

## Actividad 2 (FORO)

---

En esta actividad usted deberá aportar con 1 respuesta principal a la o las preguntas enunciadas por el profesor y comentar y/o intervenir 2 respuestas de sus compañeros (en otros hilos de conversación).

Además, junto a la respuesta principal, usted deberá subir un informe del trabajo realizado, el cual también **será considerado en la evaluación del foro**. Se permite 1 solo archivo adjunto (tamaño máximo 800MB).

**Considere los datos 'salary.csv', que contiene información recabada en una empresa, sobre potenciales clientes y cuya descripción se muestra en la siguiente tabla.**

Variable	Descripción
trabajo	Tipo de trabajo
educacion	Nivel educacional
estado civil	Estado civil
ocupacion	Ocupación
familiar	Rol en su hogar
raza	Raza
sexo	Sexo
pais	País de origen
clase	Indica si tiene un sueldo anual mayor a US\$ 50.000

Realice las siguientes actividades:

- 1) Cargue el conjunto de datos en la sesión de trabajo de R usando la función `read.table`.
- 2) Seleccione de manera aleatoria 2/3 de los datos para crear sus datos de entrenamiento y guarde el tercio restante para objeto de validación. Para esto, simule 9.768 valores 1 y 2 en proporciones 2/3 a 1/3 a través de la función `sample()`. Utilice aquellas tuplas de la base de datos asociadas al valor 1 para la base de entrenamiento, y las restantes para validación. Utilice la semilla 1, mediante el comando `set.seed(1)`.
- 3) Construya un árbol de decisión para la variable `clase`, utilizando como criterio el índice de Gini. Realice el procedimiento completo, incluyendo la poda del árbol, usando los comandos `rpart()`, `cptable()` y `prune()` de la librería `rpart`. Use la opción `cp = 0.012` en el proceso de poda.
- 4) Utilizando la función `naiveBayes()` de la librería `e1071` construya un clasificador de Bayes Ingenuo. Utilice la corrección de Laplace en caso de haber celdas vacías, `laplace = 1`.
- 5) Utilizando los datos de validación, calcule:

- a. La sensibilidad del procedimiento de clasificación, definida como el porcentaje de personas para las que el modelo predice un sueldo anual mayor a US \$ 50.000, dentro de todas aquellas que en realidad tienen un sueldo mayor a dicha cantidad.
- b. La especificidad del procedimiento de clasificación, definida como el porcentaje de personas para las que el modelo predice un sueldo anual menor o igual a US \$ 50.000, dentro de todas aquellas que en realidad tienen un sueldo menor o igual a dicha cantidad.
- c. La precisión del procedimiento de clasificación, definida como el porcentaje de personas clasificadas correctamente.

Interprete cada uno de los resultados obtenidos y compare los clasificadores. Utilice la función `predict()` para realizar las predicciones; puede utilizar la función `mmetric()`, de la librería `rminer` para verificar sus resultados.

- 6) Discuta sus resultados con sus compañeros mediante el foro
-