EJERCICIO 1- Preparación de Datos

1. Importa en RM ambos datasets (descarga el archivo “ereader.zip” de la webasignatura.
   a) Agrega ambos datasets a un nuevo proceso, y renombra los operadores Retrieve en forma apropiada
   b) Conecta las salidas, ejecuta el modelo y analiza las características y estadísticas de los atributos de los datasets
      a. Tipos de datos
      b. Outliers y faltantes
      c. Otros….
      d. ¿Cómo afectan?
   c) En el modelo, observar que el atributo **ID** no debería formar parte de los atributos usados para entrenar el modelo, ¿por qué? Para retirarlo podemos hacer dos cosas:
      a. Quitarlo del dataset
      b. Utilizar un operador “Set Role”… ver cómo hacerlo
   d) Recuerda indicar el atributo “label” (se requiere para el AD)
   e) Selecciona un operador “Decision Tree” básico y agrégalo al proceso; conéctalo al canal de entrenamiento.
   f) Observa los parámetros que tiene el operador, y cada una de sus alternativas. ¿Qué algoritmo de base utiliza? ¿cuáles son los criterios de división que admite? ¿cómo funcionan?
   g) Cambia el parámetro “máxima profundidad” a 4 (para poder ver mejor el árbol) y ejecuta el modelo y cambia a la pestaña de “Tree” en los resultados.
   h) En el árbol generado, podemos distinguir nodos internos y hojas. Los primeros son atributos que resultan apropiados, en cada llamada recursiva del algoritmo, para hacer una división.
   i) Podemos observar que el atributo “ActividadWebsite” es el mejor predictor en el modelo
   j) Analiza los caminos a las hojas, y observa las reglas correspondientes
      a. ¿tienen un sentido intuitivo?
      b. Cambia la profundidad máxima y observa los caminos de evaluación
   k) Analiza las diversas formas de visualizar el árbol y los nodos
   l) Al posicionarse sobre una hoja, observa la información desplegada (cantidad de ejemplos referenciados en esa hoja, y proporciones de la clase). Vemos que en algunas hojas existen varias posibilidades, y la hoja refiere a la clase mayoritaria. Al aplicar el modelo, veremos que esto se traduce en factores de confianza.