**El práctico**

Esta competencia ha sido creada a partir de la competencia Spaceship Titanic ([Spaceship Titanic | Kaggle](https://www.kaggle.com/competitions/spaceship-titanic/overview)). Los datos fueron tomados originalmente de la competencia y se generaron los datasets de entrenamiento y evaluación.

**Requerimientos del práctico**

1. Realizar un análisis exploratorio de datos, y utilizar lo aprendido para generar su mejor modelo.
2. Superar el puntaje del baseline (público) en la competencia kaggle.
3. Usar al menos 3 modelos distintos al del árbol de decisión que se explora como baseline (se puede explorar más profundamente el árbol de decisión, pero aún así deberán explorar otros modelos).
4. Entregar un notebook con el análisis exploratorio de datos y el código con los 3 mejores modelos entregados en la competencia (de acuerdo a los resultados obtenidos).

**Pasos a seguir**

1. Crear una cuenta en kaggle.com.
2. Sumarse a la competencia acá: [Kaggle: Your Home for Data Science](https://www.kaggle.com/account/login?returnUrl=%2Ft%2Fd7dfd142ac1042168d052624b3dee531)

* Hacer click en "Join Competition".
* Aceptar las reglas.

1. Crear un equipo (Team): El trabajo se evaluará en los grupos asignados.
2. Pueden descargar los datos (Data), aunque también están incluidos en este repo.
3. Una vez realizada una predicción (ver ejemplo abajo), subir los resultados a kaggle haciendo click en "Submit Predictions" en la página principal de la competencia. Ahí deberán subir el archivo csv generado y describir (para sus registros) qué están subiendo.

**Un ejemplo**

Adjuntamos una implementación que tiene por objetivo:

* Levantar los datos que usaremos.
* Analizar de una manera simple los datos.
* Preparar los datos para procesarlos con un modelo en particular.
* Crear un baseline para la competencia.
* Generar el archivo que se subirá a kaggle para su evaluación.

**Subir una predicción a Kaggle**

En el ejemplo de baseline que se entrega, se genera un archivo en el path data/submission.csv. Tal archivo es un csv con un formato en particular, que asigna a números de visita en el conjunto de test, una predicción de si fue teletransportado o no a otra dimensión alternativa. Ese archivo debe ser subido a kaggle como lo explicamos arriba: haciendo click en "Submit Predictions" en la página principal de la competencia.

**Algunas consideraciones**

* En el baseline solo se utiliza cross-validation (mediante GridSearchCV) para evaluar el modelo, son libres de generar un subconjunto de validación aparte del conjunto de entrenamiento si así lo desean.
* Los features escogidos no tienen ningún análisis y son casi por defecto. Parte del proceso de encontrar un buen modelo es ver cómo trabajar dichos features.
* La métrica a optimizar será el accuracy score.