

Experimento Propio

Zoe Borrone

Luca Mazzarello

Ignacio Pardo

2023-08-21

Introducción y Metodología

Para desarrollar este experimento se planteo testear el comportamiento de los modelos al aplicarle ruido a los valores de los atributos numericos en distintas proporciones, viendo si esto afecta el desempeño de los modelos.

Para esto se modificó el script de ejemplo `sample_exp.R` para aplicar ruido a cada dataset pasado a la función `run_experiment` en distintas proporciones. El ruido se aplica de la siguiente manera:

Se setean las proporciones de ruido a aplicar `prop_noise` $\in 0, 0.2, 0.4, 0.6, 0.8$.

1. Para cada dataset dado: se obtienen las columnas de atributos numericos.
2. Para cada columna de atributos numericos: se obtiene el valor minimos y el valor maximo.
3. Para cada proporción `prop_noise`,
4. Para cada valor de cada atributo numerico,
5. Se genera un valor aleatorio r entre el minimo y el maximo con una distribución uniforme.
6. Se genera un segundo valor aleatorio uniforme p entre 0 y 1.
7. Si $p < \text{prop_noise}$, se reemplaza el valor original por r .

Luego, cada valor numérico va a haber sido reemplazado por un ruido entre los valores conocidos de la columna, con una probabilidad `prop_noise`.

Como notamos en el experimento anterior (opción 3), los datasets cuentan con distintas proporciones de variables numéricas, el dataset de *Heart* cuenta con 6 variables numéricas de 11 atributos totales, el dataset de *Churn* cuenta con 10 variables numéricas de 13 atributos totales y el dataset de *Student* cuenta con 13 variables numéricas de 30 atributos totales. Si los modelos priorizan los atributos numéricos, esperamos que el ruido afecte más a los modelos de *Heart* y *Churn* que al de *Student*.

Resultados

Como

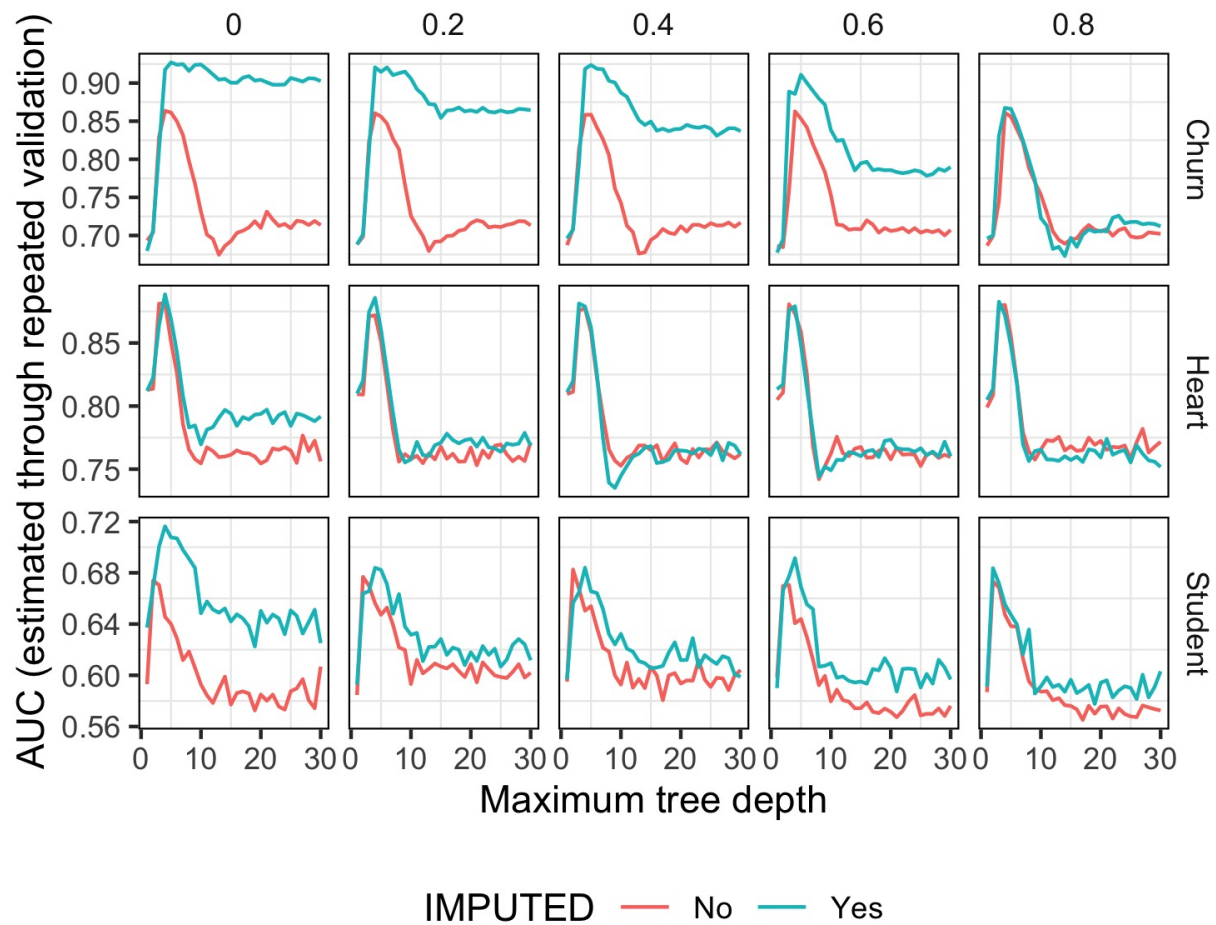


Figure 1: Resultados $\text{prop_noise} \in 0, 0.2, 0.4, 0.6, 0.8$