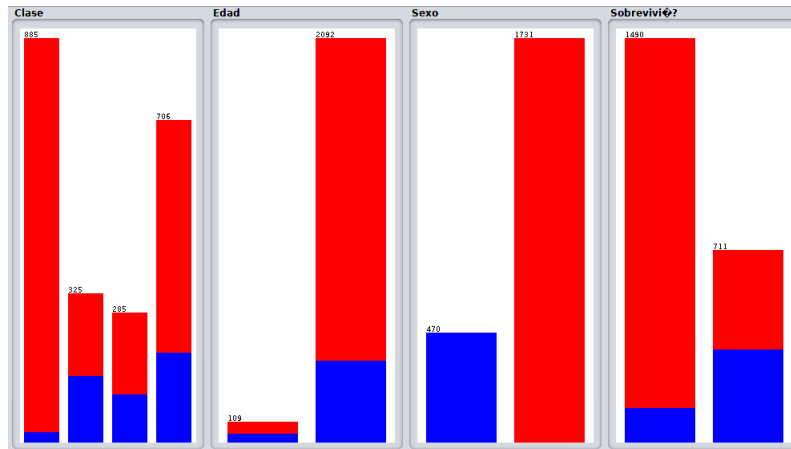
- **Bicho rojizo que cojea:** Como no cumple ninguna de las reglas con condiciones, pasaría por la sexta regla, por lo que se clasificaría como hobexa.
- **Bicho de piel escamosa:** Como no cumple ninguna de las reglas con condiciones, pasaría por la sexta regla, por lo que se clasificaría como hobexa.
- **Bicho de piel suave, rojizo y con cuatro patas:** Cumple la tercera regla, por lo que se clasificaría como un wacka.

5. Pregunta 4

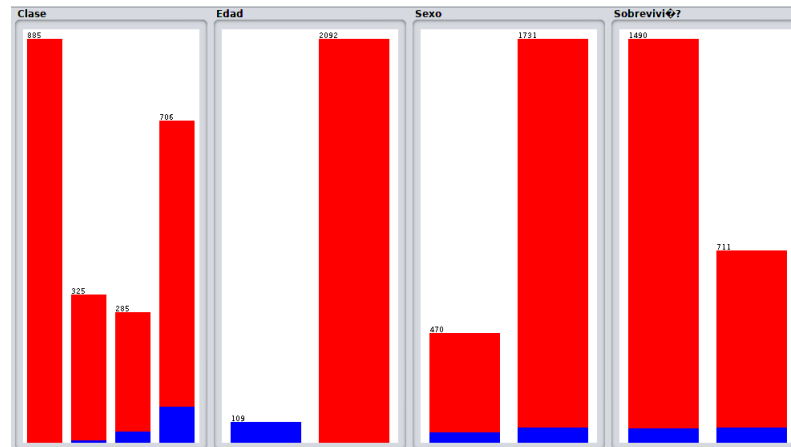
¿Cuántos varones viajaban en el Titanic? ¿Cuántas mujeres? ¿Cuántos menores de edad? ¿Cuántos viajeros en primera clase? Modifica los parámetros del algoritmo para que aprenda 26 reglas de asociación con una confianza de 0.85, e interpreta el significado de las cinco últimas.

5.1. Solución proporcionada

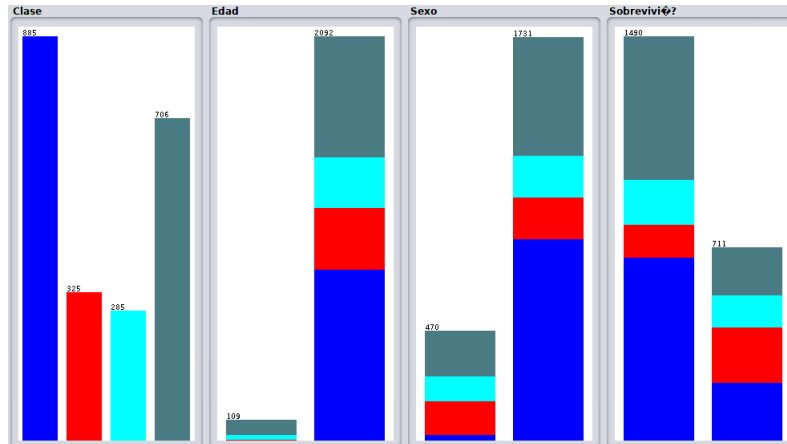
Clasificados por sexo



Clasificados por edad



Clasificados por clase



- Hombres que viajaban: 1731
- Mujeres que viajaban: 470
- Menores que viajaban: 109
- Viajeros en primera clase: 325

