UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

ALMACÉN Y MINERÍA DE DATOS

PRÁCTICA 12

- **CALDERÓN FERNÁNDEZ GABRIEL**
- **FLORES GONZÁLEZ LUIS BRANDON**
- **SANTAELLA MARÍN HÉCTOR**

1) Liste algunas de las ventajas que ofrece un árbol de decisión y qué los hace tan populares con respecto a otros métodos de predicción.

-Plantean el problema para que todas las opciones sean analizadas.
-Permiten analizar totalmente las posibles consecuencias de tomar una decisión.
-Proveen un esquema para cuantificar el costo de un resultado y la probabilidad de que suceda.
-Nos ayuda a realizar las mejores decisiones sobre la base de la información existente y de las mejores suposiciones.

2)En un árbol de decisión ¿un atributo puede ser revisado en más de una ocasión?, ¿existen atributos que, por el contrario, sean revisados una sola vez? (Justifique)

Si utiliza demasiados atributos de predicción o de entrada al diseñar un modelo de minería de datos, el modelo puede tardar mucho tiempo en procesarse o incluso quedarse sin memoria. Algunos métodos para determinar si hay que dividir el árbol son las métricas estándar del sector para la *entropía* y las redes Bayesianas. Para obtener más información sobre los métodos que se usan para seleccionar los atributos significativos y, después, puntuarlos y clasificarlos por lo que esto conlleva a que no se vuelva a utilizar el atributo para la siguiente prueba. En un árbol de decisión un atributo puede ser revisado más de una vez para determinar un conjunto de datos.

3)¿Cómo y por qué puede utilizarse un árbol de decisión como método de reducción de datos?

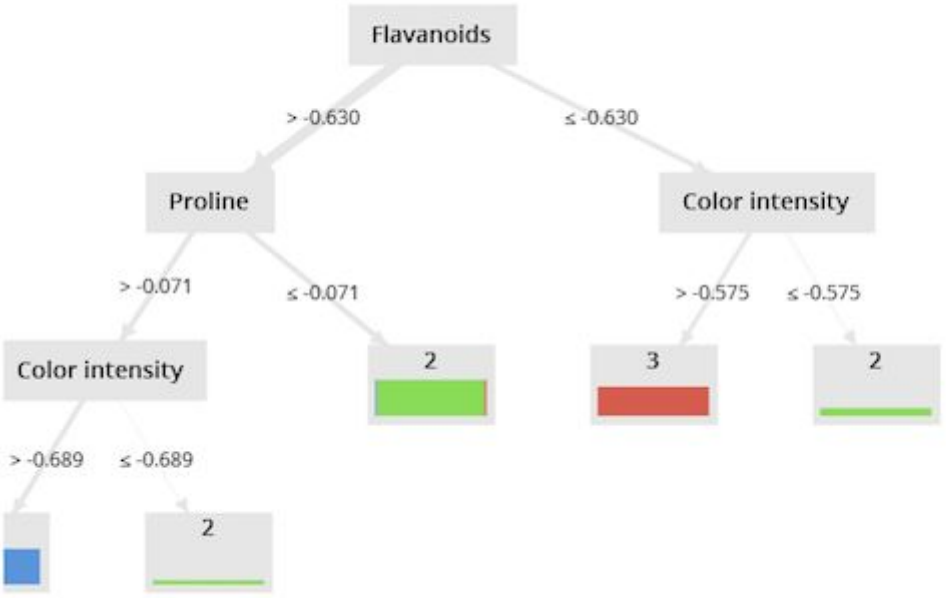
Depende del método que se use para generar el árbol se puede reducir de datos. Por ejemplo en ID3, se reduce al mínimo la información necesaria para clasificar las tuplas en las particiones resultantes y refleja menos aleatoriedad o “impureza” en estas particiones. Donde el atributo para reducir al mínimo es el tiene mayor ganancia de información.

4) Dado un conjunto de datos ¿el árbol de decisión construido a partir de este es único?

