

Trabajo Práctico 1: Reservas de Hotel

Checkpoint 2: Árboles de decisión

75.06 / 95.58 Organización de Datos - FIUBA

Ing. Rodríguez - 1°C 2023



Grupo 31 - "Datazo":

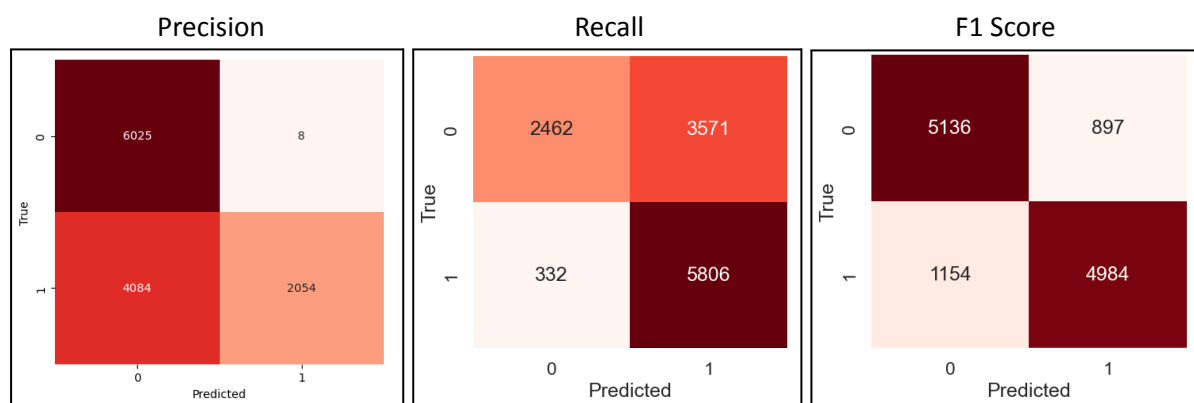
- 106203 - Kisinovsky, Diego Andrés
- 102685 - Mena Giraldo, Michael Gustavo
- 104256 - Brocca, Pablo Martín

Introducción

En este checkpoint construimos un modelo clasificador de tipo árbol de decisión para comenzar a hacer predicciones sobre el dataset de test.

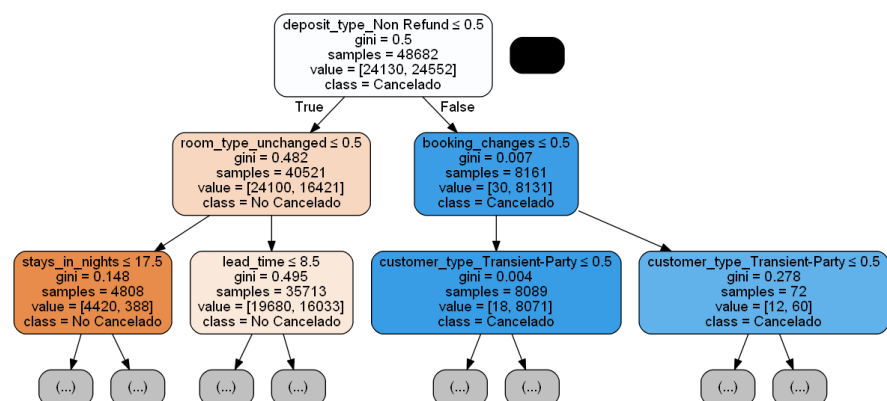
Observaciones del análisis

Elegimos la técnica Random Search Cross-Validation (CV) para encontrar los mejores hiperparámetros del árbol de decisión. Probamos las métricas precision, recall y F1 Score, obteniendo sus correspondientes matrices de confusión:



Se puede ver que con precision y recall se obtiene mayor cantidad de True Negatives (TN) y True Positives (TP) respectivamente, pero el modelo no minimiza los False Negatives (FN) en el primer caso ni los False Positives (FP) en el segundo. Elegimos F1 Score porque tiene en cuenta tanto precision como recall, maximizando ambas métricas y minimizando los falsos positivos y negativos.

Por otro lado, probando distintos niveles de profundidad del árbol, determinamos límites inferior y superior de $6,5e-5$ y $7e-5$ para el parámetro *cpp_alpha*, lo que mejora el score de la métrica elegida pero complejiza el árbol considerablemente.



Notamos que cuando el árbol era más simple (*cpp_alpha* más alto) la cantidad de combinaciones del Random Search CV no variaba significativamente; por el contrario, aumentar el número de combinaciones con un árbol más complejo (*cpp_alpha* más bajo) mejoró las métricas.