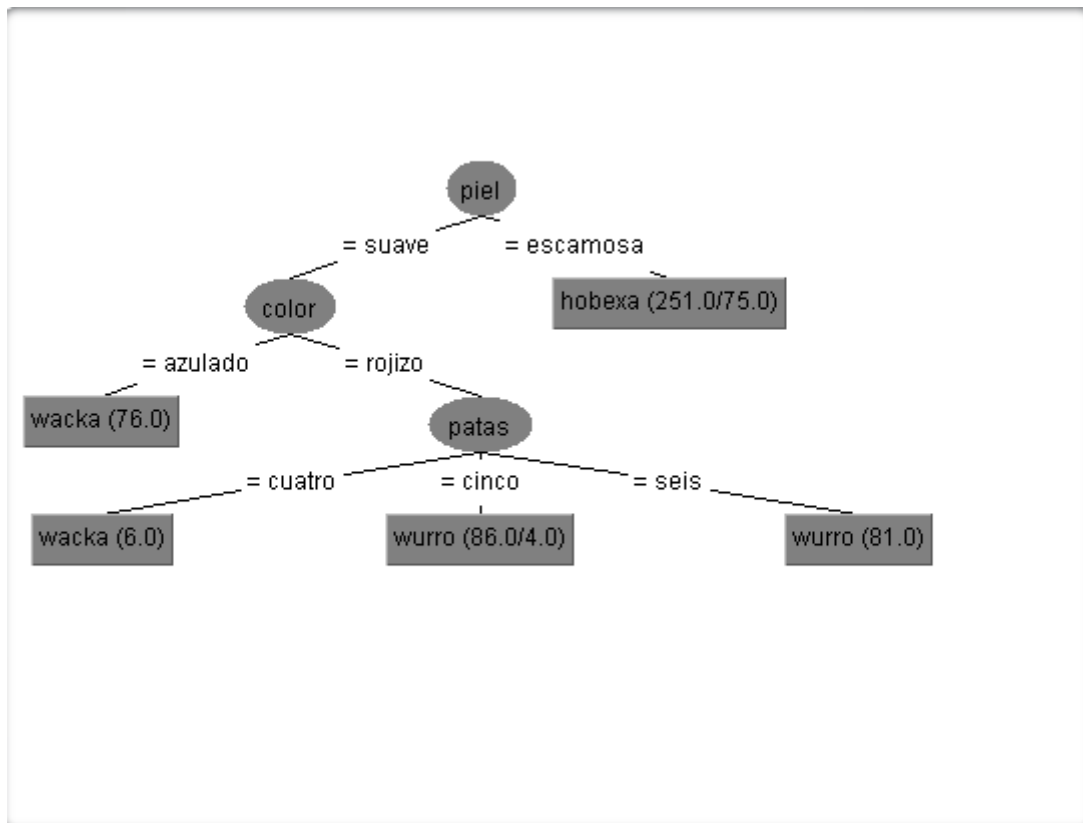


## Práctica 5 – Introducción al lenguaje computacional con WEKA

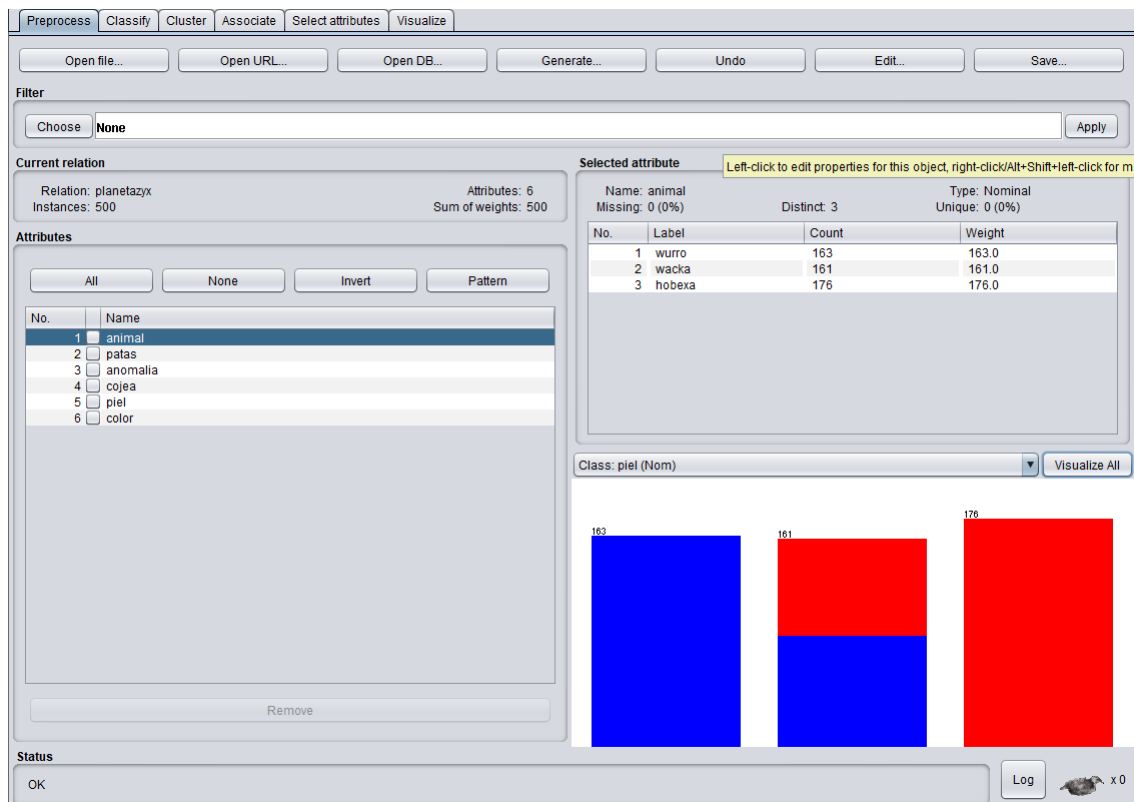
**Pregunta 1:** ¿Cómo clasificaríamos a un bicho rojizo que cojea según este modelo? ¿Y a un bicho de piel escamosa? ¿Y a un bicho de piel suave, rojizo y con cuatro patas?



Dado este modelo de, un bicho rojizo que cojea no podría ser clasificado como ningún tipo de animal, ya que cojea no existe como nodo en el árbol y además este modelo está preparado para clasificar animales empezando por su tipo de piel, dato que no tenemos. Según el modelo obtenido un bicho de piel escamosa puede ser clasificado como hobexa. Y un bicho con piel suave, rojizo y con cuatro patas será clasificado como wacka.

**Pregunta 2:** ¿Encuentras alguna explicación razonable a que los errores de clasificación se cometan con las wackas?

Según este modelo de decisión, si un animal tiene piel escamosa, directamente concluye que se trata de una hobexa, sin embargo, si miramos los datos en el histograma podemos observar que hay un gran número de wackas que tienen piel escamosa, con lo cual estas serán clasificadas como hobexas.



Veamos qué ocurre para que se confundan wackas por wurros. Según el modelo, un animal con piel suave, rojizo y con cinco patas es un wurro, pero vemos que hay wackas que tienen estas mismas características y que al ser una minoría son descartados en el proceso de decisión.

**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?

Este modelo va comprobando una serie de reglas de manera descendente y clasificando los animales. Si suponemos un bicho rojizo que cojea, como no cumple ninguna de las proposiciones de la primera a la quinta, este modelo lo clasifica como hobexa. Si un bicho tiene piel escamosa, ocurre como el caso anterior y se clasificaría como hobexa. Si un bicho tiene piel suave, rojizo y tiene cuatro patas entonces usando la tercera regla, el modelo lo clasifica como wacka.

