

# Práctica 5 - TAA

Pablo Martín de Benito

---

## Ejercicio 1

Algoritmos con opciones por defecto

| Conjunto de Datos → contact-lenses |      |       |        |      |
|------------------------------------|------|-------|--------|------|
| TASAS DE ERROR                     |      |       |        |      |
| J48                                | 1R   | PRISM | RIPPER | PART |
| 0,166667                           | 0,29 | 0,29  | 0,25   | 0,17 |

Si vemos la generación de las reglas con el Weka, podemos ver que efectivamente quedan las mismas reglas que en los ejemplos plasmados en teoría por lo que serían válidos.

Así pues, obtenemos las tasas de error registradas en la tabla en donde podemos que el algoritmo J48 es el que mejores resultados nos da junto al algoritmo PART, como segundas opciones, tenemos el RIPPER y 1R y por último es el PRISM, que además de tener la peor tasa de error, vemos que no clasifica todas las instancias.

---

## Ejercicio 2

Mismos algoritmos con varios conjuntos de datos.

| Iris           |      |       |        |      |
|----------------|------|-------|--------|------|
| TASAS DE ERROR |      |       |        |      |
| J48            | 1R   | PRISM | RIPPER | PART |
| 0,04           | 0,08 | 0,07  | 0,05   | 0,06 |

| soybean        |     |       |        |      |
|----------------|-----|-------|--------|------|
| TASAS DE ERROR |     |       |        |      |
| J48            | 1R  | PRISM | RIPPER | PART |
| 0,084919       | 0,6 | -     | 0,08   | 0,08 |

| vote           |      |       |        |      |
|----------------|------|-------|--------|------|
| TASAS DE ERROR |      |       |        |      |
| J48            | 1R   | PRISM | RIPPER | PART |
| 0,036782       | 0,04 | -     | 0,05   | 0,05 |

| thoracic_surgery |      |       |        |      |
|------------------|------|-------|--------|------|
| TASAS DE ERROR   |      |       |        |      |
| J48              | 1R   | PRISM | RIPPER | PART |
| 0,155319         | 0,17 | 0,17  | 0,15   | 0,21 |

| Statlog (Vehicle Silhouettes) |      |       |        |      |
|-------------------------------|------|-------|--------|------|
| TASAS DE ERROR                |      |       |        |      |
| J48                           | 1R   | PRISM | RIPPER | PART |
| 0,275414                      | 0,48 | 0,25  | 0,31   | 0,28 |

### ***Conclusión de resultados***

Por lo que podemos ver, en general, el algoritmo J48 parece ser el que mejor funciona, nótese también que los algoritmos RIPPER, PART funcionan bien y obtenemos resultados parecidos a los de J48.

También debemos mencionar que el algoritmo PRISM, solo admite variables nominales/catógicas como variable objetivo, por lo que debemos discretizar la variable continua para poder ejecutarlo.

No obstante, en los conjuntos de datos *soybean* y *vote*, tenemos valores *missing* en los datos por lo que tampoco podemos ejecutar el algoritmo PRISM por ese motivo.

Cabe destacar que podríamos eliminar los valores *missing* con Weka pero observamos que el conjunto de datos tiene gran cantidad de estos valores por lo que estaríamos suprimiendo parte del data set, por lo que podría no ser muy creíble.

### ***Conclusión de práctica***

En esta práctica he aprendido a manejar en mejor medida Weka, a utilizar los distintos algoritmos basados en reglas e interpretar su tasa de error.

Además he aprendido a discretizar una variable con el propio Weka utilizando los filtros, lo cual me ha sido de gran utilidad para poder ejecutar el algoritmo PRISM.