=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: iris

Instances: 150

Attributes: 5

sepallength

sepalwidth

petallength

petalwidth

class

Test mode: 10-fold cross-validation

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

J48 pruned tree

------------------

petalwidth <= 0.6: Iris-setosa (50.0)

petalwidth > 0.6

| petalwidth <= 1.7

| | petallength <= 4.9: Iris-versicolor (48.0/1.0)

| | petallength > 4.9

| | | petalwidth <= 1.5: Iris-virginica (3.0)

| | | petalwidth > 1.5: Iris-versicolor (3.0/1.0)

| petalwidth > 1.7: Iris-virginica (46.0/1.0)

Number of Leaves : 5

Size of the tree : 9

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 144 96 %

Incorrectly Classified Instances 6 4 %

Kappa statistic 0.94

Mean absolute error 0.035

Root mean squared error 0.1586

Relative absolute error 7.8705 %

Root relative squared error 33.6353 %

Total Number of Instances 150

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0,980 0,000 1,000 0,980 0,990 0,985 0,990 0,987 Iris-setosa

0,940 0,030 0,940 0,940 0,940 0,910 0,952 0,880 Iris-versicolor

0,960 0,030 0,941 0,960 0,950 0,925 0,961 0,905 Iris-virginica

Weighted Avg. 0,960 0,020 0,960 0,960 0,960 0,940 0,968 0,924

=== Confusion Matrix ===

a b c <-- classified as

49 1 0 | a = Iris-setosa

0 47 3 | b = Iris-versicolor

0 2 48 | c = Iris-virginica

**Pregunta 1:** Describe los datos del problema indicando:

* Número de clases indicando que representan dichas clases y el tipo de valor que toman.

El conjunto de datos contiene 3 clases de 50 instancias cada una, donde cada una representa a la planta de iris, tomando un tipo de valor numérico.

* Número de instancias en total.

150

* Número de instancias pertenecientes a cada clase.

Iris-setosa = 50

Iris-versicolor = 50

Iris-virginica = 50

* Número de atributos de entrada, su significado y tipo.

Atributos = 5

|  |  |  |
| --- | --- | --- |
| Atributo | Significado | Tipo |
| Sepallength | Largo del sépalo en cm | Numérico |
| Sepalwidth | Ancho del sépalo en cm | Numérico |
| Petallength | Largo del petalo en cm | Numérico |
| Petalwidth | Ancho del petalo en cm | Numérico |
| class | clases | Nominal |

* ¿Hay algún valor de atributo desconocido?

Ninguno

**Pregunta 2:** ¿Por qué crees que no se puede seleccionar el clasificador ID3?

Por qué el algoritmo ID3 se utiliza para atributos de tipo nominal y este ejemplo son de tipo numérico.

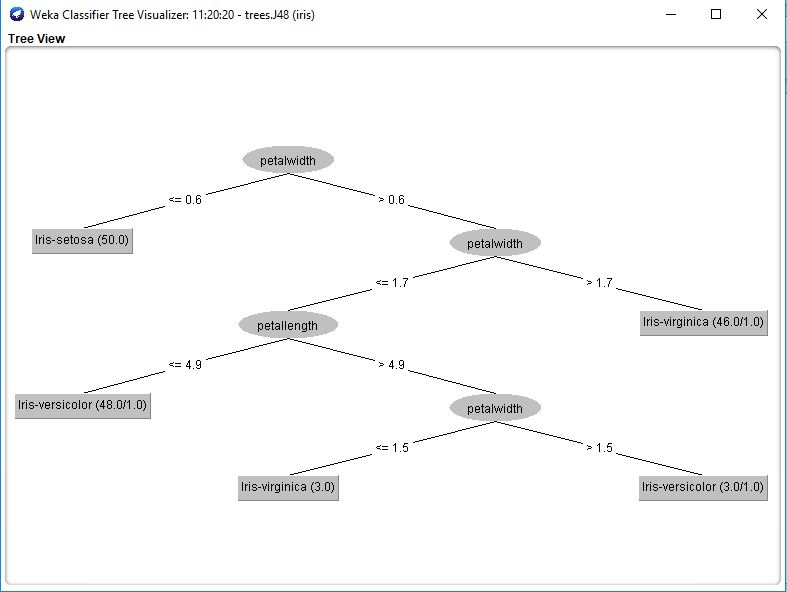
**Pregunta 3:** Dibuja el árbol resultante de ejecutar el algoritmo C4.5 con Weka en el paso 6.

Ilustración 1: árbol de decisión del dataset iris de weka

**Pregunta 4**: Una instancia nueva con los siguientes valores de atributos, según el árbol de decisión generado, ¿en qué clase se clasificaría? Indica cómo has efectuado la clasificación de la instancia.

Petalwidth= 1.6 cm

Petallength= 4 cm

Sepalwidth= 3.4 cm

Sepallength= 6 cm

Se clasificaría en una iris-versicolor, el valor de petalwidth = 1.6, cuando ingresa al nodo raíz (petalwidth) puede irse solo por la rama (>0.6), luego ingresa el siguiente nodo (petalwidth) y solo puede ingresar en la rama (<=1.7), luego ingresa en el nodo (petallength) y solo puede moverse por la rama 4cm<4.9 e ingresa al siguiente nodo clase (iris-versicolor)

**Pregunta 5:** ¿Qué significan los números que Weka muestra entre paréntesis junto a los nombres de las clases en las hojas del árbol generado?

El número de instancias que pertenecen a la clase, separadas por una barra (/) de las clases que no pertenecen a la clase.

**Pregunta 6:** De acuerdo al resumen de resultados de la validación cruzada con 10 iteraciones, que es una medida de los errores que se esperarían si se aplicara un conjunto de datos de prueba independiente, ¿se puede considerar que el árbol generado en el paso 6 (utilizando el conjunto completo de datos de entrenamiento) es optimista respecto a la precisión en la predicción? ¿Por qué?

Siguiendo los pasos que se plantean, las instancias correctamente clasificadas dan un 96% que es optimista en la precisión de la predicción.

**Pregunta 7:** De acuerdo a la matriz de confusión obtenida, de las instancias que pertenecen a la clase Iris-setosa, ¿cuántas de ellas son clasificadas correctamente como de esa clase?¿Cuántas son incorrectamente clasificadas y en qué clase son incorrectamente clasificadas?

Iris-setosa = 49 son clasificadas como verdaderas y 1 es clasificada como iris-versicolor