

Trabajo Práctico 1 : Reservas de Hotel

Conclusiones

Modelos

A lo largo del trabajo se evaluó una gran cantidad de modelos. Según los resultados obtenidos, destacan como los mejores estimadores los modelos Random Forest y XGBoost.

Feature engineering

Según los modelos entrenados y los resultados obtenidos, mi apreciación es que el feature engineering es el punto mayor relevancia a la hora de obtener una buena predicción.

Se han entrenado modelos con cross validation, optimizando los distintos hiperparámetros, sin obtener mejoras significativas. Sin embargo, al volver a analizar el dataset y aplicar un filtrado o imputación de datos distintos, se observaron mejoras significativas en la predicción.

Por ejemplo, en un modelo RF con f1-score de 0,844, se volvió a analizar el dataset y se modificó la imputación de la columna 'country'. Ésta tenía una gran cantidad de valores. Se decidió mantener aquellos valores que representan más de un 1% sobre el total.

Aplicando esta imputación, se llegó a un valor de 0,878

Otras opciones a explorar

Sin dudas que otras opciones que hubiera explorado en este trabajo es la reducción de dimensionalidad y un análisis más intensivo sobre el dataset, a fin de comprenderlo mejor y aplicar mejores estrategias en el feature engineering.