# 01 Actividad práctica: Evaluación realista en el Titanic: Riesgos del sobreajuste y la evaluación poco realista

### Objetivo:

Los estudiantes aprenderán cómo un modelo de Machine Learning puede tener un rendimiento sobreestimado cuando el conjunto de prueba no es independiente y cómo este tipo de error puede afectar la toma de decisiones en proyectos reales.

### Requisitos:

- **Python** (con bibliotecas como **pandas**, **scikit-learn**, **matplotlib**)
- **Jupyter Notebook** o entorno similar
- Conjunto de datos de Titanic disponible en [Kaggle Titanic Dataset].

## Descripción de la actividad:

1. **Cargar y preprocesar los datos**: Cargar el conjunto de datos Titanic desde un archivo CSV o Kaggle, preprocesar los datos (tratamiento de valores faltantes, codificación de variables categóricas, etc.).

2. **Entrenar un modelo**: Usar un modelo de clasificación (por ejemplo, **Árbol de Decisión** o **Regresión Logística**) para predecir la supervivencia de los pasajeros basándose en características como la clase, el sexo, la edad, el puerto de embarque, etc.

3. **Evaluación del modelo**:

    - **Evaluación inicial (métrica con datos de prueba no independientes)**: Evaluar el modelo usando un conjunto de prueba que tenga algunos datos duplicados o filtrados del conjunto de entrenamiento.
    - **Evaluación con datos de prueba verdaderamente independientes**: Evaluar el modelo en un conjunto de prueba completamente independiente (datos que no se han utilizado en el entrenamiento).
4. **Observación de las métricas**: Los estudiantes compararán las métricas de rendimiento (precisión, recall, F1-score, etc.) obtenidas de ambas evaluaciones (no independiente e independiente) y observarán cómo se inflan las métricas en el primer caso.

5. **Conclusión**: Los estudiantes documentarán cómo un conjunto de prueba no independiente puede dar lugar a una evaluación sobreestimada y cómo esto puede afectar la implementación de modelos en la práctica.