No, ya que existen diversos métodos para crear un árbol de decisión y en cada uno puede generar uno diferente dependiendo del enfoque que se use. Un enfoque para probar esto sería en ID3 donde se garantiza encontrar un árbol de forma simple pero no la más simple.

Árbol de decisión construido por RapidMiner.

Para la creación de este árbol se necesita cargar el conjunto de datos en el repositorio local, tendrás que realizar el proceso , para realizar el proceso se tiene que cargar el dataset wine , posteriormente la normalización de los valores de los atributos, ya que hay diversas unidades de medida y por último la creación del clasificador usando un árbol de decisión.



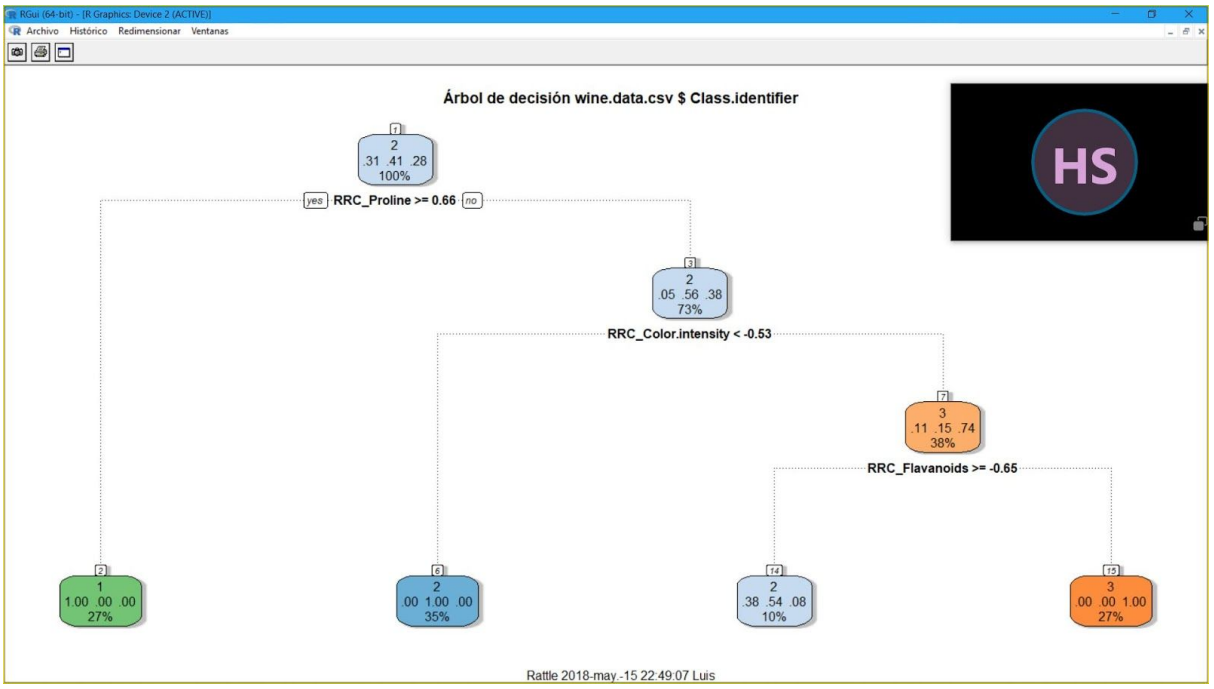
PerformanceVector

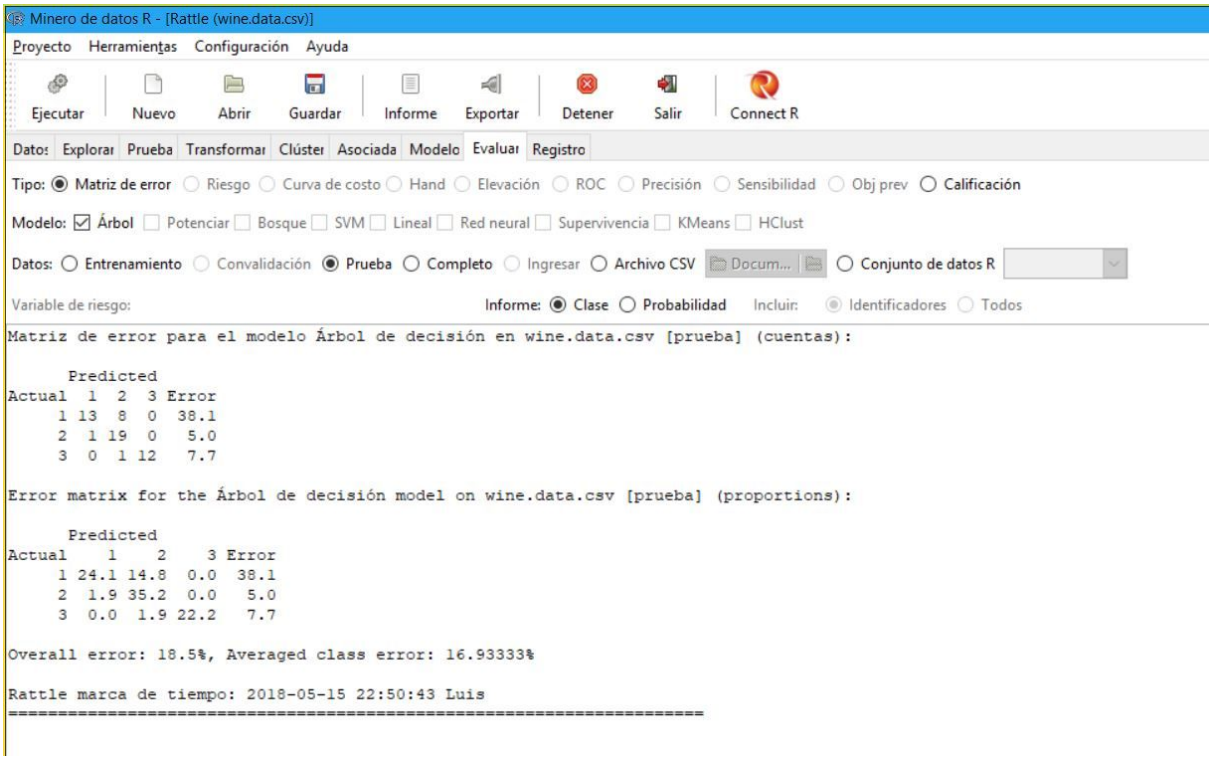
```
PerformanceVector:
accuracy: 88.68%
ConfusionMatrix:
True:   1       2       3
1:      16      2       1
2:      0       18      0
3:      0       3       13
```

Árbol de decisión construido por R.

