

# exp\_4.Rmd

Massri

2023-08-21

Se realizó un experimento en el que se analiza el impacto que tiene el ruido en la variable a predecir en un dataset para un modelo de Árbol de decisión. Para ello, se entrenaron múltiples modelos para cada uno de los datasets, lo que nos permite estimar y evaluar la performance de cada uno.

Las variables independientes del experimento son:

- Profundidad del árbol (1-30) - `maxdepth`
- Proporción de switch en la variable a predecir (0-1) - `PROP_SWITCH_Y`

Así, se entrena un modelo para cada profundidad y con diferentes valores para la proporción. Para un rango de valores de `PROP_SWITCH_Y` y cada dataset, se obtiene la estimación de AUC del modelo promediando las 30 repeticiones, y guarda el modelo con mejor performance de todas las profundidades.

En los 3 datasets se obtiene resultados similares que indican que la performance de los modelos seleccionados disminuye a medida que se agrega ruido a la variable a predecir. La peor performance se alcanza con ruido máximo en `PROP_SWITCH_Y=0.5` (ya que en 1 la performance es igual solo que invierte los labels).

```
source("exp_4.R")
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

## Loading required package: lattice

##
## Attaching package: 'caret'

## The following objects are masked from 'package:Metrics':
##
##   precision, recall

## Loading required package: iterators
```

```
## Loading required package: parallel
```

