

LABORATORIO N°1

Ejercicio 3

Explique y detalle como implementaría **Gain Information** para el caso en el que existan missing values (sección 3.7.4 de [1]). Use y de al menos un ejemplo. Puede crear un notebook en R para documentar su respuesta.

Hay varias maneras posibles de lidiar con “missing values”. Una opción es asignar a dichas variables el valor más común del conjunto de entrenamiento en el nodo n. Una posibilidad más compleja es asignar una probabilidad a cada una de los posibles valores en lugar de simplemente asignar el valor más común. Otra opción es por ejemplo usar la media, cuando los valores son numéricos, o la moda, para los casos en que los valores sean nominales.

Ejemplo:

Para el dataset de Tennis.csv si para el atributo Outlook faltaran valores, y el valor más repetido (osea la moda) fuera Sunny, a los valores faltantes podríamos asignarle el valor “Sunny”