

Investigación de Operaciones

Control 5: Árboles de Clasificación

Instrucciones:

- El escrito debe ser realizado en \LaTeX o en Word
- Todo árbol debe ser incluido en el escrito (no adjuntar archivos extra).
- JUSTIFIQUE TODAS sus respuestas. En caso de no presentar análisis y justificación escrita en una pregunta, su puntaje será de cero puntos (aunque adjunte un árbol relacionado a la pregunta).
- Una vez realizada la evaluación, subir el escrito en formato '.pdf'.

1. Vamos a tomar una cerveza - Parte 2 (100 puntos)

B. Gomez vuelve a tener un problema y necesita de su ayuda para resolverlo. Con los últimos acontecimientos nacionales/mundiales, B. Gomez decidió implementar un servicio de suscripción para sus clientes. Esto implica que por una cantidad monetaria mensual, cada cliente recibe una caja de especiales cervezas en la primera semana de cada mes. Durante el mes pasado, un programa piloto se lanzó en su página web y ha registrado la información generada en su web. El objetivo es intentar predecir el comportamiento de sus clientes con la intención de identificar de mejor forma su público objetivo y así sacarle más provecho al sistema de suscripciones. Los atributos capturados durante el mes pasado son los siguientes:

1. Cantidad de veces que una persona ha comprado en la web (en los últimos 2 años)
2. Gasto promedio de dinero en esas compras
3. Cantidad de Likes promedio en Instagram
4. Cantidad de Likes promedio en Facebook
5. Cantidad Promedio de minutos navegando en la Web de la cervecería
6. Edad (años)
7. Ciudad del cliente (1-10: Santiago, Valparaíso, Concepción, Arica, Iquique, Rancagua, Talca, Chillán, Punta Arenas, Puerto Montt).
8. Sexo (0: M, 1:F)
9. Variable objetivo: Cliente se suscribe (0 o 1)

Utilizando el software **R** y **seteando la semilla con el día de su nacimiento**¹, conteste las siguientes preguntas:

1. (40 puntos) Generar un árbol considerando el 75 % de los datos entregados para Training. Conteste lo siguiente:
 - (10 p) Describa los datos utilizados, explicando el tipo de dato de cada atributo.
 - (15 p) Describa el árbol obtenido: cantidad de niveles obtenido, cantidad de hojas del árbol obtenido, entre otros atributos.
 - (15 p) ¿Qué variables no generan particiones?, explique porqué éstas variables no participan en el árbol generado.
2. (30 puntos) Nuevamente, considerando el 75 % de los datos entregados para Training, ¿Qué sucede si las variables cantidad de Likes promedio en Instagram y Facebook no son consideradas en el modelo? describa detalladamente:
 - (20 p) Compare con el árbol obtenido en la pregunta (1) considerando cantidad de niveles del árbol, cantidad de hojas del árbol obtenido, entre otros atributos.
 - (10 p) Considerando el Testing set, ¿Cuál es la precisión del árbol?
3. (20 puntos) ¿Qué sucede si utiliza un 60 % y 90 % de los datos entregados como Training? Evaluar e imprimir los árboles obtenidos. Explique.
4. (10 puntos) Considerando el árbol de la pregunta (1), ¿cuál sería el resultado asignado al siguiente cliente: (8,91,82, 0,35,6,68,4,1)

NOTA: Datos disponibles en Aula. El escrito debe contener el código (texto, no imágenes) para obtener cada resultado.

¹Por ejemplo, si usted nació el 32 de Diciembre del 1995: `set.seed(32121995)`