Esta es la salida inicial del algoritmo:

```
Apriori
=====

Minimum support: 0.35 (770 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 13

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4
Size of set of large itemsets L(2): 5
Size of set of large itemsets L(3): 2

Best rules found:

1. Clase=0 885 ==> Edad=1 885 <conf:(1)> lift:(1.05) lev:(0.02) [43] conv:(43.83)
2. Clase=0 Sexo=1 862 ==> Edad=1 862 <conf:(1)> lift:(1.05) lev:(0.02) [42] conv:(42.69)
3. Sexo=1 Sobrevivió?=0 1364 ==> Edad=1 1329 <conf:(0.97)> lift:(1.03) lev:(0.01) [32] conv:(1.88)
4. Clase=0 885 ==> Sexo=1 862 <conf:(0.97)> lift:(1.24) lev:(0.08) [165] conv:(7.87)
5. Clase=0 Edad=1 885 ==> Sexo=1 862 <conf:(0.97)> lift:(1.24) lev:(0.08) [165] conv:(7.87)
6. Clase=0 885 ==> Edad=1 Sexo=1 862 <conf:(0.97)> lift:(1.29) lev:(0.09) [191] conv:(8.95)
7. Sobrevivió?=0 1490 ==> Edad=1 1438 <conf:(0.97)> lift:(1.02) lev:(0.01) [21] conv:(1.39)
8. Sexo=1 1731 ==> Edad=1 1667 <conf:(0.96)> lift:(1.01) lev:(0.01) [21] conv:(1.32)
9. Edad=1 Sobrevivió?=0 1438 ==> Sexo=1 1329 <conf:(0.92)> lift:(1.18) lev:(0.09) [198] conv:(2.79)
10. Sobrevivió?=0 1490 ==> Sexo=1 1364 <conf:(0.92)> lift:(1.16) lev:(0.09) [192] conv:(2.51)
```

Tras modificar el algoritmo, el resultado es el siguiente:

Best rules found:

```
1. Clase=0 885 ==> Edad=1 885 <conf:(1)> lift:(1.05) lev:(0.02) [43] conv:(43.83)
2. Clase=0 Sexo=1 862 ==> Edad=1 862 <conf:(1)> lift:(1.05) lev:(0.02) [42] conv:(42.69)
3. Clase=0 Sobrevivió?=0 673 ==> Edad=1 673 <conf:(1)> lift:(1.05) lev:(0.02) [33] conv:(33.33)
4. Clase=0 Sexo=1 Sobrevivió?=0 670 ==> Edad=1 670 <conf:(1)> lift:(1.05) lev:(0.02) [33] conv:(33.18)
5. Clase=0 Sobrevivió?=0 673 ==> Sexo=1 670 <conf:(1)> lift:(1.27) lev:(0.06) [140] conv:(35.93)
6. Clase=0 Edad=1 Sobrevivió?=0 673 ==> Sexo=1 670 <conf:(1)> lift:(1.27) lev:(0.06) [140] conv:(35.93)
7. Clase=0 Sobrevivió?=0 673 ==> Edad=1 Sexo=1 670 <conf:(1)> lift:(1.31) lev:(0.07) [160] conv:(40.82)
8. Clase=1 325 ==> Edad=1 319 <conf:(0.98)> lift:(1.03) lev:(0) [10] conv:(2.3)
9. Sexo=1 Sobrevivió?=0 1364 ==> Edad=1 1329 <conf:(0.97)> lift:(1.03) lev:(0.01) [32] conv:(1.88)
10. Clase=0 885 ==> Sexo=1 862 <conf:(0.97)> lift:(1.24) lev:(0.08) [165] conv:(7.87)
11. Clase=0 Edad=1 885 ==> Sexo=1 862 <conf:(0.97)> lift:(1.24) lev:(0.08) [165] conv:(7.87)
12. Clase=0 885 ==> Edad=1 Sexo=1 862 <conf:(0.97)> lift:(1.29) lev:(0.09) [191] conv:(8.95)
13. Sobrevivió?=0 1490 ==> Edad=1 1438 <conf:(0.97)> lift:(1.02) lev:(0.01) [21] conv:(1.39)
14. Sexo=1 1731 ==> Edad=1 1667 <conf:(0.96)> lift:(1.01) lev:(0.01) [21] conv:(1.32)
15. Edad=1 Sobrevivió?=0 1438 ==> Sexo=1 1329 <conf:(0.92)> lift:(1.18) lev:(0.09) [198] conv:(2.79)
16. Sexo=1 Sobrevivió?=1 367 ==> Edad=1 338 <conf:(0.92)> lift:(0.97) lev:(-0) [-10] conv:(0.61)
17. Sobrevivió?=1 711 ==> Edad=1 654 <conf:(0.92)> lift:(0.97) lev:(-0.01) [-21] conv:(0.61)
18. Sexo=0 Sobrevivió?=1 344 ==> Edad=1 316 <conf:(0.92)> lift:(0.97) lev:(-0) [-10] conv:(0.59)
19. Clase=3 Sexo=1 Sobrevivió?=0 422 ==> Edad=1 387 <conf:(0.92)> lift:(0.96) lev:(-0.01) [-14] conv:(0.58)
20. Clase=2 285 ==> Edad=1 261 <conf:(0.92)> lift:(0.96) lev:(-0) [-9] conv:(0.56)
21. Sobrevivió?=0 1490 ==> Sexo=1 1364 <conf:(0.92)> lift:(1.16) lev:(0.09) [192] conv:(2.51)
22. Clase=3 Sexo=1 510 ==> Edad=1 462 <conf:(0.91)> lift:(0.95) lev:(-0.01) [-22] conv:(0.52)
23. Sexo=0 470 ==> Edad=1 425 <conf:(0.9)> lift:(0.95) lev:(-0.01) [-21] conv:(0.51)
24. Clase=3 Sobrevivió?=0 528 ==> Edad=1 476 <conf:(0.9)> lift:(0.95) lev:(-0.01) [-25] conv:(0.49)
25. Sobrevivió?=0 1490 ==> Edad=1 Sexo=1 1329 <conf:(0.89)> lift:(1.18) lev:(0.09) [200] conv:(2.23)
26. Clase=3 706 ==> Edad=1 627 <conf:(0.89)> lift:(0.93) lev:(-0.02) [-44] conv:(0.44)
```

Regla 22

La probabilidad de que la persona sea de tercera clase, hombre y que muriera es del 89%.

Regla 23

La probabilidad de que la persona sea mujer y mayor de edad es del 90%.

Regla 24

La probabilidad de que la persona sea de tercera clase, mayor de edad y que muriera es del 90%.

Regla 25

La probabilidad de que la persona hombre mayor de edad y muriera es del 89%.

Regla 26

La probabilidad de que la persona sea de tercera clase y mayor de edad es del 89%.

6. Pregunta 5

A la vista de los datos relativos a cada cluster, ¿qué grupo crees que representa mejor a los estudiantes que van a aprobar la asignatura? ¿Y a los que van a suspenderla?

6.1. Solución proporcionada

Esta es la salida del algoritmo:

```

Number of iterations: 6
Within cluster sum of squared errors: 9.03426043493763

Initial starting points (random):

Cluster 0: 0,2,0,0,0,0,0,998
Cluster 1: 11,3,9,7,2,374,4601,4826

Missing values globally replaced with mean/mode

Final cluster centroids:

```

Attribute	Full Data (20.0)	Cluster#	
		0 (10.0)	1 (10.0)
n_assignment	7.4	4.1	10.7
n_posts	0.45	0.2	0.7
n_quiz	3.5	1.4	5.6
n_quiz_a	2.75	0.5	5
n_quiz_s	0.75	0.9	0.6
total_time_assignment	503.25	294.6	711.9
total_time_quiz	1498.85	702.3	2295.4
total_time_forum	1317.6	919.5	1715.7

Ya que los resultados son mayores en el segundo cluster (Cluster #1), se podría pensar que es el que mejor representa a los estudiantes que aprobarán, siendo el primer cluster (Cluster #0), el de los estudiantes que suspenderán.

Referencias

- [1] Información oficial de GeNIe, <https://www.bayesfusion.com>.