Para la creación del árbol de decisión en R se necesita una herramienta llamada Rattle , una vez en el programa Rattle seleccionamos el conjunto de datos y los insertamos y ejecutamos. Seleccionamos la la pestaña “Categorica” y ejecutamos nuevamente.

En la partición manejamos los valores 70/30 donde 70 será el porcentaje y 30 la prueba posteriormente ejecutamos para particionar el conjunto normalizado.
Después en la pestaña transformar lo dejamos por default , seleccionamos todos los elementos y normalizamos.
Nos vamos a la pestaña modelo y dibujamos el árbol.





El árbol generado con la herramiento de rattle es mejor ya que al momento de evaluar ambos árboles podemos notar en la matriz que este tiene menor margen de error.

Diccionario de Datos.

1) Alcohol tipo:REAL
2) Malic acid tipo:REAL
3) Ash tipo:REAL
4) Alcalinity of ash tipo:REAL
5) Magnesium tipo:Integer
6) Total phenols tipo:REAL
7) Flavanoids tipo:REAL
8) Nonflavanoid phenols tipo:REAL
9) Proanthocyanins tipo:REAL
10)Color intensity tipo:REAL
11)Hue tipo:REAL
12)OD280/OD315 of diluted wines tipo:REAL
13)Proline tipo:Integer
14)Class identifier tipo:Polyniminal