

## Trabajo Práctico 1 : Reservas de Hotel

### Checkpoint 2

Inicialmente revisamos los tipos de dato y modificamos los valores de tipo objeto.

- Id: fue removida por tener un valor único para cada registro
- Reservation\_status\_date: Fue removida porque revelaba información sobre la cancelación.
- Country: Fue removida porque consideramos que toma muchos valores, requiere mucho poder computacional y no aporta mucho valor.

Codificamos variables categóricas de tipo objeto numéricamente (0, 1 2, etc), en el caso de las binarias como (Yes, No) y utilizamos OneHotEncoding para el resto.

Luego armamos el armamos el árbol de decisión, buscando los mejores hiperparámetros.

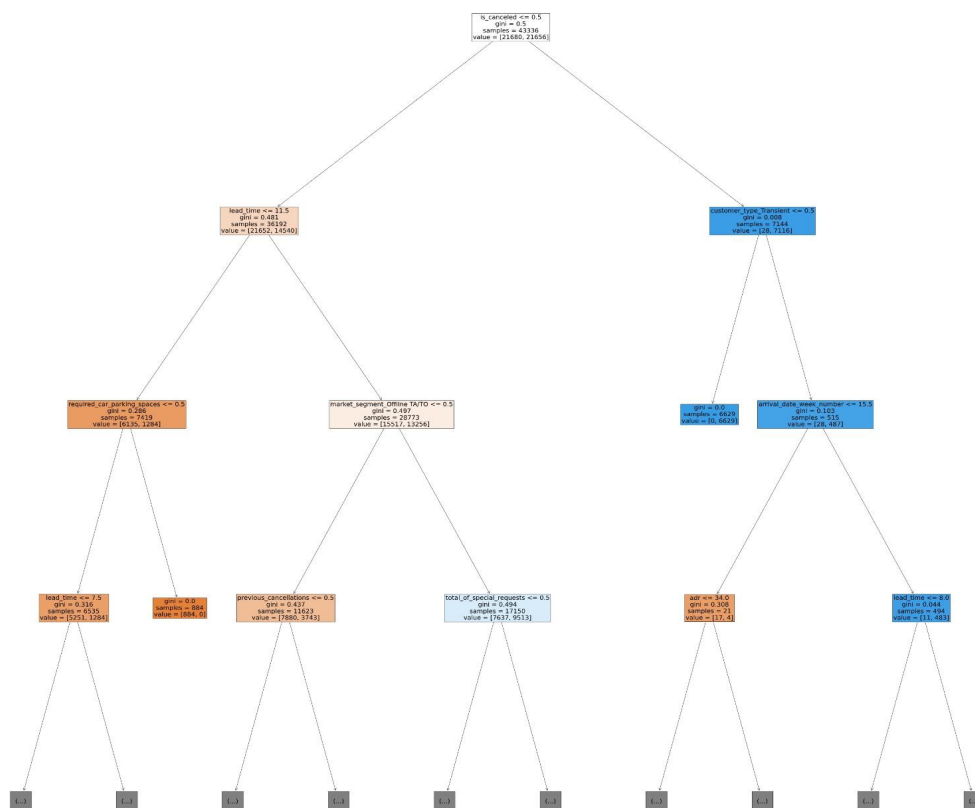
Los testeamos con folds entre 5 y 10.

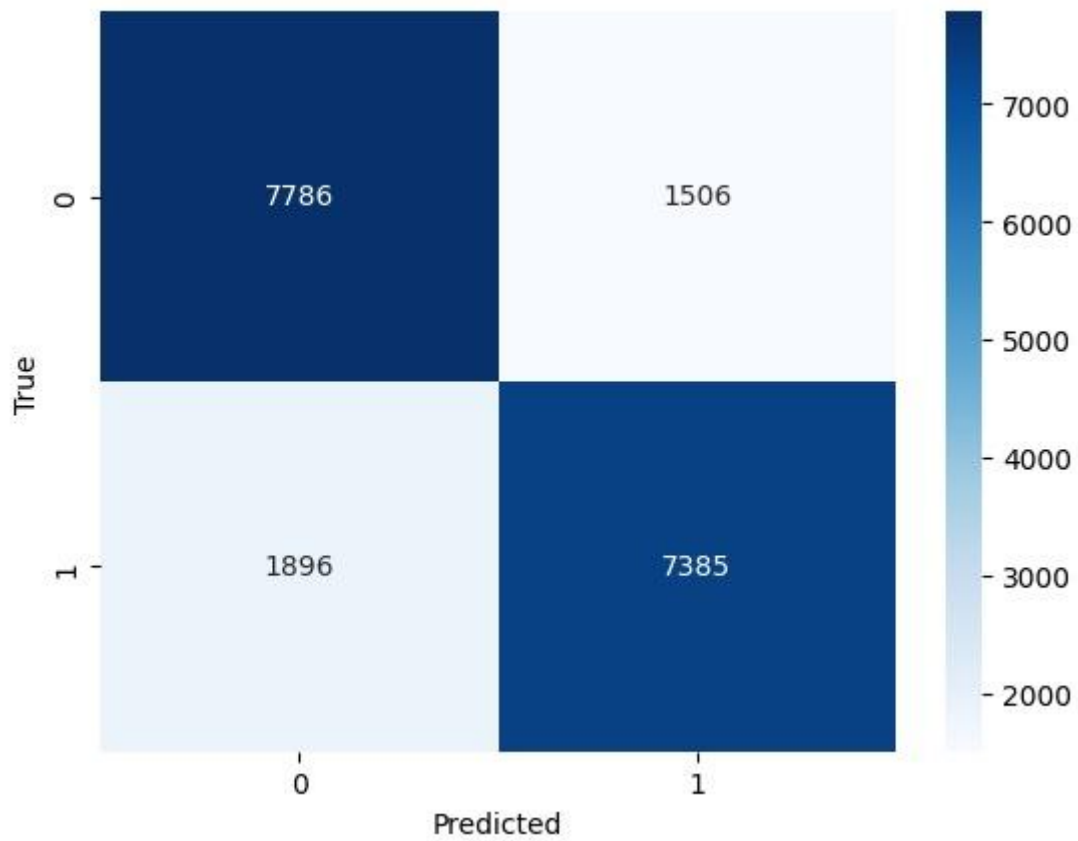
Por ser una gran cantidad de datos, consideramos que es necesario particionarlo al menos 5 folds.

Se supuso que la cantidad de datos computacional requerida para 10 o más folds era excesiva

Realizamos un cross validation y encontramos los siguientes hiperparámetros que resultaron ser los mejores:

- Criterio: Gini
- Ccp\_alpha:  $7,7 \cdot 10^{-5}$
- Max depth: 22





Utilizando los parámetros anteriormente mencionados, el modelo alcanzó las siguientes métricas:

- Accuracy: 0.8168
- Recall: 0.7957
- Precision: 0.8306
- f1 score: 0.8127