

Checkpoint 2 - Grupo 25

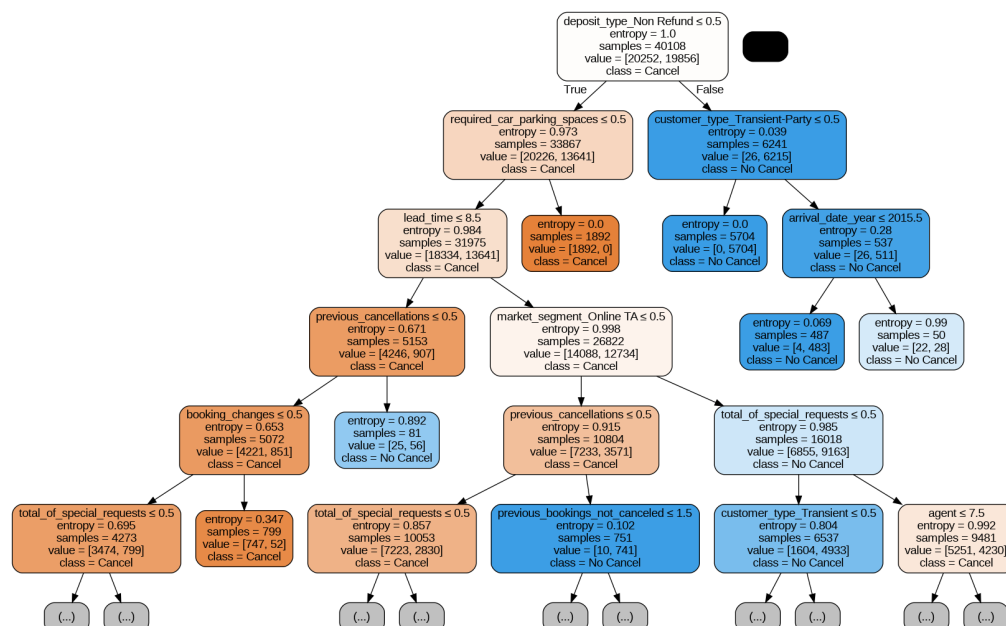
Introduccion

En el desarrollo de esta etapa del TP N°1, realizamos *One Hot Encoding* para las variables cualitativas y buscamos los mejores hiper parámetros con el algoritmo de *Random Search* ya que nos resultó más rápido en cuestión de ejecución y además notamos que hay cambios que se ven más explícitos a la hora de hacer las n combinaciones.

Construcción del modelo

Optimizamos los hiperparámetros de `'criterion'`, `'min_samples_leaf'`, `'min_samples_split'`, `'ccp_alpha'`, `'max_depth'`. Utilizamos K-fold Cross Validation, consiguiendo nuestro mejor modelo a 20 folds. Para buscar los hiper parámetros utilizamos F1-Score, que aumentó su score desde un valor de 0.71 hasta 0.8 aproximadamente al final del desarrollo.

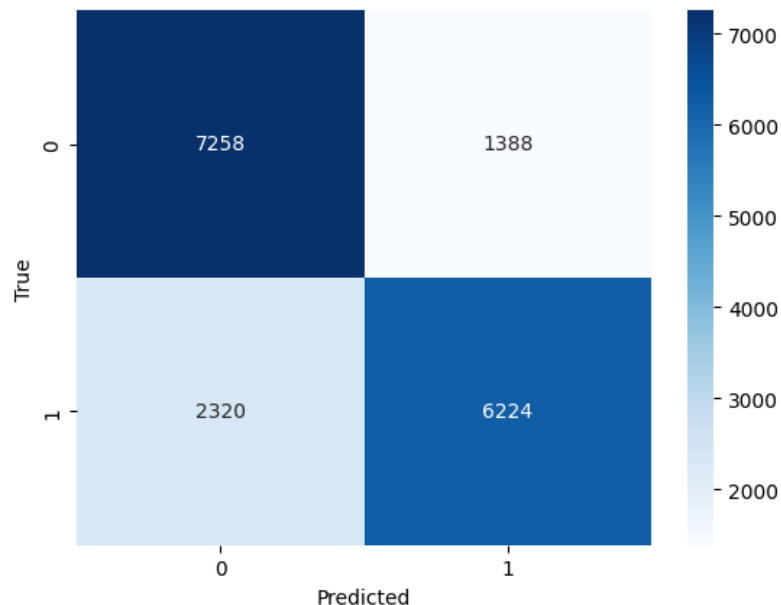
Además, realizamos una observación acerca de la métrica elegida y llegamos a la conclusión de que la entropía es preferible ya que tiende a dar más peso a la igualdad de clases y puede ser más sensible a las diferencias, siendo nuestro conjunto de datos, también equilibrado. No hay diferencia muy amplia contra gini, pero con entropy obtenemos leves mejoras en los resultados.



Cuadro de Resultados

Modelo	F1-Test	Presicion Test	Recall Test	Metrica X	Kaggle
modelo_1	0.7159	0.72	0.72	Entropy	0.72223
modelo_2	0.7614	0.79	0.76	Entropy	0.76261
modelo_3	0.7898	0.79	0.78	Entropy	0.79292

Matriz de Confusion



Encontramos un total de:

- Verdaderos negativos hay 7258 casos. Las predicciones acerca de las reservas no canceladas hay un total de 7258 predicciones acertadas.
- Verdaderos positivos hay 6224 casos. Estos son los casos en los que el modelo predice correctamente que una reserva es cancelada.
- Falsos positivos hay 1388 casos. Estos son los casos en los que el modelo predice incorrectamente que una reserva se cancelará cuando, en realidad, no se cancela.

- Falsos negativos hay 2320 casos. Estos son los casos en los que el modelo predice incorrectamente que una reserva no se cancelará cuando, de hecho, se cancela.

Maximizar los verdaderos negativos y positivos es importante. Tener esta información es muy útil para el hotel para poder planificar la asignación de recursos y gestionar sus operaciones de manera eficiente.

Podemos detectar que hay casi el doble de casos en los que se comete un error de tipo falso positivo, el cual tiende a ser más desastroso, ya que estaríamos “cancelando” una reserva cuando en realidad esa persona está dispuesta a llevarla a cabo. Minimizar este caso, y el de los falsos negativos es importante. Pueden llevar a problemas como la sobre reserva de habitaciones y la pérdida de ingresos para el hotel, ya que se asignarían recursos adicionales innecesariamente. Por ejemplo, podría llevar a la falta de disponibilidad de habitaciones cuando los clientes las necesiten.

Tareas Realizadas

Integrante	Tarea
Mariana Juarez Goldemberg	Armado de Reporte Construcción de árboles de predicción. Optimización de parámetros.
Miranda Marenzi	Armado de Reporte Construcción de árboles de predicción. Optimización de parámetros. Análisis del modelo.
Lisandro Román	Armado de Reporte Análisis del modelo.