



75.06 / 95.58

ORGANIZACIÓN DE DATOS

## Reporte TP N°1: Reservas de Hotel

### Entrega Final

Alen Davies Leccese: 107084

Luca Mauricio Lazcano: 107044

MAYO 2023

## 1 Conclusiones generales

A lo largo de este trabajo práctico, se tuvo la oportunidad de explorar y aplicar distintas técnicas de ciencia de datos, primero para preprocesar un conjunto de información, y luego para entrenar modelos y generar predicciones con nuevos datos.

Se entrenaron distintos modelos de Machine Learning para realizar una clasificación binaria en el conjunto de datos. En cada uno de los casos se buscó optimizar los hiperparámetros para mejorar la performance de los modelos. Los modelos explorados y entrenados fueron los siguientes:

- Árbol de decisión.
- K nearest neighbors.
- Support Vector Machine.
- Redes Neuronales.

Además, se realizaron ensambles de modelos, buscando mejorar las predicciones utilizando el poder de predicción de distintos modelos trabajando en conjunto.

- XGBoost (boosting).
- Random Forest (bagging).
- Voting (híbrido).
- Stacking (híbrido).

Se pudo observar que de los modelos entrenados, aquellos con mayor poder de predicción fueron los ensambles híbridos (voting y stacking) y las redes neuronales.

Existen innumerables variantes y parametrizaciones posibles de los modelos vistos, explorarlos todos llevaría un tiempo poco razonable.

Se debe destacar que en esta oportunidad, el problema a resolver era de clasificación binaria, que es sólo uno de los distintos problemas que se podría buscar resolver. Con otro problema, y otros datos, se podrían realizar soluciones de clasificación con múltiples clases: regresión (para predecir algún valor); agrupamiento (clustering), donde no se cuenta con un target, y el objetivo es que el modelo encuentre alguna relación entre las variables (unsupervised learning).

Sentimos que quedaron pendientes en este trabajo poder buscar y encontrar hiperparámetros y arquitecturas óptimas, sobre todo en los modelos más pesados computacionalmente.

Como apreciación personal de lo aprendido hasta ahora, consideramos fascinante el campo de la ciencia de datos, y cómo teniendo un conjunto de información, se pueden extraer predicciones tan precisas. Es muy interesante entender los mecanismos y algoritmos mediante los cuales un sistema computacional puede encontrar patrones e información que a simple vista puede parecer caótica e indescifrable.