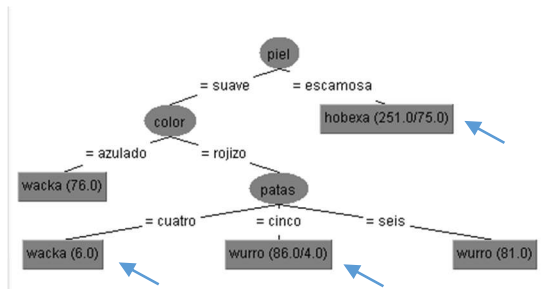


Práctica 5 WEKA

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?



- No podríamos clasificar correctamente dicho bicho rojizo que cojea, pues no tenemos un nodo que represente la posibilidad de que un bicho cojee. Tampoco tenemos especificada información acerca de la piel o el número de patas que tiene el bicho analizado. Sin embargo, como vemos en el árbol, la mayoría de instancias se consideran Wurros en el nodo hoja marcado, aunque deberemos tener en cuenta que 4 Wackas fueron incorrectamente clasificadas como Wurros.
- Un bicho de piel escamosa podríamos clasificarlo como una Hobexa, como se ve señalado en el árbol. Deberemos tener en cuenta que no será una clasificación exacta, pues 75 Wackas fueron incorrectamente instanciadas como Hobexas.
- Un bicho de piel suave, color rojizo y con cuatro patas, podemos clasificarlo con exactitud como una Wacka. Vemos que no hay instancias incorrectas, por lo que estaremos clasificando correctamente el animal.

Pregunta 2. ¿Encuentras alguna explicación razonable a que los errores de clasificación se cometan con las Wackas?

```
Correctly Classified Instances      421          84.2 %
Incorrectly Classified Instances    79           15.8 %
Kappa statistic                    0.7612
Mean absolute error                 0.1507
Root mean squared error             0.2749
Relative absolute error             33.9366 %
Root relative squared error         58.3427 %
Total Number of Instances          500
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1,000	0,012	0,976	1,000	0,988	0,982	0,995	0,986	wurro
	0,509	0,000	1,000	0,509	0,675	0,643	0,842	0,776	wacka
	1,000	0,231	0,701	1,000	0,824	0,734	0,866	0,658	hobexa
Weighted Avg.	0,842	0,085	0,887	0,842	0,830	0,786	0,900	0,803	

=== Confusion Matrix ===

```
a  b  c  <-- classified as
163  0  0 | a = wurro
 4 82 75 | b = wacka
 0  0 176 | c = hobexa
```

Esto se debe a que hay una gran variedad de Wackas con distintos atributos. Hay Wackas con piel rojiza, que cojean etc. Entonces no hay una distinción clara según el modelo entre los distintos tipos de especies según si son Wackas con atributos parecidos a otras especies.

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?

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

- Para clasificar un bicho rojizo que cojea sólo podríamos aplicar las reglas que no contradigan las instancias. Aplicando la regla 3 podríamos llegar a que se trata de una Wacka. Aplicando la regla 5 llegaríamos a la conclusión de que se trata de un Wurro, con una exactitud de 167 instancias sobre 4 erróneas. En otro caso podríamos aplicar la regla 6 con opción de que sea una Hobexa, con una exactitud de 134 instancias sobre 15, que son incorrectas.

- Un bicho de piel escamosa podríamos clasificarlo de diferentes maneras. Aplicando la regla 2 podríamos llegar a que se trata de una Wacka. Aplicando la regla 4, se trata de una Wacka, con una exactitud de 53 instancias sobre 26 erróneas. En otro caso podríamos aplicar la regla 6 con opción de que sea una Hobexa, con una exactitud de 134 instancias sobre 15, que son incorrectas.

- Un bicho de piel suave, color rojizo y con cuatro patas, podemos clasificarlo aplicando la regla 3 como una Wacka, aunque sólo encontramos 6 instancias con algunas de las características mencionadas. Aplicando la regla 5 llegaríamos a la conclusión de que se trata de un Wurro, con 167 instancias correctas sobre 4 erróneas. En otro caso podríamos aplicar la regla 6 con opción de que sea una Hobexa, con una exactitud de 134 instancias sobre 15, que son incorrectas.

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.

Name: Sexo		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
No.	Label	Count	Weight
1	0	470	470.0
2	1	1731	1731.0

Viajaban 1731 varones en el Titanic y 470 mujeres

Name: Edad		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
No.	Label	Count	Weight
1	0	109	109.0
2	1	2092	2092.0

Viajaban 109 menores de edad.

Name: Clase		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
No.	Label	Count	Weight
1	0	885	885.0
2	1	325	325.0
3	2	285	285.0
4	3	706	706.0

Viajaban 321 de primera clase.

minMetric	0.85
numRules	26

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)

```

Las 5 últimas reglas (22, 23, 24, 25, 26)

- R22: Los viajeros de tercera clase, varones, que son 510. Pues 462 eran adultos, con una confianza del 91%
- R23: Las mujeres, que eran 470. Pues 421 eran adultas, con una confianza de 90%.
- R24: Los viajeros de tercera clase que no sobrevivieron, que eran 528. Pues 476 eran adultos, con una confianza del 90%
- R25: Los que no sobrevivieron, que eran 1490. Pues 1329 eran adultos varones, con una confianza del 89%
- R26: Los viajeros de tercera clase, que eran 706. Pues 627 eran adultos, con una confianza del 89%

Pregunta 5: A la vista de los datos relativos a cada clúster, ¿qué grupo crees que representa mejor a los estudiantes que van a aprobar la asignatura? ¿Y a los que van a suspenderla?

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

Los que representa mejor a los estudiantes que van a aprobar la asignatura es **total_time_assignment**

Los que representa mejor a los estudiantes que van a suspenderla es **n_quiz**