

Árboles de Decisión

- 1 Ajustar un modelo para el conjunto de datos. El conjunto de datos contiene información de una población con la cual se intenta predecir si puede o no cobrar más de 50.000 dólares anuales.
 - 1.1.1 Separe las poblaciones en entrenamiento y validación en forma aleatoria en 70/30 (genere una semilla aleatoria e indique el valor de la semilla en el software que utilizó). Indique que cantidad de casos quedaron para cada ambiente.
 - 1.1.2 Ajuste el mejor modelo posible . Indique el modelo con todas sus variables.
 - 1.1.3 Calcular la performance y el AUC y el gráfico ROC en entrenamiento y validación indicando también el total de casos de cada una de las clases.
 - 1.1.4 Seleccione el 25% de los individuos en el ambiente de validación de acuerdo a la siguiente lógica. Entregue los resultados indicados:
 - Al azar e indique la cantidad de individuos que cobran más de 50.000 dólares.
 - Utilizando el modelo desarrollado en el punto 1.1.2 e indique la cantidad de individuos que cobran más de 50.000 dólares.
 - 1.1.5 Entregue un conjunto de datos en formato texto con el siguiente formato:
Caso
Indicador de Entrenamiento o Validación
Clase
Probabilidad calculada con el modelo utilizado para resolver el punto 1.1.4