

Trabajo Práctico 1

Reservas de Hotel

Integrantes

Apellido	Nombre	Grupo
Lopez	Francisco	17
Corn	Franco	17
Queirolo Dominguez	Cristian Daniel	17

Informe

Introducción

En esta segunda parte del tp realizaremos el checkpoint 2, además resolveremos las correcciones hechas de la primera entrega y el punto d (Valores atípicos) .

Objetivos

- a) Construir árboles de decisión y optimizar sus hiper parámetros mediante k-fold Cross Validation para obtener la mejor performance. ¿Cuántos folds utilizaron? ¿Qué métrica consideran adecuada para buscar los parámetros?
- b) Graficar el árbol de decisión con mejor performance encontrado en el punto anterior. Si es muy extenso mostrar una porción representativa.
- c)Analizar el árbol de decisión seleccionado describiendo los atributos elegidos, y decisiones evaluadas (explicar las primeras reglas obtenidas).
- d) la performance del modelo en entrenamiento y validación, explicar todas las métricas y mostrar la matriz de confusión. e. Generar predicciones con el conjunto de test y realizar los submits correspondientes en la competencia de Kaggle.

Desarrollo

Como primera parte del tp resolveremos las correcciones hechas de la primera etapa (no hay mucho para decir de esto, esperemos que hayan quedado resueltas de forma correcta).

Para la creación de árboles empezaremos desde lo más simple, creando un árbol completo con todos nuestros valores. Graficamos el árbol, su métricas, matriz de confusión y lo evaluaremos, vemos rápidamente que, si bien el árbol tiene unas buenas métricas, vemos que está usando una cantidad de datos exagerada (de los cuales seguro muchos no son necesarios).

Realizando una poda en el árbol, vemos que si bien bajan nuestras métricas, llegamos a tener unas buenas métricas usando una cantidad de datos mucho mejor. Lo que haremos entonces, será buscar hiper parámetros, para tratar de llegar a un equilibrio.

Al realizar la búsqueda de los hiper parámetros, llegamos a un árbol más equilibrado, elevando un poco el f1 score de nuestro árbol podado y pudiendo ver las variables que más útiles son a la hora de realizar una predicción. (dejamos más detallado nuestro paso a paso en el tp, junto con qué criterios realizamos el punto d de la primera etapa)