

Proyecto 4

“Árboles de Decisión”

Intelligent Systems

Descripción del problema

● Data Set

- [Churn.csv](#) contiene información de 3333 clientes de diferentes compañías móviles. Fuente: [BigML](#)
- Cada registro contiene los siguientes atributos:

State	Account Length	Area Code	<u>Phone</u>
Int'l Plan	VMail Plan	VMail Message	
Day Mins	Day Calls	Day Charge	
Eve Mins	Eve Calls	Eve Charge	
Night Mins	Night Calls	Night Charge	
Intl Mins	Intl Calls	Intl Charge	
CustServ Calls	<u>Churn</u>		

Crea una aplicación web con Shiny (1/3)

- **File -> New File -> Shiny web app**
 - Documentación: <http://shiny.rstudio.com/>
- **La aplicación debe permitir lo siguiente:**
 - Subir un fichero
 - Leer el fichero (**eliminar la columna "phone"**)
 - Solicitar los parámetros para definir el árbol de decisión:
 - Minsplit: (integer) 1... 1000
 - Maxdepth: (integer) 1...30
 - Cp: (real) 0...1

Crea una aplicación web con Shiny (2/3)

- **Cuando la columna a predecir esté seleccionada**
 - Realiza un proceso [5-fold cross validation](#)
 - Para cada partición:
 - Divide los ejemplos en training-set y test-set
 - Crea un árbol de decisión para predecir “Churn” con cada training-set.
 - Haz la predicción usando cada test-set
 - Calcula el % de aciertos (compara valor real vs. predicción)
 - Calcula el % medio de aciertos de las 5 particiones.

Crea una aplicación web con Shiny (3/3)

- **Después de terminar todo el proceso:**
 - Muestra el % de aciertos medio
 - Muestra un gráfico de barras con % de aciertos de cada ejecución.
 - Muestra la representación gráfica de uno de los 5 árboles de decisión.

Ayuda para la implementación en R

- **Para crear el “fold cross validation”**
 - Usa la función **createMultiFolds** como en el proyecto 3.
- **Para crear el árbol de decisión**
 - `model <- rpart(formula=Churn~., data=training.data)`
 - **minsplrit, cp y maxdepth con parámetros de la función rpart**
- **Para realizar predicciones**
 - `prediction <- predict(model, test.data, type="class")`
- **Para calcular el % de aciertos**
 - `sum(prediction == test.data$Churn) / nrow(test.data)`
- **Para generar una representación visual del árbol**
 - `rpart.plot(x=model)` ó `prp(x=model)`

Preguntas de teoría

- Con los valores $\text{maxdepth} = 30$ y $\text{cp} = 0$
 - ¿Cuál es el comportamiento de la precisión (% de aciertos) durante el entrenamiento y el test a medida que el parámetro “minsplit” aumenta? ¿Por qué?
 - ¿La precisión crece o decrece?
 - ¿Hay un valor máximo o mínimo para la precisión?
 - ¿Cómo afecta el incremento del “minsplit” a la representación gráfica del árbol de decisión
- Justifica tus respuestas

Formato de la entrega

- **Formato**

- ZIP con el código en R. (nombre de grupo)
 - Script en R (con el código de grupo en el nombre)
 - Documento con las respuestas a la teoría

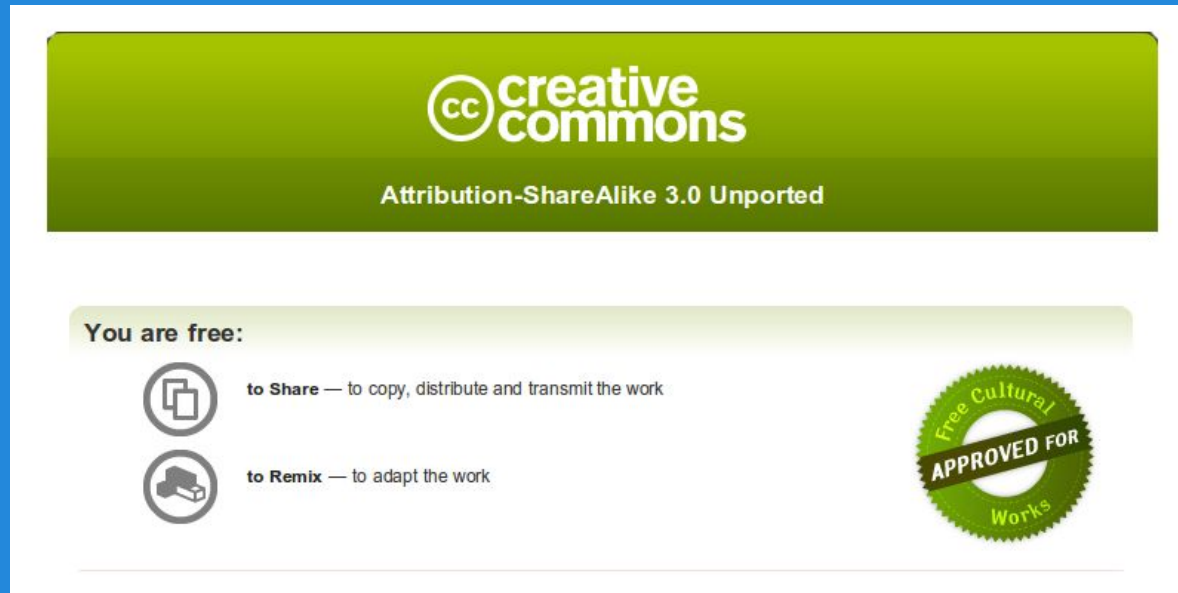
- **Calificación - 10%**

- **Criterios de corrección**

- Corrección de la implementación y la teoría (8,0%).
- Documentación y limpieza del código (2,0%).

Copyright (c) 2016 University of Deusto

This work (but the quoted images, whose rights are reserved to their owners*) is licensed under the Creative Commons “Attribution-ShareAlike” License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/3.0/>



Intelligent Systems