**Proyecto integrador etapa 1**

**Método de clasificación de I.A.**

|  |  |
| --- | --- |
|  | *Fecha:\*\*/\*\*/\*\*\*\** |
| *Nombre del estudiante:* | |
| *Nombre del docente:* | |

*Bienvenidos a la primera parte del proyecto integrador que tiene por objetivo elegir y relacionar un método* *de clasificación de la I.A. desarrollado en R.*

*Instrucciones: Lee atentamente el caso que se pide y desarróllalo según corresponda; para este ejemplo trabaja con la base de datos Iris disponible en la librería* datasets*.*

**Caso:** En el *datasets* la base de datos se encuentra compuesta por 150 observaciones de flores de la planta iris, las variables o atributos que se miden de cada flor son:

* El tipo de flor como variable categórica, cuyas categorías son: virginica, versicolor y setosa.
* El largo y el ancho del pétalo en cm como variables numéricas.
* El largo y el ancho del sepalo en cm como variables numéricas.

Deberás llevar a cabo la clasificación de árbol:

* Primero se procederá a predecir el tipo de flor a partir de las variables sepal.width y petal.width. Después, se debe realizar la elección de los datos de entrenamiento, para este caso al considerar la validación cruzada tomaremos un αα de 0.7, luego extraemos una muestra proporcional a αα del conjunto de datos Iris y los datos restantes son los de prueba.

Realiza la programación en R que debe mostrar:

1. Las probabilidades con las cuales es probable que una observación se clasifique en cierta categoría.
2. Realiza en R las predicciones de los datos de prueba a partir del algoritmo e interpreta los resultados (utiliza Predicción.obs y *head*).
3. Ilustra el árbol de clasificación e interpreta los resultados
4. Poda el árbol para evitar el sobreajuste. La función prune.tree() permite elegir cuántas hojas queremos que tenga el árbol y devuelve el mejor árbol con el tamaño necesario, realiza el programa en R para 4 hojas.