

# Aprendizaje automático

## 1. ¿Qué atributos son más importantes según el árbol de decisión?

Cuando se entrena el modelo con J48, por lo general, petal length o petal width son los valores que se presentan como nodo raíz en el árbol resultante. Esto quiere decir que son las características más estándar para distinguir las clases de iris.

- El ancho del pétalo: es la característica que mejor diferencia entre Iris-setosa y las otras dos especies. Cuando se tiene un tamaño pequeño (menor de 0.6 cm), por lo general se clasifica como Setosa.
- Longitud del pétalo: contribuye a distinguir entre Virginica y Versicolor.

- La longitud y el ancho del sépalo se encuentran en los niveles más bajos del árbol o no se utilizan para nada, lo que indica que contribuyen menos a la clasificación.

En resumen: los atributos de longitud y anchura del pétalo son los más relevantes, mientras que los atributos del sépalo juegan un rol secundario.

## 2. ¿Qué ocurre si eliminas un atributo y vuelves a entrenar el modelo?

Depende de qué atributo se quite:

- Si se elimina la longitud o el ancho de los pétalos, la exactitud del modelo disminuye considerablemente. El árbol pierde su habilidad de distinguir las clases de manera precisa y surgen más equivocaciones entre Virginica y Versicolor.
- Si se elimina el ancho del sépalo, el impacto es pequeño, ya que esta característica tiene una correlación baja con la clase. El modelo continúa operando con una exactitud aproximada del 95%.
- Si se elimina la longitud del sépalo, disminuye ligeramente la precisión, aunque no tanto como al eliminar las propiedades de los pétalos.

Esto demuestra que no todos los rasgos tienen el mismo valor en la categorización. Los atributos de los pétalos son esenciales, pero los de los sépalos son complementarios.

3. ¿Cuál de los modelos utilizados consideras que funciona mejor con este conjunto de datos (balance entre precisión y simplicidad)?

Todos los clasificadores consiguen buenos resultados con el Iris, aunque cada uno tiene sus puntos fuertes. J48 consigue una precisión alta y es fácil de interpretar, ya que muestra un árbol claro con los atributos más importantes. NaiveBayes es simple y rápido, pero tiene menos capacidad de explicación. Aunque SMO es complicado y difícil de entender, generalmente proporciona mayor precisión. Aunque depende del valor de  $k$  y no produce un modelo explícito, IBk también es eficaz.

En conclusión, J48 ofrece un balance más óptimo entre simplicidad y rendimiento, ya que combina una buena exactitud con la capacidad de ser interpretado, aunque SMO puede ser el más preciso.