



Docente: Rodriguez, Juan Manuel.

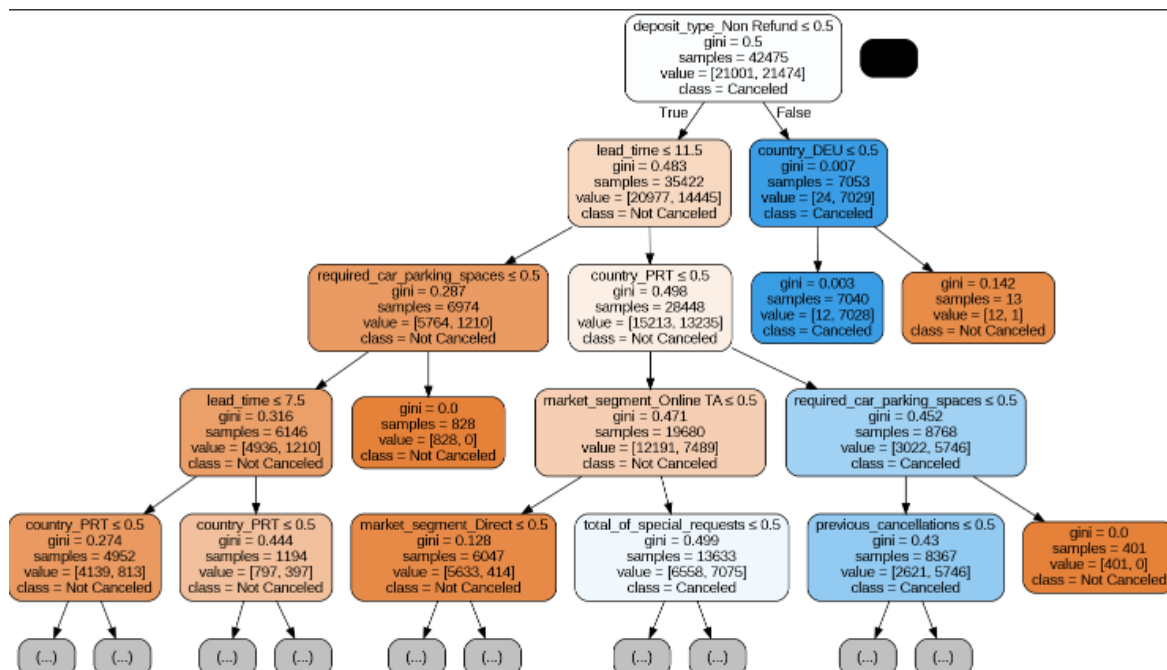
Estudiantes:

- Mendoza Hernandez, Sabrina Scarlet (108524)
- Valeriani, Matias Gabriel (108570)
- Jang, Lucas (109151)

Checkpoint 2: Árbol de decisión

Para esta segunda etapa del trabajo práctico realizamos las primeras predicciones, mediante los árboles de decisión. Para ellos, tuvimos que realizar modificaciones en los diferentes dataframes, como agrupar variables, convertir variables de tipo objeto a numéricas para poder predecir (convirtiendo a variables dummies). Luego modificamos el test de modo que ambos tengan la misma cantidad de columnas para poder aplicar nuestro modelo. Antes de elegir el mejor modelo, realizamos la búsqueda de los mejores hiper parámetros con una búsqueda random de cross validation, obteniendo un modelo con especificaciones de tipo: gini como criterio, 19 de profundidad máxima y 0,0001 de poda, que luego guardamos en un archivo joblib. Los resultados obtenidos fueron una precisión de 0,87 para los casos que no eran cancelados y 0,84 para los caso que era efectivamente cancelados, además llegamos a F1 Score de 0.8589 en la notebook y de 0.8420 en los submission de Kaggle.

Árbol de decisión



Matriz de confusión

	precision	recall	f1-score	support
0	0.87	0.84	0.85	9000
1	0.84	0.87	0.86	9204
accuracy			0.85	18204
macro avg	0.86	0.85	0.85	18204
weighted avg	0.86	0.85	0.85	18204

F1-Score: 0.8589729902850431

Text(50.72222222222214, 0.5, 'True')

