

Investigación de Operaciones

Informe Árboles de Clasificación

Profesor: Nicolás Rojas M.

Ayudantes: Felipe Vega V. - Francisca Ramirez E.

12 de Julio del 2019

1. Objetivo

El objetivo de este trabajo es utilizar una herramienta para construir árboles de clasificación, evaluando el efecto de utilizar diferentes conjuntos de datos, atributos y tipos de atributos. Además, permitirá conocer y aprender a utilizar parte del software R.

2. Enunciado

El Departamento de Informática necesita crear un sistema que ayude a determinar qué estudiantes podrían aprobar una asignatura, considerando distintas características sobre ellos. El objetivo es clasificar a los estudiantes entre aquellos que podrían obtener un promedio final sobre o bajo 55. Además, los estudiantes fueron encuestados y entregaron información sobre su estado actual.

1. Describir el conjunto de datos: Cantidad de datos, Atributo Predictor/Clasificador, Tipo de Dato por atributo, Valores posibles.
2. Construir e incluir un árbol de clasificación con los datos. Evalúe el árbol, interprete los resultados. Entregue el error de clasificación
3. Contestar las siguientes preguntas:
 - a) ¿Cuáles fueron las variables relevantes en la construcción del árbol? Explique por qué una variable es más relevante que otra.
 - b) ¿Qué sucede si se modifica el tipo de dato de *VTR*?
 - c) Si la variable *VTR* no es incluida en la construcción del árbol, ¿Qué sucede con el árbol? Evalúe los cambios
 - d) Si la variable *Inasistencias* no es incluida en la construcción del árbol, ¿Qué sucede con el árbol? Evalúe los cambios
 - e) ¿Qué sucede si utiliza un 30, 50 y 70 % de los datos entregados como Training? Evaluar e imprimir los árboles obtenidos. Explique.

3. Requerimientos e Informaciones

- En cada pregunta, mencione qué comandos y funciones debió utilizar para realizar cada requerimiento.
- **El trabajo es individual.**
- Los informes deben ser enviados vía mail a nicolas.rojasm@usm.cl el día Jueves 8 de Agosto a las 11:59 AM. **No se aceptan informes atrasados.**
- El software R está disponible en <https://www.r-project.org/>
- Es obligación detallar y fundamentar adecuadamente cada respuesta

- Debe incorporar imágenes **LEGIBLES** de los árboles obtenidos en cada pregunta.
- **Los datos del trabajo están disponibles en Moodle**
- El set de datos considera los siguientes atributos:
 - Sexo - Masculino (M) / Femenino (F)
 - Horas de Estudio Semanales - < 2 horas, $[2 - 5]$ horas, $[5 - 10]$ horas, > 10 horas
 - VTR - Veces que ha tomado el Ramo: 0, 1 o 2.
 - Tiempo Libre: Nada, Poco, Normal, Mucho o Demasiado
 - Carrete: Nada, Poco, Normal, Mucho o Demasiado
 - Salud: Muy Mala, Suficiente, Normal, Buena, Muy Buena
 - Inasistencias - cantidad de inasistencias actual: $[0, 10]$
 - Nota Final: < 55 o ≥ 55

Cualquier duda relacionada sobre el enunciado o con los contenidos favor escribir a francisca.ramirez.12@sansano.usm.cl, felipe.vega.14@sansano.usm.cl o nicolas.rojasm@usm.cl.