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

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
```

**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.

Viajaban 1731 varones según la regla 8, ya que Sexo=1 (masculino). Como tenemos que el número de instancias es 2201, entonces el número de mujeres que viajaban en el Titanic es  $2201 - 1731 = 470$ . Mirando en el histograma vemos que viajaban 109 menores de edad y había 325 viajeros en primera clase.

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)

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Generate... Undo Edit... Save...

Filter: Choose **None** Apply

Current relation: Relation: titanic.bt Instances: 2201 Attributes: 4 Sum of weights: 2201

Attributes: All None Invert Pattern

No.	Name
1	Clase
2	Edad
3	Sexo
4	Sobrevivió?

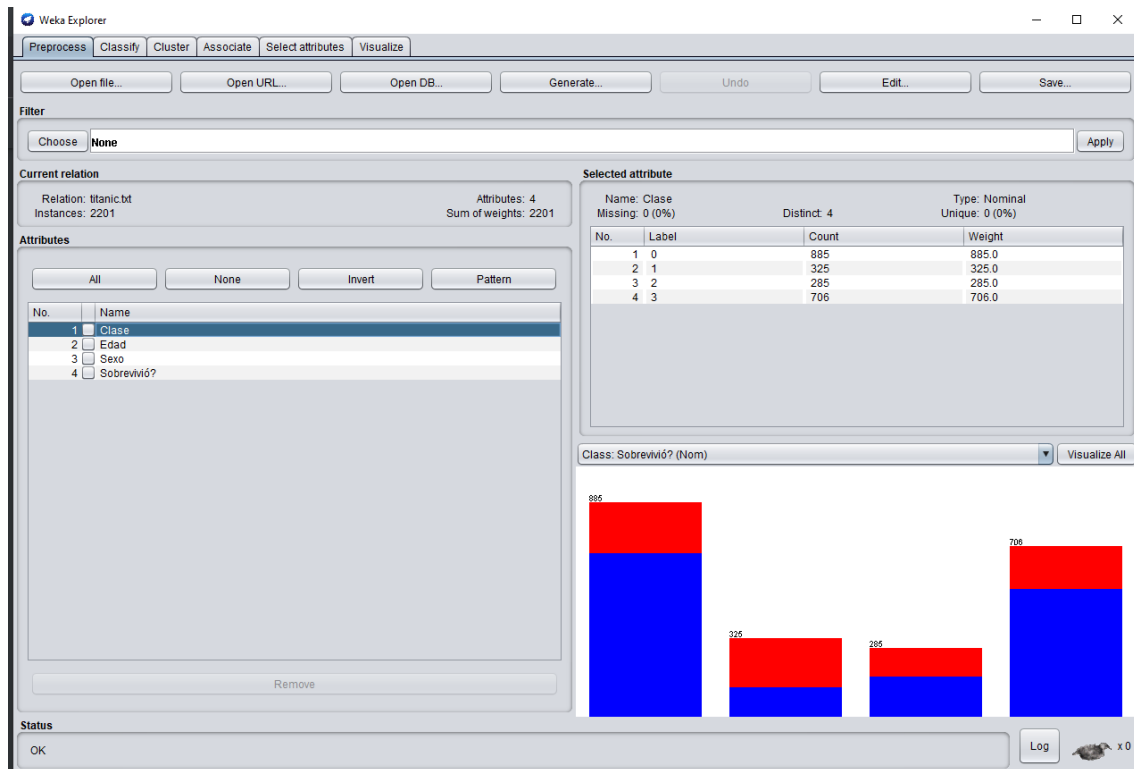
Remove

Selected attribute: Name: Edad Missing: 0 (0%) Distinct: 2 Type: Nominal Unique: 0 (0%)

No.	Label	Count	Weight
1	0	109	109.0
2	1	2092	2092.0

Class: Sobrevivió? (Nom) Visualize All

Status: OK Log x0



Tras modificar los parámetros del algoritmo, obtenemos nuevas reglas. Las 5 últimas tienen la siguiente interpretación:

22. Si viaja en clase 3 y es varón entonces muy probablemente es asulto.
23. La mayoría de mujeres son adultas.
24. Si viajaba en clase 3 y no sobrevivió entonces muy probablemente era adulto.
25. La mayoría de personas que no sobrevivieron eran varones adultos.
26. La mayoría de personas que viajaban en clase 3 eran adultas.

**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?

El cluster 1 es el que representa mejor a los estudiantes que van a aprobar, ya que son los que han dedicado más tiempo total a realizar las tareas, los tests y estar en el foro. Los que más probabilidad tendrán de suspenderla serán los del cluster 0.