

# Trabajo Práctico 1 - Reservas de Hotel

## Informe Final

Integrantes:

Ignacio Latorre - Padrón: 101305

Nicolas Ronchese - Padrón: 108169

Gastón Avila - Padrón: 104482

### Introducción

El objetivo principal de este trabajo práctico fue estudiar técnicas de análisis exploratorio de datos y aprendizaje automático para desarrollar modelos predictivos precisos que comprendan los factores que influyen en la cancelación de reservas de hotel. El trabajo práctico se dividió en cuatro checkpoints, cada uno de los cuales abordó aspectos específicos del problema. A continuación, se presenta una breve descripción de cada uno de ellos:

#### **Checkpoint 1:** Análisis Exploratorio de Datos

En este checkpoint, se realizó una exploración detallada del conjunto de datos de reservas de hotel. Se identificaron las variables relevantes y se examinó su distribución, se analizó la relación entre las variables y el target (cancelación de la reserva) y se detectaron posibles valores atípicos. Además, se buscaron patrones o tendencias que pudieran ser útiles en la etapa de modelado. Se destacó la correlación positiva entre la variable "lead\_time" y la cancelación de reservas.

#### **Checkpoint 2:** Árboles de Decisión

En este checkpoint, se aplicó la técnica de árboles de decisión para predecir la cancelación de reservas.

Se realizaron ajustes en el conjunto de datos, como la codificación de las variables categóricas como "room\_type" y "reservation\_channel" en variables numéricas mediante técnicas como one-hot encoding, se trató el problema de valores faltantes y se dividió el conjunto de datos en conjuntos de entrenamiento y prueba en una proporción de 70:30. Luego, se entrenaron distintos modelos de árboles de decisión ajustando los hiperparámetros para mejorar su rendimiento. Por ejemplo, se probó con diferentes valores de profundidad máxima y se encontró que una profundidad de 7 producía mejores resultados, donde la puntuación F1 fue del 81%.

#### **Checkpoint 3:** Ensamblados

En este checkpoint, se exploraron diferentes técnicas de ensamble de modelos para mejorar la precisión de las predicciones. Se utilizaron algoritmos como KNN, SVM, Random Forest y XGBoost, y se optimizaron sus hiperparámetros mediante técnicas de búsqueda exhaustiva. El mejor rendimiento obtenido fue el caso de XGBoost. Se realizaron diversas corridas optimizando los hiperparámetros y se obtuvo un puntaje de 0.82 en la competencia utilizando un learning rate de 0.5. También se exploraron técnicas de stacking y voting utilizando los resultados de los modelos anteriores. Sin embargo, al realizar el envío de los resultados en la competencia, los puntajes obtenidos fueron más bajos de lo esperado, lo

cual refleja que se requiere una revisión exhaustiva del procesamiento de los datos y los algoritmos utilizados.

**Checkpoint 4:** Redes neuronales

Se generaron 3 modelos de redes neuronales probando distintas combinaciones de cantidad de capas, funciones de activación, optimizadores y funciones de error. Nos quedamos con las 3 mejores combinaciones que logramos con las cuales se repitió 2 veces el puntaje de 0,83 y uno de 0,84. Una opción podría haber sido correr un gridsearch o random search para probar la mayor cantidad de parámetros posibles.